Asset valuation methods and productivity–based regulation

Report prepared for
Commerce Commission

27 April 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

Executive summary .................................................................................................................................................. ii

1 Introduction ......................................................................................................................................................... 1

2 The use of asset valuation in traditional productivity analysis ................................................................. 2

2.1 The quantity of capital input ......................................................................................................................... 2

2.2 The annual user cost of capital inputs ......................................................................................................... 3

2.3 The value of the capital stock ....................................................................................................................... 5

3 CPI–X incentive regulation ............................................................................................................................... 6

3.1 Price cap incentive regulation ..................................................................................................................... 6

3.2 The building blocks method ......................................................................................................................... 8

3.3 Traditional productivity–based regulation ................................................................................................. 10

3.4 Productivity–based regulation in the presence of sunk costs ................................................................... 13

4 Evaluation criteria for alternative asset valuation methods ........................................................................ 22

4.1 Proposed and existing evaluation criteria .................................................................................................. 22

4.2 Economic efficiency as a fundamental and encompassing criterion ....................................................... 24

4.3 Assessment of proposed and existing criteria ............................................................................................ 27

4.4 Preferred criteria for assessing asset valuation methodologies ............................................................... 30

5 Assessing the alternative asset valuation methods ....................................................................................... 32

5.1 Introduction ..................................................................................................................................................... 32

5.2 Optimised deprival value ............................................................................................................................. 33

5.3 Depreciated historic cost .............................................................................................................................. 41

5.4 Indexed historic cost ..................................................................................................................................... 45

5.5 Conclusion .................................................................................................................................................... 47

6 Transitional issues ............................................................................................................................................. 49

6.1 Electricity–specific issues ............................................................................................................................... 49

Appendix A: The theory of network regulation in the presence of sunk costs .................................................. 51

A1 Overview ....................................................................................................................................................... 51

A2 An introduction to optimal one period regulation .................................................................................... 54

A3 The comparative statics properties of the one period regulatory model .................................................... 57

A4 A continuous time approach to the determinants of welfare change ...................................................... 75

A5 Optimal one period regulation with freely variable capital services ....................................................... 81

A6 Multiple period regulation with sunk costs ................................................................................................. 93

A7 Practical regulation using only information on the regulated firm ................................................................ 108

A8 Practical regulation of a single firm using TFP growth and economy wide information ................................ 112

A9 What is the output of a regulated energy distribution firm? ................................................................. 119

A10 Financial capital maintenance and sunk costs ......................................................................................... 122

A11 Conclusions ............................................................................................................................................... 133

Appendix B: The rationale for and history of the ‘building blocks’ methodology ........................................... 135

B1 Background .................................................................................................................................................... 135

B2 The origins of FCM and CPI–X as regulatory concepts .............................................................................. 136

B3 The development of the building blocks methodology .............................................................................. 145

B4 Conclusions ............................................................................................................................................... 152

References ........................................................................................................................................................... 155
EXECUTIVE SUMMARY
1 INTRODUCTION

The Commerce Commission (‘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. Specifically, the report is to evaluate the optimised deprival value (ODV), depreciated historic cost (DHC) and indexed depreciated historic cost (IHC) asset valuation methods where CPI–X incentive regulation uses productivity analysis.

The terms of reference ask Economic Insights to:

1) identify interrelationships between asset valuation methods and CPI–X price paths set using productivity analysis, including:
   • a review of how asset valuation is used in productivity analysis, and the assumptions underpinning such use (eg the treatment of sunk costs), including (but not limited to) the relevance of the monopolistic mark–up term;
   • a discussion on the rationale and assumptions behind CPI–X incentive regulation, including a comparison of similar assumptions between that wider concept and the analytical frameworks of productivity analysis and building blocks analysis; and
   • a discussion on the rationale, assumptions and the historic development of building blocks analysis;

2) evaluate the pros and cons of ODV, IHC and DHC where CPI–X incentive regulation uses productivity analysis, including:
   • proposing appropriate principles and criteria for evaluating the relative merits of the different asset valuation methods, in light of the principle of financial capital maintenance;
   • identifying any transitional issues should there be a change in asset valuation approach; and
   • identifying the impact and relevance of any issues specific to the 2004 ODV revaluation of electricity distribution businesses (EDBs), and suggesting how these might be addressed in the context of setting default price–quality paths for EDBs in accordance with Part 4 of the Commerce Act 1986, taking into account any restrictions on the approach and information the Commission would be able to use in setting default price–quality paths.

Major limitations of the theory underlying productivity–based regulation to date has been that it has not recognised the sunk cost nature of network assets nor adequately allowed for the principle of financial capital maintenance (FCM). The sunk cost characteristic of network assets and the desirability of ensuring FCM both have important implications for how productivity analysis is used in network regulation. Similarly, the theory of network regulation has evolved in a relatively piecemeal way that has not adequately addressed key economic welfare issues. Consequently, a large part of this project has involved developing a unified theory of network regulation using productivity analysis.
The following section of the report reviews the use of asset valuation in traditional productivity analysis. Section 3 then discusses CPI–X incentive regulation using both productivity analysis and the building blocks approach. The section draws on appendix A where we develop the theory of network regulation in the presence of sunk costs and appendix B which reviews the history and evolution of the building blocks method. In section 4 we review and develop evaluation criteria for the alternative asset valuation methods before assessing the methods against these criteria in section 5. In section 6 we review transitional issues associated with moving to the preferred asset valuation method and in section 7 we review the issues specific to the situation of the electricity distribution industry.

2 THE USE OF ASSET VALUATION IN TRADITIONAL PRODUCTIVITY ANALYSIS

Productivity indexes are formed by aggregating output quantities into a measure of total output quantity and aggregating input quantities into a measure of total input quantity. The productivity index is then the ratio of total outputs to total inputs or, if forming a measure of productivity growth, the change in the ratio of total outputs to total inputs. To form the total output and total input measures we need a price and quantity for each output and each input, respectively. The quantities enter the calculation directly as it is changes in output and input quantities that we are aggregating. The prices are used to weight together changes in all output quantities and all input quantities into measures of total output quantity and total input quantity using revenue and cost measures, respectively.

Like other inputs and outputs, we thus need a cost and quantity for capital inputs.

2.1 The quantity of capital input

There are a number of different approaches to measuring both the quantity and cost of capital inputs. How the quantity of annual capital input to the production process should be measured depends on the relevant physical depreciation profile for the asset. Hotelling (1925) described the pattern of annual capital input quantities to the production process as the ‘service potential’ of the asset. An asset’s service potential relates to its physical deterioration or decline in its effective capacity (and/or service quality) over time. If the physical depreciation profile is thought to be best proxied by so-called ‘one hoss shay’ depreciation (ie the amount of annual input quantity the asset can provide remains relatively constant over its lifetime), then the quantity of capital inputs can be measured directly in quantity terms (eg using a measure of line or transformer capacity). If the physical depreciation profile is thought to be best proxied by so-called ‘geometric’ depreciation (ie the amount of annual input quantity the asset can provide falls by a given percentage each year of its lifetime), then the quantity of capital inputs can be measured indirectly using a constant dollar measure of the depreciated value of assets. Using this approach the proxied quantity of annual capital input for an asset falls relatively quickly from the first year leading to higher estimated productivity growth than under the assumption of one hoss shay physical depreciation.

For long–lived network assets such as poles, wires and transformers the physical depreciation profile is likely to be closer to ‘one hoss shay’ than it is to either declining balance
(geometric) or straight line. In this case using a measure of physical asset quantity is likely to be the best proxy. If this approach is adopted then the input quantity (which is the primary driver of productivity results) will be unaffected by which asset valuation method is used. Rather, the asset value will only affect the secondary driver of what weight is allocated to the capital quantity change in forming the productivity measure.

Most productivity studies do not have access to adequate physical data to implement this approach and so tend to use the deflated, depreciated asset value as the capital input quantity proxy. Obviously, in this case the choice of asset valuation method will have a much larger impact on the end productivity result.

2.2 The annual user cost of capital inputs

The annual cost of using capital inputs can be measured either directly by forming a user cost measure based on an estimated depreciation rate, a rate of return reflecting the opportunity cost of capital, a deduction for the estimated rate of capital gains (or addition for capital losses) and the asset value or indirectly as the residual of revenue less operating costs. The user cost measure used in the direct approach recognises that there has to be a return of capital over the asset’s lifetime (ie the firm has to recoup its original investment) and a return on capital to compensate for holding the asset over its lifetime as opposed to using the funds for an alternative investment. Following Hotelling (1925) it is important to recognise that an asset’s value will reflect both the remaining service potential of the asset and the remaining life of the asset. That is, if the asset’s service potential declines over time then its value will also fall reflecting its physical deterioration. However, even for so-called ‘one hoss shay’ assets whose service potential remains constant over their lifetime, their asset value will progressively fall over time reflecting the fact that as each year passes they have one less year of productive life left.

The direct user cost approach to measuring the annual cost of capital inputs is also described as an ex ante approach as it specifies what producers expect to happen when making decisions at the start of the period. The indirect approach, on the other hand, is often referred to as an ex post approach as it uses the results of what actually happened during the production period. If the indirect approach is adopted then the firm’s realised profitability (excluding capital gains/losses) can be determined by forming the ratio of the residual return to capital (net of estimated depreciation) to the asset value.

The traditional direct approach to measuring the annual user cost of capital in productivity studies uses the Jorgenson (1963) user cost method. This approach multiplies the value of the capital stock by the sum of the depreciation rate plus the opportunity cost rate minus the rate of capital gains (ie annual change in the asset price index). The before tax user cost, \( u \), can be represented as follows (Diewert 1993):

\[
(2.1) \quad u = rP + \delta(1 + \rho)P - \rho P
\]

where:

- \( r \) is the nominal interest rate;
- \( \delta \) is the depreciation rate;
- \( \rho \) is the inflation rate of capital items; and
$P$ is the purchase price of capital.

That is, capital gains resulting from an increase in the price of the asset reduce the cost of holding (and using) the asset over the year. Thus, if revenue is being set equal to total costs then it will be reduced by the extent of capital gains between two years, all else equal. In this sense, the productivity approach is somewhat analogous to the building blocks approach where ‘revaluation gains’ resulting from asset revaluation exercises (the equivalent of capital gains in the productivity context) are treated as income and thus reduce the return to the business from its allowable service charges.

The practical problem with this approach, however, is that if we use ex post capital gains then the user cost becomes quite volatile and, in periods of rapid asset inflation, user costs can become negative, particularly for long-lived assets which have low depreciation rates. This has led to productivity analysts smoothing the pattern of ex post capital gains in an effort to proxy ex ante expectations of capital gains which are what actually drive producers’ decision making.

A further step down this path is to use a relatively constant ‘real’ interest rate to cover both the nominal opportunity cost rate and the expected rate of capital gains. This approach effectively credits the producer with less ability to anticipate differing rates of capital gains through time. It is this approach that was used in the Lawrence (2003) electricity lines business study where an opportunity cost rate of 8 per cent was used to cover these two terms. It should be noted that this rate effectively assumes a significant rate of capital gains as 8 per cent is the net result of the sum of the standard 10 year bond risk free rate plus the market risk premium less the rate of capital gains.

The difference between the Jorgenson (1963) user cost method and the periodic adjustment for accumulated revaluation gains, that is sometimes seen in building block regulation using periodic recalculation of the ODV, is that the productivity approach allows for capital gains annually rather than in periodic lumps that then have to be either digested all at once or else spread back over a number of years. However, the standard productivity approach would not typically allow for ‘found’ assets or more detailed differentiation of operating environment conditions (eg rocky ground in gas pipeline laying) that is often seen in periodic ODV recalculations.

