The theory of network regulation in the presence of sunk costs

Technical report prepared for
Commerce Commission

5 August 2009

Erwin Diewert, Denis Lawrence and John Fallon

Economic Insights Pty Ltd
6 Kurundi Place, Hawker, ACT 2614, AUSTRALIA
Ph +61 2 6278 3628 Fax +61 2 6278 5358
Email denis@economicinsights.com.au
WEB www.economicinsights.com.au
ABN 52 060 723 631

CONTENTS

1 OVERVIEW

The Commerce Commission has engaged Economic Insights Pty Ltd ('Economic Insights') to prepare a report which considers the interrelationship between the choice of asset valuation method and CPI–X price paths set using productivity analysis. To adequately address this topic it has been necessary to revisit the theory of regulation and fill in some important gaps which have existed to date. This report presents the technical detail of the theory developed to allow a robust underpinning of productivity–based regulation in the presence of sunk costs.

This technical report accompanies our report to the Commerce Commission specifically addressing asset valuation and implementation issues (Economic Insights 2009).

The current theory of regulation – and its application to CPI–X price cap regulation in particular – has evolved in a relatively piecemeal way that has neither adequately addressed key economic welfare issues nor recognised some of the key characteristics that distinguish infrastructure industries and make them natural monopolies.

Most previous contributions have relied on partial equilibrium models that only model aspects of the industry in question and not the interactions between that industry, consumers and factors of production. For example, in the seminal paper by Bernstein and Sappington (1999), their objective is to define a regulatory regime which will lead to the smallest possible rate of proportional growth in the prices of regulated products, while maintaining the solvency of the regulated firm. It seems intuitively obvious that this is a welfare enhancing activity but is it *optimal* to reduce all regulated prices by the same proportion? We have no way of answering this question using their partial equilibrium methodology.

To provide rigorous guidance to regulators on the courses of action that will enhance economic welfare we need to move beyond partial equilibrium analysis to general equilibrium analysis. This report extends the theory of regulation by embedding the regulated firm in a small general equilibrium model of an open economy. The role of the regulator in this model is to improve the welfare of households in the economy. This approach inevitably involves the use of more demanding mathematical analysis but provides a much higher level of rigour.

Previous contributions to regulatory theory have also not explicitly recognised that capital inputs in most regulated infrastructure industries have the character of *sunk costs*; ie once the investment in these assets has been made, the firm is generally stuck with these assets and cannot readily vary the service capacity of these assets during their useful lives. The existence of sunk cost assets greatly complicates the regulator's responsibilities and changes the nature of some key regulatory theory findings.

1.1 Main findings

The main findings of this analysis are:

- To improve economic welfare regulators need to move regulated prices closer to their corresponding marginal costs and provide incentives for the regulated firm to improve its productivity performance.

- The information required to implement optimal regulation is difficult to obtain and so simpler methods of regulation that are not fully optimal, like price cap regulation, will have to be used in practice.
- Price cap regulation can be modified to accommodate both sunk costs and financial capital maintenance.
- In the presence of sunk costs, full price cap regulation requires information on opex price changes, changes in the amortisation schedule for sunk costs allowed by the regulator, the rate of technical progress, the deviation of prices from marginal costs, the deviation of allowed amortisation charges from corresponding user benefits, changes in outputs and sunk assets, costs and revenue, and the desired change in excess profits (equation 278).
- Allowed amortisation charges replace the capital goods price index in the price cap formula when there are sunk costs.
- If excess profits are close to zero, implementation of the price cap can be simplified to the sum of the rate opex price change weighted by the share of opex costs in revenue and the change in approved amortisation charges weighted by the share of amortisation charges in revenue less the rate of technical progress weighted by the opex share in revenue (equation 282).
- There is no guarantee that future rates of technical progress will mirror past rates.
- Extrapolations of past TFP growth are often used as a proxy for future technical change but TFP growth in the context of a regulated firm is far from being identical to technical progress. In fact, conventional TFP growth depends not only on technical progress but also on variables that are controlled by the regulator including profits, the selling prices of regulated products and allowable amortisation charges.
- Where CPI–X regulation is used, the X factor involves the difference between the firm's TFP growth weighted by its costs relative to its revenue and the economy–wide TFP growth rate plus the difference between economy–wide input price change and the sum of the firm's opex price growth and amortisation charges growth each weighted by their the respective shares of their cost in revenue plus a nonzero profits adjustment term less a rate of change in regulated profits term (equation 305).
- When dealing with the regulation of many firms using a common productivity target as part of the price cap, it is necessary to move to a method of price cap regulation that uses information that goes well beyond the use of conventional (revenue and cost weighted) TFP measures and that measures exactly what regulated services are being provided by the firms in the group, independently of the institutional factors that determine exactly how the firms are paid for providing these services.
- A range of asset valuation methodologies can be consistent with financial capital maintenance. Each methodology will generate a time–series of asset values and the series of amortisation charges are used to ensure financial capital maintenance is achieved.
- The main difference between the asset valuation methods is on the timing of revenue receipts rather than their net present value. The important requirements are that the

amount actually invested is the opening asset value in the first period and the scrap value is the closing asset value in the last period.

- Amortisation charges based on CPI indexed historic cost and the use of a real return to capital are likely to be the most consistent with the concept of user pays and intertemporal efficiency although this could be further enhanced by indexing by the degree of capacity utilisation as well as by the CPI. This effectively 'back–end loads' the profile of receipts and requires a high degree of regulatory credibility for investors to be confident that the regulatory rules will remain unchanged for a sufficiently long period for them to recover their costs.

1.2 Modelling summary

In section 2, we abstract from the problem of sunk costs and consider a simple one period general equilibrium model where we have one household, one regulated sector and one competitive sector in a small open economy. In this model, we assume that the regulator controls the prices of regulated products. We derive conditions for *optimal regulation* in this economy as well as conditions for *regulatory improvement*; ie we do not assume that regulation is optimal, but rather we look at a nonoptimal equilibrium and try to work out conditions for regulation to locally improve welfare.

In section 3, we explore the comparative statics of the one period model developed in section 2. In particular, we look at how consumer welfare changes as various exogenous variables change, including primary input endowments, changes in international prices, and, most importantly for our purposes, changes in regulated prices. In this simple model we find that the regulator can improve welfare by increasing the prices of products that are being sold below their marginal costs and by decreasing the prices of products that are being sold above their marginal costs.

In section 4, we follow the example of Denny, Fuss and Waverman (1981) and Bernstein and Sappington (1999) and look at the mechanics of welfare change in a continuous time framework. The results are entirely analogous to the discrete time results obtained in section 3 except that in the continuous time framework, we can deal with technical progress in a reasonably straightforward manner. Not surprisingly, in section 4.2, we find that technical progress in both the regulated and unregulated sectors lead to welfare improvements for the household sector. In section 4.3, we follow Solow (1957) and Denny, Fuss and Waverman (1981) to show how technical progress can be measured in the competitive and regulated sectors using appropriately defined Divisia (1926) indexes. It should be noted that in the regulated sector, in order to measure technical progress, we require information on the marginal costs of producing the regulated products, information that is not generally readily available.

Section 4 relied on the regulated firm's cost function; ie the total cost of producing the vector of regulated products. Implicitly, section 4 assumed that capital inputs were not sunk cost inputs; ie capital inputs could be sold as second hand goods in the marketplace at the end of each period and hence the usual user cost of capital could be used as the price for a capital input.[1] However, in many regulated industries, substantial components of the capital stock in use have the nature of sunk costs; ie once the investment is made, the firm is stuck with the associated bundle of capital services until the assets are completely worn out so that they have no resale value on second hand markets. Thus, the usual user cost methodology is not applicable in this context and it will be necessary to work out a new methodology. In section 6, we will work out an appropriate methodology to deal with sunk cost investments which will involve *operating cost or opex cost functions* for the regulated firm. An opex cost function minimises the variable input costs associated with producing an output target, conditional on the availability of a vector of fixed capital stock components. Thus, in section 5, we prepare for section 6 by reworking the results in section 4 using the opex cost function instead of the regular total cost function. Naturally, we obtain the same results in section 5 as were derived in section 4 but the purpose of section 5 is to develop the opex cost function necessary for subsequent analysis instead of the usual total cost function.

In section 6, we consider the sunk cost problem in the context of a two period general equilibrium model. We consider a situation where the regulated firm has to choose a sunk cost capital input which will last the useful life of the input (two periods) but the service flow that the asset yields in each period is constant. In section 6.1, we set up the intertemporal analogue to the one period regulatory model studied in section 2. In section 6.2, we develop the comparative statics implications of our new model in a manner analogous to the analysis of the one period model. The intertemporal comparative statics results are entirely analogous to the one period results developed in section 3 except that the treatment of the sunk cost capital is now different. In the one period model, the negative of the partial derivative of the opex cost function could be set equal to a potentially observable user cost. In the intertemporal model, we no longer have this equality. Instead of a period t user cost, we have a period t user benefit defined as the negative of the partial derivative of the period t opex cost function with respect to the sunk cost capital stock and a (discounted) sum of these user benefit terms is set equal to the purchase price of the capital input. These partial derivatives of the opex cost functions are generally not directly observed and so must be estimated, either using econometric techniques or accounting cost allocation methods. As we shall see in later sections, this changed treatment of capital leads to substantial changes to the one period results derived under the assumption that capital is freely variable from period to period. In section 6.4, we generalise the one hoss shay depreciation model to allow for other depreciation profiles. Finally, section 6.5 introduces interest rates and discounting explicitly into our intertemporal general equilibrium model. These discounting complications lead to more complicated notation but the basic results are unchanged.

The full model of regulation developed requires too much information for the regulator to be able to implement it in its entirety. Our intertemporal regulatory model requires information on:

- The partial derivatives of the period by period opex cost functions with respect to regulated outputs; ie information on marginal costs are required;
- The partial derivatives of the period by period opex cost functions with respect to sunk cost capital stock components and

[1] See Jorgenson and Griliches (1967) for an introduction to user costs.

- Consumer intertemporal substitution matrices between regulated and unregulated products.

It will be very difficult for the regulator to obtain accurate information on the above data requirements for optimal regulation. Thus, in the remaining sections we look at second and third best methods of regulation.

In section 7, we consider a simple method of *price cap regulation* that relies on information on a single regulated firm. We derive the simple price cap formula (278) which involves a price index for variable inputs, a price index for the amortisation amounts allowed by the regulator for sunk cost capital stock components, a measure of the anticipated rate of technical progress in the regulated sector, another measure involving the deviations of regulated prices from their corresponding opex marginal costs and a final measure involving the deviations of the allowed amortisation amounts for sunk cost capital stock components from their corresponding marginal user benefits. The last two measures will be difficult for the regulator to estimate numerically.

In section 8, we consider methods of price cap regulation that are based on estimates of *Total Factor Productivity growth* (as opposed to the measure of *technical progress* which was used in section 7). Following Jorgenson and Griliches (1967), TFP growth can be defined as an index of output growth divided by an index of input growth, using market prices as weights in both indexes. We follow the continuous time derivation made by Bernstein and Sappington (1999) of a price cap formula using TFP growth as one of the major drivers of the price cap and derive the price cap formula (293). However, formula (295) shows that TFP growth is a complicated function of technical progress and other factors[2] including factors involving the deviations of regulated prices from their corresponding opex marginal costs and the deviations of the allowed amortisation amounts for sunk cost capital stock components from their corresponding marginal user benefits. We continue on with our adaptation of the analysis of Bernstein and Sappington (1999) and show how a counterpart to the traditional "CPI minus X Factor" regulatory price cap formula can be obtained in the context of sunk cost assets; see the price cap formula (304).

Our approaches to the derivation of price cap formulae have concentrated on the case where only a single firm is being regulated. In the case of a single firm, it does not really matter whether we use a price cap formula involving the rate of technical progress or the rate of TFP growth; in the end, the various formulae for the price cap should turn out to be equivalent as we show in section 8. However, when regulation involves several firms and past average rates of technical progress or of TFP growth are used in price caps going forward, then the measurement of these rates becomes critical. In particular, the use of average TFP growth rates across a number of regulated firms can create an uneven playing field since the ingredients which go into TFP growth as shown in formula (295) can contain terms which are beyond the control of the regulated firm. Thus, at the end of section 8, we caution that there are additional complications when we move from regulating a single firm to the regulation of a group of firms using peer group or yardstick methods of regulation.

In section 9, we note that when regulating a group of firms using productivity–based

[2] Our analysis here follows that of Denny, Fuss and Waverman (1981; 196-199) except that their analysis (and that of Bernstein and Sappington) did not deal with the complications due to sunk cost assets.

regulation, it is necessary to be very clear on what the definition of the regulated outputs is, since different output concepts can lead to very different estimates of both technical progress and TFP growth. As well as it being necessary to use comprehensive measures of output in this instance, it will also be necessary to use output cost share weights rather than revenue weights in forming the productivity measure.

Section 10 considers three specific methods of amortisation of a sunk cost investment that are consistent with the concept of financial capital maintenance. We show that amortisation charges based on CPI indexed historic cost and the use of a real return to capital is likely to be the most consistent with the concept of user pays and intertemporal efficiency although this could be further enhanced by indexing by the degree of capacity utilisation as well as by the CPI. This effectively ‘back–end loads’ the profile of receipts and requires a high degree of regulatory credibility for investors to be confident that the regulatory rules will remain unchanged for a sufficiently long period for them to recover their costs.

2 AN INTRODUCTION TO OPTIMAL ONE PERIOD REGULATION

In this section, we will focus on the regulator's problem in the context of a one period general equilibrium model of the economy under consideration. The model will be highly simplified in that there will only be one household in the economy, there will only be one regulated firm in the economy and there will only be one competitive sector in the economy. Another simplification is that there are no government expenditures on goods and services and no taxes in the model. A final simplification is that capital services will be regarded as just another primary input in this introductory model, so that we do not deal with the complications due to the durability of capital and the possibility of sunk costs or irreversible capital stock investment decisions. However, even with these simplifications, it will be seen that the problem of optimal regulation in this model is far from being trivial.

There are three classes of commodities in our model:

- N regulated outputs; a vector of regulated outputs will be denoted by the nonnegative vector $y \equiv [y_1,...,y_N]^T \geq 0_N$;[3]
- J unregulated outputs which are also traded internationally; a vector of unregulated outputs will be denoted by the nonnegative vector $Y \equiv [Y_1,...,Y_J]^T \geq 0_J$;
- K primary inputs (not traded internationally); the economy's endowment of primary inputs will be denoted by the strictly positive vector $v \equiv [v_1,...,v_K]^T >> 0_J$; the regulated sector's primary input demand vector will be denoted by the nonnegative vector $z \equiv [z_1,...,z_K]^T \geq 0_K$ and the unregulated sector's primary input demand vector will be denoted by the nonnegative vector $Z \equiv [Z_1,...,Z_K]^T \geq 0_K$.

We now list our assumptions on tastes and technology. We assume that there is a production possibilities set s for the regulated sector and another production possibilities set S for the unregulated sector. We make the following assumptions on these sets:

(1) For any $y \geq 0_N$, the set of inputs that can produce at least y, $\{z : (y,z) \in s\}$ is a nonempty, closed convex set.

(2) S is a nonempty closed convex cone.

Assumption (1) on the regulated sector is equivalent to the assumption that the regulated production function is a quasiconcave function in the case where there is only a single regulated output. Thus, assumption (1) is a generalisation of the quasiconcavity assumption that is routinely made in single output production theory to the case of multiple outputs. Assumption (1) does not rule out increasing returns to scale in the regulated sector; ie assumption (1) is perfectly consistent with outputs increasing faster than inputs as input scale increases.[4]

[3] Notation: all vectors are to be regarded as column vectors, y^T denotes the transpose of the column vector y, 0_N denotes an N dimensional vector of zeros, $p^Ty \equiv \sum_{n=1}^N p_n y_n$ denotes the inner product of the vectors p and y, $y \geq 0_N$ means each component of y is nonnegative, $y >> 0_N$ means each component of y is positive, $y > 0_N$ means $y \geq 0_N$ but $y \neq 0_N$ and $(Y,Z) \in S$ means that the point (Y,Z) belongs to the set S.

[4] A set S is convex if given any two points belonging to the set, then the straight line joining those two points also belongs to the set.

Assumption (2) is a standard assumption used to model competitive sectors in the economy. The assumption that the production possibilities set S for the unregulated sector is a cone means that returns to scale in this sector are constant.[5]

We assume that the household has preferences defined over nonnegative consumption vectors of regulated products, $x \geq 0_N$, and nonnegative vectors of unregulated products, $X \geq 0_J$. We further assume that these preferences can be represented by the utility function $U(x,X)$ where U is a nonnegative, increasing, continuous and concave function.[6]

Since there is only one household in the economy, it will be optimal to maximise this household's utility subject to various resource constraints that face the household and producers in the economy. Three of these constraints are (3)-(5) below.

(3) $y \geq x$;

(4) $v \geq z + Z$;

(5) $P^T[Y-X] \geq 0$.

The constraints (3) merely require that the vector of regulated outputs $y \geq 0_N$ produced by the regulated sector be equal to or greater than the corresponding household demand vector $x \geq 0_N$ for these products. The constraints in (4) require that the economy's total available *supply vector for primary inputs* $v >> 0_K$ be equal to or greater than the sum of the demand for primary inputs from the regulated sector $z \geq 0_K$ plus the demand vector $Z \geq 0_K$ from the unregulated sector. Finally, $P >> 0_J$ is an exogenously given vector of *world prices for internationally traded products* (which are also the products produced by the competitive sector); $Y \geq 0_J$ is the economy's total production of these commodities and $X \geq 0_J$ is the household sector's demand for these commodities. Thus, $Y-X$ is the *economy's net export vector* and the single inequality constraint in (5) says that the value of net exports (the value of exports less the value of imports) should be nonnegative. Thus, (5) is the *economy's balance of trade* constraint.[7]

Now we are in a position to set up the *household's constrained utility maximisation problem*, which is (6) below:

$$(6) \max_{u,y,Y,z,Z,x,X} \{u : y - x \geq 0_N ; v - z - Z \geq 0_K ; P^T[Y-X] \geq 0 ; (y,z)\in s ; (Y,Z)\in S ; U(x,X) - u \geq 0\}$$

where u is the household's utility level and all other decision variables, y,Y,z,Z,x,X, have already been defined.[8]

---

[5] The assumption that S is a cone means that if $(Y,Z)\in S$ and $\lambda \geq 0$, then $(\lambda Y,\lambda Z)\in S$.

[6] The assumption that U is concave is the only nonstandard assumption. This assumption means that U has the following property. Let (x^1,X^1) and (x^2,X^2) be nonnegative vectors and let λ be a scalar that satisfies $0 \leq \lambda \leq 1$. Then $U(\lambda x^1+(1-\lambda)x^2,\lambda X^1+(1-\lambda)X^2) \geq \lambda U(x^1,X^1) + (1-\lambda)U(x^2,X^2)$. The standard assumption on the utility function U is that it be quasiconcave rather than concave. However, Diewert (1973; 424) using some results due to Afriat (1967) showed that using a finite data set of consistent consumer choices, it is not possible to reject representing the consumer's preferences by a concave utility function if they can be represented by a quasiconcave function. Thus, the assumption of concavity is empirically harmless.

[7] We are assuming that the country is small and thus the international price vector P is fixed.

[8] The nonnegativity constraints $u \geq 0$, $y \geq 0_N$, $Y \geq 0_J$, $z \geq 0_K$, $Z \geq 0_K$, $x \geq 0_N$ and $X \geq 0_J$ should also be added to the constraints in (6) but in the interests of brevity, we have omitted these constraints.

It seems clear that the regulator's *optimal regulation problem* should be the problem of setting up a regulatory regime that would guide the economy to the solution to the household's constrained utility maximisation problem (6) above. But how exactly can this be accomplished?

At this point, it is useful to step back from the above rather formal analysis and review the early history of the theory of optimal regulation.

Some 60 to 80 years ago, economists thought that forcing a monopolist to sell each product at a price equal to the marginal cost of producing it would suffice to bring about an ideal or optimal allocation of resources.[9] The deficits that this policy would lead to for firms with an increasing returns to scale technology (or large fixed costs) were to be supported by the general taxation powers of the state. However, the price equals marginal cost solution to the control of a monopoly was successfully attacked on a number of grounds:

- Fleming (1944; 336) and Wilson (1945; 456) pointed out that there was no natural index available to *evaluate* the performance of managers, or put another way, how can the regulator determine marginal costs?
- Wilson (1945; 457) also raised a problem that was later stressed by Domar (1974; 4): how can the regulator *motivate* the manager of the regulated industry to take the "optimal" course of action?
- Wilson (1945; 458-459) also noted the fact that any deficits generated by the regulated industry were to be covered out of general government revenue and this would create a tremendous incentive for an empire building manager to expand unduly. In the end, political bargaining would determine the allocation of resources due to the imprecision of the price equals marginal cost rule in an intertemporal context.
- Coase (1945; 113) (1946; 176) pointed out that the Pigou-Hotelling-Lerner price equals marginal cost rule would redistribute income to consumers of products in which fixed costs form a high proportion of total costs.
- Finally, Hotelling (1939; 155) and Coase (1946; 179) both noted that if taxes on fixed factors could not cover the government's revenue needs (for subsidies to the regulated industry), then covering the deficits of regulated industries will have adverse efficiency effects in the rest of the economy; ie raising tax revenues to cover deficits in the regulated sector will generally lead to a considerable amount of deadweight loss in the rest of the economy.

The problems with the marginal cost pricing rule listed above are still with us today. However, the early literature reviewed above did not derive the marginal cost pricing rule in a very rigorous fashion. Hence our first task will be to show the relationship of the consumer's constrained utility maximisation problem defined by (6) above to the marginal cost pricing literature.

We will assume that a strictly positive solution to (6) exists; ie $u^* > 0$, $y^* >> 0_N$, $Y^* >> 0_J$, $z^* >> 0_K$, $Z^* >> 0_K$, $x^* >> 0_N$ and $X^* >> 0_J$ solves (6). Now let $y \geq 0_N$ be in a neighbourhood of

[9] See Walras (1980; 83), Pigou (1920; 278), Hotelling (1938; 256), Lerner (1944; 182), Meade (1944) and Fleming (1944).

y^* and consider the following constrained maximisation problem that is conditional on the choice of y:

$$(7)\ H(y) \equiv \max_{u,Y,z,Z,x,X} \{u : y - x \geq 0_N ; v - z - Z \geq 0_K ; P^T[Y-X] \geq 0 ; (y,z)\in s ;$$
$$(Y,Z)\in S ; U(x,X) - u \geq 0 ; u \geq 0, Y \geq 0_J, z \geq 0_K, Z \geq 0_K, x \geq 0_N, X \geq 0_J\}.$$

It can be verified that (7) is a concave programming problem and hence the Karlin (1959; 201-203) Uzawa (1958) Saddle Point Theorem can be applied to this problem. Using this Theorem, we can absorb some of the constraints into the objective function and it turns out that u, Y, z, Z, x, and X solutions to (7) are also solutions to the following max-min problem:[10]

$$(8)\ H(y) = \max_{u,Y,z,Z,x,X} \min_{p,w,\lambda} \{u + p^T[y - x] + w^T[v - z - Z] + \lambda P^T[Y-X] : (y,z)\in s ;$$
$$(Y,Z)\in S ; U(x,X) - u \geq 0 ; u \geq 0, Y \geq 0_J, z \geq 0_K, Z \geq 0_K, x \geq 0_N, X \geq 0_J,$$
$$p \geq 0_N, w \geq 0_K, \lambda \geq 0\}$$

where p can be interpreted as a *vector of prices for the regulated products*, w can be interpreted as a *vector of primary input prices* and λ is a Lagrange multiplier that corresponds to the balance of trade constraint (5).

At this point, we define the joint cost functions for each sector, c and C, and the consumer's expenditure e. Let $y \geq 0_N$ be a vector of output targets and $w >> 0_K$ be a vector of input prices that the regulated sector faces. Then the *regulated sector's joint cost function*, c(y,w), is defined as follows:

$$(9)\ c(y,w) \equiv \min_z \{w^Tz : (y,z)\in s\}.$$

It can be shown[11] that c(y,w) is nonnegative, nondecreasing in y, and nondecreasing, (positively) linearly homogeneous[12] and concave in w. If c(y,w) is differentiable with respect to the components of the input price vector w, then Hotelling (1932; 594) and Shephard (1953; 11) showed that the vector of cost minimising input demand functions, z(y,w), is equal to the vector of first order partial derivatives of the joint cost function; ie we have:[13]

$$(10)\ z(y,w) = \nabla_w c(y,w).$$

Let $Y \geq 0_J$ be a vector of output targets and $w >> 0_K$ be a vector of input prices that the unregulated sector faces. Then the *unregulated sector's joint cost function*, C(Y,w), is defined as follows:

[10] In order to obtain the equality of (7) and (8), we need to assume that a constraint qualification condition holds. We assume that Slater's (1950) condition holds; ie there exist $x° \geq 0_N$, $X° \geq 0_J$, $z° \geq 0_K$, $Z° \geq 0_K$ and $Y° \geq 0_J$ such that $y - x° >> 0_N$; $v - z° - Z° >> 0_K$; $P^T[Y°-X°] > 0$. This is not a restrictive assumption.

[11] For properties of joint cost functions, see McFadden (1978). A joint cost function can be regarded as the negative of a profit function and for properties of profit functions, see Samuelson (1953), Gorman (1968) and Diewert (1973) (1974; 133-141) (1982; 580-583).

[12] This property means that c satisfies the following equation for all scalars $\lambda > 0$: $c(y,\lambda w) = \lambda cy,w)$. In what follows, for brevity, we will abbreviate positively linearly homogeneous to linearly homogeneous.

[13] Notation: $\nabla_w c(y,w)$ is the column vector of first order partial derivatives of c with respect to the components of w, $[\partial c(y,w)/\partial w_1,...,\partial c(y,w)/\partial w_K]^T$, $\nabla_{ww}{}^2 c(y,w)$ is the K by K matrix of second order partial derivatives of c with respect to the components of w, $\nabla_{wy}{}^2 c(y,w)$ is the K by N matrix of second order partial derivatives of c first with respect to the components of w and then with respect to the components of y and so on.

(11) $C(Y,w) \equiv \min_Z \{w^TZ\ : (Y,z)\in S\}$.

It can be shown that C(Y,w) is nonnegative, nondecreasing, linearly homogeneous and convex in Y,[14] and nondecreasing, linearly homogeneous and concave in w. If C(Y,w) is differentiable with respect to the components of the input price vector w, then Shephard's Lemma again implies that the vector of cost minimising input demand functions for the unregulated sector, Z(Y,w), is equal to the vector of first order partial derivatives of the unregulated joint cost function:

(12) $Z(Y,w) = \nabla_w C(Y,w)$.

Let u be a utility target for the household and suppose the household faces the vector of prices $p >> 0_N$ for regulated outputs and $P >> 0_J$ for unregulated outputs. then the household's *expenditure function*, e(u,p,P), is defined as the solution to the following expenditure minimisation problem:

(13) $e(u,p,P) \equiv \min_{x,X} \{p^Tx\ + P^TX : U(x,X) \geq u\ ;\ x \geq 0_N\ ;\ X \geq 0_J\}$.

The consumer's expenditure function will be nondecreasing in all of its variables and linearly homogeneous and concave in prices (p,P).[15] If e(u,p,P) is differentiable with respect to the components of the commodity prices p and P, then Shephard's Lemma implies that the consumer's system of Hicksian demand functions for regulated commodities, x(u,p,P), is equal to the vector of first order partial derivatives of e(u,p,P) with respect to the components of p and the consumer's system of Hicksian demand functions for unregulated commodities, X(u,p,P), is equal to the vector of first order partial derivatives of e(u,p,P) with respect to the components of P; ie in the differentiable case, we have:

(14) $x(u,p,P) = \nabla_p e(u,p,P)$;

(15) $X(u,p,P) = \nabla_P e(u,p,P)$.

With the above definitions and relationships in hand, we can now return to the max-min problem defined by (8) above. Using definitions (9), (11) and (13), it can be seen that we can readily perform the maximisation of (8) with respect to z, Z, x and X. Thus, (8) is equivalent to the following max-min problem which now has no constraints other than nonnegativity constraints:

(16) $H(y) = \max_{u,Y} \min_{p,w,\lambda} \{u + w^Tv + p^Ty - c(y,w) + \lambda P^TY - C(Y,w) - e(u,p,\lambda P)$:

$u \geq 0\ ;\ Y \geq 0_J\ ;\ p \geq 0_N\ ;\ w \geq 0_K\ ;\ \lambda \geq 0\}$.

Note that the cost function for the regulated sector, c(y,w), and the cost function for the unregulated sector, C(Y,w), have suddenly made their appearance in (16). Looking at (6) and (7) and the equivalence of (7) to (16) above under our regularity conditions, it can be seen that solutions to (6) are solutions to the problem of maximising H(y) with respect to the components of y. Hence solutions to the household constrained utility maximisation problem (6) are solutions to the following *two stage max-min problem*:

(17) $\max_{y\geq 0} [\max_{u,Y} \min_{p,w,\lambda} \{u + w^Tv + p^Ty - c(y,w) + \lambda P^TY - C(Y,w) - e(u,p,\lambda P)$:

[14] The assumption that S is a convex cone means that we obtain stronger regularity conditions on C as compared to our previous regularity conditions on c.

[15] See Diewert (1982; 553-556) for material on these regularity conditions.

$$u \geq 0 \; ; \; Y \geq 0_J \; ; \; p \geq 0_N \; ; \; w \geq 0_K \; ; \; \lambda \geq 0\}].$$

Thus, for the first stage, we solve the max min problem defined by (16) and in the second stage, we solve:

(18) $\max_{y \geq 0} H(y)$.

We assume that there is a strictly positive solution to the two stage max-min problem defined by (17), say $u^* > 0$; $y^* >> 0_N$; $Y^* >> 0_J$; $p^* >> 0_N$; $w^* >> 0_K$; $\lambda^* > 0$.[16] We also assume that the two cost functions, c and C, and the consumer's expenditure function e are differentiable at this equilibrium point. It can be shown that the first order necessary conditions for the two stage max-min problem defined by (17) can be obtained by simply differentiating the objective function in (17) with respect to y, u, Y, p, w and λ and setting these partial derivatives equal to zero.[17] Thus, under our assumptions, we find that the optimal solution to the consumer's constrained utility maximisation problem (6) satisfies the following first order conditions:

(19) $1 = \partial e(u^*,p^*,\lambda^* P)/\partial u$;

(20) $p^* = \nabla_y c(y^*,w^*)$;

(21) $\lambda^* P = \nabla_Y C(Y^*,w^*)$;

(22) $v = \nabla_w c(y^*,w^*) + \nabla_w C(Y^*,w^*)$;

(23) $y^* = \nabla_p e(u^*,p^*,\lambda^* P)$;

(24) $0 = P^T[Y^* - \nabla_P e(u^*,p^*,\lambda^* P)]$.

Equation (19) sets the marginal utility of income equal to unity at the optimal equilibrium. This restriction turns out to determine the scale of prices in the economy. Any other normalisation on the overall level of prices will work just as well. We will normalise domestic prices by calibrating them to international prices; ie we will set λ^* equal to unity in what follows. Thus, with this normalisation, the remaining equations (20)-(24) become the following equations:

(25) $p^* = \nabla_y c(y^*,w^*)$;

(26) $P = \nabla_Y C(Y^*,w^*)$;

(27) $v = \nabla_w c(y^*,w^*) + \nabla_w C(Y^*,w^*)$;

(28) $y^* = \nabla_p e(u^*,p^*,P)$;

(29) $0 = P^T[Y^* - \nabla_P e(u^*,p^*,P)]$.

Equations (25) and (26) are the famous price equal marginal cost equations that characterise an optimal equilibrium (that is not a boundary solution). Equations (25) tell us that a necessary condition for an optimal solution to the consumer's constrained utility

[16] These strict positivity assumptions will simplify the analysis which follows. If we do not make these positivity conditions, then it is necessary to work with more complex Kuhn-Tucker (1951) conditions.

[17] This follows by a generalisation of Samuelson's (1947; 34) Envelop Theorem to cover the case of max-min problems.

maximisation problem is that the regulator somehow sets a vector of prices p^* such that in equilibrium, p^* is equal to the optimal vector of marginal costs for the regulated sector, $\nabla_y c(y^*,w^*) \equiv [\partial c(y^*,w^*)/\partial y_1,...., \partial c(y^*,w^*)/\partial y_N]^T$. Equations (26) have a similar interpretation, except in this case, competition should cause the unregulated sector to produce the output vector Y^* such that the vector of international prices P is equal to the optimal vector of marginal costs for the unregulated sector, $\nabla_Y c(Y^*,w^*) \equiv [\partial C(Y^*,w^*)/\partial Y_1,...., \partial C(Y^*,w^*)/\partial Y_N]^T$.

Recalling equations (10) and (12) and Shephard's Lemma, it can be seen that $z^* \equiv \nabla_w c(y^*,w^*)$ is the optimal primary input demand vector for the regulated sector and $Z^* \equiv \nabla_w C(Y^*,w^*)$ is the optimal primary input demand vector for the regulated sector. Thus, equations (27) tell us that the sum of the input demands generated by the two sectors, z^* plus Z^*, is equal to the available primary input supply, v.