Having formed a quantity of capital series (by either the deflated asset value approach or the physical measure approach) and a corresponding price of annual capital input using the direct approach, it is possible to form an estimate of the firm’s total costs if it was not earning excess profits. By comparing this series with the firm’s actual revenue, we can see whether the firm is earning positive excess profits or, using the terminology of productivity analysis, a positive monopolistic markup (ie over and above efficient costs). To form this estimate of excess profits, however, we have to make a judgement on what the firm’s true opportunity cost of capital is after allowing for risk. This would normally involve some benchmarking of realised rates of return across the industry, often across countries. In this sense the decision-making process is somewhat analogous to forming a judgement on the weighted average cost of capital (WACC) in the building block process.
2.3 The value of the capital stock

From the preceding sections it can be seen that asset values typically enter at two places in the traditional productivity framework. Firstly, when using the ex ante framework for forming input cost measures, they are used to form the user cost of capital which becomes the weight applying to the change in capital quantity when aggregating input changes to form the change in total input quantity. Secondly, in some studies the deflated, depreciated asset value is used as a proxy for the capital input quantity.

In forming the asset value for a traditional productivity study, a number of approaches can be adopted. If a sufficiently long time series of investment data is available (at least as long as the assumed asset lifetime), the perpetual inventory approach can be applied whereby investment for each ‘vintage’ year is taken and progressively depreciated over its lifetime. After allowing for depreciation, the remaining capital stocks from each vintage year are then aggregated to form a measure of the total capital stock in each year. This process is conducted in constant price terms where the asset price index is then typically used to bring values into current dollars. The capital stock can be in either gross or net terms. The gross capital stock does not deduct annual depreciation but removes the asset from the stock at the end of the asset’s life. The net capital stock deducts depreciation annually.

If the length of investment data available is less than the assumed maximum asset lifetime then the normal practice in forming a net capital stock series is to take a point estimate of the asset value and then roll this forwards (and backwards if necessary) using constant price investment (relative to the investment price index) and an assumed (geometric) depreciation rate as follows:

\[
S_{t+1} = (1 - \delta)S_t + I_{t+1} \quad \text{for } t > t_0; \quad \text{and} \\
S_t = S_{t+1} / (1 - \delta) - I_{t+1} \quad \text{for } t < t_0,
\]

where:
- \( S_{t+1} \) is the end of period real capital stock in period \( t+1 \);
- \( S_t \) is the end of period real capital stock in the period \( t \);
- \( \delta \) is the declining balance rate of economic depreciation;
- \( I_t \) is constant price investment in period \( t \); and
- \( t_0 \) is the date of the asset value point estimate.

The current price asset value is then formed by multiplying the constant price series by the relevant capital goods price index.

For traditional productivity studies with a limited history of investment data available, the asset value point estimate would typically reflect the market value of assets at that point in time. It would be standard practice to take the earliest point estimate of the capital stock available, provided there was reasonable confidence in the quality of the valuation process. Existing or, in the case of energy distribution, sunk assets and new investment have traditionally been treated symmetrically. The appropriateness of this treatment in the regulatory context where there are major network sunk costs is examined in the following section.
3  CPI–X INCENTIVE REGULATION

3.1  Price cap incentive regulation

Incentive regulation has been developed over the last 30 years in response to concerns over the performance of rate of return or cost of service regulation which had previously been the norm. Vogelsang (2002, pp.5–6) observes:

‘… the leading practice of rate–of–return regulation had been severely criticized at least since the early 1960s, when the discovery of the Averch–Johnson effect (Averch and Johnson 1962) and the empirical work of Stigler and Friedland (1962) had suggested a lack of improvement over unregulated monopoly outcomes. … The resulting incentive regulation has breathed new life into the stale public utility regulation.

‘What do we mean by incentive regulation? In particular, it means that the regulator delegates certain pricing decisions to the firm and that the firm can reap profit increases from cost reductions. Incentive regulation makes use of the firm’s information advantage and profit motive. The regulator thus controls less behavior but rather rewards outcomes.’

Although there are a few US precursors, [Yes, Caves and Christensen developed precisely (3.1) for US coal companies against the railway companies in the early 1980’s—I was working with Bill Waters and Michael Tretheway for the railroads, which was my first experience with regulatory matters. We lost the case—rightly so!] CPI–X price cap regulation was developed in the UK and has become the most common form of incentive regulation. Littlechild (1983) authored an influential report proposing a CPI–X approach for British Telecom where the approach was concerned with avoiding the pitfalls of US style rate–of–return regulation.

The principal rationale advanced for CPI–X regulation is that it mimics, as much as reasonably possible, the outcomes that would be achieved in a competitive market. Competitive markets normally have a number of desirable properties. The process of competition leads to industry output prices reflecting industry unit costs, including a normal rate of return on the value of assets after allowing for risk. Because no individual firm can influence industry unit costs, each firm has a strong incentive to maximise its productivity performance to achieve lower unit costs than the rest of the industry. This will allow it to keep the benefit of new, more efficient processes that it may develop until such times as they are generally adopted by the industry. This process leads to the industry operating as efficiently as possible at any point in time and the benefits of productivity improvements being passed on to consumers relatively quickly.

Because infrastructure industries such as the provision of energy transmission and distribution networks are often subject to decreasing costs, competition is normally limited and incentives to minimise costs and provide the cheapest and best possible quality service to users are not strong. The use of CPI–X regulation in such industries attempts to strengthen the incentive to operate efficiently by imposing similar pressures on the network operator to the process of competition. It does this by constraining the operator’s output price to track the
level of estimated efficient unit costs for that industry. The change in output prices is ‘capped’ as follows:

\[
\Delta P = \Delta W - X \pm Z
\]

(3.1)

where \( \Delta \) represents the proportional change in a variable, \( P \) is the maximum allowed output price, \( W \) is a price index taken to approximate changes in the industry’s input prices, \( X \) is the estimated total factor productivity (TFP) change for the industry and \( Z \) represents relevant changes in external circumstances beyond managers’ control which the regulator may wish to allow for. Ideally the index \( W \) would be a specially constructed index which weights together the prices of inputs by their shares in industry costs. However, this price information is often not readily or objectively available, particularly in regulatory regimes that have yet to fully mature. A commonly used alternative is to choose a generally available price index such as the consumer price index or GDP deflator.

However, the theory of network regulation has evolved in a relatively piecemeal way that has not adequately addressed key economic welfare issues. In appendix A we develop a more unified theory of network incentive regulation.

There are two common alternative ways of implementing price cap regulation – the buildings blocks method (BBM) and productivity–based regulation. BBM relies on forecasts of the firm’s own costs and draws on financial capital maintenance (FCM) concepts to set prices so that the net present values of forecast revenues and costs over the regulatory period are equal.

Productivity–based regulation, as it has been applied to date, argues that in choosing a productivity growth rate to base \( X \) on, it is desirable that the productivity growth rate be external to the individual firm being regulated and instead reflect industry trends at a national or even international level. This way the regulated firm is given an incentive to match (or better) this productivity growth rate while having minimal opportunity to ‘game’ the regulator by acting strategically. The latter can be a problem with the building blocks method for setting \( X \) which relies more heavily on information on the firm’s own costs and likely best practice for that firm. The logic behind the productivity–based approach is, however, based on the assumption that starting prices are set a level which just recovers total costs and so full implementation of the productivity–based approach will need to be done in conjunction with an initial partial building blocks study which quantifies the first period’s total costs.

However, major limitations of the theory underlying productivity–based regulation developed to date have been that it has not recognised the sunk cost nature of network assets nor adequately allowed for the principle of financial capital maintenance. The sunk cost characteristic of network assets and the desirability of ensuring FCM both have important implications for how productivity analysis is used in network regulation and will be explored further in section 3.4 and appendix A.

External factors beyond management control that the regulator may wish to allow for in the \( Z \) factor include changes in government policy such as community service obligations and tax treatment.

While the CPI–X framework can provide incentives to reduce costs, it may need to be accompanied by measures to stop firms from achieving those cost reductions by reducing quality. This may take the form of an ‘S’ factor introduced to provide incentives to maintain
3.2 The building blocks method

The BBM refers to an approach where prices or revenues are regulated by calculating forward looking, allowable cost components and summing those cost components to define allowable revenue. The cost components are described as cost building blocks and the allowable revenue is sometimes described as ‘building blocks allowable revenue’. Allowable revenue can then be defined as the target regulatory variable (after making additional adjustments based on other objectives) or allowable revenue can be converted to an average price as the target regulatory variable.

It is important to recognise that the methodology is forward looking and that normally costs are both forward looking and defined to reflect prudent expenditure and realistically achievable operational efficiencies for the utility in question. In addition, adjustments are normally made to remove asset revaluation gains and losses. The methodology is thus designed so that on an ex ante basis investors can expect that funds prudently invested in regulated assets will be fully recouped in net present value terms (based on a discount rate that reflects the opportunity cost of the investment taking risk into account) provided actual costs are expected to be comparable to allowable efficient costs. This latter condition is generally referred to as ex ante FCM which means that there is an expectation that the value of invested capital will be maintained in real terms over the life of the investment.

Ex ante FCM is intended to be achieved as opposed to ex post FCM and investors are allowed to retain realised returns in excess of those required to achieve ex ante FCM and required to bear the costs of realised returns lower than expected over a defined regulatory period. The rationale for adopting ex ante FCM as a regulatory principle is that it is consistent with ensuring efficient investment occurs.

In appendix B we review the history and development of both BBM and its key component, FCM. Stephen Littlechild, Michael Beesley and Geoff Horton developed the building blocks approach in the UK in the early 1990s for Offer’s first reset of the value of X in the electricity sector. Ian Byatt of Ofwat used a model similar to Offer’s building blocks approach almost at the same time as its application in the electricity sector.

Implementing a building blocks control regime is a very information intensive exercise and focuses on the firm’s own costs and estimates of what its efficient costs might be. It has the potential advantage of being able to focus on the specific circumstances facing each firm, to be more forward–looking and take account of FCM. However, the analysis of what the firm’s efficient costs might be is usually subjective and non–reproducible as it depends on the views of the relevant engineering consultant. The process then often appears to be a ‘black box’. The regulator invariably faces information asymmetry relative to the firm’s managers and there is a risk the regulator can be ‘gamed’ by being mislead about the true level of efficient costs and how quickly gaps can be bridged. To reduce this risk the regulator normally takes a relatively intrusive or ‘heavy handed’ approach to setting price caps. This is, in turn, a relatively resource–intensive process and one that may be subject to ‘spurious accuracy’
whereby estimated costs are specified with a degree of precision not justified by the nature of the analysis. Applying BBM to a large number of businesses in the one regulatory review may be infeasible if the regulator’s resources are relatively constrained.

Many analysts have criticised the reliance of price cap methods that rely on the firm’s own costs (as BBM does). For instance, King and Maddock (1996, p.63) note:

‘Price cap regulation was initially hailed as a radical departure from regulatory processes based on observed profits. While there are important differences between ROR [rate of return] and price cap regulation, in practice the two regimes appear to have similar consequences. However, this similarity is due to inappropriate reliance on firm costs and profits in reviewing the value of X. A review process based on yardstick comparisons and other data that is not specific to the regulated firm may offer significant advantages. If implemented correctly, price caps represent a major advance compared to traditional ‘cost–based’ regulation. By sensibly addressing the problem of asymmetric information, price caps can lead to substantial community benefit without distorting production.’

Challenges in obtaining appropriate estimates of efficient costs in a cost effective way are an increasing issue in the application of BBM. For instance, one of the first regulators to apply BBM in Australia, Victoria’s Essential Services Commission, noted the following problems arising from its last price determination for electricity distribution (ESC 2005, pp.12–13):

• tensions in a privatised industry with monopoly characteristics between the firms seeking to maximise returns and the expectations and objectives of customers
• the clear information asymmetry and reliance on the information provided by the utility with incentives to “talk up” costs and “talk down” future sales
• its underestimation, in hindsight, of the challenges in relying on reported costs (this point is not clear to me; maybe redraft?)
• restructuring of EDBs including arrangement with entities with common ownership, but which are not directly covered by the regulatory regime, and the possibility that such arrangements may not be at arm’s length, with the potential to inflate or obscure reported costs
• the challenges generally of obtaining transparent cost data and unravelling complex and changing cost allocations making comparisons and forecasts difficult
• the considerable difficulty obtaining information per se, with delays in some cases and others where information was withheld entirely.

Regulators in both the UK and Australia are now considering other approaches to determining the value of X in price or revenue cap regulation and are trying to find a more light–handed approach that could better incorporate incentives for efficient investment. The Australian Energy Market Commission (AEMC) is currently conducting a review into whether TFP–based regulation should be allowed as an alternative to the building blocks approach.
3.3 Traditional productivity–based regulation

The framework that underlies the traditional productivity–based CPI–X approach can be illustrated as follows. We start with the index number definition of TFP growth:

\[
\Delta \text{TFP} = \frac{[Y_1/Y_0]}{[X_1/X_0]} = \frac{[R_1/R_0]/[P_1/P_0]}{[C_1/C_0]/[W_1/W_0]} = \frac{[M_1/M_0]/[W_1/W_0]}{[P_1/P_0]}
\]

where the superscripts represent different time periods, \(R_t\) is revenue, \(C_t\) is cost in period \(t\), \(M_t\) is the period \(t\) markup and \(R_t = M_tC_t\). As a normal return on assets (after allowing for risk) is included in the definition of costs, a firm earning normal returns will have a markup factor of one while a firm earning excess returns will have a markup of greater than one. Rearranging the above equation gives:

\[
P_1/P_0 = \frac{[M_1/M_0]/[W_1/W_0]}{\Delta \text{TFP}}
\]

where \(W_1/W_0\) is the firm’s input price index (which includes labour, intermediate and capital inputs where the price of capital inputs is given by equation (2.1)). Equation (3.3) is approximately equal to:

\[
\Delta P = \Delta M + \Delta W - \Delta \text{TFP}
\]

Thus, the admissible rate of output price increase \(\Delta P\) is equal to the rate of increase of input prices \(\Delta W\) less the rate of TFP growth \(\Delta \text{TFP}\) provided the regulator wants to keep the monopolistic markup constant (so that \(\Delta M = 0\)). Equation (3.3) or its approximation (3.4) is the key equation for setting up an incentive regulation framework: the term \(W_1/W_0\) would be an input price index of the target firm’s peers and the term \(\Delta \text{TFP}\) would be the average TFP growth rate for the target firm’s peers. The markup growth term could be set equal to zero under normal circumstances but if the target firm was making an inadequate return on capital due to factors beyond its control, this term could be set equal to a positive number. On the other hand, if the target firm was making monopoly profits or excessive returns, then this term could be set negative. This effectively sets a ‘glide path’ to bring firms closer to earning a normal or average rate of return. [Well explained! It all looks so simple and straightforward; too bad we have to make is so complicated later!]

The next issue to be considered in operationalising (3.4) is the choice of the price index to reflect changes in the industry’s input prices, \(W\). The most common choice for this index is the consumer price index (CPI), even though this is actually an index of output prices for the economy rather than input prices. The reason commonly advanced for using the CPI rather than an industry input price index is that it is often hard to obtain reliable objective data on industry input price indexes. This is far from being a correct argument in my opinion; even if the CPI is more accurately measured, it is not measuring the right thing! The CPI is also an output index rather than an input index and hence it already has the effects of productivity improvements imbedded in it. Note that your argument here is different from your usual exposition of CPI – X, which introduces the rest of the economy and looks at differences between the regulated industry and the rest of the economy (which you do below). This latter exposition is at least logical. The present model is not. Statistical agencies typically devote fewer resources to compiling input price indexes and the samples tend to be relatively small.
As a result, industry input price indexes are often relatively erratic. This just means that regulators need to be harder on the regulated companies to provide appropriate data on a timely basis. In the age of the computer, it should be possible to construct firm specific input and output price indexes at reasonable cost. On the other hand, statistical agencies typically devote large amounts of resources to compiling the CPI and so it is relatively accurate, timely and stable. Regulators tend to prefer the use of relatively robust and timely official price indexes to reduce the scope to be gamed by the regulated business (even though this typically introduces an additional layer of measurement problems related to economy–wide variables).

Normally we can expect the economy’s input price growth to exceed its output price growth by the extent of economy–wide TFP growth (since labour and capital ultimately get the benefits from productivity growth). For convenience, we assume that the markup factors for the economy as a whole are one so that the counterpart to equation (3.2) applied to the entire economy becomes:

\[
P_P^1/P_P^0 = [W^1/W^0]/\Delta \text{TFP}_E. \tag{3.5}\]

Substituting the rate of change of the CPI for the economy–wide output price index on the left hand side of (3.5) and rearranging terms leads to the following identity:

\[
1 = [\text{CPI}^1/\text{CPI}^0]/\Delta \text{TFP}_E/[W^1/W^0]. \tag{3.6}\]

Substituting the right hand side of (3.6) into (3.2) produces the following equation:

\[
P^1/P^0 = \{[\text{CPI}^1/\text{CPI}^0] \Delta \text{TFP}_E/[W^1/W^0]\} \{[W^1/W^0][W^1/W^0]/\Delta \text{TFP}\}
= [\text{CPI}^1/\text{CPI}^0][\Delta \text{TFP}_E/\Delta \text{TFP}][\{W^1/W^0\}/(W^1/W^0)/\{W^1/W^0\}][M^1/M^0]. \tag{3.7}\]

Approximating the terms in (3.7) by finite percentage changes leads to the following:

\[
\Delta P = \Delta \text{CPI} + \Delta M + [\Delta W - \Delta W_E] - [\Delta \text{TFP} - \Delta \text{TFP}_E] \tag{3.8}\]

so that the X factor is defined as:

\[
X = [\Delta \text{TFP} - \Delta \text{TFP}_E] - [\Delta W - \Delta W_E] - \Delta M. \tag{3.9}\]

I think I can see why both regulated and the regulator might prefer the seeming more complex price cap formula (3.8) to the more straightforward formula (3.4). Let’s ignore the fudge factor $\Delta M$ for now. Looking at (3.8), we see that the main driver for the price cap is the CPI: both regulator and the regulated will typically agree that $\Delta W$ is close to $\Delta W_E$ and $\Delta \text{TFP}$ is close to $\Delta \text{TFP}_E$ so that the last two difference terms in (3.8) can be taken to be small numbers, and both sides will not have much incentive to argue a lot about these small numbers. So the main driver of the price cap is $\Delta \text{CPI}$, which as you state is pretty well measured so it is not worthwhile to argue over it either. Now go back to (3.4) where the main drivers for the price cap are the terms $\Delta W - \Delta \text{TFP}$. It is difficult to measure either of these terms accurately as you argued. Now it becomes very important for both the regulated and the regulator to get the right numbers for $\Delta W$ and $\Delta \text{TFP}$ and so they are bound to argue about these numbers. What bothers me about this is that the CPI is not the price index that should be matched up with the output price index that is used to measure economy wide TFP. But I guess I see the logic of why (3.8) is so popular (and it is truly better than nothing). If the regulator screws up using (3.8) and imposes too stringent a price cap, he can always adjust $M$ in later periods to make up for the screw up. If the regulator imposes too lax a price cap, he can always adjust $W$ in later periods to make up for the screw up.
Asset Valuation and Productivity–based Regulation

This is often referred to as the ‘differential of a differential’ X factor formula. What equation (3.9) tells us is that the X factor can effectively be decomposed into three terms. The first differential term takes the difference between the industry’s TFP growth and that for the economy as a whole while the second differential term takes the difference between the firm’s input prices and those for the economy as a whole. Thus, taking just the first two terms, if the regulated industry has the same TFP growth as the economy as a whole and the same rate of input price increase as the economy as a whole then the X factor in this case is zero. If the regulated industry has a higher TFP growth than the economy then X is positive, all else equal, and the rate of allowed price increase for the industry will be less than the CPI. Conversely, if the regulated industry has a higher rate of input price increase than the economy as a whole then X will be negative, all else equal, and the rate of allowed price increase will be higher than the CPI. However, the input price index used needs to allow for the presence of sunk costs.

As noted above, the markup growth term could be set equal to zero under normal circumstances but if the target firm was making excessive returns, then this term could be set negative (leading to a higher X factor). Conversely, if the target firm was making less than normal returns, then this term could be set positive (leading to a lower X factor).

This general approach to setting the X factor was used in New Zealand’s thresholds regime for electricity distribution regulation applying from 2004 (see Lawrence 2003).

The traditional productivity–based approach uses observable information on the performance of a number of firms to set regulatory parameters. It has the advantage of being objective and transparent as it relies on observable data and a clearly specified methodology which can be readily reproduced by other analysts. It can also be implemented relatively economically for a large number of firms. It has the potential disadvantage that it may not be able to take adequate account of firm–specific circumstances or, as currently applied, allow for FCM.

This is an interesting point; i.e., does productivity based regulation allow for capital maintenance? This is a tricky business. Initially, before the price cap is imposed, the allowable cost base will be consistent with financial capital maintenance provided the user costs (or user charges for sunk cost assets) use the appropriate opportunity cost of capital as we show in the Appendix (for any consistent scheme of depreciation). But then when we impose the price cap and we include all types of capital services, both contestable and sunk, into the cost base, then if the firm does not beat the productivity factor, its capital will be impaired to some extent. I do not think that we can leave capital services out of the cost base and apply productivity based regulation.

It can thus be seen that there are a number of broad similarities between the traditional productivity analysis approach and the building blocks approach to determining regulated firm costs and revenues. For instance, the building blocks approach treats revaluation gains as income while the productivity analysis typically treats capital gains as income true; insofar as there are revaluation terms in the user cost or charge formula—so what is the difference here? Not clear to me, although these may be smoothed to varying degrees. However, there are some important differences. The traditional productivity approach typically rolls the asset
base forward using the asset price index rather than the general inflation rate. While the concept of financial capital maintenance (FCM) has received some consideration in productivity analysis, this has typically been in the context of the definition of income in macro level studies (see Diewert 2008). The concept of FCM needs to be integrated with productivity analysis at the firm level if we are to reconcile productivity analysis and building blocks regulation satisfactorily.

While the productivity–based approach to regulation incorporates allowance for adjustment for excess profits (in the form of the M or monopolistic markup factor), there is currently no accepted theoretical structure around this term or how it should be calculated. And existing productivity analysis has not addressed the issue of how to treat sunk costs. Sunk assets are assumed to be interchangeable with new investment and the framework is, in many ways, more consistent with perfect contestability than one where sunk costs are present. In particular, little work has been done to integrate the concept of FCM into productivity analysis at the regulated firm level and to assess the implications of this for the treatment of sunk costs and the calculation of excess returns.