On the left hand side of equations (28), we have the optimal supply of regulated commodities, y^*. Recalling the definition of the consumer's system of Hicksian demand functions for regulated products x(u,p,P) equal to $\nabla_p e(u,p,P)$ given by (14), it can be seen that the optimal household demand for regulated products appears on the right hand side of (28). Finally, recalling (15), which set the consumer's Hicksian demands for unregulated products X(u,p,P) equal to the vector of first order partial derivatives of the expenditure function with respect to the components of P, $\nabla_P e(u,p,P)$, it can be seen that equation (29) is equivalent to the following equations:

(30) $P^T[Y^* - \nabla_P e(u^*,p^*,P)] = P^T[Y^* - X(u^*,p^*,P)] = 0.$

The vector $Y^* - X(u^*,p^*,P)$ is the economy's net export vector and so (29) is equivalent to the economy's balance of trade restriction.

Equations (25)-(29) can be regarded as 1 + 2N + J + K equations in the 2 + 2N + J + K endogenous variables u, p, y, Y and w. *These equations characterise an optimal regulatory equilibrium under our assumptions*. The exogenous variables are v (the vector of factor endowments) and P (the vector of international prices).

At this point, it is useful to clarify the role of increasing or constant returns to scale in production in determining whether the regulated sector must be subsidised or not in a first best regulatory solution. A *measure of (reciprocal) returns to scale* for the regulated sector, $\rho(y,w)$, can be defined as follows, using the joint cost function for the regulated sector:

(31) $\rho(y,w) \equiv \partial \ln c(\lambda y,w)/\partial\lambda\ _{|\lambda=1} = y^T\nabla_y c(y,w)/c(y,w).$

Thus, $\rho(y,w)$ gives us the percentage change in cost due to a small proportional increase in all outputs. If there are *increasing returns to scale*, then $\rho(y,w)$ will be less than unity; ie we have decreasing costs and

(32) $y^T\nabla_y c(y,w)/c(y,w) < 1$ or $y^T\nabla_y c(y,w) < c(y,w).$

Now assume that the regulated sector has increasing returns to scale in a neighbourhood of the optimal equilibrium characterised by equations (25)-(29). Premultiply both sides of equations (25) by y^{*T} and we obtain the following equation:

(33) $p^{*T}y^* = y^{*T}\nabla_y c(y^*,w^*)$

$< c(y^*,w^*)$

where we have used the increasing returns to scale assumption (32) to derive the inequality in (33). Thus, optimal revenues minus costs in the regulated sector, $p^{*T}y^* - c(y^*,w^*)$, are negative under these conditions and the regulated sector requires a subsidy to operate at the optimal marginal cost solution.

The situation is quite different for the unregulated sector. Define returns to scale for this sector in the same fashion as (31):

(34) $\rho(Y,w) \equiv \partial \ln C(\lambda Y,w)/\partial\lambda_{\,|\lambda=1} = Y^T\nabla_Y Y(Y,w)/C(Y,w) = 1.$

The last equality in (34) follows from the assumption that S is a cone and hence the unregulated sector joint cost function C(Y,w) is linearly homogeneous in Y; ie we have

(35) $C(\lambda Y,w) = \lambda C(Y,w)$ for all $\lambda \geq 0$.

Differentiating both sides of (35) with respect to λ and evaluating the resulting derivatives at $\lambda = 1$ leads to the following equation:

(36) $C(Y,w) = Y^T\nabla_Y C(Y,w)$.

Equation (36) implies the last equation in (34). Thus, the assumption that S is a cone implies that the unregulated sector is subject to constant returns to scale everywhere.

For the reasons mentioned by the critics of the marginal cost pricing solution to the problem of regulation, regulators are generally not able to implement first best regulatory solutions. In particular, if there are increasing returns to scale in the regulated sector, then typically, the regulator does not have access to general government revenues in order to subsidise the regulated sector as a first best solution would require. Thus, it is desirable to have a framework for modeling second best approaches to the problem of regulation and this is what we will now provide.

Suppose the regulator sets the price of regulated outputs at $p >> 0_N$ and demands that the regulated sector meet all demands $y \geq 0_N$. If the economy's vector of primary input prices is $w >> 0_K$, then the regulated sector will incur total costs of producing the output vector y of c(y,w). Thus, the vector of marginal costs for the regulated sector will be $\nabla_y c(y,w)$, the vector of first order partial derivatives of c(y,w) with respect to the components of the output vector y. Given the vector of selling prices p and the vector of marginal costs $\nabla_y c(y,w)$, we can define the *vector of deviations from marginal cost pricing* or the *vector of margins* over marginal cost, m, as follows:

(37) $m \equiv p - \nabla_y c(y,w)$.

Recall that equations (25)-(29) characterised a first best optimal regulatory policy. We now consider nonoptimal regulatory policies by replacing the price equals marginal cost equations (25) by the price equals marginal cost plus margins equations (37). Thus, a *general nonoptimal regulatory policy* is characterised by the following equations:

(38) $p = m + \nabla_y c(y,w)$;

(39) $P = \nabla_Y C(Y,w)$;

(40) $v = \nabla_w c(y,w) + \nabla_w C(Y,w)$;

(41) $y = \nabla_p e(u,p,P)$;

(42) $0 = P^T[Y - \nabla_P e(u,p,P)]$.

Equations (38)-(42) can be regarded as 1 + 2N + J + K equations in the 1 + 2N + J + K endogenous variables u, m, y, Y and w. The exogenous variables are now p (the vector of regulatory prices for regulated outputs), v (the vector of factor endowments) and P (the vector of international prices). Thus, p has moved from being an endogenous vector to being an exogenous vector but we have added an extra endogenous vector of variables, m, the vector of margins over marginal costs. Thus, in the case where the regulated industry has increasing returns to scale, the regulator can now set regulated prices high enough so that the regulated industry makes some positive margins and does not have to be subsidised by general government revenues. In the following section, we will look at the comparative statics properties of equilibria that are characterised by equations (38)-(42).

However, in subsequent sections, it will prove to be useful to replace the balance of trade equation (42) by another equivalent equation which sets household expenditure on all goods and services, e(u,p,P), equal to household income sources. We now proceed to derive this income equals expenditure equation.

Recall equations (14) and (15) and the fact that e(u,p,P) is linearly homogeneous in p,P; ie we have

(43) $e(u,\lambda p,\lambda P) = \lambda e(u,p,P)$ for all $\lambda \geq 0$.

Differentiating both sides of (43) with respect to λ and evaluating the resulting derivatives at $\lambda = 1$ leads to the following equation:

(44) $e(u,p,P) = p^T\nabla_p e(u,p,P) + P^T\nabla_P e(u,p,P)$.

Take p^T times both sides of (41) and add to (42). Using (44), we obtain the following equation:

(45) $e(u,p,P) = p^Ty + P^TY$.

Recall that c(y,w) and C(Y,w) are both linearly homogeneous in w. Using the same method of derivation as was used in deriving (44), we can show that the following two equations are satisfied by c and C:

(46) $c(y,w) = w^T\nabla_w c(y,w)$;

(47) $C(Y,w) = w^T\nabla_w C(Y,w)$.

Premultiply both sides of (40) by w^T. Using (46) and (47), the resulting equation becomes:

(48) $w^Tv - c(y,w) - C(Y,w) = 0$.

Now add (48) to the right hand side of (45) and we obtain the following equation which can be used to replace equation (42):

(49) $e(u,p,P) = w^Tv + p^Ty - c(y,w) + P^TY - C(Y,w)$.

Thus, on the left hand side of (49), we have household expenditure on all goods and services, e(u,p,P). On the right hand side, we have sources of household income: factor income, w^Tv, plus regulated sector net profits (or losses), $p^Ty - c(y,w)$, plus unregulated sector net profits,

$P^TY - C(Y,w)$. It is straightforward to show that unregulated sector profits are zero in our model. To see this, premultiply both sides of (39) by Y^T. Using the fact that C(Y,w) is linearly homogeneous in Y so that (36) holds, we obtain the following equation:

(50) $P^TY = C(Y,w)$.

Using (50), (49) can be replaced by the following equation:

(51) $e(u,p,P) = w^Tv + p^Ty - c(y,w)$.

The income equals expenditure equation (49) illustrates a limitation in our modeling: because we have only one consumer class in the model, we cannot look at the *distributional effects* of monopoly profits on households that hold shares in the regulated sector versus households who do not hold such shares. Given the complexity of the present highly simplified model, extending the analysis to allow for different classes of consumers will be left to future research.

Recall that the exogenous variables in our model of a nonoptimal regulatory equilibrium are v (the vector of available primary inputs for the economy), P (the vector of world prices for internationally traded goods that the economy faces) and p (the vector of prices for regulated products that the regulator imposes on the regulated firm). In the following section, we will show how utility u or welfare reacts to small changes in these exogenous variables.

3 THE COMPARATIVE STATICS PROPERTIES OF THE ONE PERIOD REGULATORY MODEL

3.1 Introduction

Before we begin our comparative statics analysis, we require two additional assumptions.

Our first additional assumption is that consumer preferences satisfy *local money metric utility scaling*; ie we assume that the following equation holds at the initial equilibrium:

(52) $\partial e(u,p,P)/\partial u = 1$.

It can be seen that (52) is an implication of the following equation:

(53) $e(u,p,P) = u$ for all $u > 0$.

Equation (53) turns out to be a method for cardinalising utility. Samuelson (1974; 1262) referred to the utility scaling assumption implied by (53) as *money metric utility scaling*. In order to explain this method for cardinalising the consumer's utility scale, suppose the consumer faces the prices p, P at an initial equilibrium and spends income Y^* on goods and services at this initial equilibrium. This amount of income will determine the initial utility level, say u^*. Thus, we have:

(54) $Y^* = e(u^*,p,P)$
$\quad\quad = u^*$

where the second equation above follows using (53). Thus, the economy's initial utility level u^* is set equal to the income generated by the economy, Y^*. If we hold prices constant at their initial levels, p and P, and change household income to λY^* where λ is a positive proportional scaling of the initial income, the new budget constraint for the consumer will be:

(55) $B(\lambda) \equiv \{(x,X) : p^T x + P^T X = \lambda Y^* ; x \geq 0_N ; X \geq 0_J\}$.

Thus, if λ is greater than 1, the initial budget set expands upwards in a parallel fashion, whereas if is λ less than 1, the initial budget set contracts towards the origin in a parallel fashion. In either case, the new budget surface defined by (55) will touch a highest indifference surface, $u(\lambda)$ say. Money metric utility scaling will set the height of the indifference surface indexed by $u(\lambda)$ equal to λY^*. It can be seen that this is a very reasonable way of cardinalising utility. Assumption (52) just imposes this global method of utility scaling around the initial equilibrium.

We now introduce our second additional assumption. Recall that equations (38)-(41) and (49) determine the endogenous variables u, y, Y, w and m as functions of the exogenous variables v, P and p. The response of the endogenous variables to changes in the exogenous variables can be determined by differentiating equations (38)-(41) and (49) totally with respect to the components of v, P and p in turn and this is what we will do in sections 3.2 (comparative statics with respect to v, the economy's primary input vector), 3.3 (comparative statics with respect to P, the vector of international prices that the economy faces) and 3.4 (comparative statics with respect to p, the vector of regulated prices). In order to rigorously justify our analysis in what follows, we need to assume that the matrix A defined below has an inverse:

$$(56)\ A \equiv \begin{bmatrix} 1 & -m^T & 0_J^T & 0_K^T \\ -\nabla_{pu}^2 e & I_N & 0_{N\times J} & 0_{N\times K} \\ 0_J & 0_{J\times N} & \nabla_{YY}^2 C & \nabla_{Yw}^2 C \\ 0_K & \nabla_{wy}^2 c & \nabla_{wY}^2 C & \nabla_{ww}^2 C + \nabla_{ww}^2 c \end{bmatrix}$$

where m is the vector of regulated sector markups over marginal costs at the initial equilibrium, 0_J is a vector of zeros of dimension J, etc., I_N is an N by N identity matrix, $0_{N\times J}$ is a matrix of zeros with N rows and J columns, etc., $\nabla_{YY}{}^2C$ is $\nabla_{YY}{}^2C(Y,w)$ (the matrix of second order partial derivatives of the unregulated sector's cost function with respect to the components of Y and Y again, evaluated at the initial equilibrium), etc., $\nabla_{wy}{}^2c$ is $\nabla_{wy}{}^2c(y,w)$ (the matrix of second order partial derivatives of the regulated sector's cost function with respect to the components of w first and y second, evaluated at the initial equilibrium), etc., and $\nabla_{pu}{}^2e = \nabla_{pu}{}^2e(u,p,P)$ is the vector of second order partial derivatives of the household expenditure function with respect to the vector of regulated prices p and then with respect to utility u, evaluated at the initial equilibrium. Since this vector will appear fairly frequently in what follows, we will define this vector as b:

$$(57)\ b \equiv \nabla_{pu}{}^2e(u,p,P)$$
$$= \partial x(u,p,P)/\partial u$$

where the second line in (57) follows from Shephard's Lemma, (14); ie the consumer's vector of Hicksian demands for regulated outputs, x(u,p,P), is equal to the vector of first order partial derivatives of the consumer's expenditure function e(u,p,P) with respect to p, $\nabla_p e(u,p,P)$. Thus, b is the vector of derivatives of regulated demand responses to changes in the consumer's real income, u. We expect that, for the most part, the components of the b vector will be positive.[18]

3.2 The effects on household welfare of changes in the availability of primary inputs

Recall that the equations that define our regulatory equilibrium are (38)-(41) and (49). Differentiate equation (49) with respect to the components of the primary input availability vector v, regarding the components of u, y, Y, w and m as functions of v, so that these functions are u(v), y(v), Y(v), w(v) and m(v) and $\nabla_v u(v)$, $\nabla_v y(v)$, $\nabla_v Y(v)$, $\nabla_v w(v)$ and $\nabla_v m(v)$ are 1 by K, N by K, J by K, K by K and N by K matrices of derivatives respectively of these endogenous variables with respect to the components of v. After differentiating (49) with respect to the components of v and using equations (38)-(41) and (52), we obtain the following equation:

$$(58)\ \nabla_v u(v) = w(v) + [\nabla_v y(v)]^T m(v).$$

Thus, if at the initial equilibrium, the regulator has set prices p so that they turn out to equal

[18] If the nth component of b, b_n, is negative, then the consumption of the nth regulated product will *decrease* as real income increases; ie the nth regulated product is an *inferior good*.

the vector of regulated sector marginal costs, $\nabla_y c(y,w)$, then the vector of markups over marginal costs, m(v), will equal 0_N and equations (58) collapse down to

(59) $\nabla_v u(v) = w(v) >> 0_K$.

Thus, in this case of *first best optimal regulation*, we see that the value to the consumer of an extra unit of any primary input is equal to the initial equilibrium price of the primary input, which is an intuitively plausible result. In the general case of nonoptimal regulation, from (58), it can be seen that if the markup vector m at the initial equilibrium is small in magnitude, then $\nabla_v u(v)$ will generally be close to w(v) so that even in the nonoptimal case, it is likely that consumer welfare will increase as primary input availability increases.

The matrix of derivatives of regulated output responses to changes in primary input availability, $\nabla_v y(v)$, which appears in (58) is endogenous and so it should be possible to work out a formula for this matrix of derivatives in terms of exogenous variables. The desired formula can be derived by differentiating both sides of (41) with respect to the components of v. Thus, rewrite equation (41) as follows:

(60) $y(v) = \nabla_p e(u(v),p,P)$.

Differentiating both sides of (60) with respect to the components of v, leads to the following matrix equation:

(60) $\nabla_v y(v) = \nabla_{pu}^2 e(u(v),p,P)[\nabla_v u(v)]^T$

$\quad = b[\nabla_v u(v)]^T$

where the last equation in (60) follows using definition (57). Now substitute (60) into (58). Collecting terms, we find that the resulting equation becomes the following one:

(61) $[1 - m^T b]\nabla_v u(v) = w$.

Thus, if

(62) $1 - m^T b > 0$

then $\nabla_v u(v)$ equals $(1 - m^T b)^{-1} w >> 0_K$, a positive vector and thus increases in the economy's primary input vector will lead to increases in welfare.

Although we cannot definitively guarantee that (62) holds, it is extremely likely that this condition holds. To explain why this is the case, we need to define the vector a of derivatives of the Hicksian demand functions for internationally traded goods, X(u,p,P), with respect to utility u:

(63) $a \equiv \nabla_{Pu}^2 e(u,p,P) = \partial X(u,p,P)/\partial u$.

Thus, (63) is a counterpart to our earlier definition of b equal to $\nabla_{pu}^2 e(u,p,P)$. As was the case with b, most of the components of a will be positive, since negative components correspond to inferior goods, which are typically not widespread. Since e(u,p,P) is linearly homogeneous in the components of p and P, it can be seen that the utility derivative of this function, $\partial e(u,p,P)/\partial u$, will also be linearly homogeneous in the components of p,P; ie $\partial e(u,p,P)/\partial u$ will satisfy the following equation:

(64) $\partial e(u,\lambda p,\lambda P)/\partial u = \lambda \partial e(u,p,P)/\partial u$ for all $\lambda > 0$.

Differentiating both sides of (64) with respect to λ and then setting λ equal to 1 leads to the following equation:

$$(65)\ \partial e(u,p,P)/\partial u = \nabla_{up}^{2} e(u,p,P)p + \nabla_{uP}^{2} e(u,p,P)P$$
$$= b^{T}p + a^{T}P \qquad \text{using definitions (57) and (63)}$$
$$= 1 \qquad \text{using (52).}$$

Using (48), we have m equal to the difference between the consumer price vector p and the vector of regulated sector marginal costs, $\nabla_y c(y,w) \geq 0_N$:

$$(66)\ m = p - \nabla_y c(y,w).$$

Thus

$$(67)\ 1 - m^{T}b = 1 - [p - \nabla_y c(y,w)]^{T}b \qquad \text{using (66)}$$
$$= P^{T}a + \nabla_y c(y,w)^{T}b \qquad \text{using (65).}$$

The vector of international prices P is strictly positive and the vector of marginal costs for the regulated sector $\nabla_y c(y,w)$ is nonnegative. The income derivative vectors of demand, a and b, will have mostly positive components so (67) implies that it is very likely that (62) is satisfied. In what follows, we will assume that (62) is satisfied.

Thus, under our assumptions, welfare improves as factor endowments increase; ie when (62) is satisfied, we can convert (61) into (68):

$$(68)\ \nabla_v u(v) = [1 - m^{T}b]^{-1}w \gg 0_K.$$

3.3 The effects on household welfare of changes in the prices of internationally traded goods (terms of trade effects)

In this subsection, we examine the effects on welfare of changes in the prices of internationally traded goods and services. The analysis in this subsection largely parallels the analysis in section 3.2.

Differentiate equation (49) with respect to the components of the vector of international prices P, regarding the components of u, y, Y, w and m as functions of P, so that these functions are u(P), y(P), Y(P), w(P) and m(P) and $\nabla_P u(P)$, $\nabla_P y(P)$, $\nabla_P Y(P)$, $\nabla_P w(P)$ and $\nabla_P m(P)$ are 1 by J, N by J, J by J, K by J and N by J matrices of derivatives respectively of these endogenous variables with respect to the components of P. After differentiating (49) with respect to the components of P and using equations (38)-(41) and (52), we obtain the following equation:

$$(69)\ \nabla_P u(P) = Y(P) - \nabla_P e(u(P),p,P) + [\nabla_P y(P)]^{T}m(P)$$
$$= Y(P) - X(u(P),p,P) + [\nabla_P y(P)]^{T}m(P) \qquad \text{using (15).}$$

Thus, if at the initial equilibrium, the regulator has set prices p so that they turn out to equal the vector of regulated sector marginal costs, $\nabla_y c(y,w)$, then the vector of markups over marginal costs, m(P), will equal 0_N and equations (69) collapse down to

$$(70)\ \nabla_P u(P) = Y - X$$

where Y is the economy's initial production vector for unregulated outputs and X is the initial consumer demand vector for these outputs. Thus, Y – X is the economy's initial *net export vector*; ie if $Y_j - X_j$ is positive, then at the initial equilibrium, the jth unregulated commodity is exported while if $Y_j - X_j$ is negative, then at the initial equilibrium, the jth unregulated commodity is imported. Thus, if $m = 0_N$ at the initial equilibrium and $Y_j - X_j$ is positive so that j is exported initially, then a marginal increase in P_j of one unit will lead to an increase in (money metric) welfare of $Y_j - X_j$ units. Conversely, if $m = 0_N$ and $Y_j - X_j$ is negative so that j is imported initially, then a marginal increase in P_j of one unit will lead to an decrease in welfare of $|Y_j - X_j|$ units. These are more or less standard results in the pure theory of trade.

In the general case of nonoptimal regulation, from (69), it can be seen that if the markup vector m at the initial equilibrium is small in magnitude, then $\nabla_P u(P)$ will generally be close to Y – X so that even in the nonoptimal case, it is likely that consumer welfare will *increase* as the prices of exported goods and services increase and *decrease* as the prices of imported goods and services increase.

The matrix of derivatives of regulated output responses to changes in the prices of internationally traded goods, $\nabla_P y(P)$, which appears in (69) is endogenous and it is possible to work out a formula for this matrix of derivatives in terms of exogenous variables. The desired formula can be derived by differentiating both sides of (41) with respect to the components of P. Thus, rewrite equation (41) as follows:

(71) $y(P) = \nabla_p e(u(P),p,P)$.

Differentiating both sides of (71) with respect to the components of P, leads to the following matrix equation:

(72) $\nabla_P y(P) = \nabla_{pu}^2 e(u(P),p,P)[\nabla_P u(P)]^T + \nabla_{pP}^2 e(u(P),p,P)$

$= b[\nabla_P u(P)]^T + \nabla_{pP}^2 e(u(P),p,P)$ using definition (57).

Now substitute (72) into (69). Collecting terms, we find that the resulting equation becomes:

(73) $[1 - m^T b]\nabla_P u(P) = Y - X + \nabla_{Pp}^2 e(u,p,P)m$

Thus, if $1 - m^T b$ is positive, then the signs of the vector of derivatives of utility u with respect to changes in the prices of internationally traded commodities, $\nabla_P u(P)$, will be equal to the signs of the vector on the right hand side of (73). This right hand side vector is equal to the net export vector at the initial equilibrium, Y – X, plus the vector $\nabla_{Pp}^2 e(u,p,P)m$. Now the J by N *consumer substitution matrix* $\nabla_{Pp}^2 e(u,p,P)$ describes how consumer demand for internationally traded commodities changes as the vector of regulated prices increases. There will be a strong tendency for the entries in this matrix to be positive.[19] There will also be a strong tendency for the entries in the markup over marginal cost vector m to be positive, particularly in the case of a strongly increasing returns to scale technology for the regulated sector. Putting these two tendencies together, it is likely that the last term on the right hand side of (73) will have mostly positive entries and thus this last term will be an additive augmentation to the usual net export vector Y – X which occurs in the case of a first best

[19] Hicks (1946; 311-312) noticed this tendency and proved it must hold if N = 1 and J = 1.

equilibrium.[20]

We now turn our attention to the comparative statics results which will be of most interest to regulators.

3.4 The effects on household welfare of changes in the prices of regulated commodities

In this subsection, we examine the effects on welfare of changes in the prices of regulated goods and services.

Differentiate equation (49) with respect to the components of the vector of regulated prices p, regarding the components of u, y, Y, w and m as functions of p, so that these functions are u(p), y(p), Y(p), w(p) and m(p) and $\nabla_p u(p)$, $\nabla_p y(p)$, $\nabla_p Y(p)$, $\nabla_p w(p)$ and $\nabla_p m(p)$ are 1 by N, N by N, J by N, K by N and N by N matrices of derivatives respectively of these endogenous variables with respect to the components of p. After differentiating (49) with respect to the components of p and using equations (38)-(41) and (52), we obtain the following equation:

(74) $\nabla_p u(p) = [\nabla_p y(p)]^T m(p)$.

Thus, if at the initial equilibrium, the regulator has set prices p so that they turn out to equal the vector of regulated sector marginal costs, $\nabla_y c(y,w)$, then the vector of markups over marginal costs, m(p), will equal 0_N and equations (74) collapse down to:

(75) $\nabla_p u(p) = 0_N$.

Note that equations (75) are the first order necessary conditions for p to maximise consumer welfare. Thus, when the vector of markups over marginal costs m equals 0_N, then the vector of regulated prices p is equal to the corresponding vector of marginal costs of producing these regulated products $\nabla_y c(y,w)$[21] and in this case of *first best optimal regulation*, we attain the first best equilibrium that was discussed in section 2.

In the general case of nonoptimal regulation, from (74), it can be seen that if the markup vector m at the initial equilibrium is small in magnitude, then $\nabla_p u(p)$ will generally be close to 0_N so that even in the nonoptimal case, it is likely that consumer welfare will be close to the first best optimal level of welfare if the components of m are small in magnitude. However, in the nonoptimal case, we would like to have an explicit formula which could help guide the regulator to a better allocation of resources in the economy.

The matrix of derivatives of regulated output responses to changes in the prices of regulated commodities, $\nabla_p y(p)$, which appears in (74) is endogenous and, as usual, it is possible to work out a formula for this matrix of derivatives in terms of exogenous variables. The desired formula can be derived by differentiating both sides of (41) with respect to the components of p. Thus, rewrite equation (41) as follows:

(76) $y(p) = \nabla_p e(u(p),p,P)$.

Differentiating both sides of (76) with respect to the components of p, leads to the following

[20] This generally positive augmentation effect occurs if the price of an internationally traded good increases but if the price decreases, then this effect becomes generally a negative one.

[21] Recall that equations (38) hold in our regulatory model. These equations are $p = m + \nabla_y c(y,w)$.

matrix equation:

$$(77)\ \nabla_p y(p) = \nabla_{pu}^2 e(u(p),p,P)[\nabla_p u(p)]^T + \nabla_{pp}^2 e(u(p),p,p)$$
$$= b[\nabla_p u(p)]^T + \nabla_{pp}^2 e(u(p),p,P) \qquad \text{using definition (57).}$$

Now substitute (77) into (74). Collecting terms, we find that the resulting equation becomes:

$$(78)\ [1 - m^T b]\nabla_p u(p) = \nabla_{pp}^2 e(u,p,P)m$$

Thus, if $1 - m^T b$ is positive, then the signs of the vector of derivatives of utility u with respect to changes in the prices of the regulated commodities, $\nabla_p u(p)$, will be equal to the signs of the vector on the right hand side of (78). This right hand side vector is equal to $\nabla_{pp}^2 e(u,p,P)m$. Now the N by N *consumer substitution matrix* $\nabla_{pp}^2 e(u,p,P)$ describes how consumer demand for regulated commodities changes as the vector of regulated prices increases. There will be a strong tendency for the off diagonal entries in this matrix to be positive.[22]

Equations (78) can be used to guide the economy to a higher level of welfare. We will illustrate how this could be done in the case where there are only 2 regulated outputs. The two equations in (78) can be rewritten as follows in this N = 2 case:

$$(79)\ [1 - m^T b]\partial u(p_1,p_2)/\partial p_1 = m_1\partial^2 e(u,p_1,p_2,P)/\partial p_1\partial p_1 + m_2\partial^2 e(u,p_1,p_2,P)/\partial p_1\partial p_2\ ;$$

$$(80)\ [1 - m^T b]\partial u(p_1,p_2)/\partial p_2 = m_1\partial^2 e(u,p_1,p_2,P)/\partial p_2\partial p_1 + m_2\partial^2 e(u,p_1,p_2,P)/\partial p_2\partial p_2\ .$$

We will assume that assumption (62) is satisfied so that $1 - m^T b = 1 - m_1 b_1 - m_2 b_2$ is greater than zero. The concavity of $e(u,p_1,p_2,P)$ in the prices p_1,p_2,P will imply that the derivatives $e_{11} \equiv \partial^2 e(u,p_1,p_2,P)/\partial p_1\partial p_1$ and $e_{22} \equiv \partial^2 e(u,p_1,p_2,P)/\partial p_2\partial p_2$ are nonpositive but we will make the stronger assumption that they are negative so that the demand for each regulated product will fall as its price increases. We will also assume that the two regulated products are *substitutes* so that the demand for regulated product 1 will increase as the price of the second regulated product p_2 increases (this means $e_{12} \equiv \partial^2 e(u,p_1,p_2,P)/\partial p_1\partial p_2 = \partial x_1(u,p_1,p_2,P)/\partial p_2 > 0$) and the demand for regulated product 2 will increase as the price of the first regulated product p_1 increases (this means $e_{21} \equiv \partial^2 e(u,p_1,p_2,P)/\partial p_2\partial p_1 = \partial x_2(u,p_1,p_2,P)/\partial p_1 > 0$). Young's Theorem in calculus implies that $e_{12} = e_{21}$ so our assumptions on consumer preferences can be summarised as follows:

$$(81)\ e_{11} < 0\ ;\ e_{12} = e_{21} > 0\ ;\ e_{22} < 0.$$

There are several alternative assumptions that could be made about the two markups over marginal cost variables, m_1 and m_2. We have already considered the case where these markups are zero: in this case, the regulator has set optimal prices p_1 and p_2 for the regulated products. We will consider some additional cases below:

Case 1: $m_1 < 0\ ;\ m_2 > 0$.

In this case, equations (79) and (80) become:

$$(82)\ \partial u(p_1,p_2)/\partial p_1 = [1 - m^T b]^{-1}[m_1 e_{11} + m_2 e_{12}] = [+][(-)(-) + (+)(+)] > 0\ ;$$

$$(83)\ \partial u(p_1,p_2)/\partial p_2 = [1 - m^T b]^{-1}[m_1 e_{21} + m_2 e_{22}] = [+][(-)(+) + (+)(-)] < 0\ .$$

Thus, under our Case 1 assumptions, increasing the price p_1 of the first regulated output or

[22] Again, Hicks (1946; 311-312) noticed this tendency and proved it must hold if N = 2.

decreasing the price of the second regulated output will improve household welfare.

Case 2: $m_1 > 0$; $m_2 < 0$.

In this case, equations (79) and (80) become:

(84) $\partial u(p_1,p_2)/\partial p_1 = [1 - m^T b]^{-1}[m_1 e_{11} + m_2 e_{12}] = [+][(+)(-) + (-)(+)] < 0$;

(85) $\partial u(p_1,p_2)/\partial p_2 = [1 - m^T b]^{-1}[m_1 e_{21} + m_2 e_{22}] = [+][(+)(+) + (-)(-)] > 0$.

Thus, under our Case 2 assumptions, decreasing the price p_1 of the first regulated output or increasing the price of the second regulated output will improve household welfare.

In both of these cases, the regulator can improve welfare by raising the price of the product where the markup over marginal cost is negative and by decreasing the price of the product where the markup over marginal cost is positive. Following this strategy, the regulator will be guiding the economy towards a first best price equals marginal cost equilibrium. Note that in these cases, all the regulator has to know is whether marginal costs are above or below the corresponding regulated prices.[23]

In remaining cases where both markups are positive or both markups are negative, we cannot obtain the rather straightforward results that we obtained in Cases 1 and 2; ie in order to determine how regulated prices should be changed, the regulator would have to know not only the magnitudes of m_1 and m_2, but the regulator would also have to know the magnitudes of the consumer demand derivatives, e_{11}, e_{21} and e_{22}. It is unlikely that the regulator will be able to determine these parameters with any degree of accuracy.

If profits in the regulated sector are negative at the initial equilibrium, the above analysis assumes that the regulator has access to public funds so that in this case, the regulated sector deficit can be covered by lump sum taxes on the household sector. This is not a realistic assumption: regulators generally do not have access to tax instruments. Thus, when the regulator chooses a vector of regulated product prices p, the regulator has to make these prices high enough so that the regulated firm will be able to earn enough revenue to cover its full costs and not just its marginal costs. Thus, in the case of increasing returns to scale in the regulated sector, the regulator cannot freely choose p so that the first best equilibrium could be attained: there will be an additional constraint on the regulator's choices, namely that the regulated firm remain solvent.

Recall that our formal regulatory model consisted of equations (38)-(42), which were 1 + 2N + J + K equations in the 2 + 2N + J + K endogenous variables u, m, y, Y and w. The exogenous variables in those equations were p (the vector of prices for regulated outputs), v (the vector of factor endowments) and P (the vector of international prices). The solvency constraint on the regulator's choices can be imposed on our model if we add the following single constraint to equations (38)-(42):

(86) $p^T y = c(y,w)$.

Equation (86) simply sets the regulated firm's revenues, $p^T y$, equal to its total costs, $c(y,w)$.

The addition of equation (86) to the previous equations (38)-(42) means that we need to find one additional variable that will be endogenous in the new model. Thus, if (86) holds in the

[23] Unfortunately, it will not be easy for the regulator to determine this in most cases.

initial equilibrium, and the regulator wants this constraint to continue to hold, then the regulator will not be able to freely vary all N regulated prices; ie one of these prices will have to be chosen as an endogenous variable in the new model. Alternatively, we can treat all of the regulated prices in a symmetric manner by differentiating equation (86) with respect to the components of p and force the p_n's to change in a way that will respect the constraint (86) to the first order. Regard y and w as functions of p, y(p) and w(p), that are determined by our previous model (38)-(42) and differentiate the resulting rearranged equation (86), which we rewrite as follows:

$$(87)\ G(p) \equiv p^T y(p) - c(y(p),w(p)) = 0.$$

The vector of first order partial derivatives of the G(p) defined by (87) is:

$$\begin{aligned}(88)\ \nabla_p G(p) &= y + [\nabla_p y(p)]^T p - [\nabla_p y(p)]^T \nabla_y c(y(p),w(p)) - [\nabla_p w(p)]^T \nabla_w c(y(p),w(p)) \\ &= y + [\nabla_p y(p)]^T p - [\nabla_p y(p)]^T [p - m(p)] - [\nabla_p w(p)]^T z(p) \quad \text{using (38) and (10)} \\ &= y + [\nabla_p y(p)]^T m - [\nabla_p w(p)]^T z.\end{aligned}$$

The matrices of partial derivatives $\nabla_p y(p)$ and $\nabla_p w(p)$ which appear in (88) above can be determined by the comparative statics properties of the original model defined by (38)-(42). Define dp as a vector of small changes in the vector of regulated prices. Then to the accuracy of a first order approximation, the regulator can choose to change the regulated prices by the incremental vector dp where the increments satisfy the following equation, which is a linearisation of (86):

$$(89)\ [\nabla_p G(p)]^T dp = 0$$

where $\nabla_p G(p)$ is defined by the right hand side of (88).