One of the key articles addressing the link between productivity analysis and regulation and the role of non–marginal cost pricing – Denny, Fuss and Waverman (1981)¹ – was written before price cap regulation was introduced and concentrated on telecommunications where sunk costs are likely to be less of an issue than in energy networks. It is thus an opportune time to update and extend this analysis to address the issues currently confronting regulators.

3.4 Productivity–based regulation in the presence of sunk costs

The traditional analysis presented in section 3.3 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 given by equation (2.1) could be used as the price for a capital input. However, in many regulated network 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. The usual user cost methodology is thus not applicable in this context. The existence of sunk cost assets greatly complicates the regulator’s responsibilities and changes the nature of some key regulatory theory findings.

In appendix A we develop the theory of regulation in the presence of sunk costs. Most previous regulatory theory 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 and have ignored the sunk cost nature of network investments. For example, in the seminal paper by Bernstein and Sappington (1999), their objective was to define a regulatory regime which would lead to the smallest possible rate of proportional growth in the prices of regulated products, while maintaining the solvency of the

¹ Denny, Fuss and Waverman (1981) show how the productivity index can be decomposed into effects due to departures from marginal cost pricing, nonconstant returns to scale, technical change, and effective rate–of–return regulation.
regulated firm. While this seems likely to be a welfare enhancing activity, 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 approach inevitably involves the use of more demanding mathematical analysis but provides a much higher level of rigour. In appendix A we embed the regulated firm in a small general equilibrium model of an open economy where the role of the regulator in the model is to improve the welfare of households in the economy. In this section we will not attempt to summarise the extensive and complicated mathematical analysis presented in the appendix. Rather, we will concentrate on the key findings of the analysis.

Starting with a relatively simple general equilibrium model, we are able to verify some key regulatory theory findings including that 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.

Introducing sunk costs means that we can no longer use the standard user cost equation (2.1) and the total cost function in deriving parameters for optimal regulation. Rather, it is necessary to use operating expenditure (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 fixed quantity of capital stock components. In other words, we need to recognise that the firm’s relevant decision making options each period are to alter its level of opex given the quantity of sunk investments it has that period. Good! It can opt to change the level of sunk investments gradually over time by undertaking additional investment or allowing the existing stock to run down but it cannot treat capital stocks as freely variable from period to period as has been the implication of past theory developed in this area.

Instead of the equation (2.1) user cost playing a key role, we now have a user benefit defined as the negative of the partial derivative of the opex cost function with respect to the sunk cost capital stock playing an analogous role. The (discounted) sum of these anticipated 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.

More generally, the full model of optimal regulation developed requires too much information for the regulator to be able to implement it in its entirety. In addition to information on the partial derivatives of the period by period opex cost functions with respect to sunk cost capital stock components, we would also require information on the partial derivatives of the period by period opex cost functions with respect to regulated outputs (ie marginal opex costs) and consumer intertemporal substitution matrices between regulated and unregulated products.

Since the information required to implement optimal regulation is difficult to obtain, simpler methods of regulation that are not fully optimal, like price cap regulation, will have to be used in practice. Fortunately, price cap regulation can be modified to accommodate both sunk costs and financial capital maintenance.
We now look at second best regulatory solutions which are less informationally demanding, starting with a simple method of price cap regulation for a single firm that relies on information on that firm only.

3.4.1 Price cap regulation of a single firm

The approach we take in appendix A involves obtaining an expression for the rate of change of the regulated firm’s excess profits, $\Pi'(t)$, which has a term involving the rate of change of prices for regulated outputs, $p'(t)$. For simplicity, we initially move all of these prices in a proportional manner and determine this rate of proportional movement by setting $\Pi'(t)$ equal to a target rate of change which determines the price cap for the following period.

We derive a simple price cap formula which involves a price index and quantity for opex (or variable inputs), a price index for the amortisation amounts allowed by the regulator for sunk cost capital stock components and the quantity of sunk cost capital, a measure of the anticipated rate of opex 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.

The full formula for the price cap or allowable rate of increase in regulated prices, $\alpha'(t)$, is as follows:

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

where $\beta$ is a scalar relating to the desired rate of change of excess profits, $w$ is the opex price, $z$ is the opex quantity, $P_k$ is the per unit amortisation charge for sunk cost capital allowed by the regulator, $k$ is the quantity of sunk cost capital, $\tau$ is the rate of technical progress for opex inputs, $C_z$ is the cost of opex inputs, $\mu$ is marginal cost, $y$ is the output quantity, $\pi$ is the marginal user benefit of sunk cost capital, $R$ is revenue and $t$ refers to time.

We can simplify (3.10) by neglecting the last two terms and assuming that excess profits are currently close to zero and the regulator wants to keep them close to zero (so that $\beta=0$). By using the Divisia index method we can further simplify the price cap formula to arrive at the following:

\[
\alpha'(t) = \left[ C_z(t)/R(t) \right]w'(t) + \left[ C_k(t)/R(t) \right]P_k'(t) - \left[ C_z(t)/R(t) \right]\tau(t)
\]

where the subscript D refers to a Divisia index of the relevant variable.

Thus, if excess profits are close to zero, implementation of the price cap can be simplified to the sum of the rate of opex price change weighted by the share of opex costs in revenue and

---

2 The notation adopted is $f'(t)$ is the derivative of the function $f$ with respect to time which is its rate of growth.

3 With increasing returns to scale in the regulated sector, we would expect the components of $p(t)-\mu(t)$ to be predominantly positive (ie prices exceed marginal costs) 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 fixed capital stock components $k(t)$ are likely to remain roughly constant and hence $k'(t)$ is likely to be small.
the change in approved amortisation charges weighted by the share of amortisation charges in revenue less the rate of technical progress for opex weighted by the opex share in revenue.

This is broadly equivalent to the output price cap being an index of input prices less the rate of technical change (or IPI–X). However, note that in this case we have to focus on technical change applying to opex inputs and weight this by the share of opex in total revenue in recognition of the sunk cost nature of capital inputs. The other major difference between the price cap in (3.11) and traditional price cap formulae is that allowed per unit amortisation charges replace the capital goods price index in the price cap formula when there are sunk costs. This is because the conventional user cost formula in (2.1) cannot be applied to sunk costs because there is no fluid second-hand market where these assets are freely traded.

The price cap formula (3.11) is simple enough to be implementable provided that the regulator can make forecasts for the overall rate of increase in variable input prices and for the anticipated rate of opex technical progress. Note that the regulator will be able to construct an index of allowable amortisation charges since the regulator determines these allowable charges. Typically, forecasts for opex technical progress would be made on the basis of past rates of opex technical progress in the industry. However, there is no guarantee that future rates of opex technical progress will mirror past rates. Note also that formula (3.11) requires estimates of future opex technical progress rather than either total or opex partial factor productivity.

While formula (3.11) is implementable, it needs to be borne in mind that it is only a rough approximation to the full price cap formula in (3.10) because it assumes excess profits are close to zero and it ignores the deviation of output prices from marginal costs and of allowed amortisation charges from marginal user benefits.

3.4.2 Price cap regulation of a single firm using productivity estimates and the CPI

As noted in section 3.3, most applications of price cap regulation use productivity estimates rather than estimates of the rate of technical change in determining the X factor and use the CPI rather than an index of the firm’s input prices.

TFP growth of the regulated firm, $\tau'(t)$, is traditionally defined as an index of output growth minus an index of input growth. Input growth is defined to be a share weighted sum of the indexes of variable input and sunk cost capital services input indexes using (observable) amortisation prices as weights so that TFP growth is:

\[
\tau'(t) = y_D'(t) - s_z(t)z_D'(t) - s_k(t)k_D'(t)
\]

where $y_D$ is an output index formed using revenue weights and the input cost share weights, $s_z(t)$ and $s_k(t)$, are the shares of opex and allowable amortisation charges in total cost, respectively.

In their seminal article on price cap regulation, Bernstein and Sappington (1999) showed that if excess profits were zero and the regulator wished to keep them at zero then the price cap for the regulated firm simplifies to:

\[
\alpha'(t) = s_z(t)w_D(t) + s_k(t)p_D(t) - \tau'(t).
\]

This is the well known result that the price cap should equal an index of the growth in input
prices less an index of the growth in TFP (broadly equivalent to equation 3.4 above and analogous to equation 3.11 which used technical change rather than TFP).

However, as we show in appendix A, where excess profits are not zero, the price cap formula using TFP becomes more complicated and is given by:

\[
(3.14) \alpha'(t) = \left[ \frac{C(t)}{R(t)} \right] \left[ s_D(t)w_Z(t) + s_k(t)P_kD'(t) - T'(t) \right] + \left[ \frac{\Pi'(t)}{R(t)} \right] - \left[ \frac{\Pi(t)}{R(t)} \right] y_D'(t).
\]

Equation (3.14) recognises that if excess profits are non-zero then total revenue will not equal total cost and this has to be allowed for in setting the price cap.

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.

More generally, where there are non-zero excess profits and sunk costs then TFP growth may not be a good proxy for technical change. The relationship between TFP and opex technical change – the measure required for price cap regulation in the presence of sunk costs – is as follows:

\[
(3.15) T'(t) = \tau(t)s_Z(t) + \left[ p(t) - \mu(t) \right] y'(t)/C(t) + \left[ P_k(t) - \pi(t) \right] k'(t)/C(t) - \left[ \frac{\Pi(t)}{C(t)} \right] y_D'(t).
\]

The first term on the right hand side of (3.15) shows that opex technical progress is definitely a contributor to the rate of TFP growth, T'(t). But the remaining terms on the right hand side of (3.15) show that TFP growth encompasses more than just technical progress. The term \( [p(t) - \mu(t)] y'(t)/C(t) \) depends on the deviations of the output prices p(t) from the corresponding marginal costs \( \mu(t) \) and these interact with output growth rates, \( y'(t) \). Similarly, the term \( [P_k(t) - \pi(t)] k'(t)/C(t) \) depends on the deviations of the allowed amortisation charges \( P_k(t) \) from the corresponding marginal user benefits \( \pi(t) \) and these interact with capital stock growth rates, \( k'(t) \). It will be difficult to project past contributions to TFP growth that are due to these terms into the future. 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 the regulator can then set an appropriate price cap using formula (3.13) or the more accurate formula (3.14). The factors beyond the firm’s control in equation (3.15) which relate TFP growth for a single firm to technical progress are likely to remain relatively constant for a single firm. These factors relate to differences between prices and marginal costs and differences between allowable amortisation charges and marginal user benefits of sunk cost capital. If they are not constant then, in the case of a single firm, the regulator can make adjustments to the price cap to take this into account. As will be seen in the following section, more problems arise with the use of the TFP proxy for technical change where multiple firms are being regulated.
As noted in section 3.3, regulators often use the CPI as the inflation measure in price cap setting rather than a direct estimate of the firm’s input prices. Where CPI–X regulation is used and there are sunk costs, 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 allowed amortisation charges growth each weighted by their respective shares of their cost in revenue plus a nonzero profits adjustment term less a rate of change in regulated profits term. As explained in detail in appendix A, the equivalent to the price cap formula (3.8) becomes the following:

\[
(3.16) \alpha'(t) = P_E'(t) + \left\{\left[\frac{C(t)}{R(t)}\right]\left[s_z(t)w_D'(t) + s_k(t)P_kD'(t)\right] - W_E'(t)\right\}
- \left\{\left[\frac{C(t)}{R(t)}\right]\left[T'(t) - T_E'(t)\right] + \left[\Pi'(t)/R(t)\right] - \left[\Pi(t)/R(t)\right]\right\}y_D'(t)
= P_E'(t) - X(t)
= CPI(t) - X(t)
\]

where the subscript E denotes an economy–wide variable, \(P_E'(t)\) is the change in economy–wide output prices (often approximated by the CPI), \(W_E'(t)\) is the change in economy–wide input prices and \(T_E'(t)\) is economy–wide TFP growth. Equation (3.16) is thus the sunk costs counterpart to the traditional ‘CPI minus X Factor’ regulatory price cap formula and the X factor equivalent to the ‘differential of a differential’ formula (3.9) becomes:

\[
(3.17) X(t) = \left\{\left[\frac{C(t)}{R(t)}\right]\left[T'(t) - T_E'(t)\right] + \left[\Pi'(t)/R(t)\right] - \left[\Pi(t)/R(t)\right]\right\}y_D'(t)
+ [\Pi(t)/R(t)]y_D'(t) - \Pi'(t)/R(t)
= TFP differential growth rate term + input price differential growth rate term
+ nonzero profits adjustment term – rate of change of regulated profits term.
\]

The first term in (3.17) is the differential rate of TFP 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 costs, C(t), to its revenues, R(t). The second term is the differential rate of growth of input prices in the rest of the economy, \(W_E'(t)\), less \(C(t)/R(t)\) times a share weighted rate of the growth of opex 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, 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 ratio of variable cost to total cost and the ratio of allowable amortisation costs to total cost, respectively.