It can be seen that adding the solvency constraint (86) to our model complicates the analysis to a considerable degree and so we will not pursue this extended model in more detail in the present analysis.[24]

However, it could be argued that the solvency constraint is not necessary because the regulator will be forced to give the regulated firm some monopoly profits in order to induce the firm to make productivity improvements. Thus, our original model will be adequate for regulatory purposes: the regulator will propose changes in regulated prices that will improve consumer welfare but the regulator will also keep an eye on how changes in regulated prices will affect the profitability of the regulated firm. If the regulator had information on consumer preferences, producer technologies and on expected movements in primary inputs and international prices, the model defined by equations (38)-(42) could be used by the regulator to propose welfare improving changes in regulated prices.

We conclude this section with an observation on the consequences of a constant returns to scale technology in the regulated sector. Suppose that the regulated sector has a constant returns to scale technology. Recall equation (36) where we showed that if the unregulated sector had a constant returns to scale technology, then C(Y,w) was equal to the inner product

[24] In the solvency constrained model, it is useful to look for optimal directions of change rather than just characterising the necessary conditions for optimality. This involves the use of rather different techniques; see Diewert (1978) and Diewert, Turunen-Red and Woodland (1989) for explanations of these techniques.

of the output vector Y with the vector of marginal costs, $\nabla_Y C(Y,w)$. With constant returns in the regulated sector, the regulated sector's cost function will satisfy a similar constraint; ie the following equation will hold:

$$(90)\ c(y,w) = y^T \nabla_y c(y,w).$$

Now premultiply both sides of (38) by y^T. Using (90), the resulting equation is equivalent to:

$$(91)\ p^T y - c(y,w) = m^T y.$$

Thus, under the assumption that the regulated sector has a constant returns to scale technology, the regulated sector's pure profits, $p^T y - c(y,w)$, will be equal to the inner product of the vector of margins of prices over marginal costs, m, with the vector of outputs of the regulated sector, y. The significance of this equation is that in the constant returns to scale case, the regulator does not need to worry about the solvency constraint (86): the regulator need only choose a vector of regulated prices p which will drive the margin vector m down to a vector of zeros 0_N and the solvency constraint (86) will automatically be satisfied.

We will conclude section 3 by noting a second order approximation to the loss of welfare due to nonoptimal regulation. With information on various consumer and producer elasticities, this approximate loss of welfare formula can give us a rough idea about the magnitude of nonoptimal regulation.

Note that equations (78) can be rearranged to provide a formula for the vector of first order partial derivatives of the utility function u(p) with respect to the vector of regulated prices p at an observed (nonoptimal) regulatory equilibrium:

$$(92) \nabla_p u(p) = [1 - m^T b]^{-1} \nabla_{pp}{}^2 e(u,p,P)m.$$

We also know that if the regulator sets the vector of regulated prices at the optimal level p^* where the corresponding markup vector m^* is equal to 0_N, then using (75), we have:

$$(93)\ \nabla_p u(p^*) = 0_N.$$

A quadratic approximation to the loss of welfare moving from the optimal utility level $u(p^*)$ to the observed nonoptimal utility level u(p) is given by the following formula:[25]

$$
\begin{aligned}
(94)\ u(p^*) - u(p) &\approx (1/2)[\nabla_p u(p^*) + \nabla_p u(p)]^T [p^* - p] \\
&= (1/2)[1 - m^T b]^{-1} m^T \nabla_{pp}{}^2 e(u,p,P)[p^* - p] && \text{using (92) and (93)} \\
&\approx -(1/2)[\ 1 - m^T b]^{-1} m^T \nabla_{pp}{}^2 e(u,p,P)m && \text{using } p \approx p^* + m \\
&\geq 0
\end{aligned}
$$

where the inequality follows from assumption (62) and from the negative semidefiniteness of the N by N consumer substitution matrix $\nabla_{pp}{}^2 e(u,p,P)$.

We now turn our attention to a continuous time approach to the sources of welfare gain.

[25] The first line in (94) follows using Diewert's (1976; 118) Quadratic Identity.

4 A CONTINUOUS TIME APPROACH TO THE DETERMINANTS OF WELFARE CHANGE

4.1 Introduction

Denny, Fuss and Waverman (1981) adapted Solow's (1957) continuous time model of technical progress to the regulated context. Denny, Fuss and Waverman made several important contributions to the literature on productivity measurement in the regulated context, including the introduction of the cost function for the regulated sector, allowing for multiple outputs and nonconstant returns to scale and allowing for regulated prices that were not necessarily equal to marginal costs. However, their focus was on the measurement of Total Factor Productivity whereas our focus is on regulation and welfare improvements. Thus, our task in this section is to extend the analysis of Denny, Fuss and Waverman from the analysis of the determinants of productivity growth to the determinants of welfare growth using the continuous time approach pioneered by Solow.

4.2 The determinants of welfare change

Recall that our regulatory model consists of equations (38)-(41) and either (42) or (49). For our present purposes, equation (49) will prove to be more useful than equation (42).

We begin our analysis by regarding all exogenous variables, p, P and v, as differentiable functions of time, p(t), P(t) and v(t), where t is time.[26] We assume that equations (38)-(41) and (49) hold at each instant of time and that the endogenous variables in our regulated model are also differentiable functions of time, u(t), m(t), y(t), Y(t) and w(t). However, we now allow for *technical progress* in our regulatory model. Recall that the cost function for the unregulated sector was defined as C(Y,w) and the cost function for the regulated sector was defined as c(y,w). We now assume that these cost functions are also functions of time, say C(Y,w,t) and c(y,w,t). If there is ongoing technical progress in the two sectors, then as time marches on, costs will fall, holding outputs and input prices fixed;[27] ie in the case of *technical progress* in the two sectors, the partial derivatives of these cost functions with respect to time will be negative:

(95) $\partial C(Y,w,t)/\partial t < 0$; $\partial c(y,w,t)/\partial t < 0$.

If there is *technological regress* in the two sectors, then the two partial derivatives in (95) will be positive; if there is *no technical progress* at time t, then the two partial derivatives in (95) will be zero. The usual case in real life economies will be the case of technical progress so that the inequalities in (95) will hold.

Now we can begin our analysis of the determinants of welfare growth over time. Differentiate both sides of equation (49), regarding all variables as functions of time, and replacing c(y,w) by c(y(t),w(t),t) and C(Y,w) by C(Y(t),w(t),t). We obtain the following equation:

[26] This continuous time approach to economic measurement was pioneered by Divisia (1926).
[27] For the relationship of the cost function measure of technical progress to the corresponding direct production function measure of technical progress, see Denny, Fuss and Waverman (1981; 196-197).

(96) $u'(t) = w(t)^T v'(t) + [Y(t) - X(t)]^T P'(t) + m(t)^T y'(t) - \partial c(y,w,t)/\partial t - \partial C(Y,w,t)/\partial t$

where $u'(t)$, $v'(t)$, $P'(t)$ and $y'(t)$ denote the time derivatives of the functions $u(t)$, $v(t)$, $P(t)$ and $y(t)$ respectively. In order to derive equation (96), we used equations (38)-(41) in order to simplify the equation and we also used the local money metric utility scaling assumption, (52), at time t. Equation (96) tells us that the rate of growth of money metric welfare at time t, $u'(t)$, is equal to the sum of five terms. The first term is $w(t)^T v'(t)$, the rate of growth of primary inputs $v'(t)$ in the economy, weighted by the vector of time t input prices $w(t)$. The second term is the inner product of the economy's net export vector at time t, $Y(t) - X(t)$, with the vector of rates of change in the prices of internationally traded goods and services, $P'(t)$ (this is a measure of the rate of change in the economy's terms of trade). The third term is the inner product of the vector of time t markups over marginal costs in the regulated sector, $m(t)$, with the vector of growth rates in regulated outputs, $y'(t)$. We will defer analyzing this term for now except to note that if $m(t)$ equals 0_N so that prices equal the corresponding marginal costs in the regulated sector, then this term vanishes. The fourth term is $-\partial c(y,w,t)/\partial t$ and it can be seen that this term is the rate of technical progress in the regulated sector and the fifth term is $-\partial C(Y,w,t)/\partial t$, which in turn is equal to the rate of technical progress in the unregulated sector. Thus, loosely speaking, the rate of welfare growth is equal to primary input growth plus terms of trade effects plus a term involving markups over marginal costs plus technical progress in the regulated and unregulated sectors.

A problem with the growth of welfare decomposition (96) is that the vector of growth rates for regulated outputs, $y'(t)$, is endogenous rather than exogenous. Hence, it what follows, we will obtain an expression for $y'(t)$ in terms of exogenous variables.

Equations (41) can be rewritten in our new continuous time framework as follows:

(97) $y(t) = \nabla_p e(u(t),p(t),P(t))$.

Define the time t vector of income derivatives of demand for regulated outputs, $b(t)$, as follows:

(98) $b(t) \equiv \nabla_{pu}{}^2 e(u(t),p(t),P(t)) = \partial x(u(t),p(t),P(t)/\partial u$.

Differentiating equation (97) with respect to t leads to the following expression for $y'(t)$:

(99) $y'(t) = \nabla_{pu}{}^2 e(u(t),p(t),P(t))u'(t) + \nabla_{pp}{}^2 e(u(t),p(t),P(t))p'(t) + \nabla_{pP}{}^2 e(u(t),p(t),P(t))P'(t)$

$\qquad = b(t)u'(t) + \nabla_{pp}{}^2 e(u(t),p(t),P(t))p'(t) + \nabla_{pP}{}^2 e(u(t),p(t),P(t))P'(t) \qquad$ using (98).

Substitute (99) into (96) and rearranging terms, we obtain the following expression for the growth of welfare in terms of exogenous variables:

(100) $[1 - m(t)^T b(t)]u'(t) = w(t)^T v'(t) + [Y(t) - X(t)]^T P'(t) + m(t)^T \nabla_{pP}{}^2 e(u(t),p(t),P(t))P'(t)$

$\qquad + m(t)^T \nabla_{pp}{}^2 e(u(t),p(t),P(t))p'(t) - \partial c(y,w,t)/\partial t - \partial C(Y,w,t)/\partial t$.

As usual, we will assume that $m(t)^T b(t)$ is less than 1:

(101) $1 - m(t)^T b(t) > 0$.

Using (101) in (100), we have our final expression for the rate of growth of welfare as a function of the rates of growth of the exogenous variables in our model:

(102) $u'(t) = [1 - m(t)^T b(t)]^{-1}\{w(t)^T v'(t) + [Y(t) - X(t)]^T P'(t) + m(t)^T[\nabla_{pP}{}^2 e]P'(t)$

$$+ m(t)^T[\nabla_{pp}{}^2 e]p'(t) - \partial c(y,w,t)/\partial t - \partial C(Y,w,t)/\partial t\}.$$

Recall our comparative statics results for the change in utility due to changes in v, equation (68) in section 3.2 above, for the changes in utility due to changes in P, equations (73) in section 3.3 above and for the changes in utility due to changes in p, equations (78) in section 3.4 above. For convenience, we will rewrite these equations below as equations (103)-(105).

(103) $\nabla_v u(v) = [1 - m^T b]^{-1} w$;

(104) $\nabla_P u(P) = [1 - m^T b]^{-1}\{Y - X + \nabla_{Pp}{}^2 e(u,p,P)m\}$;

(105) $\nabla_p u(p) = [1 - m^T b]^{-1}\nabla_{pp}{}^2 e(u,p,P)m$.

It can be seen that the inner products of the vectors on the right hand sides of equations (103)-(105) with the vectors $v'(t)$, $P'(t)$ and $p'(t)$ respectively make their appearance on the right hand side of (102); ie equation (102) can be regarded as a first order approximation to the discrete time comparative statics changes in welfare that we studied in section 3.[28] However, our new equation (102) represents a generalisation of the results in section 3 since we now allow for technical progress in both sectors.

Formula (102) is an important result for regulatory purposes. The regulator cannot control changes in the economy's endowment of primary inputs, $v'(t)$, or changes in the vector of prices for internationally traded goods and services, $P'(t)$, or the amount of cost saving technical progress in the competitive sector, $-\partial C(Y,w,t)/\partial t$. However, the regulator *can* determine changes in the vector of regulated product prices, $p'(t)$, and by allowing the regulated firm to keep some excess profits (above and beyond its cost of capital) if it undertakes cost saving changes in its technology, it *can* induce a certain amount of cost saving technical progress in the regulated sector, $-\partial c(y,w,t)/\partial t$. Thus, if the regulator has some knowledge of the vector of markups of regulated prices over marginal costs, $m(t)$, and if the regulator has some knowledge of the household substitution matrix for regulated products, $\nabla_{pp}{}^2 e(u(t),p(t),P(t))$, then the regulator can attempt to determine a vector of price changes for the regulated products, $p'(t)$, such that $m(t)^T[\nabla_{pp}{}^2 e]p'(t)$ is positive and such that the firm's profits are positive (but not excessive). Essentially, the regulator has two primary tasks:

- To guide regulated prices along a path which will optimally adjust the gaps between the prices of regulated products and their marginal costs.
- In the case of an increasing returns to scale technology for the regulated sector, the regulator needs to set prices so that margins remain sufficiently positive to ensure the solvency of the regulated firm and also to induce it to undertake cost saving productivity improvements.

We will spell out in a little more detail how this optimal determination of regulated prices might proceed. Suppose that the regulator is able to determine the vector of marginal costs in the regulated sector and hence determine the vector of markups over marginal costs, $m(t)$. Suppose further that the regulator changes the regulated prices according to the following rule:

[28] Equation (102) cannot provide a second order approximation to the corresponding discrete time change in welfare in general so this limitation of the continuous time approach should be kept in mind.

(106) $p'(t) = -\alpha m(t)$

where α is a positive constant. Thus, if the nth markup over marginal cost $m_n(t)$ is positive, the regulator *decreases* the price $p_n(t)$ of the nth regulated output and if $m_n(t)$ is negative, then the regulator *increases* the price $p_n(t)$ of the nth regulated output. Substituting (106) into (102) changes the term $m(t)^T[\nabla_{pp}{}^2e]p'(t)$ in (102) to:

(107) $m(t)^T[\nabla_{pp}{}^2e]p'(t) = -\alpha\, m(t)^T[\nabla_{pp}{}^2e]m(t) \geq 0$

where the inequality in (107) follows from the negative semidefiniteness of $\nabla_{pp}{}^2e$. Thus, the pricing rule (106) will generally lead to an improvement in welfare.[29]

It should be noted that our equation (102), which indicates that technical progress in either sector translates directly into a welfare improvement (after adjusting for the multiplicative factor $[1 - m^Tb]^{-1}$), is quite similar to the results obtained by Basu and Fernald (2002) in their model of a distorted economy. Their model was somewhat different in its details and they used primal optimisation techniques as opposed to our use of dual cost functions but roughly the same type of result emerged: productivity improvements translated more or less directly into welfare improvements, even in an economy with distortions. However, it must be kept in mind that their results and our equation (102) offer only first order approximations to the corresponding welfare changes in discrete time.

We now turn our attention to the problems associated with obtaining observable approximations to the technical change parameters, $\partial c(y,w,t)/\partial t$ and $\partial C(Y,w,t)/\partial t$.

4.3 Continuous time approaches to the determination of technical progress

We begin by attempting to determine an expression that will allow us to estimate the rate of technical progress in the competitive, unregulated sector. Recall that the cost function for this sector at time t is $C(Y(t),w(t),t)$. The vector of input demands at time t is $Z(t)$ defined via Shephard's Lemma as follows:

(108) $Z(t) \equiv \nabla_w C(Y(t),w(t),t)$.

We also know that the competitive sector faces the international price vector $P(t)$ for its outputs at time t and $Y(t)$ is adjusted so that the following equation is satisfied:

(109) $P(t) = \nabla_Y C(Y(t),w(t),t)$.

Using the linear homogeneity of $C(Y,w,t)$ in both Y and w, we have:

(110)
$$\begin{aligned} C(Y(t),w(t),t) &= w(t)^T\nabla_w C(Y(t),w(t),t) \\ &= Y(t)^T\nabla_Y C(Y(t),w(t),t) \\ &= w(t)^T Z(t) && \text{using (108)} \\ &= P(t)^T Y(t) && \text{using (109).} \end{aligned}$$

Equations (110) imply that $C(Y(t),w(t),t)$ equals $w(t)^T Z(t)$. Differentiating this equation with respect to t leads to the following equations:

[29] However, a limitation of the pricing rule (106) must be noted here: it ignores the solvency constraint for the regulated firm. The solvency constraint must also be taken into account in setting regulated prices.

(111) $w(t)^T Z'(t) + w'(t)^T Z(t)$

$$= \nabla_Y C(Y(t),w(t),t)^T Y'(t) + \nabla_w C(Y(t),w(t),t)^T w'(t) + \partial C(Y,w,t)/\partial t$$

$$= P(t)^T Y'(t) + Z(t)^T w'(t) + \partial C(Y,w,t)/\partial t \qquad \text{using (108) and (109).}$$

Equations (111) imply the following expression for the rate of technical progress in the unregulated sector:

(112) $-\partial C(Y,w,t)/\partial t = P(t)^T Y'(t) - w(t)^T Z'(t)$.

Define the *Divisia* (1926) *index of aggregate unregulated sector output growth* at time t, $Y_D'(t)$, and the corresponding *Divisia index of aggregate input growth* at time t, $Z_D'(t)$, as follows:

(113) $Y_D'(t) \equiv \sum_{j=1}^{J} [P_j(t)Y_j(t)/P(t)^T Y(t)][Y_j'(t)/Y_j(t)]$;

(114) $Z_D'(t) \equiv \sum_{k=1}^{K} [w_k(t)Z_k(t)/w(t)^T Z(t)][Z_k'(t)/Z_k(t)]$.

Now divide both sides of (112) by the unregulated sector's cost at time t, C(Y,w,t). Then using the equations in (110), we obtain the following expression for the logarithmic rate of technical progress in the unregulated sector at time t:

(115) $[-\partial C(Y,w,t)/\partial t]/C(Y,w,t) = Y_D'(t) - Z_D'(t)$;

ie the logarithmic rate of technical progress in the unregulated sector is equal to the Divisia index of output growth less the Divisia index of input growth. This is the many output counterpart to Solow's (1957) famous formula for measuring technical progress in the case of one output.[30]

Good discrete time approximations to the continuous time Divisia indexes defined by (113) and (114) can be obtained using superlative indexes; see Diewert (1976) for definitions of various superlative indexes.[31]

The bottom line for the above algebra is that for the case of a constant returns to scale competitive industry, it is relatively straightforward to obtain fairly accurate estimates of technical progress. However, the situation is much more difficult in the case of a regulated firm as we shall soon see.

Recall that the cost function for the regulated sector at time t is c(y(t),w(t),t). The vector of input demands at time t is z(t) defined via Shephard's Lemma as follows:

(116) $z(t) \equiv \nabla_w c(y(t),w(t),t)$.

We also know that the following relationships hold between the time t vector of regulated output prices p(t), the vector of markups over marginal costs m(t) and the vector of regulated sector marginal costs $\nabla_y c(y(t),w(t),t)$:

(117) $p(t) = m(t) + \nabla_y c(y(t),w(t),t) \equiv m(t) + \mu(t)$

where for notational convenience, we have defined $\mu(t)$ to be the vector of marginal costs, $\nabla_y c(y(t),w(t),t)$.

[30] The one output case was generalised to many outputs by Jorgenson and Griliches (1967).

[31] More sophisticated discrete time estimates for productivity growth indexes can be obtained using the techniques explained in Caves, Christensen and Diewert (1982) and Diewert and Morrison (1986).

Using the linear homogeneity of c(y,w,t) in w, we have:

(118) $c(y(t),w(t),t) = w(t)^T\nabla_w c(y(t),w(t),t) = w(t)^T z(t)$ using (116).

Differentiating (118) with respect to t leads to the following equations:

(119) $w(t)^T z'(t) + w'(t)^T z(t)$

$= \nabla_y c(y(t),w(t),t)^T y'(t) + \nabla_w c(y(t),w(t),t)^T w'(t) + \partial c(y,w,t)/\partial t$

$= \mu(t)^T y'(t) + z(t)^T w'(t) + \partial c(y,w,t)/\partial t$ using (116).

Equations (119) imply the following expression for the rate of technical progress in the regulated sector:

(120) $-\partial c(y,w,t)/\partial t = \mu(t)^T y'(t) - w(t)^T z'(t)$.

Thus, the rate of technical progress in the regulated sector, $-\partial c(y,w,t)/\partial t$, is equal to the inner product of the output growth rate vector at time t, $y'(t)$, with the vector of marginal costs at time t, $\mu(t)$, minus the inner product of the price vector for primary inputs at time t, $w(t)$, with the vector of input growth rates for the regulated sector, $z'(t)$. It can be seen that equation (120) is a regulated sector counterpart to the corresponding expression for the rate of technical progress in the unregulated sector, (112). What makes the regulated sector estimation problem much more difficult is that in order to implement (120), we require a knowledge of the marginal cost vector for the regulated sector, $\mu(t)$, whereas in the unregulated sector estimation of technical progress, we required only a knowledge of observable output prices for the unregulated sector, P(t).[32]

It is also possible to develop a counterpart to (115) for the regulated sector; ie we can obtain an expression for the logarithmic rate of cost reduction (but the resulting expression will involve the difficult to measure vector of marginal costs). Recall expression (31), which defined the (reciprocal) degree of returns to scale, ρ, in the regulated sector. In the present context, this definition becomes the following one:

(121) $\rho(t) \equiv y(t)^T\nabla_y c(y(t),w(t),t)/c(y(t),w(t),t) = y(t)^T\mu(t)/c(y(t),w(t),t)$.

Equation (121) can be rearranged to give us the following equation:

(122) $\mu(t)^T y(t) = \rho(t)c(y(t),w(t),t)$.

Define the *Divisia* (1926) *index of aggregate regulated sector output growth using marginal cost weights* at time t, $y_D'(t)$, and the corresponding *Divisia index of aggregate input growth* at time t, $z_D'(t)$, as follows:

(123) $y_D'(t) \equiv \sum_{n=1}^N [\mu_n(t)y_n(t)/\mu(t)^T y(t)][y_n'(t)/y_n(t)]$;

(124) $z_D'(t) \equiv \sum_{k=1}^K [w_k(t)z_k(t)/w(t)^T z(t)][z_k'(t)/z_k(t)]$.

Now divide both sides of (120) by the regulated sector's cost at time t, c(y,w,t). Then using (118) and (122), we obtain the following expression for the logarithmic rate of technical

[32] It must be emphasised that equation (120) uses the vector of marginal costs at time t, $\mu(t)$, to weight the vector of regulated sector output growth derivatives, $y'(t)$, and *not* the vector of output prices p(t); ie marginal cost weights must be used in (120) rather than revenue weights.

progress in the regulated sector at time t:[33]

(125) $[-\partial c(y,w,t)/\partial t]/c(y,w,t) = \rho(t)y_D'(t) - z_D'(t)$;

ie the logarithmic rate of technical progress in the regulated sector is equal to the product of the degree of reciprocal returns to scale in the regulated sector at time t, $\rho(t)$, times the marginal cost weighted Divisia index of output growth, $y_D'(t)$, less the Divisia index of input growth, $z_D'(t)$.[34]

[33] The decomposition (125) is equivalent to a decomposition obtained by Denny, Fuss and Waverman (1981; 196).

[34] More sophisticated discrete time estimates for technical progress indexes than that given by (125) (which is only a first order approximation to the underlying technical progress measure) can be obtained using the techniques explained in Diewert and Fox (2008).

5 OPTIMAL ONE PERIOD REGULATION WITH FREELY VARIABLE CAPITAL SERVICES

5.1 The determination of a regulatory equilibrium when capital is freely variable

We now single out capital inputs for special attention. The reason why capital inputs require special treatment as compared to regular flow inputs like labour and materials is twofold:

- Capital inputs are durable and so their purchase price must be decomposed into components that represent the contribution of the capital input during the time periods that it is used; and
- Some capital inputs are irreversible; ie once an irreversible input is installed, its service contributions are more or less fixed until the asset is retired. Examples of such *sunk cost investments* are electricity poles and wires, pipeline networks, telecommunication networks and roads.

The first reason listed above can be dealt with if we can assume that the capital input can readily be sold (or purchased) on second hand markets; ie used capital inputs can be varied freely as time marches on. Examples of this type of capital input are transportation equipment and many types of machines. In this freely variable case, the purchase price of a durable capital input can be decomposed into a sum of discounted period by period rental prices or user costs of capital and in principle, there are no particular difficulties: in each period, we simply use the appropriate user cost of capital[35] as the input price and this price would appear in our input price vector w which appeared in previous sections. However, in the case of irreversible investments or sunk cost investments, the argument in the previous sentence does not work. Instead, it is necessary to take an intertemporal approach to the problem of sunk costs.

In the present section, we will prepare the way for the intertemporal model in the following section where we will address the sunk cost problem. In the present model, we will assume that the regulated firm uses a single durable capital input k but it is freely variable and hence there is a well defined user cost for it, P_k. This is the price that the regulated firm faces for the use of this input during the period under consideration.

We again make the same assumptions that were made in section 2 with a few changes. There are now four classes of commodities in our model instead of the previous three classes since we have added capital services $k \geq 0$ to the regulated sector as an additional input.[36] The remaining three classes are the N regulated outputs, $y \geq 0_N$, the J internationally traded outputs produced by the unregulated sector, $Y \geq 0_J$, and the K primary inputs with the

[35] For a sampling of material on the user cost of capital concept, the reader is referred to Hall and Jorgenson (1967), Jorgenson and Griliches (1967), Christensen and Jorgenson (1969), Diewert (1980) (2005a) (2005b), Harper, Berndt and Wood (1989), Jorgenson (1989) (1996a) (1996b), Hulten (1990) (1996), Diewert and Lawrence (2000) and Schreyer (2001) (2009).

[36] For simplicity, we have introduced only one type of capital service but there are no difficulties in extending the analysis to a vector of capital services.

economy's noncapital primary input endowment vector being $v >> 0_K$ where v equals z (noncapital primary inputs used by the regulated sector) plus Z (noncapital primary inputs used by the unregulated sector).

We will assume that the regulated sector imports units of its capital input k at the positive user cost price P_k.[37]

As in section 2, we assume that there is a production possibilities set s for the regulated sector and another production possibilities set S for the unregulated sector. We continue to assume that S satisfies assumption (2); ie S is a closed convex cone. Our assumptions on the regulated sector technology set are as follows:

(126) For any $y \geq 0_N$, the set of inputs (z,k) that can produce at least y, $\{(z,k) : (y,z,k) \in s\}$ is a nonempty, closed convex set.

As in section 2, we assume that the household has preferences defined over nonnegative consumption vectors of regulated products, $x \geq 0_N$, and nonnegative vectors of unregulated products, $X \geq 0_J$. We further assume that these preferences can be represented by the utility function U(x,X) where U is a nonnegative, increasing, continuous and concave function.

As in section 2, since there is only one household in the economy, it will be optimal to maximise this household's utility subject to various resource constraints that face the household and producers in the economy. The counterparts to our old constraints (3)-(5) are now:

(127) $y \geq x$;

(128) $v \geq z + Z$;

(129) $P^T[Y-X] - P_k k \geq 0.$

The constraints (127) and (128) are exactly the same as (3) and (4) but the new balance of trade constraint (129) says that the value of net exports of outputs has to be equal to or greater than the value of imported capital services.

The *household's constrained utility maximisation problem* is now (130):

(130) $\max_{u,y,Y,z,Z,x,X,k} \{u : y - x \geq 0_N ; v - z - Z \geq 0_K ; P^T[Y-X] - P_k k \geq 0 ; (y,z,k) \in s ;$

$$(Y,Z) \in S ; U(x,X) - u \geq 0\}$$

where u is the household's utility level and all other decision variables, y,Y,z,Z,x,X,k are nonnegative and have been defined above.

The joint cost function, C(Y,w), defined earlier by (11) continues to hold for the present model as does the consumer's expenditure function, e(u,p,P), defined by (13). However, we now need to define a new joint cost function c for the regulated sector since we have added imported capital services to this sector. Let $k \geq 0$ be capital services input, let $y \geq 0_N$ be a vector of feasible output targets given this k and let $w >> 0_K$ be a vector of (noncapital) input prices that the regulated sector faces. Then the *regulated sector's variable cost function* or

[37] An alternative assumption is that the regulated sector purchases its capital inputs from the domestic economy. This alternative assumption can readily be accommodated but at the cost of more notational complexity so for simplicity, we have assumed that the capital input is imported.

operating cost function or *opex function*, c(y,w,k), is defined as follows:[38]

(131) $c(y,w,k) \equiv \min_z \{w^T z : (y,z,k) \in s\}$.

It can be shown[39] that c(y,w,k) is nonnegative, nondecreasing in y, nonincreasing in k and nondecreasing, (positively) linearly homogeneous and concave in w. If c(y,w,k) is differentiable with respect to the components of the input price vector w, then we can adapt the arguments of Hotelling (1932; 594) and Shephard (1953; 11) and show that the vector of cost minimising input demand functions, z(y,w), is equal to the vector of first order partial derivatives of the joint cost function; ie we have:[40]

(132) $z(y,w,k) = \nabla_w c(y,w,k)$.

We will assume that a strictly positive solution to (130) exists; ie $u^* > 0$, $y^* >> 0_N$, $Y^* >> 0_J$, $z^* >> 0_K$, $k^* > 0$, $Z^* >> 0_K$, $x^* >> 0_N$ and $X^* >> 0_J$ solves (130). Now let $y \geq 0_N$ be in a neighbourhood of y^* and consider the following constrained maximisation problem that is conditional on the choice of y:

$$(133)\ H(y) \equiv \max_{u,Y,z,Z,x,X,k} \{u : y - x \geq 0_N ; v - z - Z \geq 0_K ; P^T[Y - X] - P_k k \geq 0 ; (y,z,k) \in s ;$$
$$(Y,Z) \in S ; U(x,X) - u \geq 0 ; u \geq 0, Y \geq 0_J, z \geq 0_K, Z \geq 0_K, x \geq 0_N, X \geq 0_J\}.$$

It can be verified that (133) is a concave programming problem and hence the Karlin (1959; 201-203) Uzawa (1958) Saddle Point Theorem can be applied to this problem. Using this Theorem, we can absorb some of the constraints into the objective function and it turns out that u, k, Y, z, Z, x, and X solutions to (133) are also solutions to the following max-min problem:[41]

$$(134)\ H(y) = \max_{u,Y,z,Z,x,X,k} \min_{p,w,\lambda} \{u + p^T[y - x] + w^T[v - z - Z] + \lambda P^T[Y - X] - \lambda P_k k ;$$
$$(y,z,k) \in s ; (Y,Z) \in S ; U(x,X) - u \geq 0 ; u \geq 0, Y \geq 0_J,$$
$$z \geq 0_K, Z \geq 0_K, x \geq 0_N, X \geq 0_J, p \geq 0_N, w \geq 0_K, \lambda \geq 0\}$$

where p can be interpreted as a *vector of prices for the regulated products*, w can be interpreted as a *vector of (noncapital) primary input prices* and λ is a Lagrange multiplier that corresponds to the balance of trade constraint (129).

Using definitions (11), (13) and (131), it can be seen that we can readily perform the maximisation of (134) with respect to z, Z, x and X. Thus, (134) is equivalent to the following max-min problem which now has no constraints other than nonnegativity constraints:

$$(135)\ H(y) = \max_{u,Y,k} \min_{p,w,\lambda} \{u + w^T v + p^T y - c(y,w,k) + \lambda P^T Y - C(Y,w) - \lambda P_k k$$

[38] If k is very low relative to the output targets y, then it may be the case that there is no vector of variable inputs z such that $(y,z,k) \in s$; ie in this case the vector of output targets is infeasible. If this case occurs, then we define c(y,w,k) to be plus infinity.

[39] For properties of joint cost functions, see McFadden (1978). A joint cost function can be regarded as the negative of a profit function and for properties of profit functions, see Samuelson (1953), Gorman (1968) and Diewert (1973) (1974; 133-141) (1982; 580-583).

[40] We also require that c(y,w,k) be finite in a neighbourhood around the point (y,w,k).

[41] In order to obtain the equality of (133) and (134), we need to assume that a constraint qualification condition holds. We assume that Slater's (1950) condition holds; ie there exist $x° \geq 0_N$, $X° \geq 0_J$, $z° \geq 0_K$, $Z° \geq 0_K$ and $Y° \geq 0_J$ such that $y - x° >> 0_N$; $v - z° - Z° >> 0_K$; $P^T[Y° - X°] > 0$ and $k° > 0$. This is not a restrictive assumption.

$$- e(u,p,\lambda P) : u \geq 0 \; ; \; Y \geq 0_J \; ; \; p \geq 0_N \; ; \; w \geq 0_K \; ; \; \lambda \geq 0\}.$$

Note that the opex cost function for the regulated sector, c(y,w,k), and the cost function for the unregulated sector, C(Y,w), have suddenly made their appearance in (135). Looking at (130) and (133) and the equivalence of (133) to (135) above under our regularity conditions, it can be seen that solutions to (130) are solutions to the problem of maximising H(y) with respect to the components of y. Hence solutions to the household constrained utility maximisation problem (130) are solutions to the following *two stage max-min problem*:

$$(136)\ \max_{y\geq 0} [\max_{u,Y} \min_{p,w,\lambda} \{u + w^T v + p^T y - c(y,w) + \lambda P^T Y - C(Y,w) - \lambda P_k k$$
$$- e(u,p,\lambda P) : u \geq 0 \; ; \; Y \geq 0_J \; ; \; p \geq 0_N \; ; \; w \geq 0_K \; ; \; \lambda \geq 0\}].$$

Thus, for the first stage, we solve the max min problem defined by (134) and in the second stage, we solve:

$$(137)\ \max_{y\geq 0} H(y).$$

We assume that there is a strictly positive solution to the two stage max-min problem defined by (136), say $u^* > 0$, $k^* > 0$, $y^* >> 0_N$, $Y^* >> 0_J$, $p^* >> 0_N$, $w^* >> 0_K$, $\lambda^* > 0$.[42] We also assume that the two cost functions, c and C, and the consumer's expenditure function e are differentiable at this equilibrium point. It can be shown that the first order necessary conditions for the two stage max-min problem defined by (136) can be obtained by simply differentiating the objective function in (136) with respect to y, k, u, Y, p, w and λ and setting these partial derivatives equal to zero.[43] Thus, under our assumptions, we find that the optimal solution to the consumer's constrained utility maximisation problem (130) satisfies the following first order conditions:

$$(138)\quad 1 = \partial e(u^*,p^*,\lambda^* P)/\partial u \; ;$$

$$(139)\quad p^* = \nabla_y c(y^*,w^*,k^*) \; ;$$

$$(140)\ \lambda^* P = \nabla_Y C(Y^*,w^*) \; ;$$

$$(141)\quad v = \nabla_w c(y^*,w^*,k^*) + \nabla_w C(Y^*,w^*) \; ;$$

$$(142)\quad y^* = \nabla_p e(u^*,p^*,\lambda^* P) \; ;$$

$$(143)\quad P_k = -\partial c(y^*,w^*,k^*)/\partial k$$

$$(144)\quad 0 = P^T[Y^* - \nabla_P e(u^*,p^*,\lambda^* P)] - P_k k^* \, .$$

Equation (138) sets the marginal utility of income equal to unity at the optimal equilibrium. This restriction determines the scale of prices in the economy. Any other normalisation on the overall level of prices will work just as well. We will normalise domestic prices by calibrating them to international prices as in section 2; ie we will set λ^* equal to unity in what follows. Thus, with this normalisation, the remaining equations (139)-(144) become the following equations:

$$(145)\quad p^* = \nabla_y c(y^*,w^*,k^*) \; ;$$

[42] These strict positivity assumptions will simplify the analysis which follows. If we do not make these positivity conditions, then it is necessary to work with more complex Kuhn-Tucker (1951) conditions.

[43] This follows by a generalisation of Samuelson's (1947; 34) Envelop Theorem to cover the case of max-min problems.

(146) $P = \nabla_Y C(Y^*,w^*)$;

(147) $v = \nabla_w c(y^*,w^*,k^*) + \nabla_w C(Y^*,w^*)$;

(148) $y^* = \nabla_p e(u^*,p^*,P)$;

(149) $P_k = -\partial c(y^*,w^*,k^*)/\partial k$

(150) $0 = P^T[Y^* - \nabla_P e(u^*,p^*,P)] - P_k k^*$.

Equations (145)-(148) and (150) are the counterparts to equations (25)-(29) in section 2 and have similar interpretations. The only really new equation is (149) which says that in an optimal equilibrium, the user cost of capital, P_k, should be equal to the *negative* of the partial derivative $\partial c(y^*,w^*,k^*)/\partial k$ of the opex cost function with respect to the capital input variable, k. This optimality condition makes intuitive sense; as the regulated firm adds extra units of capital to produce the same vector of outputs, operating expenditures should decline and for optimality, this decline in opex cost should just equal the cost of the extra unit of capital that has been added.

Equations (146) are the usual price equals marginal cost equations which characterise behaviour in the unregulated sector. Equations (145) tell us that a necessary condition for an optimal solution to the consumer's constrained utility maximisation problem is that the regulator somehow sets a vector of prices p^* such that in equilibrium, p^* is equal to the optimal vector of opex marginal costs for the regulated sector, $\nabla_y c(y^*,w^*,k^*)$. Although this rule for regulatory optimality appears to be different from the corresponding rule developed in section 2 where regulated prices were set equal to *marginal costs* rather than *opex marginal costs*, it turns out that the rules are the same since it can be shown that the vector of optimal opex marginal costs will equal the vector of optimal full marginal costs (including the costs of capital).[44]

Finally, recall (15), which said that the consumer's Hicksian demands for unregulated products X(u,p,P) is equal to the vector of first order partial derivatives of the expenditure function with respect to the components of P, $\nabla_P e(u,p,P)$. Using this fact, it can be seen that equation (144) is equivalent to the following equations:

(151) $P^T[Y^* - \nabla_P e(u^*,p^*,P)] = P^T[Y^* - X(u^*,p^*,P)] = P_k k^*$.

Thus, the value of net exports of goods and services, $P^T[Y^* - X^*]$, must equal the value of capital imports, $P_k k^*$.

Equations (145)-(150) can be regarded as 2 + 2N + J + K equations in the 2 + 2N + J + K endogenous variables u, k, p, y, Y and w. *These equations characterise an optimal regulatory equilibrium under our assumptions*. The exogenous variables are v (the vector of factor endowments), P (the vector of international prices for traded goods and services) and P_k, the price for importing a unit of capital services.

Suppose the regulator sets the price of regulated outputs at $p >> 0_N$ and demands that the regulated sector meet all demands $y \geq 0_N$. Given the vector of selling prices p and the vector

[44] Let the full cost function equal $c^*(y,w,P_k)$. In equilibrium, we will have $c^*(y^*,w^*,P_k) \equiv \min_{k \geq 0} \{c(y^*,w^*,k) + P_K k\} = c(y^*,w^*,k^*) + P_K k^*$. Equation (149) and the Envelop Theorem will imply that $\nabla_y c^*(y^*,w^*,P_k) = \nabla_y c(y^*,w^*,k^*)$. .

of opex marginal costs $\nabla_y c(y,w,k)$, we can define the *vector of deviations from opex marginal cost pricing* or the *vector of margins* over opex marginal cost, m, as follows:

(152) $m \equiv p - \nabla_y c(y,w,k)$.

Recall that equations (145)-(150) characterised a first best optimal regulatory policy. We now consider nonoptimal regulatory policies by replacing the price equals opex marginal cost equations (145) by the price equals opex marginal cost plus margins equations (152). Thus, a *general regulatory policy* is characterised by the following equations:

(153) $p = m + \nabla_y c(y,w,k)$;

(154) $P = \nabla_Y C(Y,w)$;

(155) $v = \nabla_w c(y,w,k) + \nabla_w C(Y,w)$;

(156) $y = \nabla_p e(u,p,P)$;

(157) $P_k = -\partial c(y,w,k)/\partial k$

(158) $0 = P^T[Y - \nabla_P e(u,p,P)] - P_k k$.

Equations (153)-(158) can be regarded as 2 + 2N + J + K equations in the 2 + 2N + J + K endogenous variables u, k, m, y, Y and w. These equations determine a general equilibrium for the economy under our assumptions. The exogenous variables are now p (the vector of regulatory prices for regulated outputs), v (the vector of factor endowments), P (the vector of international prices) and P_k (the price of imported capital). Thus, p has moved from being an endogenous vector to being an exogenous vector but we have added an extra endogenous vector of variables, m, the vector of margins over opex marginal costs. Thus, in the case where the regulated industry has increasing returns to scale, the regulator can set regulated prices high enough so that the regulated industry makes some positive margins and does not have to be subsidised by general government revenues.

As in section 2, we are assuming competitive price taking behaviour in the unregulated sector. We are also assuming competitive cost minimising behaviour on the part of the regulated firm.

Using the same techniques that were used in section 2, it can be shown that the balance of trade constraint, (158), can be replaced by an income equals expenditure constraint on the household. This replacement equation is (159) below and it is a counterpart to equation (49) in section 2.

(159) $e(u,p,P) = w^T v + p^T y - c(y,w,k) - P_k k + P^T Y - C(Y,w)$.

In the following subsections, we will briefly look at the comparative statics properties of the present model, using the same techniques that were used in sections 3 and 4 above.

5.2 The comparative statics properties of the freely variable capital model

The exogenous variables in our one period, freely variable capital model are p (the vector of regulatory prices for regulated outputs), v (the vector of factor endowments), P (the vector of international prices) and P_k (the price of imported capital services). Using the same techniques that were used in section 3,[45] it can be shown that we obtain the same comparative statics results as were obtained in section 3 for the changes in welfare or utility with respect to small increases in the components of v, P and p; ie we have the following results:

(160) $\nabla_v u(v) = [1 - m^T b]^{-1} w$;

(161) $\nabla_P u(P) = [1 - m^T b]^{-1}\{Y - X + \nabla_{Pp}{}^2 e(u,p,P)m\}$;

(162) $\nabla_p u(p) = [1 - m^T b]^{-1}\nabla_{pp}{}^2 e(u,p,P)m$.

In order to determine how utility will change if the price of capital services P_k increases, recall equation (159) above and regard u, w, y, Y, P and k as functions of P_k. Now differentiate (159) with respect to P_k and using equations (153)-(158), we find that:

(163) $u'(P_k) = -P_k + y'(P_k)^T m(P_k)$.

Thus, if at the initial equilibrium, the regulator has set prices p so that they turn out to equal the vector of regulated sector marginal costs, $\nabla_y c(y,w)$, then the vector of markups over marginal costs, $m(P_k)$, will equal 0_N and equations (163) collapse down to

(164) $u'(P_k) = -P_k$.

Thus, in this case of *first best optimal regulation*, we see that welfare will decrease as the price of imported capital services increases and this rate of decrease will be equal to the negative of the price of capital services in the initial equilibrium, $-P_k$, which is an intuitively plausible result. In the following paragraph, we will work out a formula for $u'(P_k)$ when regulation is not first best optimal.

The vector of derivatives of regulated output responses to changes in the price of capital services, $y'(P_k)$, which appears in (163) is endogenous and as usual, we can work out a formula for this vector of derivatives in terms of exogenous variables. The desired formula can be derived by differentiating both sides of (156) with respect to the components of P_k. Thus, rewrite equation (156) as follows:

(165) $y(P_k) = \nabla_p e(u(P_k),p,P)$.

Differentiating both sides of (165) with respect to P_k, leads to the following equation:

(166) $y'(P_k) = \nabla_{pu}{}^2 e(u(P_k),p,P)u'(P_k)$

$= b\, u'(P_k)$

where the last equation in (166) follows using definition (57). Now substitute (166) into (163). Collecting terms, we find that the resulting equation becomes the following one:

(167) $[1 - m^T b]u'(P_k) = -P_k$.

Thus, if $1 - m^T b$ is positive (which will generally be the case), then $u'(P_k)$ equals $-(1 - m^T b)^{-1}P_k < 0$, and thus an increase in the price of imported capital services will lead to a decrease in welfare.

[45] Recall that in section 3, in order to justify our comparative statics results, we required that the matrix A defined by (56) had an inverse. A similar assumption is required in this section but the counterpart to the A matrix will have an extra row (due to the addition of the extra equation (157) which ensures that the regulated firm is cost minimising with respect to its capital input) and A will have an extra column (due to the addition of the endogenous capital variable k).

In the following section, we rework the analysis that was presented in section 4.2 above; ie we look at welfare growth in a model where the opex cost function replaces the total cost function.

5.3 The determinants of welfare change in continuous time for the opex cost function model with freely variable capital

Recall equation (159) which set household expenditures on goods and services, e(u,p,P), equal to household sources of income from primary inputs and profits. We now allow the two cost functions to be once differentiable, continuous functions of time; ie c(y,w,k,t) is the opex cost function for the regulated sector and C(Y,w,t) is the cost function for the competitive sector. As in section 4.2, we assume that all variables are differentiable functions of time. Thus, equation (159) is now rewritten as follows:

$$(168)\ e(u(t),p(t),P(t)) = w(t)^T v(t) + p(t)^T y(t) - c(y(t),w(t),k(t),t) - P_k(t)k(t) + P(t)^T Y(t) - C(Y(t),w(t),t).$$

We also assume that counterparts to equations (153)-(158) also hold in continuous time and as usual, we assume that there is money metric utility scaling. Differentiating both sides of equation (168) with respect to time t and using (153)-(158) and money metric scaling, we obtain the following equation:

$$(169)\ u'(t) = w(t)^T v'(t) + [Y(t) - X(t)]^T P'(t) - P_k'(t)k(t) + m(t)^T y'(t) - \partial c(y,w,k,t)/\partial t - \partial C(Y,w,t)/\partial t$$

where $u'(t)$, $v'(t)$, $P'(t)$, $P_k'(t)$ and $y'(t)$ denote the time derivatives of the functions $u(t)$, $v(t)$, $P(t)$, $P_k(t)$ and $y(t)$ respectively. Equation (169) is a counterpart to equation (96) and it tells us that the rate of growth of money metric welfare at time t, $u'(t)$, is equal to the sum of six terms. The first term is $w(t)^T v'(t)$, the rate of growth of primary inputs $v'(t)$ in the economy, weighted by the vector of time t input prices $w(t)$. The second term is the inner product of the economy's net export vector at time t (excluding capital imports), $Y(t) - X(t)$, with the vector of rates of change in the prices of internationally traded goods and services, $P'(t)$ (this is a measure of the rate of change in the economy's terms of trade). The fourth term, $-P_k'(t)k(t)$, can also be interpreted as a terms of trade term: if the foreign price of capital services increases by $P_k'(t)$, then welfare will decrease by this amount of change times the vector of time t capital services, $k(t)$. The fourth term is the inner product of the vector of time t markups over marginal costs in the regulated sector, $m(t)$, with the vector of growth rates in regulated outputs, $y'(t)$. The fifth term is $-\partial c(y,w,k,t)/\partial t$ and it can be seen that this term is a measure of the rate of technical progress in the regulated sector and the sixth term is $-\partial C(Y,w,t)/\partial t$, which in turn is equal to the rate of technical progress in the unregulated sector. Thus, loosely speaking, the rate of welfare growth is equal to primary input growth plus terms of trade effects plus a term involving markups over marginal costs plus technical progress in the regulated and unregulated sectors.

As usual, a problem with the growth of welfare decomposition (169) is that the vector of growth rates for regulated outputs, $y'(t)$, is endogenous rather than exogenous. Hence, it what follows, we will obtain an expression for $y'(t)$ in terms of exogenous variables.

Equations (156) can be rewritten in our continuous time framework as follows:

(170) $y(t) = \nabla_p e(u(t),p(t),P(t))$.

Define the time t vector of income derivatives of demand for regulated outputs, b(t), by (98). Differentiating equation (170) with respect to t leads to the expression (99) for $y'(t)$. Substituting this expression into (169) and rearranging terms leads to the following expression for the growth of welfare in terms of exogenous variables:

$$(171)\ u'(t) = [1 - m(t)^T b(t)]^{-1}\{w(t)^T v'(t) + [Y(t) - X(t)]^T P'(t) + m(t)^T[\nabla_{pP}{}^2 e]P'(t)$$
$$- P_k'(t)k(t) + m(t)^T[\nabla_{pp}{}^2 e]p'(t) - \partial c(y,w,k,t)/\partial t - \partial C(Y,w,t)/\partial t\}$$

where we assume that (101) holds; ie that $1 - m(t)^T b(t) > 0$. Note that (171) is the counterpart to our previous expression for the rate of growth of welfare, (102), that we obtained in section 4.2 above. There are only two differences between (171) and (102):

- The new decomposition has added the extra terms of trade term, $- P_k'(t)k(t)$, and
- The measure of *opex technical progress*, $-\partial c(y,w,k,t)/\partial t$, appeared in (171) whereas the regulated sector *total cost function measure of technical progress*, $-\partial c(y,w,t)/\partial t$, appeared in (102).

The second difference means that we must rework our algebra that was developed in section 4.3 (which was used to obtain an index number estimate for the total cost measure of technical progress) in order to obtain an index number estimate for the operating cost measure of technical progress that appears in (171). We will do this in section 5.4 below.

Formula (171) is an important result for regulatory purposes. The regulator cannot control changes in the economy's endowment of primary inputs, $v'(t)$, or changes in the vector of prices for internationally traded goods and services, $P'(t)$, or changes in the price of capital, $P_k'(t)$, or the amount of cost saving technical progress in the competitive sector, $-\partial C(Y,w,t)/\partial t$. However, the regulator *can* determine changes in the vector of regulated product prices, $p'(t)$, and by allowing the regulated firm to keep some excess profits (above and beyond its cost of capital) if it undertakes cost saving changes in its technology, it *can* induce a certain amount of operating cost saving technical progress in the regulated sector, $-\partial c(y,w,k,t)/\partial t$. Thus, as was explained in section 4.2, if the regulator has some knowledge of the vector of markups of regulated prices over marginal costs, m(t), and if the regulator has some knowledge of the household substitution matrix for regulated products, $\nabla_{pp}{}^2 e(u(t),p(t),P(t))$, then the regulator can attempt to determine a vector of price changes for the regulated products, $p'(t)$, such that $m(t)^T[\nabla_{pp}{}^2 e]p'(t)$ is positive and such that the firm's profits are positive (but not excessive).

In the following subsection, we will look for an opex counterpart to equation (125) in section 4.3, which showed how Divisia index number techniques could be used in order to obtain an estimate for the rate of technical progress in the regulated sector using the regulated sector total cost function, $[-\partial c(y,w,t)/\partial t]/c(y,w,t)$.

5.4 Continuous time approaches to the determination of operating cost technical progress

Our task in the present section is to obtain index number estimates for the percentage change in operating cost for the regulated sector that is due to technical progress,[46] $[-\partial c(y,w,k,t)/\partial t]/c(y,w,k,t)$.

Recall that the opex cost function for the regulated sector at time t is $c(y(t),w(t),k(t),t)$. The vector of (noncapital) input demands at time t is $z(t)$ defined via Shephard's Lemma as follows:

$$(172)\ z(t) \equiv \nabla_w c(y(t),w(t),k(t),t).$$

Using equations (153), we know that the following relationships hold between the time t vector of regulated output prices $p(t)$, the vector of markups over marginal costs $m(t)$ and the vector of regulated sector marginal costs, $\mu(t) \equiv \nabla_y c(y(t),w(t),k(t),t)$:

$$(173)\ p(t) = m(t) + \mu(t).$$

The continuous time counterpart to (157) is:

$$(174)\ P_k(t) = -\partial c(y(t),w(t),k(t),t)/\partial k.$$

Using the linear homogeneity of $c(y,w,k,t)$ in w, we have:

$$(174)\ c(y(t),w(t),k(t),t) = w(t)^T\nabla_w c(y(t),w(t),k(t),t) = w(t)^T z(t) \qquad \text{using (172).}$$

Differentiating (174) with respect to t leads to the following equations:

$$\begin{aligned}(175)\ w(t)^T z'(t) + w'(t)^T z(t) &= \nabla_y c(y(t),w(t),k(t),t)^T y'(t) + \nabla_w c(y(t),w(t),k(t),t)^T w'(t) + [\partial c(y,w,k,t)/\partial k]k'(t) \\ &\quad + \partial c(y,w,k,t)/\partial t \\ &= \mu(t)^T y'(t) + z(t)^T w'(t) - P_k(t)k'(t) + \partial c(y,w,t)/\partial t \qquad \text{using (172) and (174).}\end{aligned}$$

Equations (175) imply the following expression for the opex rate of technical progress in the regulated sector:

$$(176)\ -\partial c(y,w,k,t)/\partial t = \mu(t)^T y'(t) - w(t)^T z'(t) - P_k(t)k'(t).$$

Thus, the rate of technical progress in the regulated sector is equal to the marginal cost weighted sum of time t output growth derivatives, $\sum_{n=1}^N \mu_n(t)y_n'(t)$, minus a time t input price weighted sum of (noncapital) input growth derivatives, $\sum_{k=1}^K w_k(t)z_k'(t)$, minus the price of capital services at time t, $P_k(t)$, times the rate of growth of capital services at time t, $k'(t)$. It can be seen that (176) is an exact counterpart to our old result (120), which used the regulated firm's total cost function instead of the operating cost function as in the present section. As was the case in section 4.3, what makes the estimation of the rate of technical progress for the regulated sector so difficult is that we require a knowledge of marginal costs in order to form estimates for the right hand side of (176), or equivalently, we need estimates for the selling prices of regulated products, $p(t)$ (not a problem in principle), as well as for the vector of markups over marginal costs at time t, $m(t)$, (and this is a problem).

[46] This measure of technical progress can also include changes in variables that reflect exogenous operating conditions for the regulated firm.

We now turn our attention to the problem of obtaining an expression for the logarithmic rate of operating cost reduction. In order to accomplish this task, we need to define the opex counterpart to our earlier measure of the degree of returns to scale in the regulated sector that was defined by (31) and later applied in (121).

Using the operating cost function for the regulated sector, a new *measure of (reciprocal) returns to scale* for the regulated sector, $\rho(y,w,t)$, can be defined as follows:

$$(177)\ \rho(y,w,k,t) \equiv \partial \ln c(\lambda y,w,\lambda k,t)/\partial\lambda\,_{|\lambda=1}$$
$$= [y^T\nabla_y c(y,w,k,t) + k\partial c(y,w,k,t)/\partial k]/c(y,w,k,t).$$

Thus, $\rho(y,w,k,t)$ gives us the percentage change in opex cost due to a small proportional increase in all outputs and in capital input. If there are *increasing returns to scale* in the regulated sector, then $\rho(y,w,k,t)$ will be less than unity; ie we have decreasing costs and $\rho(y,w,k,t)$ will be less than one. If there are constant returns to scale in the regulated sector so that $c(\lambda y,w,\lambda k,t)$ equals $\lambda c(y,w,k,t)$ for all λ greater than zero, then using Euler's Theorem on homogeneous functions, it can be shown that $\rho(y,w,k,t)$ will equal one. Now regard y, w and k as the functions of t, y(t), w(t) and k(t) and define the *degree of reciprocal opex returns to scale as a function of t*, $\rho(t)$, as follows:

$$(178)\ \rho(t) \equiv \rho(y(t),w(t),k(t),t)$$
$$= [y(t)^T\nabla_y c(y(t),w(t),k(t),t) + k(t)\partial c(y(t),w(t),k(t),t)/\partial k]/c(y(t),w(t),k(t),t)$$
$$= [\mu(t)^T y(t) - P_k(t)k(t)]/c(y(t),w(t),k(t),t)$$

where we have used (174) and the fact that $\mu(t)$ is the vector of opex marginal costs at time t, $\nabla_y c(y(t),w(t),k(t),t)$, in order to establish the last equation in (178). Note that (178) implies:

$$(179)\ \mu(t)^T y(t) - P_k(t)k(t) = \rho(t)c(y(t),w(t),k(t),t).$$

Define the *Divisia index of aggregate regulated sector output growth using opex marginal cost weights* at time t, $y_D'(t)$, and the *Divisia index of opex input growth* at time t, $z_D'(t)$, as follows:

$$(180)\ y_D'(t) \equiv \sum_{n=1}^N [\mu_n(t)y_n(t)/\mu(t)^T y(t)][y_n'(t)/y_n(t)]\ ;$$

$$(181)\ z_D'(t) \equiv \sum_{k=1}^K [w_k(t)z_k(t)/w(t)^T z(t)][z_k'(t)/z_k(t)]\ .$$

Now divide both sides of (176) by the regulated sector's operating cost at time t, $c(y(t),w(t),k(t),t)$. Then using (118) and (122), we obtain the following expression for the logarithmic rate of technical progress in the regulated sector at time t:[47]

$$(182)\ [-\partial c(y,w,k,t)/\partial t]/c(y(t),w(t),k(t),t)$$
$$= [\mu(t)^T y(t)/c(y(t),w(t),k(t),t)]y_D'(t) - z_D'(t) - [P_k(t)k(t)/c(y(t),w(t),k(t),t)][k'(t)/k(t)];$$

ie the logarithmic rate of opex technical progress in the regulated sector is equal to the product of the ratio of (hypothetical) revenue at marginal cost prices, $\mu(t)^T y(t)$, to opex cost, $c(y(t),w(t),k(t),t)$, times the marginal cost weighted Divisia index of output growth, $y_D'(t)$,

[47] Note that (182) is not quite as simple as its counterpart in section 4.3, which was equation (125). It can be seen that the sum of the weights associated with the terms $y_D'(t)$ and $k'(t)/k(t)$ in (182) sum up to the degree of reciprocal opex returns to scale $\rho(t)$ defined by (178).

less the Divisia index of opex input growth, $z_D'(t)$, *less* the ratio of capital services cost, $P_k(t)k(t)$, to opex cost, $c(y(t),w(t),k(t),t)$, times the Divisia index of capital services growth, $k'(t)/k(t)$.[48]

It can be seen (using (176) in particular) that the one period model of regulation developed in sections 2 and 3 (which did not mention capital services explicitly) is more or less applicable if capital services are in the list of primary inputs, provided that the capital services can be freely varied from one period to the next. However, in the regulated context, many types of capital services are sunk costs; ie once the capital investment has been made, the investment is more or less irreversible until the capital input completely wears out and must be scrapped or discarded. In order to model the effects of sunk costs, we need to generalise our one period model of regulation into an intertemporal model and this is the topic of the following section.

[48] The generalisation to many types of freely available capital services is straightforward. More sophisticated discrete time estimates for technical progress indexes than that given by (182) (which is only a first order approximation to the underlying technical progress measure) can be obtained using the techniques explained in Diewert and Fox (2008).

6 MULTIPLE PERIOD REGULATION WITH SUNK COSTS

6.1 A two period intertemporal model

The regulator's problem becomes much more complex once we recognise that some capital stock components in the regulated sector cannot be readily adjusted going from one accounting period to the next. Thus, once an infrastructure investment has been made by a regulated firm, the carrying capacity of that investment typically cannot be varied; ie the investment has the character of a sunk cost. Examples of such irreversible investments are:

- electricity networks;
- natural gas pipelines;
- water supply systems;
- railway lines;
- roads; and
- telecommunications networks.

In order to model this sunk cost situation, it is necessary to take an intertemporal perspective. Thus, we need to assume that the households in the economy have preferences over goods and services that are consumed not only in the current period but also over future periods. In the interests of notational simplicity, we will only consider the case of two periods but the complexities of the regulator's problem already emerge in this simple framework.

As usual, we will assume that there is a competitive sector in each period. The feasible set of period t output vectors $Y^t \geq 0_J$ and input vectors $Z^t \geq 0_K$ for the competitive sector is the production possibilities set S^t for t = 1,2. As in the one period models studied in the previous sections, we will assume that there are constant returns to scale in the competitive sector and that S^t is a nonempty, closed convex cone for t = 1,2.

There is also a regulated sector in the economy in each period t which produces a vector of regulated outputs $y^t \geq 0_N$ and uses a vector of variable inputs $z^t \geq 0_K$ along with a capital input scalar $k^t \geq 0$ for t = 1,2. The production possibilities set for the regulated sector in period t is s^t for t = 1,2. We assume that s^t satisfies assumptions (126) in the previous section for t = 1,2.[49]

As in section 5.1, we need to define various cost functions for the regulated and competitive sectors. Thus, let $y \geq 0_N$ be a vector of output targets, k a capital input that is available for the sector and $w >> 0_K$ a vector of input prices that the regulated sector faces. Then the *regulated sector's period t joint operating cost (or opex) function*, $c^t(y,w,k)$, is defined as follows:

$$(183)\quad c^t(y,w,k) \equiv \min_z \{w^T z : (y,z,k) \in s^t\} ; \qquad t = 1,2.$$

[49] Generally, the production possibilities sets will grow over time; ie usually S^1 will be a subset of S^2 and s^1 a subset of s^2 so that there is technical progress in both sectors.

As in section 5.1, it can be shown that $c^t(y,w)$ is nonnegative, nondecreasing in y, and nondecreasing, (positively) linearly homogeneous and concave in w. Moreover, if c(y,w,k) is differentiable with respect to the components of the input price vector w, then adapting the arguments of Hotelling (1932; 594) and Shephard (1953; 11), we can show that the period t vector of opex cost minimising input demand functions, $z^t(y,w,k)$, is equal to the vector of first order partial derivatives of the joint cost function; ie we have:[50]

(184) $z^t(y,w,k) = \nabla_w c^t(y,w,k)$; $\quad t = 1,2.$

Let $Y \geq 0_J$ be a vector of output targets and $w >> 0_K$ be a vector of input prices that the competitive sector faces. Then the *competitive sector's period t joint cost function*, $C^t(Y,w)$, is defined as follows:

(185) $C^t(Y,w) \equiv \min_Z \{w^T Z : (Y,Z) \in S^t\}$; $\quad t = 1,2.$

It can be shown that $C^t(Y,w)$ is nonnegative, nondecreasing, linearly homogeneous and convex in Y, and nondecreasing, linearly homogeneous and concave in w. If $C^t(Y,w)$ is differentiable with respect to the components of the input price vector w, then Shephard's Lemma again implies that the vector of period t cost minimising input demand functions for the unregulated sector, $Z^t(Y,w)$, is equal to the vector of first order partial derivatives of the unregulated joint cost function:

(186) $Z^t(Y,w) = \nabla_w C^t(Y,w)$; $\quad t = 1,2.$

We now turn our attention to describing consumer preferences. Recall that the one period utility function U(x,X) was defined over nonnegative consumption vectors of regulated products, $x \geq 0_N$, and nonnegative vectors of unregulated products, $X \geq 0_J$. We now assume that the household sector has preferences defined over regulated and unregulated products for two periods and that these preferences can be represented by the utility function $U(x^1,X^1,x^2,X^2)$ where x^t and X^t are the period t consumption vectors of regulated and unregulated products respectively for t = 1,2. As usual, we assume that U is a nonnegative, increasing, continuous and concave function in the components of x^1,X^1,x^2,X^2.

Let u be a utility target for the household and suppose the household faces the vector of prices $p^t >> 0_N$ for regulated outputs and $P^t >> 0_J$ for unregulated outputs for periods t = 1,2. Then the household's *intertemporal expenditure function*, $e(u,p^1,P^1,p^2,P^2)$, is defined as the solution to the following expenditure minimisation problem:

(187) $e(u,p^1,P^1,p^2,P^2) \equiv \min_{x^1,X^1,x^2,X^2} \{p^{1T}x + P^{1T}X + p^{2T}x + P^{2T}X : U(x^1,X^1,x^2,X^2) \geq u$;

$x^1 \geq 0_N$; $X^1 \geq 0_J$; $x^2 \geq 0_N$; $X^2 \geq 0_J\}$.

The consumer's expenditure function will be nondecreasing in all of its variables and linearly homogeneous and concave in the prices (p^1,P^1,p^2,P^2). If $e(u,p^1,P^1,p^2,P^2)$ is differentiable with respect to the components of the commodity prices p^t and P^t, then Shephard's Lemma implies that the consumer's period t system of Hicksian demand functions for regulated commodities, $x^t(u, p^1,P^1,p^2,P^2)$, is equal to the vector of first order partial derivatives of $e(u,p^1,P^1,p^2,P^2)$ with respect to the components of p^t and the consumer's system of Hicksian demand functions for

[50] We also require that $c^t(y,w,k)$ be finite in a neighbourhood around the point (y,w,k) in order to derive this result.

unregulated commodities in period t, $X^t(u,p^1,P^1,p^2,P^2)$, is equal to the vector of first order partial derivatives of $e(u,p^1,P^1,p^2,P^2)$ with respect to the components of P^t for t = 1,2; ie in the differentiable case, we have:

(188) $x^t(u,p^1,P^1,p^2,P^2) = \nabla_{p^t} e(u,p^1,P^1,p^2,P^2)$; t = 1,2 ;

(189) $X^t(u,p^1,P^1,p^2,P^2) = \nabla_{P^t} e(u,p^1,P^1,p^2,P^2)$; t = 1,2.

The resource constraints that face the household and producers in the economy are as follows:

(190) $y^t \geq x^t$; t = 1,2 ;

(191) $v^t \geq z^t + Z^t$; t = 1,2 ;

(192) $P^{1T}[Y^1 - X^1] + P^{2T}[Y^2 - X^2] - P_k k \geq 0$.

The constraints (190) impose the restriction that period t production vector of the regulated commodities, y^t, be equal to or greater than period t household consumption vector of these regulated commodities for t = 1,2. The constraints (191) impose the constraint that the period t vector of (noncapital) primary input demands from the regulated sector, z^t, plus the vector of period t primary input demands from the competitive sector, Z^t, be less than or equal to the economy's vector of available resources in period t, v^t, for t = 1,2. The new intertemporal balance of trade constraint (192) says that the value of net exports of outputs in period 1, $P^{1T}[Y^1 - X^1]$, plus the value of net exports in period 2, $P^{2T}[Y^2 - X^2]$, has to be equal to or greater than the value of capital imports at the beginning of period 1, $P_k k$, where P_k is the beginning of period 1 (stock) price for a unit of capital (which we assume is imported or made from imported materials for simplicity[51]) and $k \geq 0$ is the number of units of capital purchased at the beginning of period 1. Note that P^t is the vector of internationally determined prices for the outputs of the competitive sector, Y^t is the period t output vector for the competitive sector and X^t is the period t household consumption vector for unregulated products for t = 1,2.