The last two terms on the right hand side of (3.17) involve the level of excess profits of the regulated firm, \(\Pi(t)\), and the rate of change of excess profits, \(\Pi'(t)\). These two terms are also present in the simpler price cap formula (3.14) (which did not involve the rest of the economy). If the excess profits of the regulated firm are not close to zero, then if excess profits were markedly 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 excess profits 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 excess...
profits are 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. In the case where excess profits are positive, typically the regulator will set \( \Pi'(t) \) in the price cap formula (3.16) equal to a negative number, which will cause the proportional change in regulated prices, \( \alpha'(t) \), to become smaller, i.e. under these conditions, the price cap will become more stringent.

There are a number of differences between the traditional ‘differential of a differential’ \( X \) factor formula (3.9) which does not allow for sunk costs and the \( X \) factor formula (3.17) which allows for sunk costs. The main difference is the replacement of the Jorgenson user cost of capital (which incorporates the capital price index) by the regulator–allowed per unit capital amortisation charge when calculating input prices for the regulated firm. Other differences are adjusting the regulated firm’s productivity and input prices by the ratio of costs to revenue and the inclusion of more structured change in ‘monopolistic markup’ terms in (3.17).

We turn now to look at the complications introduced by regulating several firms using a common \( X \) factor.

3.4.3 Price cap regulation of multiple firms using a common \( X \) factor

The approaches to the derivation of price cap formulae discussed in this section up till now have concentrated on the case where only a single firm is being regulated. In the case of a single firm it is less critical whether we use a price cap formula involving the rate of opex technical progress or the rate of TFP growth – as discussed above, the factors included in TFP other than opex technical change are likely to remain relatively constant for a single firm or else changes can relatively easily be adjusted for. 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 (3.15) can contain terms which are beyond the control of the regulated firm.

As shown in equation (3.15) TFP growth for a single firm depends not only on the firm’s rate of technological improvement (which is presumably an industry wide effect) but it also depends on the firm excess profits and factors which are largely beyond its control, namely the gaps between the regulated prices that the firm faces and its corresponding marginal costs and the gaps between the allowable amortisation costs for sunk cost capital stock components and the corresponding marginal user benefits of sunk capital. Thus, while basing a price cap on a forecast of future industry wide rates of technological progress (i.e using equation (3.10) or the simplified (3.11)) seems appropriate, caution will be required in basing a price cap on a forecast of future industry wide rates of TFP growth for all of the regulated firms. This is because there will generally be substantial differences in the last three factors on the right hand side of (3.15) across the firms – namely, the extent to which prices exceed marginal costs, the extent to which allowed amortisation charges differ from user benefits for sunk capital and excess profits. If there are differences in these three factors across the individual
firms then application of a ‘one size fits all’ rate of TFP growth may not be appropriate. The single firm focus has also 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 and lower productivity levels. For example, if the group being regulated are electricity distribution businesses and some distribution businesses are located in areas of high 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 ideally either group the regulated firms into peer groups who face roughly similar operating environments or adjust the price caps for each firm (or groups of firms) according to differences in operating environments.

If a common rate of productivity growth is to be used in setting the price cap when regulating a group of firms using productivity–based regulation, then output specification becomes critical since different output concepts can lead to very different estimates of both technical progress and TFP growth. The output concept used is less 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. In particular, it is necessary to move beyond the use of TFP measures based on revenue weighted outputs. Rather, it is necessary for the output measure to capture exactly what regulated services are being provided by the firms in the group, independently of the institutional and historical factors that determine how the firms happen to charge consumers. 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.

In appendix A we illustrate how it is important in measuring TFP growth for the regulation of multiple network businesses to use an output measure that captures the system capacity provided as well as elements of peak demand and throughput. This is consistent with the approach to productivity measurement used in the New Zealand EDB thresholds regime (see Lawrence 2003).

3.4.4 Productivity–based regulation and financial capital maintenance

As noted in section 3.2, an important element of building blocks regulation has been the use of ex ante FCM in setting the price cap. The rationale for adopting ex ante FCM as a regulatory principle is that it is consistent with ensuring efficient investment occurs as there can be an expectation that the value of invested capital will be maintained in real terms over the life of the investment.

Most previous productivity studies have indexed the capital stock by the capital goods price index rather than the CPI and have used the Jorgenson user cost formula from equation (2.1) in calculating the annual cost of using capital inputs. As such, this approach has been more
consistent with the criterion of operational capability maintenance than FCM. [Again, you should note that the Jorgensonian approach is consistent with FCM if the right opportunity cost of capital is used and the depreciation allowances “add up”.] Under operational capability maintenance the emphasis is on being able to maintain the capacity of the asset to contribute to the production process rather than on maintaining the value of invested capital in real terms.

In appendix A we demonstrate that in the presence of sunk costs the Jorgenson user cost no longer applies because sunk assets, by definition, cannot be freely traded in a second-hand market. Rather, the appropriate annual cost of capital inputs becomes the series of amortisation charges for the capital good approved by the regulator. These approved amortisation charges should ideally be the marginal user benefits from the sunk capital. They can be readily structured to achieve FCM.

A range of asset valuation methodologies can be consistent with financial capital maintenance, provided that the allowed cost of capital interest rates are equal to the firm’s opportunity cost of financial capital. 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 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. Right on!

In the following section we turn to examine evaluation criteria that might be used to assess the three alternative asset valuation methodologies.
4 EVALUATION CRITERIA FOR ALTERNATIVE ASSET VALUATION METHODS

4.1 Proposed and existing evaluation criteria

A number of criteria have previously been proposed for assessing network asset valuation methodologies in New Zealand. In a report for the Ministry of Commerce, Ernst and Young (1994, p.1) used the following criteria to assess alternative valuation methodologies for regulatory purposes:

• ability to reveal monopoly pricing behaviour;
• consistency in providing benchmarked comparisons of value; and
• low compliance cost.

Ernst and Young (1994, pp.4–5) went on to list additional criteria that should be considered in a more general business context as follows:

• business sustainability (ie allow the firm to set prices at a level sufficient to allow it to maintain its operations);
• a reasonable return (commensurate with perceived risk and sufficiently attractive for new investment where business expansion is warranted); and,
• efficiency (ie the firm should have incentives to utilise the lowest cost asset base and the most efficient configuration of assets);

Carpenter and Lapuerta (2000, pp.2–3) list a number of criteria they suggest are relevant for the pricing of monopoly infrastructure services in New Zealand. These criteria have direct implications for asset valuation and are as follows:

• ‘A ‘fair and reasonable’ price for monopoly services, whether negotiated between parties or enforced by formal regulation, is one that is low enough to prevent the earning of monopoly profits, but high enough to generate a revenue stream that fully recovers the invested (or acquired) capital over the life of the assets involved, plus a normal return on that investment consistent with the business risks of the enterprise.

• ‘To ensure that prices do not reflect monopoly charges, any pricing methodology must be designed to satisfy the ‘net present value’ or ‘NPV Test’ for the relevant assets over time. That is, the present value of capital charges collected in prices over time should not exceed the present value of the capital outlays associated with the construction or purchase of the assets involved.

• ‘An important component in the pricing methodology is the opening asset value. The most popular methods for determining the opening asset value worldwide include historical/original costs, replacement costs and what is variously referred to as acquisition, establishment or flotation value. The key
to determining which of these methods is most appropriate in a given case depends on the context in which investor expectations of future cash flows were formed.

- ‘In cases where the assets have been previously privatised, international practice has been to rely on establishment or flotation value for determining the opening asset value for pricing purposes. This has become standard practice in the UK for natural gas, electricity and water assets. Where such assets have always been in private hands, such as the North American utility industries, the practice has been to rely on historical costs or its variants, such as ‘trended original cost.’ In each context these different methods establish ‘fair and reasonable’ prices because the asset valuation method is consistent with the reasonable expectations of shareholders.’

- ‘Using optimized depreciated replacement cost is acceptable as long as the optimized replacement cost values are consistent with shareholder expectations at privatization (i.e., that they bear some relationship to acquisition or flotation value). Tying the initial valuation to acquisition cost or flotation value eliminates the highly subjective problem of assessing the replacement costs of certain assets that either do not depreciate (such as easements or land), or that have extremely long lives and no meaningful alternative uses, and therefore would never be replaced (such as breakwaters and runways).

- ‘Frequently, as we have observed in New Zealand, the owners of infrastructure assets attempt to revalue them upward (above their value at privatization or acquisition) in order to justify higher prices. This is inappropriate. Such upward revaluations constitute additional returns on the assets in the form of capital gains. To avoid violating the NPV test, any upward revaluations must be offset by an adjustment to the depreciation charge when determining prices. This principle applies whether the assets in question tend to depreciate over relatively short lives, or whether they are extremely long–lived or tend not to depreciate (such as land or easements).’

The Commerce Commission (2004, p.22) has subsequently set out four key principles for assessing asset valuation methodologies in the context of information disclosure requirements. Two of these related to the regulatory framework:

- Efficiency: the method should support outcomes that are allocatively, productively and dynamically efficient; and

- Profit measurement: the method should facilitate the identification of excessive profits on a basis that reflects sound regulatory economic principles and practices.

And two related to implementation:

- Cost effectiveness: the method should achieve regulatory valuation objectives at a reasonable cost; and
Asset Valuation and Productivity–based Regulation

• Consistency and accuracy: the method should be consistent and accurate to the extent appropriate to the relevant function under (the then) Part 4A (ie information disclosure, resetting the thresholds and post–breach inquiries or control).

An earlier Commerce Commission paper (2002, p.23) also proposed efficiency, identification of any excessive profits and cost effectiveness as evaluation criteria.

In a recent discussion paper the Commerce Commission (2008c, pp.46–56) has reinforced the importance of economic efficiency, workable competition, adoption of the NPV=0 or FCM principle and the importance of consistency, cost effectiveness and predictability balanced by flexibility as factors evolve over time as implementation principles.