The *household's intertemporal constrained utility maximisation problem* is now the problem of maximising utility u subject to the constraints (190)-(192) and the following constraints:

(193) $U(x^1,X^1,x^2,X^2) - u \geq 0$;

(194) $(y^t,z^t,k) \in s^t$; t = 1,2 ;

(195) $(Y^t,Z^t) \in S^t$; t = 1,2.

What is fundamentally different about the new intertemporal utility maximisation problem as opposed to the one period utility maximisation problem considered in section 5.1 above is the treatment of capital in the constraints (194); ie note that *the capital variable k is constrained to be the same over the two periods*. It is this fixity that captures the nature of the sunk cost problem in the regulation of utilities. When a utility makes an infrastructure investment, it typically has to plan ahead over a long horizon because once the infrastructure has been built,

[51] If the capital good is made domestically, then its price becomes an endogenous variable in the model. This endogeniety can readily be accommodated but at the cost of extra notational complexity (it would be necessary to introduce a new class of products into our model, namely intermediate inputs). The overall structure of the problem would not change materially.

it lasts for a long time and its maximum carrying capacity cannot readily be varied. Thus, at the beginning of period 1, we are assuming that the regulated firm must make a capital investment which will determine the capacity of the network not only for period 1, but also for subsequent periods. Of course, we have simplified the problem by assuming a horizon of only two periods but this limitation of our model can readily be generalised. However, the case of two periods will suffice to illustrate the complexities of the regulator's optimal regulation problem.

We will assume that a strictly positive solution to the household's intertemporal constrained utility maximisation problem exists; ie $u^* > 0$, $y^{t*} >> 0_N$, $Y^{t*} >> 0_J$, $z^{t*} >> 0_K$, $k^* > 0$, $Z^{t*} >> 0_K$, $x^{t*} >> 0_N$ and $X^{t*} >> 0_J$ for t = 1,2 solves the household's constrained intertemporal utility maximisation problem. Now let $y^t \geq 0_N$ be in a neighbourhood of y^{t*} and consider the following constrained maximisation problem that is conditional on the choice of y^1 and y^2:

$$\begin{aligned}(196)\ H(y^1,y^2) \equiv \max_{u,Y\text{'s},z\text{'s},Z\text{'s},x\text{'s},X\text{'s},k} \{&u : y^t - x^t \geq 0_N ,\ t = 1,2 ;\ v^t - z^t - Z^t \geq 0_K ,\ t = 1,2 ;\\ &P^{1T}[Y^1 - X^1] + P^{2T}[Y^2 - X^2] - P_k k \geq 0 ;\ (y^t,z^t,k)\in s^t ;\ (Y^t,Z^t)\in S^t ,\ t = 1,2 ;\\ &U(x^1,X^1,x^2,X^2) - u \geq 0 ;\ u \geq 0,\ Y^t \geq 0_J,\ z^t \geq 0_K,\ Z^t \geq 0_K,\ x^t \geq 0_N,\ X^t \geq 0_J ,\ t = 1,2\}.\end{aligned}$$

It can be verified that (196) is a concave programming problem and hence the Karlin (1959; 201-203) Uzawa (1958) Saddle Point Theorem can be applied to this problem. Using this Theorem, we can absorb some of the constraints into the objective function and it turns out that u, k, Y^t, z^t, Z^t, x^t, and X^t solutions to (196) are also solutions to the following max-min problem:[52]

$$\begin{aligned}(197)\ H(y^1,y^2) = \max_{u,Y\text{'s},z\text{'s},Z\text{'s},x\text{'s},X\text{'s},k} \min_{p\text{'s},w\text{'s},\lambda} \{&u + p^{1T}[y^1 - x^1] + w^{1T}[v^1 - z^1 - Z^1]\\ &+ p^{2T}[y^2 - x^2] + w^{2T}[v^2 - z^2 - Z^2] + \lambda\{P^{1T}[Y^1 - X^1] + P^{2T}[Y^2 - X^2] - P_k k\} :\\ &(y^t,z^t,k)\in s^t,\ t = 1,2 ;\ (Y^t,Z^t)\in S^t,\ t = 1,2 ;\ U(x^1,X^1,x^2,X^2) - u \geq 0 ;\ u \geq 0,\\ &\lambda \geq 0,\ Y^t \geq 0_J,\ z^t \geq 0_K,\ Z^t \geq 0_K,\ x^t \geq 0_N,\ X^t \geq 0_J,\ p^t \geq 0_N,\ w^t \geq 0_K,\ t = 1,2\}\end{aligned}$$

where p^t can be interpreted as a *vector of prices for the regulated products in period t*, w^t can be interpreted as a *vector of (noncapital) primary input prices for period t*, t = 1,2 and λ is a Lagrange multiplier that corresponds to the balance of trade constraint (192).

Using definitions (183), (185) and (187), it can be seen that we can readily perform the maximisation of (197) with respect to the z^t, Z^t, x^t and X^t. Thus, (197) is equivalent to the following max-min problem which now has no constraints other than nonnegativity constraints:

$$\begin{aligned}(198)\ H(y^1,y^2) = \max_{u,Y\text{'s},k} \min_{p\text{'s},w\text{'s},\lambda} \{&u + w^{1T}v^1 + p^{1T}y^1 - c^1(y^1,w^1,k) + \lambda P^{1T}Y^1\\ &- C^1(Y^1,w^1) + w^{2T}v^2 + p^{2T}y^2 - c^2(y^2,w^2,k) + \lambda P^{2T}Y^2 - C^2(Y^2,w^2) - \lambda P_k k\\ &- e(u,p^1,\lambda P^1,p^2,\lambda P^2) : u \geq 0,\ \lambda \geq 0 ;\ Y^t \geq 0_J,\ p^t \geq 0_N,\ w^t \geq 0_K,\ t = 1,2\}.\end{aligned}$$

Note that the period t opex cost functions for the regulated sector, $c^t(y^t,w^t,k)$, and the period t cost functions for the unregulated sector, $C^t(Y^t,w^t)$, have made their appearance in (198). It can be seen that solutions to the household's initial constrained utility maximisation problem

[52] In order to obtain the equality of (196) and (197), we need to assume that a constraint qualification condition holds.

are solutions to the problem of maximising $H(y^1,y^2)$ with respect to the components of y^1,y^2. Hence solutions to the household's initial constrained utility maximisation problem are solutions to the following *two stage max-min problem*:

$$(199)\ \max_{y\text{'s}} [\max_{u,Y\text{'s},k} \min_{p\text{'s},w\text{'s},\lambda} \{u + w^{1T}v^1 + p^{1T}y^1 - c^1(y^1,w^1,k) + \lambda P^{1T}Y^1 - C^1(Y^1,w^1) + p^{2T}y^2 - c^2(y^2,w^2,k) + \lambda P^{2T}Y^2 - C^2(Y^2,w^2) - \lambda P_k k - e(u,p^1,\lambda P^1,p^2,\lambda P^2) : u \geq 0, \lambda \geq 0 ; Y^t \geq 0_J, p^t \geq 0_N, w^t \geq 0_K, t = 1,2\}].$$

Thus, for the first stage, we solve the max min problem defined by (198) and in the second stage, we solve:

$$(200)\ \max_{y^1\geq 0_N, y^2 \geq 0_N} H(y^1,y^2).$$

We assume that there is a strictly positive solution to the two stage max-min problem defined by (199), say $u^* > 0$, $\lambda^* > 0$, $y^{t*} >> 0_N$, $Y^{t*} >> 0_J$, $z^{t*} >> 0_K$, $k^* > 0$, $Z^{t*} >> 0_K$, $x^{t*} >> 0_N$, $X^{t*} >> 0_J$, $p^{t*} >> 0_N$, $w^{t*} >> 0_K$ for $t = 1,2$.[53] We also assume that the four cost functions, c^t and C^t for $t = 1,2$ and the consumer's expenditure function e are finite and differentiable in a neighbourhood of this equilibrium point.[54] It can be shown that the first order necessary conditions for the two stage max-min problem defined by (199) can be obtained by simply differentiating the objective function in (199) with respect to the y^t, k, u, Y^t, p^t, w^t and λ and setting these partial derivatives equal to zero.[55] Thus, under our assumptions, we find that the optimal solution to the consumer's constrained utility maximisation problem satisfies the following first order conditions:

$$(201)\quad 1 = \partial e(u^*,p^{1*},\lambda P^1,p^{2*},\lambda P^2)/\partial u ;$$

$$(202)\quad p^{t*} = \nabla_y c^t(y^{t*},w^{t*},k^*) ; \qquad t = 1,2 ;$$

$$(203)\quad \lambda^* P^t = \nabla_Y C^t(Y^{t*},w^{t*}) ; \qquad t = 1,2 ;$$

$$(204)\quad v^t = \nabla_w c^t(y^{t*},w^{t*},k^*) + \nabla_w C^t(Y^{t*},w^{t*}) ; \qquad t = 1,2 ;$$

$$(205)\quad y^{t*} = \nabla_{p^t} e(u^*,p^{1*},\lambda P^1,p^{2*},\lambda P^2) ; \qquad t = 1,2 ;$$

$$(206)\quad P_k = -\partial c^1(y^{1*},w^{1*},k^*)/\partial k - \partial c^2(y^{2*},w^{2*},k^*)/\partial k ;$$

$$(207)\quad 0 = P^{1T}[Y^{1*} - \nabla_{p^1} e(u^*,p^{1*},\lambda P^1,p^{2*},\lambda P^2)] + P^{2T}[Y^{2*} - \nabla_{p^2} e(u^*,p^{1*},\lambda P^1,p^{2*},\lambda P^2)] - P_k k^*.$$

Equation (201) sets the marginal utility of income equal to unity at the optimal equilibrium. This restriction determines the scale of prices in the economy. Any other normalisation on the overall level of prices will work just as well. We will normalise domestic prices by calibrating them to international prices as in section 2; ie we will set λ^* equal to unity in what follows. Thus, with this normalisation, the remaining equations (202)-(207) become the following equations:

$$(208)\quad p^{t*} = \nabla_y c^t(y^{t*},w^{t*},k^*) ; \qquad t = 1,2 ;$$

[53] As usual, if we do not make these positivity conditions, then it is necessary to work with more complex Kuhn-Tucker (1951) conditions.

[54] As will be seen in section xx below, it is not always the case that these functions are differentiable.

[55] Again, this follows by a generalisation of Samuelson's (1947; 34) Envelop Theorem to cover the case of max-min problems.

(209) $P^t = \nabla_Y C^t(Y^{t*},w^{t*})$; $t = 1,2$;

(210) $v^t = \nabla_w c^t(y^{t*},w^{t*},k^*) + \nabla_w C^t(Y^{t*},w^{t*})$; $t = 1,2$;

(211) $y^{t*} = \nabla_{p^t} e(u^*,p^{1*},P^1,p^{2*},P^2)$; $t = 1,2$;

(212) $P_k = -\,\partial c^1(y^{1*},w^{1*},k^*)/\partial k - \partial c^2(y^{2*},w^{2*},k^*)/\partial k$;

(213) $0 = P^{1T}[Y^{1*}-\nabla_{P^1} e(u^*,p^{1*},P^1,p^{2*},P^2)] + P^{2T}[Y^{2*}-\nabla_{P^2} e(u^*,p^{1*},P^1,p^{2*},P^2)] - P_k k^*$.

Equations (208)-(211) and (213) are the counterparts to equations (145)-(148) and (150) in section 5.1 and have similar interpretations. However, equation (212) is different from its counterpart equation in the one period model, equation (149). Equation (212) says that in an optimal equilibrium, the (stock) price of a unit of capital that cannot be varied over its useful life, P_k, should be equal to the *negative* of the sum of the partial derivatives $\partial c^1(y^{1*},w^{1*},k^*)/\partial k$ plus $\partial c^2(y^{2*},w^{2*},k^*)/\partial k$ of the period 1 and 2 opex cost functions with respect to the (fixed) capital input variable, k. In section 5.1, when capital was freely variable from period to period, the optimality condition was that $-\partial c^1(y^{1*},w^{1*},k^*)/\partial k$ be equal to the relevant user cost of capital (a flow price) for period 1. However, when capital is fixed over multiple periods, this old optimality condition no longer holds and the new optimality condition is the intertemporal condition (212). Additional implications of equation (212) for the measurement of productivity will be discussed later in section 6.3.

Recall equations (189) which showed that the consumer's period t Hicksian demands for unregulated products $X^t(u,p^1,P^1,p^2,P^2)$ is equal to the vector of first order partial derivatives of the expenditure function with respect to the components of P^t. Using this fact, it can be seen that equation (213) is equivalent to the following equation:

(214) $P^{1T}[Y^{1*}-\nabla_{P^t} e(u^*,p^{1*},P^1,p^{2*},P^2)] + P^{2T}[Y^{2*}-\nabla_{P^2} e(u^*,p^{1*},P^1,p^{2*},P^2)] - P_k k^*$

$$= P^{1T}[Y^{1*}-X^{1*}] + P^{2T}[Y^{2*}-X^{2*}] - P_k k^* = 0.$$

Thus, (214) says the (discounted)[56] value of net exports of goods and services over the two periods, $P^{1T}[Y^{1*}-X^{1*}] + P^{2T}[Y^{2*}-X^{2*}]$, must equal the value of purchases of the fixed capital imports at the beginning of period 1, $P_k k^*$.

Equations (208)-(213) can be regarded as $2 + 4N + 2J + 2K$ equations in the $2 + 4N + 2J + 2K$ endogenous variables u, k, p^1, p^2, y^1, y^2, Y^1, Y^2, w^1 and w^2. *These equations characterise an optimal regulatory equilibrium under our assumptions*. The exogenous variables are v^1 and v^2 (the two vectors of factor endowments), P^1 and P^2 (the two vectors of international prices for traded goods and services) and P_k, the (stock) price for importing a unit of capital at the beginning of period 1.

We now generalise the above optimal model to a model of regulation that falls short of attaining a first best equilibrium. Suppose as in the one period model, the regulator sets the price of regulated outputs in period t at $p^t >> 0_N$ and demands that the regulated sector meet all period t demands $y^t \geq 0_N$ for t = 1,2. Given the vector of selling prices p^t and the vector of opex marginal costs $\nabla_y c^t(y^t,w^t,k)$, we can define the *period t vector of deviations from opex*

[56] In a later section, we will make clear the role of discounting in our model but for now, bringing in discounting would lead to a great deal of additional notation.

marginal cost pricing or the *period t vector of margins* over opex marginal cost, m^t, as follows:

(215) $m^t \equiv p^t - \nabla_y c^t(y^t,w^t,k)$; $t = 1,2.$

Recall that equations (208)-(213) characterised a first best optimal regulatory policy. We now consider nonoptimal regulatory policies by replacing the price equals opex marginal cost equations (208) by the price equals opex marginal cost plus margins equations (215). Thus, a *general regulatory policy equilibrium* is characterised by equations (215) and the following equations:

(216) $P^t = \nabla_Y C^t(Y^t,w^t)$; $t = 1,2$;

(217) $v^t = \nabla_w c^t(y^t,w^t,k) + \nabla_w C^t(Y^t,w^t)$; $t = 1,2$;

(218) $y^t = \nabla_{p^t} e(u,p^1,P^1,p^2,P^2)$; $t = 1,2$;

(219) $P_k = -\partial c^1(y^1,w^1,k)/\partial k - \partial c^2(y^2,w^2,k)/\partial k$;

(220) $0 = P^{1T}[Y^1 - \nabla_{P^1} e(u,p^1,P^1,p^2,P^2)] + P^{2T}[Y^2 - \nabla_{P^2} e(u,p^1,P^1,p^2,P^2)] - P_k k$.

These equations can be regarded as 2 + 4N + 2J + 2K equations in the 2 + 4N + 2J + 2K endogenous variables u, k, m^t, y^t, Y^t and w^t for t = 1,2. These equations determine a general equilibrium for the economy under our assumptions. The exogenous variables are now p^1 and p^2 (the two vectors of regulatory prices for regulated outputs), v^1 and v^2 (the two vectors of factor endowments), P^1 and P^2 (the two vectors of international prices) and P_k (the purchase price of a unit of imported fixed capital). Thus, the p^t have moved from being endogenous vectors to being exogenous vectors but we have added two extra endogenous vectors of variables, m^1 and m^2, the vectors of margins over the opex marginal costs. Thus, in the case where the regulated industry has increasing returns to scale, the regulator can set regulated prices high enough so that the regulated industry has some positive margins and does not have to be subsidised by general government revenues.

As in section 2, we are assuming competitive price taking behaviour in the unregulated sector. We are also assuming competitive cost minimising behaviour on the part of the regulated firm, both with respect to operating costs and also with respect to minimising long run costs with respect to the initial level of the fixed capital input.[57]

Using the same techniques that were used in section 2, it can be shown that the balance of trade constraint, (220), can be replaced by an intertemporal income equals expenditure constraint on the household. This replacement equation is (221) below and it is a counterpart to equation (159) in section 5.1.

$$\begin{aligned}(221)\ e(u,p^1,P^1,p^2,P^2) = {} & w^{1T}v^1 + w^{2T}v^2 + p^{1T}y^1 - c^1(y^1,w^1,k) + p^{2T}y^2 - c^2(y^2,w^2,k) \\ & + P^{1T}Y^1 - C^1(Y^1,w^1) + P^{2T}Y^2 - C^2(Y^2,w^2) - P_k k.\end{aligned}$$

It should be mentioned that the model defined by equations (215)-(220) does not assume that

[57] Given output target vectors y^1 and y^2 and variable input price vectors w^1 and w^2, the regulated cost minimising producer will want to minimise the sum of discounted expected costs, $c^1(y^1,w^1,k) + c^2(y^2,w^2,k) - P_k k$, with respect to the choice of the fixed capital input, k. The first order necessary condition for solving this intertemporal cost minimisation problem is (219).

period by period income is equal to household expenditure; only the equality of (discounted) income with (discounted) expenditure is assumed. Any gaps between income and expenditure for a given period are made up by foreign borrowing or lending.

In section 6.2, we will briefly look at the comparative statics properties of the present model, using the same techniques that were used in section 5.2 above. In section 6.3, we will look at a special case of the general model presented in this section that is simple enough so that we can obtain a version of the one period model that was studied in section 5 above.

6.2 The comparative statics properties of the intertemporal regulatory model

The same techniques that were used to derive comparative statics properties of the model in section 5 can be used in order to derive the properties of the model defined by (215)-(220).

As usual, we will impose money metric utility scaling on the household utility function at the initial equilibrium prices. Thus, we assume that:

(222) $e(u,p^1,P^1,p^2,P^2) = u$ for all $u \geq 0$.

Differentiating (222) with respect to u and evaluating the resulting equation at the initial equilibrium gives us the following condition:

(223) $\partial e(u,p^1,P^1,p^2,P^2/\partial u = 1$.

We begin by looking at how welfare or utility changes as the regulator varies the components in the vector of period 1 regulated prices, p^1. Thus, regard u, k, m^t, y^t, Y^t and w^t for t = 1,2 as vector valued functions of the vector of prices p^1 and differentiate both sides of equation (221) with respect to the components of p^1. Using (223) and equations (215)-(220), we find that the resulting equation simplifies to:

(224) $\nabla_p u(p^1) = [\nabla_p y^1(p^1)]^T m^1(p^1) + [\nabla_p y^2(p^1)]^T m^2(p^1)$.

Note that if the regulator has somehow chosen the regulated price vectors p^1 and p^2 so that the period 1 and 2 markup vectors m^1 and m^2 are both equal to 0_N (so that we have marginal cost pricing in each period), then

(225) $\nabla_p u(p^1) = 0_N$;

ie the first order necessary conditions for a first best equilibrium are satisfied. This is to be expected; if we have marginal cost pricing in both periods, then we can expect to be at a first best equilibrium.

The matrices of derivatives $[\nabla_p y^1(p^1)]$ and $[\nabla_p y^2(p^1)]$ are endogenous to the model and so it will be useful to derive formulae for these matrices in terms of exogenous variables. Recall equations (188), which defined the Hicksian demand functions for regulated products, $x^t(u,p^1,P^1,p^2,P^2)$, in terms of partial derivatives of the consumer's expenditure function, $e(u,p^1,P^1,p^2,P^2)$. Differentiating both sides of equations (188) with respect to u leads to the following equations (and definitions):

(226) $\partial x^t(u,p^1,P^1,p^2,P^2)/\partial u = \nabla^2_{p^t u} e(u,p^1,P^1,p^2,P^2) \equiv b^t$; t = 1,2

Now regard y^1, y^2 and u as functions of p^1 and differentiate both sides of equations (218) with respect to the components of p^1. We obtain the following matrix equations:

(227) $\nabla_{p^1} y^1(p^1) = \nabla^2_{p^1u} e(u,p^1,P^1,p^2,P^2)\nabla_{p^1} u(p^1) + \nabla^2_{p^1p^1} e(u,p^1,P^1,p^2,P^2)$;

(228) $\nabla_{p^1} y^2(p^1) = \nabla^2_{p^2u} e(u,p^1,P^1,p^2,P^2)\nabla_{p^1} u(p^1) + \nabla^2_{p^2p^1} e(u,p^1,P^1,p^2,P^2)$.

Differentiating the Hicksian demand functions for regulated products defined by (188) with respect to the components of the vectors p^1 and p^2 leads to the following negative semidefinite and symmetric consumer substitution matrix, Σ:

(229) $\Sigma \equiv \begin{bmatrix} \nabla_{p^1} x^1 & \nabla_{p^2} x^1 \\ \nabla_{p^2} x^2 & \nabla_{p^2} x^2 \end{bmatrix} = \begin{bmatrix} \nabla^2_{p^1p^1} e & \nabla^2_{p^1p^2} e \\ \nabla^2_{p^2p^1} e & \nabla^2_{p^2p^2} e \end{bmatrix} = \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix}$.

Since Σ is a negative semidefinite symmetric matrix, so are the submatrices, Σ_{11} and Σ_{22}. Moreover, the symmetry of Σ implies $\Sigma_{21}{}^T$ equals Σ_{12}. Substitute (227) and (228) into (224). Using (226) and (229), we obtain the following expression (after some simplifications) for the partial derivatives of utility u with respect to the components of p^1:

(230) $(1 - m^{1T}b^1 - m^{2T}b^2)\nabla_p u(p^1) = \Sigma_{11}\, m^1 + \Sigma_{12}\, m^2$.

Typically, $m^{1T}b^1 + m^{2T}b^2$ will be much smaller than 1 so that the vector of derivatives $\nabla_p u(p^1)$ of utility with respect to the components of the vector of first period vector of regulated prices p^1 will be equal to the vector $\Sigma_{11}\, m^1 + \Sigma_{12}\, m^2$ times the positive number $(1 - m^{1T}b^1 - m^{2T}b^2)^{-1}$. If the regulator has some information on the magnitude of the current and future period vectors of markups over marginal cost, m^1 and m^2, and estimates of consumer substitution derivatives, Σ_{11} and Σ_{12}, then the regulator could in principle adjust the vector of period 1 regulated prices in order to improve consumer welfare, while preserving an intertemporal solvency constraint for the regulated firm.

The same type of computations can be used in order to derive the vector of derivatives of u with respect to the components of the period 2 vector of regulated prices, p^2. The counterpart to (230) turns out to be:

(231) $(1 - m^{1T}b^1 - m^{2T}b^2)\nabla_p u(p^2) = \Sigma_{21}\, m^1 + \Sigma_{22}\, m^2$.

Equations (230) and (231) can be combined into the following matrix equation:

(232) $(1 - m^{1T}b^1 - m^{2T}b^2)\begin{bmatrix} \nabla_{p^1} u(p^1,p^2) \\ \nabla_{p^2} u(p^1,p^2) \end{bmatrix} = \Sigma \begin{bmatrix} m^1 \\ m^2 \end{bmatrix}$.

Note that the intertemporal comparative statics results derived in this section are very similar to the results derived in section 3.4 above which worked out expressions for the effects on household welfare of changes in the price of regulated commodities. Our advice in section 3.4 was (roughly speaking) that the regulator should decrease the regulated price for a commodity where the markup of price over marginal cost was unusually large and increase the price where the markup was low, while keeping an eye on the solvency constraint of the regulated firm. The same advice will *probably* lead to an improvement in welfare in the present intertemporal context as well but it is difficult to establish rigorous rules due to the fact that the regulator will have very imperfect information on the current period markup vector m^1 and even less information on the future period markup vector m^2 and on the consumer's intertemporal substitution matrix for regulated products sold in the present and future periods, Σ.

We will not work out the details of how welfare changes as other exogenous variables are changed; the results are similar to the results we established for the one period model of regulation. However, in the following section, we will focus on the implications of equation (219), which set the purchase price of the sunk cost capital input equal to minus a sum of partial derivatives of the period by period opex cost function derivatives with respect to the fixed capital input.

6.3 The implications of sunk costs for the measurement of capital services

Recall equation (219) in the previous section, which set P_k, the purchase price of a new unit of capital, which when installed or built cannot be varied for its useful life, equal to $-\partial c^1(y^1,w^1,k)/\partial k \geq 0$ (which is the marginal user benefit that this fixed capital stock will generate in period 1) plus $-\partial c^2(y^2,w^2,k)/\partial k \geq 0$ (which is the marginal user benefit that this capital stock will generate in period 2). These partial derivatives of the opex cost functions play a crucial role in the determination of the rate of opex technical progress as we saw in section 5.4 above; ie recall equation (157) in section 5.4. However, in section 5.4, because we assumed that the capital input was variable, we could argue that the derivative $-\partial c^1(y^1,w^1,k)/\partial k$ could be closely approximated by its observable user cost. In the present context, we cannot make the same argument due to the fixity of the capital. This fact creates problems for the measurement of technical progress.

Define the *user benefit price* for the use of one unit of the fixed capital input in period t, P_k^t, as follows:

$$(233)\ P_k^1 \equiv -\partial c^1(y^1,w^1,k)/\partial k\ ;\ P_k^2 \equiv -\partial c^2(y^2,w^2,k)/\partial k.$$

Using the above definitions, it can be seen that equation (219) becomes the following equation:

$$(234)\ P_k = P_k^1 + P_k^2$$

where $P_k > 0$ is the purchase price of one unit of the fixed capital input at the beginning of period 1. Thus, the purchase price of a unit of fixed capital, P_k, should be equal to the sum of the period by period user benefit prices, $P_k^1 + P_k^2$.

It can be seen that definitions (233) and equation (234) are equivalent to the first order necessary conditions for the regulated firm to choose k at the beginning of period 1 in a way that will minimise the firm's expected (discounted) sum of costs over time. Thus, the firm's expected intertemporal cost minimisation problem (conditional on forecasts for future expected input prices w^1 and w^2 and future demands for its products y^1 and y^2) is the following problem:

$$(235)\ \min_{k\geq 0} \{c^1(y^1,w^1,k) + c^2(y^2,w^2,k) + P_k k\}.$$

The first order necessary condition for an interior solution to (235) is:

$$(236)\ \partial c^1(y^1,w^1,k)/\partial k + \partial c^2(y^2,w^2,k)/\partial k + P_k = 0.$$

It can be seen that (236) is equivalent to (234). The important implication of all of this is that condition (219) is consistent with the regulated firm choosing a sunk cost capital input to solve an intertemporal cost minimisation problem and this problem determines the period by period user benefit terms that can be used in place of the usual user cost terms that apply to capital inputs that can be varied from period to period.[58]

Recall equation (176) in section 5.4 which allowed us to obtain an estimator for the rate of technical progress in the regulated sector. We repeat this equation for convenience:

(237) $-\partial c(y,w,k,t)/\partial t = \mu(t)^T y'(t) - w(t)^T z'(t) - P_k(t)k'(t)$

where $\mu(t) \equiv \nabla_y c(y(t),w(t),t)$ is the time t vector of regulated sector marginal costs, w(t) and z(t) are the vectors of time t variable input prices and quantities, k(t) is the quantity of fixed capital in use at time t and $P_k(t)$ is now interpreted as the period t user benefit price defined as in (233) above. The problem with expression (237) is that both $\mu(t)$ and $P_k(t)$ are not generally observable. Thus, econometric techniques or "reasonable" guesses will have to be used in order to determine these variables. It can be seen that the existence of sunk cost capital inputs makes the regulator's measurement problems more difficult as compared to the situation where all capital inputs can be freely variable.

6.4 One hoss shay depreciation versus other forms of depreciation

Recall the max-min problems defined by (197) and (198) which involved the period 1 and 2 production possibilities sets s^1 and s^2 for the regulated sector. In these max-min problems, we implicitly assumed that the capital input k chosen at the beginning of period 1 did not experience any impairment in the delivery of its services over its useful life; ie it did not deteriorate with use over time until it is retired. This type of depreciation model is known as a *one hoss shay model*[59] and we have used it up to this point because in the context of sunk capital (which is generally infrastructure capital), it seems to be a reasonable assumption for many types of sunk capital such as pipelines or power lines. However, for other types of infrastructure capital such as pumping stations, it may be more reasonable to assume that efficiency declines over time until the asset is retired and so the one hoss shay model that we have used thus far may need to be modified for these assets.

Suppose that the sunk cost capital installed at the beginning of period 1 declines in efficiency (until the asset is retired) according to the geometric depreciation rate δ where $0 < \delta < 1$ so that the efficiency of the capital input in period t is only $(1-\delta)^t$ times its efficiency of 1 in period 1. We will call this model of depreciation a *modified one hoss shay model*. In our very simple model, a new asset lasts for only 2 periods before it is retired and so in period 1, the efficiency of a newly purchased asset is 1 and then in period 2, its efficiency is $(1-\delta)$ and then in period 3, the asset is retired. Under these conditions, we can still define the opex cost

[58] Remember that our regulatory equilibrium equations had their origin in an intertemporal consumer welfare maximisation problem and these equations can be used to move the economy closer to an optimal situation. Thus, an implication of this intertemporal welfare maximisation problem is that sunk cost capital inputs should be chosen in order to minimise (discounted) costs.

[59] Solow, Tobin, von Weizsäcker and Yaari (1966; 81) used the term "one hoss shay" to describe this model of depreciation. See Diewert (2005a) (2004b) for extensive discussions and references to the literature on alternative depreciation models.

function for the regulated sector in period t, $c^t(y,w,k)$, for t = 1,2 by (183) where k is understood to be newly purchased units of capital. Thus, if k units of capital are purchased by the regulated firm at the beginning of period 1, the period 1 opex cost function is $c^1(y,w,k^1)$ where $k^1 = k$ and the period 2 opex cost function is $c^2(y,w,k^2)$ where $k^2 = (1-\delta)k$. Under these conditions, the max-min problem defined by (197) becomes the following problem:

$$(238)\ H(y^1,y^2) = \max_{u,Y\text{'s},z\text{'s},Z\text{'s},x\text{'s},X\text{'s},k} \min_{p\text{'s},w\text{'s},\lambda} \{u + p^{1T}[y^1-x^1] + w^{1T}[v^1-z^1-Z^1] + p^{2T}[y^2-x^2] + w^{2T}[v^2-z^2-Z^2] + \lambda\{P^{1T}[Y^1-X^1] + P^{2T}[Y^2-X^2] - P_k k\} : (y^1,z^1,k)\in s^1 ; (y^2,z^2,(1-\delta)k)\in s^2, t = 1,2 ; (Y^t,Z^t)\in S^t, t = 1,2 ; U(x^1,X^1,x^2,X^2) - u \geq 0 ; u \geq 0, \lambda \geq 0, Y^t \geq 0_J, z^t \geq 0_K, Z^t \geq 0_K, x^t \geq 0_N, X^t \geq 0_J, p^t \geq 0_N, w^t \geq 0_K, t = 1,2\}$$

where as before, p^t can be interpreted as a *vector of prices for the regulated products in period t*, w^t can be interpreted as a *vector of (noncapital) primary input prices for period t*, t = 1,2 and λ is a Lagrange multiplier that corresponds to the balance of trade constraint (192). Using the same arguments as were made in section 6.1, it can be shown that (238) is equivalent to the following max-min problem which now has no constraints other than nonnegativity constraints:

$$(239)\ H(y^1,y^2) = \max_{u,Y\text{'s},k} \min_{p\text{'s},w\text{'s},\lambda} \{u + w^{1T}v^1 + p^{1T}y^1 - c^1(y^1,w^1,k) + \lambda P^{1T}Y^1 - C^1(Y^1,w^1) + w^{2T}v^2 + p^{2T}y^2 - c^2(y^2,w^2,(1-\delta)k) + \lambda P^{2T}Y^2 - C^2(Y^2,w^2) - \lambda P_k k - e(u,p^1,\lambda P^1,p^2,\lambda P^2) : u \geq 0, \lambda \geq 0 ; Y^t \geq 0_J, p^t \geq 0_N, w^t \geq 0_K, t = 1,2\}.$$

We can now follow the rest of the analysis in section 6.1 and derive counterparts to equations (215)-(220), which characterised a *regulatory equilibrium*. The new equations are:

$$(240)\ m^t \equiv p^t - \nabla_y c^t(y^t,w^t,(1-\delta)^{t-1}k) ; \qquad t = 1,2 ;$$

$$(241)\ P^t = \nabla_Y C^t(Y^t,w^t) ; \qquad t = 1,2 ;$$

$$(242)\ v^t = \nabla_w c^t(y^t,w^t,(1-\delta)^{t-1}k) + \nabla_w C^t(Y^t,w^t) ; \qquad t = 1,2 ;$$

$$(243)\ y^t = \nabla_{p^t} e(u,p^1,P^1,p^2,P^2) ; \qquad t = 1,2 ;$$

$$(244)\ P_k = -\partial c^1(y^1,w^1,k)/\partial k - (1-\delta)\partial c^2(y^2,w^2,(1-\delta)k)/\partial k ;$$

$$(245)\ 0 = P^{1T}[Y^1 - \nabla_{p^1} e(u,p^1,P^1,p^2,P^2)] + P^{2T}[Y^2 - \nabla_{p^2} e(u,p^1,P^1,p^2,P^2)] - P_k k .$$

Comparing the above equations with the corresponding one hoss shay equilibrium equations (215)-(220) shows that not much has changed except equation (244) is different from its counterpart in section 6.1, equation (219), which we later rewrote as equation (234) in section 6.3.[60] We will now obtain counterparts to equations (233) and (234) in section 6.3. Define the capital stock in use in constant efficiency units in period t as k^t for t = 1,2. If the regulated firm purchases k units of the sunk cost capital at the beginning of period 1 and there is geometric depreciation, then we have the following relationships between k, k^1 and k^2:

$$(246)\ k^1 = k ; k^2 = (1-\delta)k.$$

[60] The new equations are exactly the same as the old ones if the geometric depreciation rate δ equals 0.

Define the *user benefit price* for the use of one (new) unit of the fixed capital input in period t, P_k^t, as follows:

(247) $P_k^1 \equiv -\partial c^1(y^1,w^1,k^1)/\partial k$; $P_k^2 \equiv -\partial c^2(y^2,w^2,k^2)/\partial k$.

Definitions (247) are the modified one hoss shay depreciation counterparts to definitions (233) in the previous section where we assumed one hoss shay depreciation. Using the above definitions, it can be seen that equation (244) becomes the following equation:

(248) $P_k = P_k^1 + (1-\delta)P_k^2$

where as usual, $P_k > 0$ is the purchase price of one unit of the fixed capital input at the beginning of period 1. Equation (248) is the counterpart to equation (234) in the previous subsection. Note that $P_k k$ is the total value of purchases of sunk cost capital at the beginning of period 1. Using definitions (246) and multiplying both sides of (248) through by k, we obtain the following decomposition of the total asset cost $P_k k$ into intertemporal components:

(249) $P_k k = P_k^1 k^1 + P_k^2 k^2$.

Thus, the beginning of period 1 cost of the sunk cost asset, $P_k k$, is equal to the sum of the period by period user benefit prices, P_k^t, times period t capital input in constant efficiency units, k^t, for t = 1,2. Note that (249) is equivalent to (244).[61]

Up to this point, there has been no explicit mention of interest rates and the role of discounting in our intertemporal model. In the following section, we will introduce discounted prices into our model.

6.5 Characterising an intertemporal regulatory equilibrium using spot prices and discounting

The notation that we have been using for prices in our intertemporal model did not recognise the role of interest rates and discounting explicitly. Our reason for not explicitly introducing interest rates at an earlier stage of our analysis is that the resulting notation becomes rather cumbersome as we shall see. However, the time has come to recognise the important role played by the cost of capital.

We will assume that the economy in question is a small open economy and the rest of the world determines the cost of capital faced by the regulated firm. Thus, we assume that the exogenous foreign lending and borrowing one period interest rate at the beginning of period t is r_t where $1+r_t > 0$ for t = 1,2. The counterpart to the intertemporal balance of trade equation (192) using discounted prices can be rewritten as follows:

(250) $-P_k k + (1+r_1)^{-1}P^{1T}[Y^1-X^1] + (1+r_1)^{-1}(1+r_2)^{-1}P^{2T}[Y^2-X^2] \geq 0$.

In equation (250), all values are discounted (using the one period sequence of nominal interest rates r_t) to the beginning of period 1. At the beginning of period 1, the regulated firm imported k units of sunk cost capital at the price P_k and the resulting beginning of period 1

[61] We have spelled out all of this algebra in great detail because in order to measure productivity growth or technical change, it is important to measure capital services input in the correct quantity units. Thus, in the one hoss shay model, the capital services input remains constant until the asset is retired whereas in the present modified one hoss shay model, capital services input declines over time until the asset is finally retired.

value, $P_k k$, is given a minus sign since this is an import. Recall that period 1 net exports (excluding the capital import) are equal to $Y^1 - X^1$ and period 2 net exports are equal to $Y^2 - X^2$. We assume that all export and import accounts are settled at the end of each period[62] so that the period 1 value of net exports (excluding the imported sunk capital), $P^{1T}[Y^1-X^1]$, is divided by 1 plus the period 1 interest rate, r_1. Similarly, we assume that the expected period 1 world vector of spot prices for internationally traded goods and services is P^2 and that the export and import accounts are settled at the end of period 2 so that the expected period 2 value of net exports, $P^{2T}[Y^2-X^2]$, discounted to the beginning of period 1 is $P^{2T}[Y^2-X^2]/(1+r_1)(1+r_2)$. Since the price vectors P^1 and P^2 and the two interest rates r_1 and r_2 are all assumed to be exogenous in our intertemporal model, it can be seen that (192) and (250) are equivalent equations; the only difference between the equations is the interpretation that we place on P^1 and P^2.

With equation (250) replacing (192), we can set up a counterpart to the consumer's intertemporal utility maximisation problem (196) and then we can use the Karlin Uzawa Saddle Point Theorem to derive a counterpart to (197). However, when we introduce the dual variables p^t and w^t, we now replace these vectors of variables by their discounted counterparts, $(1+r_1)^{-1}p^1$, $(1+r_1)^{-1}(1+r_2)^{-1}p^2$, $(1+r_1)^{-1}w^1$ and $(1+r_1)^{-1}(1+r_2)^{-1}w^2$. Thus, the counterpart to (197) becomes:

$$\begin{aligned}(251)\ H(y^1,y^2) = &\max_{u,Y's,z's,Z's,x's,X's,k} \min_{p's,w's,\lambda} \{u + (1+r_1)^{-1}p^{1T}[y^1-x^1] \\ &+ (1+r_1)^{-1}w^{1T}[v^1-z^1-Z^1] + (1+r_1)^{-1}(1+r_2)^{-1}p^{2T}[y^2-x^2] + (1+r_1)^{-1}(1+r_2)^{-1}w^{2T}[v^2-z^2-Z^2] \\ &+ \lambda\{- P_k k + (1+r_1)^{-1}P^{1T}[Y^1-X^1] + (1+r_1)^{-1}(1+r_2)^{-1}P^{2T}[Y^2-X^2]\} : (y^t,z^t,k)\in s^t,\ t = 1,2\ ; \\ &(Y^t,Z^t)\in S^t,\ t = 1,2\ ;\ U(x^1,X^1,x^2,X^2) - u \geq 0\ ;\ u \geq 0,\ \lambda \geq 0,\ Y^t \geq 0_J,\ z^t \geq 0_K,\ Z^t \geq 0_K, \\ &x^t \geq 0_N,\ X^t \geq 0_J,\ p^t \geq 0_N,\ w^t \geq 0_K,\ t = 1,2\}.\end{aligned}$$

In (197), the vectors p^t and w^t were interpreted as *discounted* future expected price vectors for regulated products and for primary inputs in period t whereas in (251), p^t and w^t are now to be interpreted as future expected *spot* price vectors for regulated products and for primary inputs in period t.

The rest of the analysis in section 6.1 proceeds in much the same manner. The counterparts to equations (215)-(220) which characterise a regulatory equilibrium using explicit discounting are the following equations:

$$(252)\ m^t \equiv p^t - \nabla_y c^t(y^t,w^t,k)\ ; \qquad t = 1,2\ ;$$

$$(253)\ P^t = \nabla_Y C^t(Y^t,w^t)\ ; \qquad t = 1,2\ ;$$

$$(254)\ v^t = \nabla_w c^t(y^t,w^t,k) + \nabla_w C^t(Y^t,w^t)\ ; \qquad t = 1,2\ ;$$

$$(255)\ y^t = \nabla_{p^t} e(u,(1+r_1)^{-1}p^1,(1+r_1)^{-1}P^1,(1+r_1)^{-1}(1+r_2)^{-1}p^2,(1+r_1)^{-1}(1+r_2)^{-1}P^2)\ ; \qquad t = 1,2\ ;$$

$$(256)\ P_k = - \partial c^1(y^1,(1+r_1)^{-1}w^1,k)/\partial k - \partial c^2(y^2,(1+r_1)^{-1}(1+r_2)^{-1}w^2,k)/\partial k\ ;$$

$$(257)\ 0 = (1+r_1)^{-1}P^{1T}[Y^1 - X^1] + (1+r_1)^{-1}(1+r_2)^{-1}P^{2T}[Y^2- X^2] - P_k k$$

[62] Note that we are assuming that all period one flow variables are realised at the end of each period. This is consistent with accounting treatments of assets at the beginning and end of the accounting period and cash flows that occur during the period; see Peasnell (1981; 56).

where $X^t \equiv \nabla_{p^t} e(u, (1+r_1)^{-1}p^1,(1+r_1)^{-1}P^1,(1+r_1)^{-1}(1+r_2)^{-1}p^2,(1+r_1)^{-1}(1+r_2)^{-1}P^2)]$ for t = 1,2. It can now be seen why we did not introduce explicit discounting earlier in our exposition of the intertemporal model.

The important equation in the above equations is (256), the first order necessary condition for the choice of the sunk cost capital input to minimise discounted cost. Using the fact that the regulatory opex cost functions are homogeneous of degree one in the input cost variables, (256) can be rewritten as follows:

(258) $P_k = -(1+r_1)^{-1}\partial c^1(y^1,w^1,k)/\partial k - (1+r_1)^{-1}(1+r_2)^{-1}\partial c^2(y^2,w^2,k)/\partial k.$

Now define the period by period *user benefit prices* for the use of one unit of the fixed capital input in period t, P_k^t, as in section 6.3 as follows:

(259) $P_k^1 \equiv -\partial c^1(y^1,w^1,k)/\partial k$; $P_k^2 \equiv -\partial c^2(y^2,w^2,k)/\partial k.$

Using definitions (259), it can be seen that equation (258) becomes the following equation:

(260) $P_k = (1+r_1)^{-1}P_k^1 + (1+r_1)^{-1}(1+r_2)^{-1}P_k^2.$

Thus, the purchase price of a unit of fixed capital, P_k, should be equal to the discounted sum of the period by period user benefit prices, P_k^t.

It can be seen that definitions (259) and equation (260) are equivalent to the first order necessary conditions for the regulated firm to choose k at the beginning of period 1 in a way that will minimise the firm's discounted expected sum of costs over time. Thus, the firm's expected intertemporal cost minimisation problem (conditional on forecasts for future expected input prices w^1 and w^2 and future demands for its products y^1 and y^2) is the following problem:

(261) $\min_{k\geq 0} \{(1+r_1)^{-1}c^1(y^1,w^1,k) + (1+r_1)^{-1}(1+r_2)^{-1}c^2(y^2,w^2,k) + P_k k\}.$

The first order necessary condition for an interior solution to (261) is (258).

Condition (256) and the equivalent conditions (258) and (260) are consistent with the regulated firm choosing a sunk cost capital input to solve an intertemporal cost minimisation problem and this problem determines the period by period user benefit terms that can be used in place of the usual user cost terms that apply to capital inputs that can be varied from period to period. We will explore additional implications of condition (260) in the Report.

The counterpart to (260) when the modified one hoss shay depreciation model is used is the following equation:

(262) $P_k k = (1+r_1)^{-1}P_k^1 k^1 + (1+r_1)^{-1}(1+r_2)^{-1}P_k^2 k^2$

where the P_k^t are defined by (259) and the k^t are defined by (246); ie $k^1 \equiv k$ and $k^2 \equiv (1-\delta)k$.

7 PRACTICAL REGULATION USING ONLY INFORMATION ON THE REGULATED FIRM

The theory of optimal regulation that has been presented in the previous sections is useful because it shows what type of information is required in order for a regulator to induce changes in the regulated firm that will improve consumer welfare. Unfortunately, even in the case where all capital inputs can be varied from period to period, the informational requirements are very high: reasonably accurate information on marginal costs by product is required as well as information on household substitution matrices. The required information is not likely to be fully available to the regulator. When we have sunk cost capital inputs, the informational requirements are even higher. In this case, information on intertemporal consumer substitution matrices is required as well as information on the derivatives of the sequence of opex cost functions with respect to the sunk cost assets. Accurate information on these derivatives is notoriously difficult to obtain (and it is not easy to estimate intertemporal consumer substitution matrices either).

What then can the regulator do? It seems clear that regulators will have to rely on approximate methods of regulation in order to improve welfare. Below, we will sketch out the main features of a price cap approach to regulation. For simplicity, we will follow the continuous time approach to regulation that was pioneered by Denny, Fuss and Waverman (1981) but the arguments presented below can be converted into discrete time index number formulae using the techniques explained in Caves, Christensen and Diewert (1982), Diewert and Morrison (1986), Diewert and Fox (2000) and Lawrence, Diewert and Fox (2006).

The approach we will take here is the following one: we will obtain an expression for the rate of change of the regulated firm's profits $\Pi'(t)$ which will have a term involving $p'(t)$, the vector of derivatives of the period t prices for regulated outputs. We will move all of these prices in a proportional manner and determine this rate of proportional movement by setting $\Pi'(t)$ equal to a constant.[63] This determines the price cap for the following period.

We use the same notation that was developed in section 5.4 above (with a few exceptions which will be explained later). Thus, let $c(y,w,k,t)$ denote the opex cost function for the regulated cost function at time t and let y, w and k be differentiable functions of time, $y(t)$, $w(t)$ and $k(t)$. We now let $k(t)$ be an M dimensional vector of sunk cost type capital inputs, whereas in previous sections, k was a scalar. Thus, opex cost, regarded as a function of t, is defined as follows:

(263) $C_z(t) = c(y(t),w(t),k(t),t)$.

The *vector of time t marginal costs,* $\mu(t)$, is defined as in section 5:

(264) $\mu(t) \equiv \nabla_y c(y(t),w(t),k(t),t)$

The vector of *time t marginal user benefits* $\pi(t)$ for the time t capital stocks $k(t)$ is defined as

[63] If the regulated firm's profits at time t, $\Pi(t)$, are close to 0, then this constant can be chosen to be a small positive number. If $\Pi(t)$ is quite negative, then the constant can be chosen to be a larger positive number. Finally, if $\Pi(t)$ is quite positive, then the constant can be chosen to be a negative number; ie the regulator could attempt to reduce the monopoly profits that the regulated firm currently enjoys.

follows:

(265) $\pi(t) \equiv - \nabla_k c(y(t),w(t),k(t),t) \geq 0_M$.

The *logarithmic time t opex rate of technical progress*, $\tau(t)$, is defined as the negative of the partial derivative of the log of the opex cost function with respect to t:[64]

(266) $\tau(t) \equiv - [\partial c(y(t),w(t),k(t),t)/\partial t]/c(y(t),w(t),k(t),t)$

Recall equation (176) in section 5.4 which explained how an estimator for $\tau(t)$ could be obtained. Using the notation in the present section, the counterpart to (176) is:

(267) $\tau(t)C_z(t) = \mu(t)^T y'(t) - w(t)^T z'(t) - \pi(t)\cdot k'(t)$.

The right hand side of equation (267) is more difficult to estimate than the right hand side of (176) since in the case where capital can be freely varied from period to period, we can use observable user costs to estimate the price of variable capital services but in the case of sunk cost capital inputs, there are no reliable user costs; instead we have endogenous user benefits defined by (265), which must be estimated using econometric techniques or accounting or engineering cost allocation methods.

As in section 5.4, the vector of opex input demands at time t is z(t) defined via Shephard's Lemma as usual:

(268) $z(t) \equiv \nabla_w c(y(t),w(t),k(t),t)$.

Differentiating period t opex cost $C_z(t)$ defined by (263) with respect to time t and using (264)-(268), we obtain the following expression for $C'(t)$:

(269) $C_z'(t) = \mu(t)^T y'(t) + z(t)^T w'(t) - \pi(t)\cdot k'(t) - \tau(t)C_z(t)$.

Thus, the rate of change of opex cost at time t, $C_z'(t)$, is equal to an opex marginal cost weighted sum of rates of change of outputs, $\mu(t)^T y'(t)$, plus an opex input weighted sum of rates of change of opex input prices, $z(t)^T w'(t)$, less a user benefit weighted sum of rates of change of sunk cost capital inputs, $\pi(t)\cdot k'(t)$, less the opex rate of technical change, $\tau(t)C_z(t)$.

The revenues earned by the regulated firm at time t are R(t) defined as the sum of the prices $p_n(t)$ of the regulated products times the corresponding quantities sold $y_n(t)$:

(270) $R(t) \equiv p(t)\cdot y(t)$.

Thus, the rate of change of time t revenues is:

(271) $R'(t) = p'(t)\cdot y(t) + p(t)\cdot y'(t)$.

We think of the regulator as controlling the prices of regulated products, p(t). The regulator can also specify how sunk cost capital investments are amortised over time. In section 10, we will consider various amortisation schemes that could be used by the regulator that will maintain the financial capital of the regulated firm. In the present section, we will take as given a vector of time t amortisation amounts, $P_k(t)$, that the regulated firm can use as time t

[64] In the interests of simplicity, we have not considered the effects of changes in operating environment variables in our analysis of the regulated sector. For example, it may be desirable to adjust the regulated firm's price cap for the adverse effects on opex cost of hurricane damage or other unusual events. This adjustment can be made explicit but in the present paper, the effects of past changes in operating environment variables will be part of technical progress.

charges to its time t net income. Thus, the *time t amount of amortisation charges for the fixed capital stock components* in use at time t is $C_k(t)$ defined as follows:

$$(272)\ C_k(t) \equiv \textstyle\sum_{m=1}^{M} P_{km}(t)k_m(t) = P_k(t)\cdot k(t)$$

where $P_k(t) \equiv [P_{k1}(t),...,P_{kM}(t)]$ is a vector of time t allowable amortisation charges and $k(t) \equiv [k_1(t),...,k_M(t)]$ is the corresponding vector of constant quality sunk cost capital inputs. The rate of change of the capital stock amortisation charges is:

$$(273)\ C_k'(t) = P_k'(t)\cdot k(t) + P_k(t)\cdot k'(t).$$

The *time t profits of the regulated firm* are $\Pi(t)$ defined as time t revenues less time t operating costs less time t allowable amortisation expenses for fixed capital stock components:[65]

$$(274)\ \Pi(t) \equiv R(t) - C_z(t) - C_k(t).$$

The rate of change of profits, $\Pi'(t)$, can be obtained by differentiating (274). Using (269), (271) and (273), we obtain the following expression for $\Pi'(t)$:

$$(275)\ \Pi'(t) = p'(t)\cdot y(t) - w'(t)\cdot z(t) - P_k'(t)\cdot k(t) + \tau(t)C_z(t)$$
$$+ [p(t)-\mu(t)]\cdot y'(t) - [P_k(t)-\pi(t)]\cdot k'(t).$$

The first three terms on the right hand side of (275) can be converted into Divisia like indexes of output price change, $p'(t)\cdot y(t)$, minus a Divisia like index of opex input price change, $w'(t)\cdot z(t)$, minus a Divisia like index of amortisation price change, $P_k'(t)\cdot k(t)$. The last three terms are difficult to measure terms: the rate of opex technical change, $\tau(t)C_z(t)$, plus a weighted sum of output quantity changes, $y'(t)$, where the weights are the difference between the market prices for regulated outputs, $p(t)$, less the (unobserved) marginal cost weights, $\mu(t)$, minus a weighted sum of sunk cost capital quantity changes, $k'(t)$, where the weights are the difference between the allowed amortisation charges, $P_k(t)$, less the (unobserved) marginal benefit charges, $\pi(t)$, defined by (265). However, note that if $p(t)$ equals $\mu(t)$ and $P_k(t)$ equals $\pi(t)$ (which is an implication of first best optimal regulation), then the last two terms on the right hand side of (275) vanish and it also becomes straightforward to measure $\tau(t)$ (at least on an ex post basis) using equation (275).

We will now assume that the regulator will force some proportional price change in the prices of regulated outputs at time t; ie we assume that the regulator changes all regulated prices according to the following formula:

$$(276)\ p'(t) = \alpha'(t)p(t)$$

where $\alpha(t)$ is set equal to 1. We also assume that the regulator has a target for the rate of change in profits for the regulated firm at time t equal to the rate $\beta R(t)$ say, where $R(t)$ is time t revenue. Thus, the regulator would like to determine a rate of change in regulated prices such that

$$(277)\ \Pi'(t) = \beta R(t).$$

[65] As is explained in section 10, the user charges or allowable amortisation amounts $P_k(t)$ are sufficiently generous so that the regulated firm can cover its cost of capital over time. Thus, the profits $\Pi(t)$ are *pure profits* or profits that are in excess of what is required to maintain the long run solvency of the regulated firm.

Substituting (275) and (276) into (277) gives us the following solution for the allowable rate of increase in regulated prices $\alpha'(t)$:[66]

$$(278)\ \alpha'(t) = \beta + \{w'(t)\cdot z(t) + P_k'(t)\cdot k(t) - \tau(t)C_z(t) - [p(t)-\mu(t)]\cdot y'(t) + [P_k(t)-\pi(t)]\cdot k'(t)\}/R(t).$$

Equation (278) is our desired approximate price cap formula. Suppose that the last two terms on the right hand side of (278) could be neglected.[67] Further suppose that the profits of the regulated firm at time t are close to 0 and the regulator wants to keep profits close to 0 in the future. Under these conditions, the regulator would set β equal to 0 and (278) would simplify to:

$$(279)\ \alpha'(t) = \{w'(t)\cdot z(t) + P_k'(t)\cdot k(t) - \tau(t)C_z(t)\}/R(t).$$

We can further simplify (279) if we define the *Divisia input price indexes* for variable inputs and for sunk cost capital inputs as follows:

Define the *Divisia index of opex input price growth* at time t, $w_D'(t)$, and the *Divisia index of amortisation prices for sunk capital stocks* at time t, $P_{kD}'(t)$, as follows:

$$(280)\ w_D'(t) \equiv \textstyle\sum_{k=1}^{K} [w_k(t)z_k(t)/w(t)^T z(t)][w_k'(t)/w_k(t)]\ ;$$

$$(281)\ P_{kD}'(t) \equiv \textstyle\sum_{m=1}^{M} [P_{km}(t)k_n(t)/P_k(t)^T k(t)][P_{km}'(t)/P_{km}(t)]\ .$$

Substituting (280) and (281) into (279) gives us the following formula for the rate of increase in regulated prices (the X factor in price cap regulation):

$$(282)\ \alpha'(t) = [C_z(t)/R(t)]w_D'(t) + [C_k(t)/R(t)]P_{kD}'(t) - [C_z(t)/R(t)]\tau(t).$$

Thus, if we can neglect the last two terms in (278), (282) tells us that the allowable rate of price increase for all regulated products, $\alpha'(t)$, should be set equal to the ratio of opex costs to revenues, $C(t)/R(t)$, times the Divisia rate of increase in opex prices, $w_D'(t)$, plus the ratio of time t allowable amortisation charges to revenues, $C_k(t)/R(t)$, times the Divisia rate of increase in these amortisation prices, $P_{kD}'(t)$, less the ratio of opex costs to revenues, $C(t)/R(t)$, times the opex logarithmic rate of technical progress, $\tau(t)$. Formula (282) for the price cap is simple enough to be implementable provided that the regulator can make forecasts for the overall rate of increase in variable input prices, $w_D'(t)$, and for the anticipated rate of technical progress, $\tau(t)$.[68]

---

[66] Note that $\tau(t)$ and the terms involving input price changes in the price cap formula should be *anticipated* rates of price change and technical progress and not the firm's *actual* rates. Bernstein and Sappington (1999; 9) explain why: "To provide incentives for productivity gains, price cap regulation should require the regulated firm's prices to vary with projected, not actual, changes in the firm's productivity and input prices. Under such a policy, the firm will gain financially if it achieves productivity growth that exceeds expectations and will suffer financially if its productivity growth falls short of expectations. Consequently, the firm will face strong incentives to operate diligently and secure productivity gains."

[67] With increasing returns to scale in the regulated sector, we would expect the components of $p(t)-\mu(t)$ to be predominantly positive and with growth in the economy, we would also expect the components of $y'(t)$ to be positive and thus the term $-[p(t)-\mu(t)]\cdot y'(t)$ is likely to be negative. The last term is likely to be small since the vector of fixed capital stock components $k(t)$ is likely to remain roughly constant and hence $k'(t)$ is likely to be small.

[68] The regulator will be able to construct a discrete time approximation to the Divisia index of allowable amortisation charges $P_D(t)$, since the regulator will determine these allowable charges. Typically, forecasts for

However, the price cap formula (282) is only a very rough approximation to the more accurate price cap formula defined by (278). Unfortunately, the endogenous variables $y'(t)$ and $k'(t)$ appear in (278)[69] along with the not easily observed weighting vectors $\mu(t)$ and $\pi(t)$.

Finally, note that the price cap exercise does not directly address the adjustment of output prices for regulated products towards their marginal costs. As we have seen in our theoretical work in the previous sections, there can be substantial welfare gains in moving regulated prices towards their marginal costs.

$\tau(t)$ are made on the basis of past rates of technical progress in the industry. Unfortunately, there is no guarantee that future rates of technical progress will mirror past rates.

[69] Strictly speaking, the vector of rates of increase in variable inputs, $w'(t)$, is also endogenous but usually the regulated sector is not big enough to really affect the rate of increase in these prices. Put another way, usually, it is not too difficult to forecast $w'(t)$.

8 PRACTICAL REGULATION OF A SINGLE FIRM USING TFP GROWTH AND ECONOMY WIDE INFORMATION

The material in the previous section does not reflect the practice of current price cap regulation in many jurisdictions which rely on what has become known as CPI minus X price cap regulation.[70] In this section, we will rework the analysis presented in the previous section into the CPI minus X factor framework using the continuous time approach that was explained in Bernstein and Sappington (1999). It should be noted that we continue to assume that only a single firm is being regulated. Problems arising from regulating multiple firms in an industry are addressed later in the section and in the following section.

We start off by defining the *Divisia indexes of output growth* $y_D'(t)$, of *variable input growth* $z_D'(t)$ and of (sunk cost) *capital services growth*, $k_D'(t)$ at time t:

(283) $y_D'(t) \equiv p(t)\cdot y'(t)/p(t)\cdot y(t) = p(t)\cdot y'(t)/R(t)$;

(284) $z_D'(t) \equiv w(t)\cdot z'(t)/w(t)\cdot z(t) = w(t)\cdot z'(t)/C_z(t)$;

(285) $k_D'(t) \equiv P_k(t)\cdot k'(t)/P_k(t)\cdot k(t) = P_k(t)\cdot k'(t)/C_k(t)$.

Note that the price weights p(t) for the output derivatives $y'(t)$ are the observable market prices for the outputs at time t (and not the unobserved marginal cost weights $\mu(t)$ which reflect the cost of producing an extra unit of each regulated output) and the price weights $P_k(t)$ for the capital services input derivatives $k'(t)$ are the observable amortisation charges for time t that are determined by the regulator (and not the unobserved reductions in opex cost $\pi(t)$ due to additional marginal units of capital at time t).

The time t *Total Factor Productivity* (TFP) *growth* of the regulated firm, $T'(t)$, is traditionally defined as the Divisia index of output growth minus the Divisia index of input growth; see Bernstein and Sappington (1999; 9). The traditional theory does not deal with the sunk cost problem and TFP growth for the regulated firm can be defined as the Divisia index of output growth less a share weighted sum of the Divisia indexes of variable input and sunk cost capital services input indexes using (observable) amortisation prices as weights:

(286) $T'(t) \equiv y_D'(t) - s_z(t)z_D'(t) - s_k(t)k_D'(t)$

where the time t *input cost share weights* $s_z(t)$ and $s_k(t)$ and total cost C(t) are defined as follows:

(287) $s_z(t) \equiv C_z(t)/C(t)$; $s_k(t) \equiv C_k(t)/C(t)$; $C(t) \equiv C_z(t) + C_k(t)$.

Recall that time t revenue for the regulated firm is R(t) equal to $p(t)\cdot y(t)$ and time t total cost (using amortisation prices for capital services) is $C_z(t)$ plus $C_k(t)$ equal to $w(t)\cdot z(t)$ plus $P_k(t)\cdot k(t)$. Time t pure profits are $\Pi(t)$ equal to R(t) less $C_z(t)$ less $C_k(t)$. Thus, the time t derivative of profits is equal to the following expression:

(288) $\Pi'(t) = p'(t)\cdot y(t) - w'(t)\cdot z(t) - P_k'(t)\cdot k(t) + p(t)\cdot y'(t) - w(t)\cdot z'(t) - P_k(t)\cdot k'(t)$.

[70] For an exposition of this theory in continuous time without the complications implied by the existence of sunk costs, see Bernstein and Sappington (1999). For an exposition of the theory in discrete time, see Lawrence (2003; 3-8) and Lawrence and Diewert (2006).

Recall our earlier expression for $\Pi'(t)$ given by (275). Equating these two expressions leads to the following equation:

$$(289)\ \tau(t)C_z(t) + [p(t)-\mu(t)]\cdot y'(t) - [P_k(t)-\pi(t)]\cdot k'(t) = p(t)\cdot y'(t) - w(t)\cdot z'(t) - P_k(t)\cdot k'(t)$$
$$= R(t)y_D'(t) - C_z(t)z_D'(t) - C_k(t)k_D'(t) \qquad \text{using (283)-(285)}$$
$$= C(t)\{[R(t)/C(t)]y_D'(t) - s_z(t)z_D'(t) - s_k(t)k_D'(t)\}$$
$$= C(t)\{[\Pi(t)/C(t)]y_D'(t) + T'(t)\} \qquad \text{using (286).}$$

Now substitute (289) into (275) and divide the resulting equation by C(t). After some rearrangement, the resulting equation becomes:

$$(290)\ p'(t)\cdot y(t)/C(t) = s_z(t)w_D'(t) + s_k(t)P_{kD}'(t) - T'(t) + [\Pi'(t)/C(t)] - [\Pi(t)/C(t)]y_D'(t)$$

where the *Divisia index of opex input price growth*, $w_D'(t)$, and the *Divisia index of allowable amortisation cost growth*, $P_{kD}'(t)$, are defined as follows:

$$(291)\ w_D'(t) \equiv w'(t)\cdot z(t)/w(t)\cdot z(t) = w'(t)\cdot z(t)/C_z(t)\ ;$$

$$(292)\ P_{kD}'(t) \equiv P_k'(t)\cdot k(t)/P_k(t)\cdot k(t) = P_k'(t)\cdot k(t)/C_k(t).$$

Equation (290) is the sunk cost generalisation of equation (2.5) in Bernstein and Sappington (1999; 9).[71] If we force all regulated prices to change by the same proportion so that we can substitute (276) into (290), we obtain the following expression for the rate of change in capped prices, $\alpha'(t)$:

$$(293)\ \alpha'(t) = [C(t)/R(t)]\{s_z(t)w_D'(t) + s_k(t)P_{kD}'(t) - T'(t)\} + [\Pi'(t)/R(t)] - [\Pi(t)/R(t)]y_D'(t)$$

Bernstein and Sappington (1999; 9) explain the intuition behind their (somewhat simpler version of) equation (293) and they note that if profits are zero at time t and the regulator wishes to keep profits at this zero level, so that both $\Pi(t)$ and $\Pi'(t)$ are zero (and C(t) also equals R(t) when $\Pi(t)$ is zero), then (293) simplifies to:

$$(294)\ \alpha'(t) = s_z(t)w_D'(t) + s_k(t)P_{kD}'(t) - T'(t).$$

Thus, under these conditions, the proportional rate of increase in all regulated prices should be set equal to a cost share weighted average of the rates of growth of the Divisia indexes of opex prices and amortisation charges, $s_z(t)w_D'(t) + s_k(t)P_{kD}'(t)$, less the anticipated rate of growth of TFP, $T'(t)$.[72]

It should be noted that in the general case where $\Pi(t)$ and $\Pi'(t)$ are not necessarily equal to zero, TFP growth is not equal to technical progress. From (289), we have the following expression for the determinants of TFP growth:

$$(295)\ T'(t) = \tau(t)s_z(t) + [p(t)-\mu(t)]\cdot y'(t)/C(t) - [P_k(t)-\pi(t)]\cdot k'(t)/C(t) - [\Pi(t)/C(t)]y_D'(t).$$

Recall definition (266), which defined $\tau(t)$ as the logarithmic time t opex rate of technical

[71] We divided both sides of (275) by C(t) to obtain our equation (290) whereas Bernstein and Sappington divided by R(t) in their derivation of equation (2.5). However, the right hand side of (293) corresponds exactly to their equation (2.5).