In addition to the above criteria it is helpful to consider relevant higher order principles for effective regulatory governance. These principles are also important to achieve the various aspects of economic efficiency. A recent World Bank report on ‘Evaluating Infrastructure Regulatory Systems’ (Brown, et al 2006, p.7) proposes three useful higher order or meta–principles for all infrastructure regulatory systems (including transitional regulatory governance systems) as follows:

• Credibility: Investors must have confidence that the regulatory system will honor its commitments.
• Legitimacy: Consumers must be convinced that the regulatory system will protect them from the exercise of monopoly power, whether through high prices or poor service, or both.
• Transparency: The regulatory system must operate transparently, so that investors and consumers “know the terms of the deal”. This last principle is needed to ensure that both investors and consumers perceive the system as fair.

4.2 Economic efficiency as a fundamental and encompassing criterion

4.2.1 Definition

Economic efficiency is an encompassing and important criterion because, when properly defined, it encompasses aspects that take account of efficient production, efficient investment, efficient outcomes for consumers and environmental effects. The achievement of economic efficiency will maximise the value of the community’s resources.

Economic efficiency is thus a critical criterion as a high level principle for evaluating alternative asset valuation methodologies. It can have implications for ‘fairness’ to producers and consumers in terms of balancing various efficiency considerations but, by itself, it does not encompass equity considerations, eg whether a particular asset valuation and associated regulated price outcome has an adverse impact on the distribution of income. Equity considerations are best addressed by separate policy instruments and are not considered further in this report. If equity considerations were to be introduced in setting prices then it is likely they would mainly concern the time profile of prices and the structure of prices for different users rather than the asset value and associated average price level that are appropriate for achieving economic efficiency.
The three dimensions of economic efficiency that are commonly discussed in the context of economic regulation and competition policy issues are productive efficiency, allocative efficiency and dynamic efficiency. Productive and allocative efficiency are often considered as static concepts, however, they will have dynamic aspects as well. But it is useful to consider the concepts separately as there are circumstances where trade-offs have to be made in order to maximise overall economic efficiency. To help understand the various dimensions of economic efficiency and how they might impact on valuation principles, a brief explanation is provided below of allocative, productive and dynamic efficiency.

### 4.2.2 Allocative efficiency

Allocative efficiency involves allocating scarce resources to their most highly valued uses. Allocative efficiency is achieved when the marginal benefit to society of the last unit produced is equal to the marginal cost to society. The marginal benefit of a product to society is reflected in the value of the market demand for the last unit produced of the product. The marginal cost of producing a product is the cost of the resources used in producing the marginal unit of production.

The main concern in relation to allocative inefficiency in utility regulation relates to the potential use of market power. In situations where there is unregulated monopoly or market power, the market price will exceed marginal cost. As a result the actual output that is sold will be smaller than is optimal from the perspective of society. This is because the optimal level of output at each point in time is one where price just equals (short run) marginal cost but, because of the use of unregulated market power, there will be some consumers prepared to pay at least the marginal cost for a unit of output but who cannot purchase it at the prevailing price. However, in situations where marginal costs are declining (which will occur when there are unexhausted scale economics) it will not be possible for the firm to recover all its costs and earn a normal commercial rate of return if regulated prices are set to equal marginal costs. To the extent that this occurred it would compromise the achievement of dynamic efficiency as reflected in the need to facilitate efficient investment decisions. Thus, typically the need to set a price to recover costs and earn a normal commercial rate of return means there is still some loss of allocative efficiency, ie there is necessarily some trade-off with the objective of providing appropriate investment incentives. This loss of allocative efficiency is considered to be justified by considerations of the need to ensure appropriate investment incentives consistent with achieving dynamic efficiency, as discussed below.

There are also dynamic aspects to allocative efficiency. As recognised by the Commerce Commission (2002, p.24–25) and as formally demonstrated in appendix A of this report, there are an infinite number of expected net cash flow streams and associated price profiles that enable the exante FCM principle to be achieved. However, provided FCM is adopted, the decision regarding different price profiles is effectively a decision about what depreciation profile to adopt. This will depend on such issues as the extent of utilisation of the network, the confidence of investors that FCM will continue to be adhered to and the intertemporal impacts on consumers. Excellent! Thus, to the extent that investors are confident the regime will maintain the principle of ex ante FCM, decisions about the pattern of cash flows can be considered a separate issue. However, this assumption points to the importance of regulatory credibility as a principle for asset valuation.
An additional aspect of allocative efficiency relates to ensuring the quality of the service meets consumers’ preferences and is reflected in the price of the service for a defined quality level. As noted by the Commerce Commission (2002, p.25) the quality of electricity lines services can also be affected by the asset valuation methodology adopted. At one extreme a methodology that does not provide some check on the prudence of investments might provide incentives to over-build and raise quality inefficiently high. While at the other extreme, a regime that entails significant uncertainty may raise the cost of capital which, if not recognised in the allowable rate of return, may lead to under-investment and quality that is lower than optimal.

4.2.3 Productive efficiency

Productive efficiency simply means that any given quantity (or, for multi-product firms, bundle of quantities) of output is produced at minimum cost. In other words, productive efficiency requires that a firm uses the lowest cost technology and also that it uses inputs in proportions that will minimise costs. Productive efficiency is also sometimes referred to as cost efficiency.

Profit incentives and competitive forces operate to help ensure productive efficiency. From a regulatory perspective, productive efficiency may be a concern where circumstances preclude the effective operation of profit incentives or competitive forces to minimise costs. Concerns about productive inefficiency can lead to the adoption of benchmarks based on relevant efficient costs in certain regulatory contexts.

4.2.4 Dynamic efficiency

Dynamic efficiency encompasses the intertemporal aspects of efficiency. Investment and innovation play an important role in dynamic efficiency issues, although they are not the only concern. Dynamic efficiency also embodies other issues to do with the allocation and use of resources across time. The concept of efficient cost reduction is an example of productive efficiency in a dynamic context as it contains an investment component. Efficient cost reduction requires equality of the marginal social costs and marginal social benefits of investments over time.

In economic regulation a key issue in achieving dynamic efficiency is to ensure there are adequate incentives to undertake optimal investment in projects with natural monopoly characteristics and to facilitate innovation in products that use these facilities. If regulatory decisions lead to too low a return on the investment, future investment and innovation may be deterred. However, setting regulated prices too high may delay or deter the development of products that use the facility which could also harm dynamic efficiency.

4.2.5 Trade-offs and supplementary principles

At times trade-offs will need to be made in considering the relative importance of allocative, productive and dynamic efficiency. In particular, the concept of excess profits derives from a consideration and balancing of both allocative and dynamic efficiency, but its exact definition and identification at least cost depends on regulatory principles and practices which can be defined as supplementary criteria. In the context of asset valuation, the supplementary criteria
includes a number of principles relating to the credibility and cost effectiveness of the regulatory process.

### 4.3 Assessment of proposed and existing criteria

This section provides a brief assessment the criteria for asset valuation suggested by Ernst and Young (1994), Carpenter and Lapuerta (2000) and the Commerce Commission (2004) recognising the importance of high order principles for effective regulatory governance and economic efficiency as outlined in the foregoing sections.

**Ernst and Young (1994)**

1. **Ability to reveal monopoly pricing behaviour.** Allocative economic efficiency requires that monopoly pricing behaviour be minimised and it is important to be able to define and implement an appropriate methodology for asset valuation that will be cost effective and reliable in detecting and helping to avoid monopoly power. This criterion is very similar to the Commerce Commission’s criterion for profit measurement, both of which can be interpreted as implementation principles for achieving economic efficiency.

2. **Consistency in providing benchmarked comparisons of value.** This principle is important for achieving economic efficiency as well as for the regulatory governance objectives of credibility and legitimacy. Inconsistent comparisons have little meaning and enable unwarranted discretion in regulatory decision making. This criterion is also adopted by the Commerce Commission as an implementation principle.

3. **Low compliance cost.** This is important for minimising the costs to society as a whole, as an aspect of economic efficiency, and is important for effective regulation. It is also reflected in the Commerce Commission’s criterion of cost effectiveness as an implementation principle.

4. **Business sustainability.** This is defined as allowing the firm to set prices at a level sufficient to allow it to maintain its operations. However, this is only economically efficient if sustainability is in the interests of society which is not always be the case. Thus, there is not an economic efficiency rationale for business sustainability per se and this criterion is not accepted as defined here.

5. **A reasonable return.** This is defined to be commensurate with perceived risk and sufficiently attractive for new investment where business expansion is warranted. This is an aspect of dynamic efficiency. However, it is effectively covered in terms of detecting excess profits or unwarranted monopoly pricing behaviour.

6. **Efficiency.** This is defined as ensuring the regulated firm has incentives to utilise the lowest cost asset base and the most efficient configuration of assets. This is consistent with achieving productive efficiency. However, care needs to be taken where its application conflicts with dynamic efficiency. For example, in considering how to value sunk assets, an approach that only looked at forward looking costs would send signals to investors that, once sunk, there is a risk that revaluations could impact adversely on the scope to recover the cost of capital.
1. **A ‘fair and reasonable’ price for monopoly services.** Carpenter and Lapuerta (2000, p.3) provide a useful definition and justification of ‘fair and reasonable’ as follows:

   ‘A “fair and reasonable” price is a price that fairly balances the interests of buyers and sellers. From the seller’s point of view, the price must be high enough to generate a revenue stream that permits it to recover its invested capital over the lifetime of its assets, plus a normal return on that investment consistent with its business risks. From the buyers’ point of view such a price must not be excessive in the sense that it permits the service provider to recover more than its costs including a normal return, or that it contains an element of monopoly profit. When speaking of the provider’s “invested capital”, we are referring to the actual investment made by the provider in constructing the facilities (“original cost”) or in acquiring them (“establishment” or “flotation value”).

   ‘This interpretation of a “fair and reasonable” price is consistent with international legal precedent and with modern economic and finance theory. Indeed, the discounted cash flow (DCF) methodology one can employ to establish prices based on a fair return on investment is the same method used commercially to evaluate whether projects or businesses are sufficiently profitable to warrant investment over other investment alternatives.’

This definition of a fair and reasonable price reflects a balancing of allocative and dynamic efficiency effects and the interests of consumers and producers. It can be implemented by applying the concept of ex ante FCM which can alternatively be expressed as an ex ante NPV test as also recommended by Carpenter and Lapuerta in their second principle.

2. **To ensure that prices do not reflect monopoly charges, any pricing methodology must be designed to satisfy ‘NPV = 0 Test’ for the relevant assets over time.** In this respect Carpenter and Lapuerta (2000, p.6) note:

   ‘Capital charges that are consistent with the “NPV test” allow the monopoly service provider to expect a “fair” return on investment, i.e., the same expected return as would apply in a competitive market from investments of equivalent risk.’

Note that the NPV = 0 test assumes a reasonable estimate of the risk–adjusted opportunity cost of capital (return on capital) for the discount rate used to discount the cash flows, ie the same expected return as would apply in a competitive market (should it exist).

3. **The key to determining which of the various asset valuation methods is most appropriate in a given case depends on the context in which investor expectations of future cash flows were formed.** This principle is supplementary to the ‘fair and reasonable’ price principle and the NPV test and recognises that the fair value from an investor’s perspective depends on what the investor expected at the time the investment was made given existing and announced government policies and regulations and
reasonable expectations about future policies and regulations. It can be implemented by recognising that an investor should be entitled to expect NPV=0 on the actual cost incurred provided the investment was a reasonably prudent expenditure at the time.

4. To avoid violating the NPV test, any upward revaluations to assets must be offset by an adjustment to the depreciation charge when determining prices. This principle follows from acceptance of the ‘fair and reasonable’ price principle and its implementation in terms of an NPV=0 test.