[72] Note that if $T'(t)$ is positive, then this *direct method of price cap regulation* is definitely not equivalent to old style rate of return regulation (which would use formula (294) with $T'(t)$ set equal to 0). The type of direct price cap regulation (without the sunk cost complications) defined by (294) dates back to Caves and Christensen (1982).

progress. Thus, the first term on the right hand side of (295) shows that technical progress is definitely a contributor to the rate of TFP growth, $T'(t)$. But the remaining terms on the right hand side of (295) show that TFP growth encompasses a lot more than technical progress. The term $[p(t)-\mu(t)]\cdot y'(t)/C(t)$ depends on the deviations of the selling prices $p(t)$ from the corresponding marginal costs $\mu(t)$ and this vector of deviations interacts with the vector of output growth rates $y'(t)$. It will be difficult to project past contributions to TFP growth that are due to this term into the future. A similar comment applies to the term $-[P_k(t)-\pi(t)]\cdot k'(t)/C(t)$. Thus, measured TFP growth is a rather complex concept in terms of its explanatory factors. *Since the regulator controls* $p(t)$ (*the vector of regulated prices*), $P_k(t)$ (*the vector of regulator approved amortisation charges for sunk capital stock components*) *and* $\Pi(t)$ (*the profits of the regulated firm that are in excess of the regulated firm's cost of capital*), *measured TFP growth will not be a "pure" measure of technical progress; it will be a blend of technical progress and improvements in managerial efficiency and other factors which are heavily influenced by the regulator*.

In the case of a single firm, the regulator can look at past TFP growth for that firm and make a judgement about whether it can be sustained, and then the regulator can set an appropriate price cap. The factors beyond the single firm's control (related to differences between marginal costs and revenue weights and differences between opex cost function derivatives and allowable amortisation charges in equation (295) which relates TFP growth for a single firm to technical progress) are likely to remain relatively constant for a single firm but if they are not constant, then the regulator can make adjustments to the price cap to take this into account.[73]

However, if there are multiple firms being regulated, we may need to move away from common TFP targets to common technical progress targets so as to not disadvantage firms in the group who have unfavourable exogenous factors. Asking a single regulated firm to make efficiency improvements according to past rates of growth of measured TFP for a group of firms may not be appropriate since the last three components of TFP growth which appear on the right hand side of (295) are actually beyond the control of each individual regulated firm and hence some firms will obtain an inappropriate advantage over other firms in the group, due to relatively favourable uncontrollable factors. In this case it will be necessary, among other things, to move away from the use of a common TFP growth rate target to a common rate of technical progress in the price cap formula, while also taking into account the other explanatory factors on the right hand side of (295) that the regulated firm cannot control.

As we noted earlier, a problem with the types of "practical" regulatory schemes discussed

[73] If a single regulated firm agrees to this type of price cap regulation, this makes the regulator's task much simpler: the regulator need only pick a positive rate of TFP growth, $T'(t)$, that is acceptable to the regulated firm and find an appropriate measure of opex input price inflation, $w_D'(t)$, and the price cap can be implemented using a discretisation of (294). Note that the index of changes in amortisation charges, $P_{kD}'(t)$, is in principle controlled by the regulator. Using this simple direct method for determining price caps will give benefits to consumers that will be bigger than what would be obtained using simple rate of return regulation (but will be smaller than the benefits that could be obtained by also trying to align prices to their marginal costs). However, it must be emphasised that this type of price cap regulation will only be appropriate if there is only a single firm being regulated, since in this case, the factors beyond the firm's control will be roughly constant over time and hence having a TFP based price cap will not disadvantage a single firm compared to its peers (because there are no peers).

here is that the structure of selling prices is frozen by this type of regulation. Freezing the structure of output prices (except for a scalar factor) will generally not be optimal from the viewpoint of maximising consumer welfare.[74] As we have seen in our earlier analysis, in order to improve consumer welfare, the regulator should try to reduce the gaps between selling prices and marginal costs.

Note that if the price vector $p(t)$ equals the corresponding marginal cost vector at time t, $\mu(t)$, and if the vector of allowable charges for sunk capital components $P_k(t)$ equals the corresponding vector of marginal opex cost reductions due $\pi(t)$ and if profits $\Pi(t)$ are zero, then (295) simplifies into the following relationship between TFP growth $T'(t)$ and logarithmic opex technical progress $\tau(t)$:

$$(296)\ T'(t) = \tau(t)s_z(t)$$

where $s_z(t)$ is the share of operating costs in total costs at time t, $C_z(t)/C(t)$. Thus, assuming that $\tau(t)$ is positive, (296) tells us that the rate of opex technical change will be reduced by the factor $s_z(t)$ in order to obtain the rate of TFP growth. This is the type of relationship that was first noticed by Domar (1961) who noted that gross output rates of TFP growth were lower than value added rates of TFP growth. This phenomenon can be explained by the fact that the input measure for TFP in the gross output framework is bigger than the input measure in the value added framework.[75]

We now follow Bernstein and Sappington (1999; 10-11) and complicate the above model by adding the rest of the economy to the regulated sector.[76] However, for convenience, we will assume that the rest of the economy behaves competitively, there are constant returns to scale and there are no sunk cost assets in the rest of the economy. Let $P(t)$ and $Y(t)$ denote the time t vectors of the rest of economy output prices and quantities and let $W(t)$ and $Z(t)$ denote the time t vectors of the rest of economy input prices and quantities. With constant returns to scale and competitive price taking behaviour in the rest of the economy, we will have the rest of the economy time t revenue, $R_E(t)$, equal to the corresponding time t cost, $C_E(t)$, and the following relationships will hold:

$$(297)\ R_E(t) = P(t)\cdot Y(t) = W(t)\cdot Z(t) = C_E(t).$$

Assuming that all prices and quantities in the rest of the economy are differentiable functions of time, we can form the following time t *Divisia indexes* of output growth, $Y_D'(t)$, input growth, $Z_D'(t)$, output price growth, $P_D'(t)$, and input price growth, $W_D'(t)$:

[74] In some cases, the price cap on regulated products will not be one that freezes the structure of regulated prices except for a proportional adjustment factor; instead the price cap will apply to a *price index* of regulated products so that the regulated firm is free to vary the prices of regulated products within this overall price cap. However, it is not clear that this less onerous price cap will lead to improvements in welfare that are superior to the proportional price cap unless the regulator applies rigorous pricing principles side constraints aimed at moving prices towards marginal costs. Otherwise the regulated firm will have an incentive with the less onerous general price cap to maximise profits subject to compliance with the general price cap and the resulting resource allocation will generally give the regulated firm higher profits as compared to the proportional price cap result with no incentive to narrow the gaps between prices and marginal costs in a welfare enhancing way.

[75] See Balk (2003) for a fuller explanation of this phenomenon. Balk called the factor $s_z(t)$ in (296) the "Domar factor".

[76] We will call the resulting formula (304) the *indirect method* for implementing the price cap formula given by (293), which is the *direct method*.

(298) $Y_D'(t) \equiv P(t)\cdot Y'(t)/P(t)\cdot Y(t) = P(t)\cdot Y'(t)/R_E(t)$;

(299) $Z_D'(t) \equiv W(t)\cdot Z'(t)/W(t)\cdot Z(t) = W(t)\cdot Z'(t)/C_E(t)$;

(300) $P_D'(t) \equiv P'(t)\cdot Y(t)/P(t)\cdot Y(t) = P'(t)\cdot Y(t)/R_E(t)$;

(301) $W_D'(t) \equiv W'(t)\cdot Z(t)/W(t)\cdot Z(t) = W'(t)\cdot Z(t)/C_E(t)$.

The time t *Total Factor Productivity* (TFP) *growth* of the rest of the economy, $T_E'(t)$, is traditionally defined as the Divisia index of output growth minus the Divisia index of input growth:

(302) $T_E'(t) \equiv Y_D'(t) - Z_D'(t)$.

Now we can repeat the analysis that allowed us to derive (290). Note that our assumption of competitive behaviour in the rest of the economy will lead to both profits and the rate of change of profits being zero. Substituting $\Pi_E'(t) = 0$ and $\Pi_E(t) = 0$ into the counterpart to (290) leads to the following equation:

(303) $P_D'(t) = W_D'(t) - T_E'(t)$.

We can add a rearrangement of (303) to the right hand side of our earlier expression (293) for the rate of proportional change in regulated prices, $\alpha'(t)$, in order to obtain the following equivalent expression:

$$\begin{aligned}(304)\ \alpha'(t) &= P_D'(t) + \{[C(t)/R(t)](s_z(t)w_D'(t) + s_k(t)P_{kD}'(t)) - W_D'(t)\} \\ &\quad - \{[C(t)/R(t)]T'(t) - T_E'(t)\} + [\Pi'(t)/R(t)] - [\Pi(t)/R(t)]y_D'(t) \\ &= P_D'(t) - X(t)\end{aligned}$$

where the *X factor* at time t, X(t) is defined as follows:

$$\begin{aligned}(305)\ X(t) &\equiv \{[C(t)/R(t)]T'(t) - T_E'(t)\} + \{W_D'(t) - [C(t)/R(t)](s_z(t)w_D'(t) + s_k(t)P_{kD}'(t))\} \\ &\quad + [\Pi(t)/R(t)]y_D'(t) - \Pi'(t)/R(t) \\ &= \text{TFP differential growth rate term + input price differential growth rate term} \\ &\quad \text{+ nonzero profits adjustment term – rate of change of regulated profits term.}\end{aligned}$$

The first term in (305) is the differential rate of Total Factor Productivity growth between the regulated firm, $T'(t)$, and the rest of the economy, $T_E'(t)$, at time t. However, the TFP growth rate of the regulated firm must be weighted by the ratio of the regulated firm's period t costs, C(t), to its period t revenues, R(t). The second term is the differential rate of growth of input prices in the rest of the economy, $W_D'(t)$,[77] less C(t)/R(t) times a share weighted rate of growth of non sunk cost input prices for the regulated firm, $w_D'(t)$, and the rate of growth of allowable amortisation charges for sunk cost capital inputs, $P_{kD}'(t)$. Total cost for the regulated firm at time t, C(t), is defined as the sum of variable input costs, $C_z(t)$, plus allowable amortisation costs, $C_k(t)$, for sunk cost capital inputs. The regulated firm input cost shares which appear in the input price differential term, $s_z(t)$ and $s_k(t)$, are defined as the period t ratio of variable cost to total cost, $s_z(t) \equiv C_z(t)/C(t)$ and the period t ratio of allowable

[77] Strictly speaking, $W_D(t)$ should correspond to the economy wide output price index $P_D(t)$ that is used in (304) so that if the CPI is used as the economy wide price index, the corresponding "input" price index should equal an aggregate index of inputs less all of the outputs that are excluded in the scope of the CPI.

amortisation costs to total cost, $s_k(t) \equiv C_k(t)/C(t)$ respectively.

The last two terms on the right hand side of (305) involve the period t level of pure profits of the regulated firm, $\Pi(t)$, and the rate of change of pure profits, $\Pi'(t)$. These two terms are also present in the simpler price cap formula (293) (which did not involve the rest of the economy). If the pure profits of the regulated firm are not close to zero, then if pure profits were excessively positive, the regulator will likely want to set $\Pi'(t)$ equal to a negative number in order to reduce these excess profits over time. On the other hand, if $\Pi(t)$ were substantially negative, then the regulator will likely want to set $\Pi'(t)$ equal to a positive number in order to maintain the financial viability of the regulated firm. Thus, when $\Pi(t)$ is substantially different from zero, the regulator will typically want to set a glide path for profitability so that either profits in excess of what is required to raise capital in the industry are eliminated or, in the case of negative profits, a glide path must be set to restore the long term solvency of the regulated firm. Thus, in the case where $\Pi(t)$ is positive, typically the regulator will set $\Pi'(t)$ in the price cap formula (304) equal to a negative number, which will cause the proportional change in regulated prices, $\alpha'(t)$, defined by (304) to become smaller; ie under these conditions, the price cap will become more stringent.

Formula (304) is completely equivalent to our earlier formula (293) under our assumptions. The perceived advantage of using (304) over (293) is that it is often hard to obtain reliable objective data on industry input price indexes. The disadvantage of using (304) as compared to (293) in a regulatory context is that if there are measurement errors in computing the rest of the economy aggregates, then these measurement errors will show up in (304).

Some obvious measurement errors that should be avoided when using the traditional price cap formula (304) are listed below:

- If the rate of change of the CPI is used as the economy wide inflation rate $P_D'(t)$ which plays a prominent role in (304), then the economy wide TFP growth rate $T_E'(t)$ should match up with this price index as should the economy wide input price index, $W_D'(t)$.[78]
- The capital services component of the economy wide input price index may be poorly measured since capital services are not a part of the regular System of National Accounts.[79]
- The labour input component of the economy wide input price index may also be poorly measured since typically, quality adjusted measures of labour input are not part of the regular SNA in most countries.

We conclude this section with a cautionary note. The theory of regulation that has been

[78] If the rate of change of the country's CPI is used as an approximation to $P_D'(t)$, then: (i) commodity taxes should be removed from the CPI price components; (ii) I + G +X – M (the components of market sector output that are not consumption) need to be subtracted from primary inputs and (iii) TFP estimates need to be recomputed using consumption as the only output. These points were made by Lawrence (2003; 6-7). All of these very important adjustments are rarely done in practice. It should be noted that Bernstein and Sappington (1999; 10) do not suggest the use of the CPI as a proxy for $P_D(t)$; they suggest that a true economy wide output price index be used: "A particularly important variable that is often calculated carefully and disseminated in a timely fashion by government agencies is the economy-wide rate of output price inflation."

[79] Different treatments of the user cost formula can give different TFP growth estimates; see Harper, Berndt and Wood (1989).

developed has only dealt with the problems associated with regulating a single firm. In practice, most regulatory regimes apply to more than one firm. Thus, when setting up a price cap regime for many regulated firms in the same industry, the TFP growth rate $T'(t)$, defined by (286) and which appears in the direct and indirect price cap formulae (293) and (304), is typically based on past measures of industry wide TFP growth. This type of industry wide price cap regulation can cause considerable difficulties for both the regulator and the regulated due to a number of factors:

- Recall equation (295) which shows that TFP growth for a single firm, $T'(t)$, depends not only on the firm's rate of technological improvement $\tau(t)$ (which is presumably an industry wide effect) but it also depends the profitability of the firm, $\Pi(t)$, and factors which are largely beyond its control, namely the gaps between the regulated prices, $p(t)$, that the firm faces and the corresponding opex marginal costs, $\mu(t)$, and the gaps between the allowable amortisation costs for sunk cost capital stock components, $P_k(t)$, and the corresponding marginal user benefits $\pi(t)$ defined by (265). Thus, while basing a price cap on a forecast of future industry wide rates of technological progress seems appropriate, basing a price cap on a forecast of future industry wide rates of TFP growth will not be appropriate for all of the regulated firms since there will generally be substantial differences in the last three factors on the right hand side of (295) across the firms, and these last three factors are largely not controlled by the individual firms.
- Our single firm focus has allowed us to abstract from *operating environment factors* beyond the control of the firm that may impact a group of regulated firms differently and affect their past and future productivity performances. Adverse operating environment conditions are likely to limit opportunities for future productivity growth as well as resulting in higher costs/lower productivity levels. For example, if the group being regulated are electricity distribution businesses, and some distribution businesses are located in areas of active storm activity while others are not, the distribution businesses in the bad weather areas will generally face higher operating costs and fewer opportunities for productivity improvements than distribution businesses in good weather areas. Thus, when regulating groups of firms using a single TFP or technical progress target across firms in a price cap regime, the regulator should either group the regulated firms into *peer groups* who face roughly similar operating environments or adjust the price caps for each firm according to differences in *operating environments*. The latter procedure can be contentious due to the fact that adjusting for different operating environments will generally involve econometric estimation, which can be difficult to replicate and contentious.
- Often the regulated firm produces or distributes both regulated and unregulated products with the price cap being applied to only regulated products. In terms of the algebra used, there are no major difficulties associated with accommodating unregulated outputs; they can be treated as inputs except that a minus sign is attached to the quantities. However, this change in the treatment of unregulated products (treating them as negative inputs) will lead to changes in the magnitude of our technical progress and TFP measures; ie treating unregulated products as negative inputs will lead to increases in our estimates of technical progress and TFP growth as opposed to traditional measures which include

unregulated outputs in the measure of aggregate output. Care must be taken to make the necessary adjustments to all firms in the group.

- There can be difficulties defining exactly what the outputs of the regulated firm are. In the context of a single firm, these difficulties are not as important as in the case of group regulation, where the choice of definition for the output can play a major role in disadvantaging some firms while giving an advantage to others.

In the following section, we will illustrate some of the problems associated with defining the output measure for energy distribution businesses in a group setting.

9 WHAT IS THE OUTPUT OF A REGULATED ENERGY DISTRIBUTION FIRM?

A major problem with all index number measures of productivity and cost function based estimates of changes in efficiency for regulated energy distribution firms is that there is a lack of agreement on exactly what the correct measure of output in this industry is. For instance, Lawrence and Diewert (2006; 215) have argued that there are several possible concepts for the measure of output in an electricity distribution business:

> "A number of distributor representatives in Australia have drawn the analogy between an electricity distribution system and a road network. The distributor has the responsibility of providing the 'road' and keeping it in good condition but it has little, if any, control over the amount of 'traffic' that goes down the road. Consequently, they argue it is inappropriate to measure the output of the distributor by a volume of sales or 'traffic' type measure. Rather, the distributor's output should be measured by the availability of the infrastructure it has provided and the condition in which it has maintained it – essentially a supply side measure.
>
> "This way of viewing the output of a network industry can be extended to a number of public utilities. For instance, a number of analysts have measured the output of public transport providers using both a 'supply side' and a 'demand side' measure of output. The supply side measure of a passenger train system, for instance, would be measured by the number of seat kilometres the system provides while the demand side output would be measured by the number of passenger kilometres. In the case of public transport this distinction is often drawn because suppliers are required to provide transport for community service obligation and other non–commercial reasons. Using the supply side measure looks at how efficient the supplier has been in providing the service required of it without disadvantaging the supplier as happens with the demand side measure because of low levels of patronage beyond its control."

Lawrence and Diewert draw the analogy between an energy distribution business and a passenger transportation company where *passenger-kilometres* and *seat-kilometres* are possible alternative measures of output in the case of transportation.

In the case of an energy distribution business, it can be seen that there are at least three alternative definitions of output that could be used:

- A *volume measure* V say, which reflects the physical volume of energy deliveries to customers over the accounting period. This is the counterpart to the gross output definition of output in retailing. It is the output measure that was used in early energy supply TFP studies (eg see Lawrence, Swan and Zeitsch 1991).
- A *measure of delivery capacity* of the distributor, which would be equal to a weighted sum of power lines of different capacities times their lengths or of pipelines of different size times their lengths. If there were only a single line or pipe size, this measure of output would be proportional to the total length or distance D of the delivery grid of pipelines times V_C, which is equal to the maximum daily carrying capacity of the pipeline times the number of days in the accounting period. This measure of output, DV_C, is the

counterpart to the seat-kilometres definition of output in the transport industry. This is similar to the MVA–kilometres output used in Lawrence (2003) and Lawrence and Diewert (2006).

- A *measure of deliveries times the delivery distance*. Using the single power line or pipeline size model, this measure of output would be equal to DV, where D is the distance of the grid and V (which will be less than V_C) is the volume of energy actually delivered over the accounting period. This is the counterpart to the passenger-kilometres definition of output in the transport industry.

In the case of a single regulated firm subject to price cap regulation, it would not matter which definition of output is used but in the context of group regulation using a common target technical progress or TFP growth rate in the price cap, it very much matters which definition of output is used. Thus, it is necessary to determine which output concept is the 'right' one for regulatory purposes. In order to answer this question, we need to determine *exactly* what service does an energy distribution business provide.

In the simplest possible case, the provider takes energy from one point (say point A) and moves a volume V of energy to a purchaser at another point (say point B). Suppose the distance between points A and B is D kilometres. The purchaser of the energy should be willing to pay a price for the delivery of the energy that is proportional to the distance D that the energy is transported. Moreover, the purchaser will generally derive utility from the energy purchase that is proportional to the volume V that is purchased. These considerations suggest that the utility to the final demander of the energy delivery service provided is proportional to both the distance D the energy travels and the volume V of the energy that is delivered. Thus, the quantity of delivery services Y should be approximately equal to the product of the volume V and the distance D; ie we should have:

(306) $Y = DV$.

In the context of an unregulated competitive industry, this would be the end of the story; the counterpart to passenger-kilometres defined by (306) appears to be the 'right' concept of the service provided by the energy distribution business and hence the "right" definition of output to be used for price cap purposes. But the energy distribution business is not competitive due to the large sunk costs for the provision of the power line or pipeline network. In the regulated case, the regulator demands that the firm meet all possible demands at the regulated prices. Thus, the regulated firm is *forced* to provide a delivery capacity V_C which is equal to or greater than *anticipated peak load demand* V_P over the life of the infrastructure investment. Thus, it appears that the 'right' measure of output for the regulated energy distribution business is

(307) $Y = DV_P > DV$

where the inequality in (307) follows since peak load demand V_P will always be less than average volume purchased over the accounting period. The output measure defined by (307) is a *fourth possible measure of the output of an energy distribution business* that could be used in a regulatory setting and it appears to be the most appropriate measure in a regulatory setting where a group productivity adjustment will be used as part of a price cap formula. It is likely that this fourth measure is reasonably close to the delivery capacity measure of output mentioned above; ie the measure defined by (307) will typically be close to the counterpart to the seat-kilometres measure of output in the transportation context. What seems clear in the above discussion is that the V measure of output is deeply flawed for regulatory purposes in a group regulatory context, the DV measure is less flawed and the DV_C measure will be the closest to the theoretically preferred DV_P.

As mentioned before, the output concept used is not critical when regulating a single firm: the price cap will by definition not *discriminate* against a single firm. But in the context of using TFP growth rates in a group setting, it is extremely important to have the right definition for the outputs of the regulated firms so that the price cap can be applied to the firms in the group in an even-handed way. Thus, it is important to define very carefully exactly what regulated services are being provided by the firms in the group, independently of the institutional factors that determine exactly how the firms are paid for providing these services. As noted in the preceding section, in this case it will also be necessary to move to a method of price cap regulation that uses information that goes well beyond the use of conventional TFP measures.

The above definitions for the output of a distribution business are of course highly simplified. In reality, each distribution business's operating conditions will be somewhat different than the operating conditions facing other distributors; ie one distribution business may have to construct its pipeline grid over much more difficult terrain than another distribution business and hence using unadjusted distance D to define the output of each distribution business will not be appropriate for the distribution business that faces the more difficult operating environment factors. A brief summary of some of the important operating environment variables are:

- Differences in energy density and customer density;
- The terrain that the grid must travel over;
- Weather, fire and earthquake risks to the grid;
- The peak load that the grid must handle and
- The reliability level the service is required to meet.

As well as resulting in higher costs/lower productivity levels, adverse operating environment conditions are likely to limit opportunities for future productivity growth and will need to be allowed for in setting price caps.

We turn now to the question of how to incorporate financial capital maintenance in productivity–based regulation in the presence of sunk costs.

10 FINANCIAL CAPITAL MAINTENANCE AND SUNK COSTS

In section 7 we noted that regulation in the presence of sunk costs requires the regulator to specify a series of approved amortisation payments over the asset's life, $P_k(t)$. In this section we examine ways of specifying these amortisation payments or user cost charges that are consistent with the concept of financial capital maintenance.

The accounting literature, starting with Peasnell (1981), has provided a useful framework for accounting for sunk costs. The regulatory literature has utilised this framework to provide some guidance to regulators on how to adapt this accounting literature into a regulatory context; eg see Schmalensee (1989) and Johnstone (2003).

The issues surrounding the problems associated with accounting for an infrastructure project in a regulated context can be explained using a simple model. Suppose that at the beginning of period 1, the project has been completed and the total cost of the project is P_1 at the beginning of period 1. This asset is expected to yield a stream of services for T periods before it is retired. The regulated firm faces the (actual) weighted average cost of capital (WACC) at the beginning of period 1, r_1, and for the subsequent periods, its anticipated nominal cost of capital for period t is r_t. Rather than charge the entire purchase cost of the asset in the first period, the firm will be allowed by the regulator to amortise the cost of the asset over T periods. The period t amortisation amount or asset charge is denoted by C_t and we assume that this charge is recovered by the firm at the end of period t for t = 1,2....,T. Note that period t users of the utility's outputs will be paying the period t charge C_t since these charges will be imbedded in output prices. Thus, C_t can also be regarded as a *period t user charge* that customers of the regulated firm will pay in period t for the services of the sunk cost asset. The firm will fully recover its purchase cost of the asset (in a present value sense) if the *period by period user charges* C_t satisfy the following fundamental equation:[80]

(308) $P_1 = (1+r_1)^{-1}C_1 + (1+r_1)^{-1}(1+r_2)^{-1}C_2 + ... + (1+r_1)^{-1}...(1+r_T)^{-1}C_T.$

In order to relate equation (308) to the notation used in the preceding sections, think of P_1 as being equal to the initial asset purchase price P_k which appeared in the left hand side of equation (260) times k and think of the C_t as the user benefit prices P_k^t defined by (259) times k. Thus, (308) is a counterpart to equation (260), which assumed one hoss shay depreciation. If the modified one hoss shay model is more applicable, then equation (262) related the initial purchase charge, P_kk, to the period t user charges, $P_k^tk^t$. Equation (262) can be put into the form defined by (308) if we set P_1 equal to P_kk and the period t user charges C_t equal to $P_k^tk^t$ for t = 1,2,...,T where the P_k^t were defined by (259) and the k^t were defined by (246); ie $k^t \equiv (1-\delta)^{t-1}k$.

Looking at equation (308) from the viewpoint of the regulator, it can be seen that if the period by period opportunity costs of capital r_t are reasonably accurate, then *the regulator can choose any pattern of periodic charges* for the use of the capital asset, $C_1,...,C_T$, which satisfy equation (308) and the regulated firm will have no grounds for complaint; ie *its financial capital will be maintained under these conditions*. A question that will be

[80] We have neglected any scrap value that the asset may have at the end of its useful life and we have neglected any disposal or decommissioning costs associated with disposing of the asset.

addressed in due course is: how exactly should the regulator choose the C_t?[81]

Before addressing the above question, it is useful to phrase the regulator's choice problem in a more familiar framework; ie the above choice of period by period charges can be converted into a choice of period by period depreciation amounts. Thus, if expectations about interest rates made at the beginning of period 1 turn out to be accurate at the beginning of period 2, the discounted stream of admissible user charges that the regulator will allow over periods 2 to T is the period 2 *Regulatory Asset Base* P_2 and it is defined as follows:

(309) $P_2 = (1+r_2)^{-1}C_2 + (1+r_2)^{-1}(1+r_3)^{-1}C_3 + ... + (1+r_2)^{-1}...(1+r_T)^{-1}C_T$.

Period 1 *depreciation* D_1 can be defined as the decline in the Regulatory Asset Base going from the beginning of period 1 to the beginning of period 2:

(310) $D_1 \equiv P_1 - P_2$.

Looking at equations (308) and (309), it can be seen that equation (308) can be rewritten as follows:

(311) $P_1 = (1+r_1)^{-1}C_1 + (1+r_1)^{-1}P_2$.

Equation (311) can now be rearranged to give us the following *user cost like formula* for the period 1 user charge C_1:[82]

(312) $C_1 = (1+r_1)P_1 - P_2$

$= (1+r_1)P_1 - [P_1 - D_1]$ using (310)

$= r_1P_1 + D_1$.

Thus, the period 1 user charge can be expressed as r_1 times the period 1 undepreciated asset value P_1 plus the period 1 depreciation charge D_1.

A formula similar to (312) can be obtained for the remaining user charges C_t. Define the *period t regulatory asset base* or the *beginning of period t undepreciated asset value* P_t and the *period t depreciation charge* D_t by equations (313) and (314) below:

(313) $P_t \equiv (1+r_t)^{-1}C_t + (1+r_t)^{-1}(1+r_{t+1})^{-1}C_{t+1} + ... + (1+r_t)^{-1}...(1+r_T)^{-1}C_T$; t = 2,3,...,T;

(314) $D_t \equiv P_t - P_{t+1}$; t = 2,3,...,T

where P_{T+1} is defined to be 0. The period t counterpart to equation (311) is:

(315) $P_t = (1+r_t)^{-1}C_t + (1+r_t)^{-1}P_{t+1}$; t = 2,3,...,T.

Equation (315) can now be rearranged to give us the following *user cost like formula* for the period t user charge C_t:

[81] This question is more complicated that it appears to be at first sight. We have just said that the regulator could choose any sequence of amortisation charges that satisfy equation (308) and the solvency of the regulated firm would not be impaired. Moreover, any sequence of amortisation charges that satisfy (308) will also be consistent with our intertemporal model of optimal regulation defined in section 6. But when we turn to price cap models of regulation, it is no longer true that the pattern of amortisation charges is irrelevant (except that they should satisfy equation (308)). In the case of price cap regulatory schemes, the amortisation charge for any period will end up being a component of total cost for that period, and thus the amortisation charges will influence the period by period price caps and they will also influence the measured TFP growth rates for the regulated firm.

[82] See Schmalensee (1989; 294) and Johnstone (2003; 4) for a similar formula.

$$(316)\ C_t = (1+r_t)P_t - P_{t+1} \qquad t = 2,3,...,T$$
$$= (1+r_t)P_t - [P_t - D_t] \qquad \text{using (314)}$$
$$= r_tP_t + D_t$$
$$= r_t\,[P_1 - D_1 - D_2 - ... - D_{t-1}] + D_t$$

where the last equation also uses (314) repeatedly.

Equations (312) and (316) show that the sequence of user charges C_t is completely determined (and these charges will satisfy equation (308)) by the initial cost of the asset P_1, the sequence of nominal costs of capital r_t and the sequence of nonnegative depreciation amounts D_t (which sum up to P_1). Thus, the regulator can specify either a sequence of period by period user charges C_t satisfying equation (308) *or* a sequence of depreciation amounts D_t which sum up to the initial asset cost P_1, which in turn can be used in order to define a sequence of C_t using (312) and (316). Note that the third equations in (312) and (316) are the *usual building blocks allowable charges* used by Australian and New Zealand regulators. Note also that it will be easier to work with equations (312) and (316) in practice (and choose the D_t) than to work with equation (308) and the subsequent counterpart equations for each period, since equations (312) and (316) require only a knowledge of the current opportunity cost of capital whereas equation (308) and its subsequent counterparts require knowledge of an entire sequence of expected nominal interest rates r_t, which may change over time.[83]

At this point, it is useful to demonstrate that the regulated firm will always recover its going opportunity cost of capital if the period by period user charges C_t satisfy equation (308) and the anticipated future period r_t actually materialise. To see this, rearrange the first equation in (312) as follows:

$$(317)\ (1+r_1)P_1 = C_1 + P_2.$$

This equation says that the beginning of period 1 regulatory asset base, P_1, times one plus the prevailing nominal weighted average cost of capital to the firm at the beginning of period one, r_1, equals the regulator's allowable period 1 charge for the use of the asset in period 1, C_1,[84] plus the regulatory asset base P_2 which the regulator will allow the firm to have at the beginning of period 2. Thus, the period 1 user charge C_1 plus the end of period 1 regulatory asset base P_2 will be just large enough for the regulated firm to recover its initial asset value P_1 plus make a nominal return of r_1 on its investment in this asset; ie the asset will make the going cost of capital during period 1. Now look at the first equation in (316) for period t = 2 and rearrange this equation as follows:

[83] However, if we use equations (312) and (316) and the r_t change over time, then the resulting user charges will not be the same as the original planned sequence of user charges which satisfied equation (308) using the original sequence of anticipated r_t.

[84] Note that we assume that all period one flow variables (like the period 1 user charge C_1) are realised at the end of period 1. This is consistent with accounting treatments of assets at the beginning and end of the accounting period and cash flows that occur during the period. "Here A_{t-1} is discounted as a flow dated t−1 and $C_t + A_t$ as a flow at t. This accords with the assumption conventional in discrete compounding that flows occur at the end of each period." K.V. Peasnell (1981; 56). A_{t-1} and A_t are Peasnell's counterparts to our P_1 and P_2 and C_t is Peasnell's counterpart to our C_1. These timing conventions are discussed in more detail in Diewert (2005b; 8) and they are consistent with the use of end of period user costs as discussed in Diewert (2005a).

(318) $(1+r_2)P_2 = C_2 + P_3$.

This equation says that the beginning of period 2 regulatory asset base, P_2, times one plus the prevailing nominal weighted average cost of capital to the firm at the beginning of period two, r_2, equals the regulator's allowable period 2 charge for the use of the asset in period 1, C_2, plus the regulatory asset base P_3 which the regulator will allow the firm to have at the beginning of period 3. Thus, the regulated firm will make its period 2 weighted average cost of capital on its beginning of period 2 regulatory asset base. And so on for the remaining periods over which the asset yields useful services.

The fact that the period by period user charges C_t are largely arbitrary at this point (they need only satisfy equation (308) in order for the firm to recover its cost of capital associated with the investment) or alternatively, that the period by period amortisation amounts D_t are arbitrary (except that they must sum to the initial asset cost P_1), creates a problem for the regulator and possibly for the regulated firm.[85] If the allowed rates of return are reasonable, then, regulatory risk considerations aside, the regulated firm has no particular incentive to contest whatever pattern of periodic user charges or depreciation charges that the regulator chooses. However, it is unlikely that the r_t will in fact correspond exactly to appropriate weighted costs of capital and thus depending on whether the allowed rates are too high or two low, the firm will have incentives to either postpone depreciation (the r^t are higher than the firm's opportunity cost of capital) or to lobby for accelerated depreciation (the r^t are below the firm's opportunity cost of capital). The regulator's choice problem is much more complex[86]. In the following paragraphs, we will examine some possible choices.

How should the regulator choose either the sequence of C_t or the sequence of D_t? We will consider three special cases and analyze the resulting patterns of user charges and depreciation charges. Simplicity is always a virtue and so the simplest thing to do at first sight is to choose the C_t to be constant over each period. For our second special case, we will choose to make the C_t constant in real terms, using the Consumer Price index as our deflator. Finally, instead of deflating by the CPI, we will consider using an asset price index as our deflator.