1. Efficiency: the method should support outcomes that are allocatively, productively and dynamically efficient. This principle is consistent with widely accepted economic theory which establishes economic efficiency as the most important high level goal, provided equity and other social goals can be effectively addressed by other policy instruments. It is considered to be the most important high level principle for economic regulation and for the valuation of assets for regulatory purposes.

2. Profit measurement: the method should facilitate the identification of excessive profits on a basis that reflects sound regulatory economic principles and practices. As noted above, it is important to be able to define and implement an appropriate methodology for asset valuation that will be cost effective and reliable in detecting and helping to avoid monopoly power.

3. Cost effectiveness: the method should achieve regulatory valuation objectives at a reasonable cost. As noted above, this is also an aspect of economic efficiency from the perspective of society as a whole and is important for effective regulation.

4. Consistency and accuracy: the method should be consistent and accurate to the extent appropriate to the relevant function. Inconsistent and inaccurate methodologies clearly reduce the credibility of the regulatory process and jeopardise the consistent achievement of economic efficiency.

Additional Criteria

Consideration of regulatory governance principles and experience with asset valuation determinations suggests some additional criteria for assessing asset valuation methods. Consideration of statistical principles, in particular those that underlie the estimation of capital for the purposes of productivity measurement, also imply an additional criterion. The additional criteria and their rationale are summarised below:

1. Regulatory commitment and predictability. Regulatory commitment and predictability in the context considered here means the extent to which the asset valuation methodology facilitates regulatory commitment to achieving higher order economic efficiency while also minimising regulatory discretionary decision making in the determination of asset values. This principle is important for signalling a strong commitment by the regulator to specified objectives and to the predictability of decisions. This is important for ensuring dynamic efficiency. In particular, it will help ensure investors are confident that
the ‘rules of the game’ will not change and, together with other principles, reduce uncertainty and risk premiums and facilitate efficient investment.

2. **Transparency.** Transparency in the asset valuation methodology refers to the extent to which asset value estimates can be readily understood from the documentation as well as being able to be replicated by an independent audit with minimal judgemental assessments. This is important for both regulatory credibility and economic efficiency. It is particularly important in terms of the higher order regulatory principle of legitimacy from the perspective of consumers. If the process or methodology is not transparent, consumers will be concerned about ‘regulatory capture’ by industry interests. Experience has also shown that transparency is important for giving all firms confidence of even-handed treatment in regulatory decisions. This will reduce uncertainty and the cost of capital.

3. **Ready conversion of asset values from current to constant prices and vice-versa.** This is so that the asset valuation data could be used in forming both the annual capital user cost (rental or factor price of capital) as well as the capital input quantity (if need be) in total factor productivity measurement.

### 4.4 Preferred criteria for assessing asset valuation methodologies

Our preferred criteria for assessing asset valuation methodologies based on the considerations and principles set out above are as follows:

1. **Supports economic efficiency.** The asset valuation methodology should support outcomes that are allocatively, productively and dynamically efficient. This is an overarching principle supplemented by the other principles set out below.

2. **Facilitates FCM for prudent investment.** The asset valuation methodology should be effective in avoiding excess profits on an exante basis which is equivalent to allowing exante FCM. This principle is considered essential for achieving overall economic efficiency and a fair ‘average' price for consumers and firms. This principle is a more precise definition of ensuring ‘excess profits' are avoided on a basis that reflects sound regulatory economic principles and practices.

3. **Cost effectiveness.** The asset valuation methodology should achieve regulatory valuation objectives at a reasonable cost.

4. **Consistency and accuracy.** The asset valuation methodology should be consistent and accurate to the maximum extent appropriate for the circumstances.

5. **Regulatory commitment and predictability.** The asset valuation methodology should help ensure regulatory commitment and predictability and minimise discretionary decision making but allow some flexibility to take account of changing circumstances.

6. **Transparency.** The asset valuation methodology should be readily understood and be capable of being independently replicated with minimal need for judgemental assessments.
7. Enables ready conversion of asset values from current to constant prices and vice-versa. This principle is relevant for facilitating measurement of capital inputs in total factor productivity measurement.
5 ASSESSING THE ALTERNATIVE ASSET VALUATION METHODS

‘If the valuations are wrong, then the line charges are wrong’
(Ministry of Economic Development 2000)

5.1 Introduction

This section applies the preferred criteria for assessing alternative asset valuation methodologies to the three main asset valuation methods we have been asked to review: optimised deprival value (ODV), depreciated historic cost (DHC) and indexed (for inflation) depreciated historic cost (IHC).

It is worth noting that alternative approaches to asset valuation and capital measurement and their implications for economic efficiency have long been under-researched as highlighted by Joskow (2005, pp. 81–2):

‘Price cap mechanisms are the most popular form of incentive regulation used around the world, in part because this mechanism has been heavily advertised as being a simple alternative to cost of service regulation. There is a lot of loose and misleading talk about the application of price caps in practice. … They are not so simple to implement because defining the relevant capital and operating costs and associated benchmarks is challenging. … Effective implementation of a good price cap mechanism with periodic ratchets requires many of the same types of accounting, auditing, capital service, and cost of capital measurement protocols as does cost of service regulation. Capital cost accounting and investment issues have received embarrassingly little attention in both the theoretical literature and applied work on price caps and related incentive mechanisms, especially the work related to benchmarking applied to the construction of price cap mechanisms. Proceeding with price caps without this regulatory information infrastructure and an understanding of benchmarking and the treatment of capital costs … can lead to serious performance problems.’ (emphasis added)

In this report we assume that ex ante financial capital maintenance (FCM) will be adopted as a regulatory principle. This is an important part of ensuring there is dynamic efficiency and adequate incentives for efficient investment. One of our preferred principles for selecting asset valuation methods is that the method used should be effective in allowing \( \text{NPV} = 0 \) to be implemented on an ex ante basis, which is equivalent supporting the implementation of ex ante FCM. However, as noted in the preceding section, there is a range of economic efficiency considerations that are not captured by this simple rule that need to be considered along with other regulatory and practical considerations.

The implications of the different methods for the time profile of prices also need to be considered. As explained in section 4 and formally demonstrated in appendix A, a range of asset valuation methodologies can be consistent with FCM, including the three methods reviewed here. Consistency with FCM requires an appropriate time profile of amortisation charges so the main difference between the asset valuation methods is likely to be on the
timing of revenue receipts rather than their net present value. However, the time profile of amortisation charges impacts on the time profile of prices and will have intertemporal economic efficiency effects.

Although analytical separation of the valuation of assets and the time profile of charges is possible, in practice each method will imply different price profiles unless major adjustments are made to standard conventions for estimating regulatory depreciation. So, in practice, the choice of asset valuation method will also depend on a range of intertemporal economic efficiency effects when the different methods are implemented using common regulatory depreciation profiles (such as straight-line) and practices. For this reason, the main intertemporal economic efficiency considerations implied by each asset valuation method when standard conventions with respect to allowable depreciation are adopted are also considered in the following assessments.

5.2 Optimised deprival value

Optimised deprival value is a methodology that determines an asset value based on value-to-the owner rules. The origins of the concept as applied to publicly owned or regulated businesses can be traced to the Sandilands (1975) and Byatt (1986) reports in the United Kingdom (Hay and Morris 1993, pp.430–2).

As noted by the Commerce Commission (2004, p.12) ODV was designed to produce valuations for network assets consistent with contestable market outcomes and was first specified in a regulatory context in New Zealand for the valuation of the fixed assets of Transpower. However, it was first used as a regulatory valuation method for lines businesses on the basis that historical book values were considered to be unreliable or unavailable and there was a need for a common methodology for benchmarking purposes. Further, the Commerce Commission (2004, p.12) noted that ODV–based valuations were not required to be used for deriving line charges but they did form the basis for determining ‘excess returns’.

Definition

ODV as applied by the Commerce Commission in relation to network assets in New Zealand is defined as the minimum of optimised depreciated replacement cost (ODRC) and economic value (EV):

\[
ODV = \min [ODRC, EV].
\]

The ODRC is defined as the depreciated cost of replicating the system using modern equivalent asset (MEA) values in the most efficient way possible from an engineering perspective, given the network’s service capability, with depreciation based on the age of the existing assets (Commerce Commission 2004, p.13). As implemented in New Zealand, the optimal network is restricted to the existing network configuration (Commerce Commission 2008b, p.120).

The EV of any network segment is defined as the maximum of the net realisable value (NRV) of the segment and the present value of the notional after–tax cash flows that would be attributable to that segment (limited by the cost of alternatives, and net of any initial investment in working capital and fixed assets other than system fixed assets associated with
the segment). Note that an issue in defining the EV is defining cash flows to determine the present value, as typically in a regulatory context these depend on regulated prices which in turn depend on allowable asset values and allowable returns thus entailing a fundamental circularity problem. However, this circularity problem was resolved for lines businesses by defining maximum prescribed tariff rates for calculating EV.

The optimised depreciated replacement cost (ODRC) method received considerable support in Australia and New Zealand as publicly owned business enterprises where being reformed through a process of corporatisation and, in some cases, privatisation in the late 1980s and through the 1990s. In Australia the approach is described as depreciated optimised replacement cost (DORC).

However, the approach is based on some strict theoretical assumptions and in practice allows considerable discretion in arriving at an asset value for regulated networks. The rationale for adopting ODRC and problems in implementation are reviewed below before assessing ODV more explicitly in terms of the preferred criteria specified in section 4.

The hypothetical efficient new entrant benchmark

The Commerce Commission (2004, p. 27) notes that the ODV method assumes a hypothetical operating environment where the relevant market is contestable and there are no material barriers to entry into that market by an alternative service provider or efficient new entrant. This assumption clearly applies to the ODRC part of ODV but does not strictly apply to the EV component where maximum prescribed tariffs are used to determine the EV and those tariffs are not directly reflective of the costs of an efficient new entrant. In practice, most New Zealand energy network ODVs rest on the ODRC rather than the EV component.

The hypothetical new entrant benchmark refers to a methodology for determining allowable costs for the purpose of regulating prices based on the costs a hypothetical efficient new entrant would face in providing the regulated service. The approach is particularly important and controversial in the determination of allowable capital costs.

Some regulatory authorities have argued that the approach is justified as it is a relevant application of the theory of contestable markets in the valuation of assets. The idea is that a valuation of assets based on an estimate of forward looking efficient capital costs to serve the regulated market will justify a price for the regulated services at which a new entrant would have the incentive to compete for the provision of the regulated services at the regulated price.

For example, the ACCC (1998a) said:

‘A return on replacement cost is the maximum that a monopoly firm could earn in a perfectly contestable market.’

The ACCC (1999, p.39) also provided the following argument in the context of formulating principles to support the depreciated optimised replacement cost (DORC) methodology for the valuation of assets for electricity transmission:
‘One interpretation of DORC is that it is the valuation methodology that would be consistent with the price charged by an efficient new entrant into an industry, and so it is consistent with the price that would prevail in long run equilibrium.

‘The second interpretation is that it is the price that a firm with a certain service requirement would pay for existing assets in preference to replicating the assets.’

The ACCC (1997, pp.28–30) has used similar arguments to justify the use of total service long run incremental cost (TSLRIC) in the context of determining access prices in telecommunications.

The hypothetical efficient new entrant benchmark has also been used in New Zealand, in initially justifying the appropriate valuation methodology for Transpower and subsequently in the Commerce Commission’s (2008b, p.120) ODV Handbook for electricity lines businesses.