---

[85] Peasnell, Schmalensee and Johnstone all recognised this indeterminacy in their models as is indicated in the following quotations: "The concept of income employed in equation (4), $P_t - i_tA_{t-1}$, is very general. Profit P_t and asset book value A_{t-1} are not restricted to a particular accounting model, for example, to economic income. The only restrictions are those concerning the valuation of the opening and closing capital stocks, A_0 and A_N respectively. Opening book capital must be valued at outlay (i.e. $A_0 = C_0$); closing book capital must be valued at the amount expected to be received from disposal of the asset(s) (i.e. $A_N = R_N$). How capital stock is valued in the periods in between (t = 2, ... , N−1) is of no consequence whatsoever. Interim capital stocks, $A_1, A_2, \ldots,$ A_{N-1}, can be valued at economic value, HC, RC, VO, or NRV: from an investment planning viewpoint, the choice is immaterial." K.V. Peasnell (1981; 54). "It is important to recognise that the invariance Proposition does not imply that all depreciation schedules are equally socially desirable. Inappropriate choice of depreciation policy can lead to an intertemporal pattern of utility rates that bears no relation to the corresponding intertemporal pattern of capital costs. But the Proposition indicates that depreciation policy can be altered to produce more efficient rates without being unfair in a present value sense to utilities or their customers". Richard Schmalensee (1989; 295-296). "For given total depreciation ($RAB_0 - RAB_T$), PV is constant regardless of the time pattern of period depreciation charges. It makes no difference over what interval $0 \le t \le T$ assets are written down, or how aggregate depreciation expense is distributed within this interval." David Johnstone (2003; 6)

[86] See Lally (2003) and Bertram and Twaddle (2005) who discuss identifying excess returns.

10.1 Case 1: Constant nominal user charges

If we choose to make the periodic charges C_t constant over time, we need only solve the following equation for C and then set all C_t equal to this common C:

$$(319)\ P_1 = C\{(1+r_1)^{-1} + (1+r_1)^{-1}(1+r_2)^{-1} + ... + (1+r_1)^{-1}...(1+r_T)^{-1}\}$$
$$= C\{\delta_1 + \delta_1\delta_2 + ... + \delta_1\delta_2...\delta_T\}$$

where the discount factors δ_t are defined as follows:

$$(320)\ \delta_t \equiv 1/(1+r_t)\ ; \qquad t = 1,2,...,T.$$

Solving equation (319) for the constant user charge C, we have

$$(321)\ C_t = C^* \equiv P_1/\{\delta_1 + \delta_1\delta_2 + ... + \delta_1\delta_2...\delta_T\}\ ; \qquad t = 1,2,...,T.$$

Once the constant C_t have been determined using equations (321), the sequence of regulatory asset values P_t and the sequence of allowable depreciation amounts D_t for t = 2,...,T can be determined using equations (313) and (314). Using equations (313), we obtain the following sequence of beginning of period t regulatory asset values P_t:

$$(322)\ P_2 \equiv C^*\{\delta_2 + \delta_2\delta_3 + ... + \delta_2...\delta_T\}\ ;$$
$$P_3 \equiv C^*\{\delta_3 + \delta_3\delta_4 + ... + \delta_3...\delta_T\}\ ;$$
$$...$$
$$P_T \equiv C^*\delta_T.$$

Now equations (314), $D_t \equiv P_t - P_{t+1}$, can be applied to the allowable asset bases P_t defined by (319) and (322) in order to determine the allowable depreciation amounts D_t.

Since equations (322) are rather complicated looking, it will be useful to consider a special case of these equations; namely, the case where the nominal weighted average costs of capital r_t are expected to remain constant over the life of the asset. Thus, assume that

$$(323)\ r_t = r \geq 0\ ; \qquad t = 1,2,...,T.$$

Under assumptions (323), the discount rates δ_t defined by (320) will all be equal to 1/(1+r) and this will greatly simplify equations (322). We need to consider two cases at this point. The *first case* is where

$$(324)\ r = 0 \text{ and hence all } \delta_t = 1/(1+r) = 1\ ; \qquad t = 1,2,...,T.$$

Under these conditions, we obtain the following solution for the constant period t charges, C_t, regulatory asset values P_t and depreciation amounts D_t:

$$(325)\ C_t = C^* = P_1/T\ ; \qquad t = 1,2,...,T;$$

$$(326)\ P_t = C^*[T+1-t]\ ; \qquad t = 1,2,...,T;$$

$$(327)\ D_t = C^*\ ; \qquad t = 1,2,...,T.$$

Thus, in this case where the cost of capital r_t is always 0, we find that depreciation is *straight line depreciation*; ie the regulatory asset value falls by a constant amount each period.

In the much more realistic *second case*, we assume that the constant nominal cost of capital is *positive* so that

(328) $r_t = r > 0$ and hence all $\delta_t = 1/(1+r) \equiv \delta$; $\quad t = 1,2,...,T$

where $0 < \delta < 1$. In this case, equation (319) becomes:

(329) $P_1 = C\{\delta_1 + \delta_1\delta_2 + ... + \delta_1\delta_2...\delta_T\}$

$= C\{\delta + \delta^2 + ... + \delta^T\}$

$= \delta C\{(1-\delta)^{-1} - \delta^T(1-\delta)^{-1}\}$

$= \delta C\{1 - \delta^T\}/(1-\delta).$

Solving equation (329) for the constant user charge C leads to the following equations for C_t:

(330) $C_t = C^* = P_1(1-\delta)/\{\delta(1 - \delta^T)\}$; $\quad t = 1,2,...,T.$

Now equations (330) for the C_t can be substituted back into equations (313) in order to obtain the following expressions for the allowable beginning of period t asset values:

(331) $P_t = \delta C^*\{1 - \delta^{T+1-t}\}/(1-\delta)$; $\quad t = 2,3,...,T.$

Using (329) and (331) for t = 2, we obtain the following expression for period 1 allowable depreciation, D_1:

(332) $D_1 \equiv P_1 - P_2$

$= [\delta C^*\{1 - \delta^T\}/(1-\delta)] - [\delta C^*\{1 - \delta^{T-1}\}/(1-\delta)]$

$= \delta^T C^*.$

Note that if r equals 0 so that δ equals one, then the D_1 defined by (332) agrees with our earlier first case solution (327). But now we are assuming that r is greater than zero so the discount rate δ is less than one so that allowable first period depreciation in this case will be less than straight line depreciation since δ^T is less than one.

Using equations (331), we can derive the following formula for allowable period t depreciation D_t under our present constant nominal weighted average cost of capital assumptions:

(333) $D_t \equiv P_t - P_{t+1}$ $\quad t = 1,2,...,T$

$= [\delta C^*\{1 - \delta^{T+1-t}\}/(1-\delta)] - [\delta C^*\{1 - \delta^{T+1-t-1}\}/(1-\delta)]$

$= \delta^{T-t+1} C^*.$

Thus, period t allowable depreciation D_t is always less than the constant user charge C^* until the last period T when D_T will equal the user charge. Note also that the (nominal) depreciation charges increase monotonically over time so that depreciation turns out to be the *opposite* of accelerated depreciation in this constant nominal user charge with a constant nominal cost of capital.

However, if there is general inflation in the economy going from period 1 to T, then it can be seen that in real terms, the constant nominal user charge will turn out to be larger in real terms in the earlier periods than in the later periods. This difficulty can be overcome as we shall see in our next special case of formula (308).

10.2 Case 2: Constant real user charges for households

In this case, we let the anticipated rate of Consumer Price Index inflation for period t be ρ_t for t = 1,...,T. In order to make the user charges constant in real terms for households, let the user charge in period t be defined as follows:

$$(334)\ C_t = C(1+\rho_1)...(1+\rho_t)\ ; \qquad t = 1,2, ... , T.$$

The logic behind (334) is this: a dollar of income that a household receives at the beginning of period 1 should be roughly equivalent to $(1+\rho_1)$ dollars of income received at the end of period 1 which in turn should be equivalent to $(1+\rho_1)(1+\rho_2)$ dollars of income received at the end of period 2 and so on. Thus, to keep the utility's capital charges constant in real purchasing power terms, the period by period charge should be allowed to grow at the rate of expected CPI inflation.

Now substitute (334) into (308) and get the following equation for the constant C:

$$\begin{aligned}(335)\ P_1 &= (1+r_1)^{-1}C(1+\rho_1) + (1+r_1)^{-1}(1+r_2)^{-1}C(1+\rho_1)(1+\rho_2) \\ &\qquad + ... + (1+r_1)^{-1}...(1+r_T)^{-1}C(1+\rho_1)...(1+\rho_T) \\ &= C\{(1+r_1^*)^{-1} + (1+r_1^*)^{-1}(1+r_2^*)^{-1} + ... + (1+r_1^*)^{-1}...(1+r_T^*)^{-1}\} \\ &= C\{\delta_1^* + \delta_1^*\delta_2^* + ... + \delta_1^*\delta_2^*...\delta_T^*\}\end{aligned}$$

where the period t *real interest rates* r_t^* are defined by equations (336) below and the period *t real discount factors* δ_t^* are defined by (337):

$$(336)\ (1+r_t^*) \equiv (1+r_t)/(1+\rho_t)\ ; \qquad t = 1,2,...,T;$$

$$(337)\ \delta_t^* \equiv 1/(1+r_t^*)\ ; \qquad t = 1,2,...,T.$$

Solving equation (335) for the constant C, we have

$$(338)\ C^{**} \equiv P_1/\{\delta_1^* + \delta_1^*\delta_2^* + ... + \delta_1^*\delta_2^*...\delta_T^*\}.$$

Once C^{**} has been determined, the sequence of nominal period t charges is determined by equations (334) which we repeat here for convenience:

$$(339)\ C_t = C^{**}(1+\rho_1)...(1+\rho_t)\ ; \qquad t = 1,2, ... , T.$$

Once the C_t have been determined using equations (338) and (339), the sequence of regulatory asset values P_t and the sequence of allowable depreciation amounts D_t for t = 2,...,T can be determined using equations (313) and (314). Using equations (313), (336) and (337), we obtain the following sequence of beginning of period t regulatory asset (nominal) values P_t:

$$\begin{aligned}(340)\ P_2 &\equiv (1+\rho_1)C^{**}\{\delta_2^* + \delta_2^*\delta_3^* + ... + \delta_2^*...\delta_T^*\}\ ; \\ P_3 &\equiv (1+\rho_1)(1+\rho_2)C^{**}\{\delta_3^* + \delta_3^*\delta_4^* + ... + \delta_3^*...\delta_T^*\}\ ; \\ &... \\ P_T &\equiv (1+\rho_1)...(1+\rho_{T-1})C^{**}\delta_T^*\end{aligned}$$

where C^{**} is defined by (338) and the real discount factors δ_t^* are defined by (337). Now equations (314), $D_t \equiv P_t - P_{t+1}$, can be applied to the allowable asset bases P_t defined by (340) in order to determine the allowable depreciation amounts D_t. This regulatory regime corresponds to an indexed (by the CPI) historical cost amortisation scheme.

It is useful to consider a special case of the above scheme where it is assumed that the firm's *real* weighted average cost of capital in period t, r_t^*, is a positive constant r^* for all periods. Thus, we assume that:

$$(341)\ r_t^* = r^* > 0 \text{ and hence all } \delta_t^* = 1/(1+r^*) \equiv \delta^* ; \qquad t = 1,2,...,T$$

where $0 < \delta^* < 1$. In this case, equation (335) becomes:

$$(342)\ P_1 = C\{\delta^* + \delta^{*2} + ... + \delta^{*T}\}$$
$$= \delta^* C\{1 - \delta^{*T}\}/(1-\delta^*).$$

Solving equation (342) for C leads to the following solution:[87]

$$(343)\ C^{**} = P_1(1-\delta^*)/\{\delta^*(1 - \delta^{*T})\}.$$

Now substitute the expression for C^{**} given by (343) into equations (339) in order to obtain the sequence of period t user charges C_t (in nominal dollars).

Now equations (339) for the C_t can be substituted back into equations (313) in order to obtain the following expressions for the allowable beginning of period t asset values:

$$(344)\ P_t = (1+\rho_1)...(1+\rho_{t-1})\delta^* C^{**}\{1 - \delta^{*T+1-t}\}/(1-\delta^*) ; \qquad t = 2,3,...,T$$

where C^{**} is defined by (343) and δ^* equals $1/(1+r^*)$, where r^* is the allowed constant real weighted cost of capital for the regulated firm.

Using equations (342)-(344), we can derive the following formula for allowable period t depreciation D_t under our present constant real weighted average cost of capital assumptions:

$$(345)\ D_t \equiv P_t - P_{t+1} \qquad t = 1,2,...,T$$
$$= (1+\rho_1)...(1+\rho_{t-1})[\delta^* C^{**}\{1 - \delta^{*T+1-t}\}/(1-\delta^*)]$$
$$- (1+\rho_1)...(1+\rho_{t-1})(1+\rho_t)\ [\delta^* C^{**}\{1 - \delta^{*T+1-t-1}\}/(1-\delta^*)]$$
$$= (1+\rho_1)...(1+\rho_{t-1})\delta^{*T-t+1}C^{**} - (1+\rho_1)...(1+\rho_{t-1})\rho_t\ [\delta^* C^{**}\{1 - \delta^{*T+1-t-1}\}/(1-\delta^*)]$$
$$\approx (1+\rho_1)...(1+\rho_{t-1})\delta^{*T-t+1}C^{**}$$

where the last approximation will be satisfactory if ρ_t is small. Thus, the constant real cost of capital model leads to much more complex expressions, (344) and (345), for the regulatory asset base P_t and depreciation D_t as compared to our earlier constant nominal cost of capital model.

10.3 Case 3: Indexation of user charges using asset price inflation rates

We do not have to use the CPI inflation rate to do the indexation of the common charge C. For example, choose the expected rate of inflation for a new asset of the type that was just built or purchased. Thus, let the anticipated rate of asset inflation for period t be i_t for t = 1,...,T. Let the user charge in period t be:

[87] It will generally be the case that the constant nominal interest rate r can be assumed to be greater than the corresponding real r^* and hence δ will be less than δ^*. Hence the C^{**} which satisfies (342) and (343) will be smaller than the corresponding C^* which satisfies (329) and (330).

$$(346)\ C_t = C(1+i_1)...(1+i_t)\ ; \qquad t = 1,2, ... , T.$$

Now substitute (346) into (308) and get the following equation for C:

$$(347)\ P_1 = (1+r_1)^{-1}C(1+i_1) + (1+r_1)^{-1}(1+r_2)^{-1}C(1+i_1)(1+i_2) + ... + (1+r_1)^{-1}...(1+r_T)^{-1}C(1+i_1)...(1+i_T).$$

Once C has been determined by (347), then the C_t are determined by (346). What is interesting about this scheme is that the resulting user charges C_t are exactly equal to the sequence of user costs one would get assuming one hoss shay depreciation. The advantage in choosing amortisation costs to line up with user costs is that we do not have to distinguish sunk capital from new capital, assuming that the one hoss shay model of depreciation is appropriate.

10.4 Efficiency and other considerations

It must be emphasised that in all three of the special case amortisation schemes just proposed above, the regulated firm's financial capital will be maintained, provided that the allowed cost of capital interest rates are equal to the firm's opportunity cost of financial capital.

The sunk cost allocation problem occurs in competitive industries as well. However, in the competitive case with a single aggregate sunk cost capital stock, clear theoretical guidance is available: amortisation charges should be proportional to the period by period cash flows that can be attributed to the aggregate sunk cost capital stock; see Diewert (2009). This methodology does not work in the regulated context because future period cash flows are not independent entities; they depend on prices that are set by the regulator.

The above three special cases do not exhaust the regulator's reasonable choices for an amortisation scheme. For example, it may be economic to install excess capacity for a new infrastructure project because population growth and/or economic growth will lead to increases in demand in future periods. Thus, intertemporal efficiency considerations might lead the regulator to impose smaller user charges for the early periods in the lifetime of the project reflecting the low marginal cost of usage and to encourage use of the infrastructure asset but these charges would progressively increase as demand growth occurred and capacity utilisation increased towards full capacity.[88] This is effectively a modified version of case 2 above with an additional user pays component. That is, the user charges are not only indexed by the CPI but also by demand with the user charges increasing proportionally with higher demand (and hence utilisation). This approach seems to be most consistent with the matching principle in accounting that suggests that sunk costs should be matched with user benefits. A potential problem with this form of 'peak load' pricing, however, is that there

[88] In our intertemporal theoretical model of regulation described in section 6, recall that equations (259) defined our optimal sequence of user charges $P_k^t \equiv -\partial c^t(y^t,w^t,k)/\partial k$ in terms of partial derivatives of the period t opex cost function with respect to the sunk cost capital variable and these user charges satisfied equation (308). There can be problems in using this sequence of "optimal" charges in the context of price cap regulation of the type described in section 8, particularly if the underlying production function is of the Leontief fixed coefficient type which does not allow for input substitution. In this case, under many reasonable specifications for the intertemporal utility function, it will be optimal to charge nothing for the capital services of the sunk cost capital stock except in periods where the economy is at peak demand. In these peak demand periods, prices of the regulated outputs would be unusually high while in other periods, prices would be unusually low.

may be a mismatch between the consumers who benefit from relatively low prices in the early years of the asset's life and those who have to pay high prices as full capacity is approached.

We have shown in this section that a range of asset valuation methodologies can be consistent with financial capital maintenance. Each methodology will generate a time–series of asset values and the series of amortisation charges are used to ensure financial capital maintenance is achieved. The main difference between the asset valuation methods is on the timing of revenue receipts rather than their net present value as recognised by Schmalensee (1989) and Johnstone (2003) in the regulatory context and earlier by Peasnell (1981) in an accounting context. The important requirements are that the amount actually invested is the opening asset value in the first period and the scrap value is the closing asset value in the last period.

We have shown above that CPI indexed historic cost and the use of a real return to capital (case 2) is likely to be the most consistent with the concept of user pays and intertemporal efficiency although this could be further enhanced by indexing by the degree of capacity utilisation as well as by the CPI. This effectively 'back–end loads' the profile of receipts and requires a high degree of regulatory credibility for investors to be confident that the regulatory rules will remain unchanged for a sufficiently long period for them to recover their costs.

Annual depreciated replacement cost revaluations would be similar to indexing by the asset price index as in case 3. This has the advantage of lining up with the current opportunity costs of adding more infrastructure but indexing by the CPI is more consistent with general investor expectations and welfare analysis.

Finally, the use of unindexed historic cost and a nominal return to capital (case 1) has the effect of 'front–end loading' the profile of receipts and appears the least consistent with efficiency considerations. However, if the degree of regulatory credibility is not high, investors may be reluctant to commit funds that will require a deferral of income. In the extreme, investors would require immediate compensation for sunk costs if they were not confident a regulatory contract could be honoured. This highlights the importance of achieving a high degree of regulatory credibility that is compatible with the more 'back–end loaded' pattern of income receipts that is consistent with efficiency considerations.

11 CONCLUSIONS

Some of the important conclusions that emerge from this analysis are as follows:

- The role of the regulator should be to improve the welfare of the households in the economy. The effects of regulation on welfare can best be modelled in the context of a small general equilibrium model.
- An important task for the regulator is to create incentives for the regulated firm to improve its productivity performance but it is also important for the regulator to move regulated prices closer to their corresponding marginal costs.
- The information required to implement optimal regulation is difficult to obtain and so simpler methods of regulation that are not fully optimal, like price cap regulation, will have to be used in practice.
- Price cap regulation can be modified to accommodate both sunk costs and financial capital maintenance.
- Allowed amortisation charges replace the capital goods price index in the price cap formula when there are sunk costs.
- There is no guarantee that future rates of technical progress will mirror past rates.
- Extrapolations of past TFP growth are often used as a proxy for future technical change but TFP growth in the context of a regulated firm is far from being identical to technical progress. In fact, conventional TFP growth depends not only on technical progress but also on variables that are controlled by the regulator including profits, the selling prices of regulated products and allowable amortisation charges.
- Where CPI–X regulation is used, the X factor involves the difference between the firm's TFP growth weighted by its costs relative to its revenue and the economy–wide TFP growth rate plus the difference between economy–wide input price change and the sum of the firm's opex price growth and amortisation charges growth each weighted by their the respective shares of their cost in revenue plus a nonzero profits adjustment term less a rate of change in regulated profits term (equation 305).
- When dealing with the regulation of many firms using a common productivity target as part of the price cap, it is necessary to move to a method of price cap regulation that uses information that goes well beyond the use of conventional (ie revenue and cost weighted) TFP measures and that measures exactly what regulated services are being provided by the firms in the group, independently of the institutional factors that determine exactly how the firms are paid for providing these services.
- A range of asset valuation methodologies can be consistent with financial capital maintenance. Each methodology will generate a time–series of asset values and the series of amortisation charges are used to ensure financial capital maintenance is achieved.
- The main difference between the asset valuation methods is on the timing of revenue receipts rather than their net present value. The important requirements are that the

amount actually invested is the opening asset value in the first period and the scrap value is the closing asset value in the last period.

- Amortisation charges based on CPI indexed historic cost and the use of a real return to capital are likely to be the most consistent with the concept of user pays and intertemporal efficiency although this could be further enhanced by indexing by the degree of capacity utilisation as well as by the CPI. This effectively 'back–end loads' the profile of receipts and requires a high degree of regulatory credibility for investors to be confident that the regulatory rules will remain unchanged for a sufficiently long period for them to recover their costs.

REFERENCES

Afriat, S.N. (1967), “The Construction of Utility Functions from Expenditure Data”, *International Economic Review* 8, 67–77.

Balk, B.M. (2003), “On the Relationship between Gross Output and Value Added Based Productivity Measures: The Importance of the Domar Factor”, Centre for Applied Economic Research Working Paper 2003/05, Faculty of Commerce and Economics, University of New South Wales, Australia.

Basu, S. and J.G. Fernald (2002), “Aggregate Productivity and Aggregate Technology”, *European Economic Review* 46, 963–991.

Bernstein, J.I. and D.E.M. Sappington (1999), “Setting the X Factor in Price Cap Regulation Plans”, *Journal of Regulatory Economics* 16, 5–25.

Bertram, G. and D. Twaddle (2005), “Price–Cost Margins and Profit Rates in New Zealand Electricity Distribution Networks Since 1994: the Cost of Light Handed Regulation”, *Journal of Regulatory Economics* 27(3), 281–307.

Caves, D.W. and L.R. Christensen (1982), *Ex Parte 290 (Sub No. 4) Railroad Cost Recovery Procedures—Productivity Adjustment*, Submission to the Interstate Commerce Commission, Washington D.C.

Caves, D.W., L.R. Christensen and W.E. Diewert (1982), “The Economic Theory of Index Numbers and the Measurement of Input, Output, and Productivity’, *Econometrica* 50, 1392–1414.

Christensen, L.R. and D.W. Jorgenson (1969), “The Measurement of U.S. Real Capital Input, 1929–1967”, *Review of Income and Wealth* 15, 293–320.

Coase, R.H. (1945), “Price and Output Policy of State Enterprises: A Comment”, *Economic Journal* 55, 112–113.

Coase, R.H. (1946), “The Marginal Cost Controversy”, *Economica* 13, 169–182.

Denny, M., M. Fuss and L. Waverman (1981), “The Measurement and Interpretation of Total Factor Productivity in Regulated Industries, with an Application to Canadian Telecommunications”, pp. 179–218 in *Productivity Measurement in Regulated Industries*, T.G. Cowing and R.E. Stevenson (eds.), New York: Academic Press.

Diewert, W.E. (1973a), “Afriat and Revealed Preference”, *The Review of Economic Studies* 40, 419–425.

Diewert, W.E. (1973b), “Functional Forms for Profit and Transformation Functions”, *Journal of Economic Theory* 6, 284–316.

Diewert, W.E. (1974), “Applications of Duality Theory,” pp. 106–171 in M.D. Intriligator and D.A. Kendrick (ed.), *Frontiers of Quantitative Economics*, Vol. II, Amsterdam: North–Holland.

Diewert, W.E. (1976), “Exact and Superlative Index Numbers”, *Journal of Econometrics* 4, 114–145.

Diewert, W.E. (1978), "Optimal Tax Perturbations", *Journal of Public Economics* 10, 138–177.

Diewert, W.E. (1980), "Aggregation Problems in the Measurement of Capital", pp. 433–528 in *The Measurement of Capital*, D. Usher (ed.), Chicago: The University of Chicago Press. http://www.econ.ubc.ca/diewert/1capital.pdf

Diewert, W.E. (1982), "Duality Approaches to Microeconomic Theory", pp. 535–599 in *Handbook of Mathematical Economics*, Volume 2, K.J. Arrow and M.D. Intriligator (eds.), Amsterdam: North–Holland.

Diewert, W.E. (2005a), 'Issues in the Measurement of Capital Services, Depreciation, Asset Price Changes and Interest Rates', pp. 479–542 in *Measuring Capital in the New Economy*, C. Corrado, J. Haltiwanger and D. Sichel (eds.), Chicago: University of Chicago Press.

Diewert, W.E. (2005b), "The Measurement of Capital: Traditional User Cost Approaches", Chapter 1 of *The Measurement of Business Capital, Income and Performance*, Tutorial presented at the University Autonoma of Barcelona, Spain, September 21–22, 2005, revised December, 2005.

Diewert, W.E. (2009), "The Aggregation of Capital over Vintages in a Model of Embodied Technical Progress", *Journal of Productivity Analysis*, forthcoming.

Diewert, W.E. and K.J. Fox (2000), "Incentive Indexes for Regulated Industries", *Journal of Regulatory Economics* 17(1), 5–24.

Diewert, W.E. and K.J. Fox (2008), "On the Estimation of Returns to Scale, Technical Progress and Monopolistic Markups", *Journal of Econometrics* 145, 174–193.

Diewert, W.E. and D. Lawrence (2000), "Progress in Measuring the Price and Quantity of Capital", pp. 273–326 in *Econometrics Volume 2: Econometrics and the Cost of Capital: Essays in Honor of Dale W. Jorgenson*, Lawrence J. Lau (ed.), Cambridge: The MIT Press. http://www.econ.ubc.ca/diewert/progress.pdf

Diewert, W.E. and C.J. Morrison (1986), "Adjusting Output and Productivity Indexes for Changes in the Terms of Trade", *Economic Journal* 96, 659–679.

Diewert, W.E., A.H. Turunen–Red and A.D. Woodland (1989), "Productivity and Pareto Improving Changes and Taxes and Tariffs", *The Review of Economic Studies* 56, 199–216.

Divisia, F. (1926), *L'indice monetaire et la theorie de la monnaie*, Paris: Societe anonyme du Recueil Sirey.

Domar, E.D. (1961), "On the Measurement of Technological Change", *Economic Journal* 71, 709–729.

Domar, E.D. (1974), "On the Optimal Compensation of a Socialist Manager", *The Quarterly Journal of Economics* 88, 1–18.

Economic Insights (2009), *Asset valuation methods and productivity–based regulation*, Report by Erwin Diewert, Denis Lawrence and John Fallon to the Commerce

Commission, Canberra, 10 June.

Fleming, J.M. (1944), "Price and Output Policy of State Enterprises: Comment", *Economic Journal* 54, 328–337.

Gorman, W.M. (1968), "Measuring the Quantities of Fixed Factors", pp. 141–172 in *Value, Capital and Growth: Papers in Honour of Sir John Hicks*, J.N. Wolfe (ed.), Chicago: Aldine.

Hall, R.E. and D.W. Jorgenson (1967), "Tax Policy and Investment Behavior", *American Economic Review* 57:3, 391–414.

Harper, M.J., E.R. Berndt and D.O. Wood (1989), "Rates of Return and Capital Aggregation Using Alternative Rental Prices", pp. 331–372 in *Technology and Capital Formation*, Dale W. Jorgenson and Ralph Landau (eds.). Cambridge, Massachusetts: MIT Press.

Hicks, J.R. (1946), *Value and Capital*, Second Edition, Oxford: Clarendon Press.

Hotelling, H. (1925), 'A general mathematical theory of depreciation', *Journal of the American Statistical Association* 20, 340–353.

Hotelling, H. (1938), "The General Welfare in Relation to Problems of Taxation and of Railway and Utility Rates", *Econometrica* 6, 242–269.

Hotelling, H. (1939), "The Relation of Prices to Marginal Costs in an Optimum System", *Econometrica* 7, 151–155.

Hulten, C.R. (1990), "The Measurement of Capital", pp. 119–152 in *Fifty Years of Economic Measurement*, E.R. Berndt and J.E. Triplett (eds.), Studies in Income and Wealth, Volume 54, The National Bureau of Economic Research, Chicago: The University of Chicago Press.

Hulten, C.R. (1996), "Capital and Wealth in the Revised SNA", pp. 149–181 in *The New System of National Accounts*, J.W. Kendrick (ed.), New York: Kluwer Academic Publishers.

Johnstone, D. (2003), *Replacement Cost, Asset Valuation and the Regulation of Energy Infrastructure Tariffs: Theory and Practice in Australia*, CRI (Centre for the Study of Regulated Industries) International Series 8, School of Management, The University of Bath, UK.

Jorgenson, D.W. (1989), "Capital as a Factor of Production", pp. 1–35 in *Technology and Capital Formation*, D.W. Jorgenson and R. Landau (eds.), Cambridge MA: The MIT Press.

Jorgenson, D.W. (1996a), "Empirical Studies of Depreciation", *Economic Inquiry* 34, 24–42.

Jorgenson, D.W. (1996b), *Investment: Volume 2; Tax Policy and the Cost of Capital,* Cambridge, Massachusetts: The MIT Press.

Jorgenson, D.W. and Z. Griliches (1967), "The Explanation of Productivity Change", *Review of Economic Studies* 34, 249–283.

Karlin, S. (1959), *Mathematical Methods and Theory in Games, Programming and Economics*, Volume 1, Reading MA: Addison–Wesley Publishing Co.

Kuhn, H.W. and A.W. Tucker (1951), "Nonlinear Programming", pp. 481–492 in *Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability*, Berkeley CA: University of California Press.

Lally, M. (2003), *The Weighted Average Cost of Capital for Electricity Lines Businesses*, School of Economics and Finance, Victoria University of Wellington, 4 August.

Lawrence, D. (2003), *Regulation of Electricity Lines Businesses, Analysis of Lines Business Performance – 1996–2003*, Report prepared by Meyrick and Associates for the New Zealand Commerce Commission, Canberra, 19 December.

Lawrence, D.A. and W.E. Diewert (2006), "Regulating Electricity Networks: The ABC of Setting X in New Zealand", pp. 207–241 in *Performance Measurement and Regulation of Network Utilities*, T. Coelli and D. Lawrence (eds.), Cheltenham, UK: Edward Elgar.

Lawrence, D.A., W.E. Diewert and K.J. Fox (2006), "The Contribution of Productivity, Price Changes and Firm Size to Profitability", *Journal of Productivity Analysis* 26, 1–13.

Lawrence, D., P. Swan and J. Zeitsch (1991), "The Comparative Efficiency of State Electricity Authorities", in P. Kriesler, A. Owen and M.R. Johnson (eds.), *Contemporary Issues in Australian Economics*, MacMillan.

Lerner, A.P. (1944), *The Economics of Control*, New York: Macmillan.

McFadden, D. (1978), "Cost, Revenue and Profit Functions", pp. 3–109 in *Production Economics: A Dual Approach*, Volume 1, M. Fuss and D. McFadden (eds.), Amsterdam: North–Holland.

Meade, J.E. (1944), "Price and Output Policy of State Enterprise", *Economic Journal* 54, 321–328.

Peasnell, K.V. (1981), 'On capital budgeting and income measurement', *Abacus* 17(1), 52–67.

Pigou, A.C. (1920), *The Economics of Welfare*, London: Macmillan.

Samuelson, P.A. (1947), *Foundations of Economic Analysis*, Cambridge MA: Harvard University Press.

Samuelson, P.A. (1974), "Complementarity—An Essay on the 40th Anniversary of the Hicks–Allen Revolution in Demand Theory", *The Journal of Economic Literature* 12, 1255–1289.

Schmalensee, R. (1989), 'An expository note on depreciation and profitability under rate of return regulation', *Journal of Regulatory Economics* 1(3), 293–298.

Schreyer, P. (2001), *OECD Productivity Manual: A Guide to the Measurement of Industry–Level and Aggregate Productivity Growth*, Paris: OECD.

Schreyer, P. (2009), *Measuring Capital: Revised Manual*, OECD Working Paper STD/NAD(2009)1, January 13 version, Paris: OECD.

Slater, M. (1950), "Lagrange Multipliers Revisited: A Contribution to Nonlinear Programming", Cowles Commission Discussion Paper, Mathematics, Number 403, November.

Solow, R.M. (1957), “Technical Change and the Aggregate Production Function”, *Review of Economics and Statistics* 39, 312–320.

Solow, R.M., J. Tobin, C.C. von Weizsäcker and M. Yaari (1966), “Neoclassical Growth with Fixed Factor Proportions”, *The Review of Economic Studies* 33, 79–115.

Uzawa, H. (1958), “The Kuhn–Tucker Theorem in Concave Programming”, pp. 32–37 in *Studies in Linear and Nonlinear Programming*, K.H. Arrow, L. Hurwicz and H. Uzawa (eds.), Stanford CA: Stanford University Press.

Walras, L. (1980), “The State and the Railways”, *Journal of Public Economics* 13, 81–100.

Wilson, T. (1945), “Price and Outlay Policy of State Enterprise”, *Economic Journal* 55, 454–461.