The approach has also been advocated based on an appeal to the economic theory underlying Tobin’s Q (Brainard and Tobin 1968 and Tobin 1969). Tobin’s Q is simply the market value of a firm relative to minimum depreciated replacement cost and in long run equilibrium in a competitive market Q should have a value of one. Where Q was in excess of one the theory was that firms would have an incentive to enter or existing firms to expand until, in a long run competitive equilibrium, Q would be driven to one. The Office of the Regulator General in Victoria (1998, p.5) used the theory underlying Tobin’s Q to justify the use of a DORC approach to asset valuation. However, the theory of Tobin’s Q is not so much a precursor to the more formal contestability theory developed by Baumol et al (1988) but rather an approach used in macroeconomics and financial theory in identifying determinants of investment. In addition, Tobin’s Q is a marginal concept relating to incremental decisions rather than a valuation methodology.

In the United Kingdom the ‘Byatt report’ (Byatt 1986, Vol II, pp.98–99) argued that the theory of contestable markets provided a unifying rationale for current cost accounting and what a new producer would have to pay to enter the market, including for assets that once invested are effectively sunk costs. However, as explained by Hay and Morris (1993, p.432), the relevance of the approach depends on the extent to which a contestable markets framework is relevant. When assets are effectively sunk so that their use is tied to a specific purpose in the regulated market and they are an important part of the cost structure then the ‘hit and run’ entry that is a defining characteristic of contestable market theory is not a valid assumption and the contestable markets theory is not a relevant theory for supporting asset valuation.

The valuation of assets based on a hypothetical new entrant’s efficient capital costs is also rationalised by interpreting such costs as relevant opportunity costs (ACCC 1997, p.30 and Ergas 2008, p.95 in criticising the ACCC’s approach in regulating Telstra). However, where assets are sunk their opportunity cost (in another use) from the perspective of both the owner and society is zero.

Criticisms of the Hypothetical New Entrant Test

The underlying theory of contestability is not relevant when there are sunk costs
The main rationale that is advanced to support the hypothetical new entrant test is the economic theory of contestable markets. In determining the relevance of the theory of contestable markets in establishing an approach to asset valuation it is important to assess the relevance of the underlying assumptions of contestability theory which are (Commerce Commission 2008, p.14, based on Baumol et al 1988):

- entry is completely free and exit is costless, which requires that entry must not require the firm to make any ‘sunk’ investments;
- entrants and incumbents compete on completely ‘symmetric’ terms (ie on a ‘level playing field’), and
- entry is not impeded by fear of retaliatory price changes.

As explained by the Commerce Commission (2008, pp.14–15) the existence of sunk costs violates all the underlying assumptions of a perfectly contestable market. Entry is far from free because there are significant sunk costs and exit is not costless and so firms will have an incentive to recover all their investment costs if possible. Entrants and incumbents do not compete on symmetric terms because the existence of sunk costs means that when considering whether to enter the entrant does not have sunk costs prior to entry, whereas the incumbent does and this creates a risk for the entrant that prices will fall as low as operating costs in the event of entry. The prospect that the incumbent can reduce prices to such a level because its costs are already sunk creates a barrier to entry.

Thus, the underlying theory of contestable markets is not applicable to network businesses because it assumes there are no sunk costs in a situation where the market or regulated service at issue involves substantial sunk costs. Furthermore, assuming the price adjustment implied by the theory of contestability is not relevant when there are significant sunk costs. This is because there needs to be a mechanism to ensure that sunk costs are recovered in an economically efficient manner and the theory of contestability does not specify such a mechanism when there are substantial sunk costs.

It is notable that Professor Stephen King (2000, p.2), one of the current ACCC Commissioners, has stated that the theory of contestable markets and the hypothetical efficient new entrant benchmark have ‘limited economic merit’ in the context of determining asset values of sunk assets for regulated businesses.

Windfall gains and losses cannot be justified

The adoption of a depreciated optimised replacement cost approach will entail windfall gains and losses. In practice, windfall gains are more likely to occur given the long lived nature of regulated energy network infrastructure.

Johnstone (2003, p.36) argues that in cases where service providers have dated but long–lived assets DORC valuations are likely to imply similar tariffs as if the assets were brand new. The implication is that this is likely to imply windfall gains to the asset owners and, as Johnstone (2003, p.36) puts it, such valuations are therefore ‘prone to be well outside the bounds of political sustainability’.

In the New Zealand context, Bertram and Terry (2000, p.ii) make the following observation:
‘Capital gains obtained from revaluing assets are a source of income. For a natural monopoly, which prices its services directly from the value of its assets, the treatment of that capital gain is of great importance. If prices are raised in line with the increased asset value without any offsetting adjustment, the owner reaps windfall gains which are far more than just a one off boost. The owner can effectively earn a return on and of capital that it has never actually invested in the business. Consumers pay for these increased earnings through higher charges while receiving no improvement in the scope or level of service. Such earnings are pure monopoly rents.’

Bertram and Terry (2000, p.5) also observe that many submissions from consumer groups in relation to the electricity industry in New Zealand in 2000 focussed on the issue of wealth transfer from consumers to shareholders of electricity line businesses as a result of the application of the optimised deprival value methodology.

The implementation of a methodology that, as a concept and in practice, is likely to entail windfall gains and losses also runs counter to the regulatory objective of obtaining an appropriate balance of the interests of producers and consumers and, more specifically, the determination of a ‘fair price’ as highlighted by Carpenter and Lapuerta (2000, p.3).

A final point is that the use of a methodology (such as DORC or ODV) that will lead to windfall gains and losses in contrast to a methodology that avoids such windfall gains and losses while also ensuring the ex ante recovery of prudent investment (as embodied in the concept of ex ante FCM) will entail a number of economic inefficiencies. If there is a bias so that on average windfall gains are realised there would be allocative inefficiency as price would exceed the actual average cost of production and there would be dynamic inefficiency as there would be an incentive to over invest. However, there would also be allocative and dynamic inefficiencies that would occur in a situation where windfall losses were realised.

Other Considerations

The second interpretation of the ACCC (1999, p. 39) set out above requires that the DORC estimate equals or approximates the amount that a new entrant would be prepared to pay for existing assets not to have to replicate the existing infrastructure. This assumes that the incumbent would be prepared to sell at that price. However, this assumption would be inappropriate in a situation where the incumbent had market power which is likely given the service is regulated.

In addition, as explained by Johnstone (2003, p.16), a new entrant in, for example, the market for energy transmission services would have to pay the full undepreciated optimised replacement cost to duplicate existing infrastructure as there is no second hand market for such a network or its individual components. Johnstone develops a present value model specified by King (2001) that defines the exclusion condition on the regulatory asset base and highlights the relevance of undepreciated optimised replacement cost as the relevant asset

---

4 The full economic inefficiency requires a comparison with marginal cost but some allocative inefficiency must occur to ensure the recovery of sunk costs. However, allocative inefficiency is worsened when prices also exceed average cost.
value for defining the price when the remaining life of the asset base is large and the entrant could expect to capture the whole market.

However, if the entrant could not capture the whole market the prospective entrant would require a higher price than implied by undepreciated (or depreciated) optimised replacement cost, recognising the importance of economies of scale in infrastructure. But capturing the whole market is unrealistic and the entrant would have to consider the pricing outcome from sharing the market including the prospect that if the incumbent decided to compete for market share it could price as low as variable costs recognising its assets were sunk costs. These considerations are clearly like to deter entry at a price for services consistent with a DORC asset value.

In addition, the hypothetical efficient new entrant cost benchmark does not take account of the full costs to society if an entirely new optimised network were to be built. This would include the full design, approval and development costs as well as the costs of disrupting existing neighbourhoods.

These considerations highlight the incredulity of the DORC concept as the entry price for a prospective entrant.

The ACCC has also argued that the efficient new entrant benchmark (in the form of TSLRIC in the regulation of telecommunications access) provides for the efficient use of the existing infrastructure by access seekers, as the implied access price signals the long term value of the resources embodied in that service. However, this proposition does not recognise the role of short run marginal cost (comprehensively defined to include relevant social costs including externalities such as congestion) in determining the efficient use of resources. This is particularly so in the case where infrastructure is not fully utilised and pricing based on long run marginal costs would lead to further under-utilisation of the asset (Bonbright et al 1988, King 1996, and Johnstone 2002).

The ACCC has also argued that the TSLRIC benchmark allows efficient access providers to fully recover the costs of the infrastructure in the long term (ACCC 1997, p.30). However, this is only true for a new investor from an ex ante perspective if the investor expects that a future TSLRIC benchmark does not differ substantially from the current benchmark or the principle of FCM is adopted from an ex post perspective. In contrast, if the ex ante FCM concept is adopted for actual prudent expenditure and applied for investment already sunk then appropriate investment incentives would still be provided and without the prospect of windfall gains or losses.

A final criticism of the hypothetical new entrant benchmark for the valuation of assets is the considerable scope it provides for discretion, arbitrary valuations and divergences of views in the valuation of network assets.

For example, Walker and Walker (2001, p.100) observe in the context of Australian regulation:

‘The CCA [current cost accounting] rules were nothing if not flexible. Managers looking at ‘current written down replacement values’ could virtually pick a number – any number. Technological changes meant that existing infrastructure could be constructed or rebuilt in different ways. Estimates of the cost of
reconstruction were always going to be crude estimates. Deductions for wear and tear also involved crude estimates. Allowances for the value of the enhanced functionality of modern technology introduced even further arbitrary adjustments.’

Walker and Walker (2001, pp.119–20) go on to observe:

‘The term ‘creative accounting’ is commonly associated with the way that private sector entrepreneurs have been able to create illusions of profitability through a combination of paper transactions and financial engineering. The same term might be applied to the way that Australian GTEs [government trading enterprises] have been required to adopt unusual accounting methods which reduces their reported profitability.

‘The intentions might have been honourable – to place managements under pressure by imposing tough performance targets, and to drive efficiency so as to achieve a better deal for the community. But in the end, those responsible may have deceived themselves, as well as shaping perceptions about the unprofitability of businesses built with the use of taxpayers’ money.’

Johnstone (2003) documents the process of acceptance of replacement cost valuation methods in the Australian utilities sector for regulatory purposes despite their earlier rejection within the private sector as being too arbitrary. He goes on to note:

‘The most effective constraint on existing asset owners’ initial DORC valuation, apart from any indirect benchmarking by the regulator, is the level to which the ‘independent’ engineering valuers, hired by asset owners to find this value, are ready to stretch. Given the alleged failures of independence of auditors in other, innately less subjective asset valuation contexts, the analogous economic incentives applying to engineering based DORC–valuers in tariff setting, and the scope for ‘creative engineering’, should be of serious concern to regulators. When seen in this light, the market discipline purportedly inherent to tariff settings based on DORC is more a product of economic sophistry than economic theory.’

Quite apart from the relative arbitrariness of replacement cost valuations, there has also been controversy over the appropriate treatment of revaluation gains when successive replacement cost valuations are undertaken. Some argue that the latest replacement cost valuation should be used as the basis of price setting because it represents the costs that would face a ‘hypothetical new entrant’. However, as discussed above, this argument is based on the assumption of perfect contestability which is clearly inappropriate for network industries with a high proportion of long–lived sunk assets. It also runs counter to the concept of FCM which requires that revaluation gains – from any source – must be treated as income.

It is also important to recognise that although the DORC and ODRC concepts may have been adopted to help determine asset values for regulated network assets, they have often been adjusted significantly or not used when determining regulated prices in Australia. The Commerce Commission (2008, pp.115–119) documents these examples for gas transmission and distribution assets in Australia, including views of the ACCC expressing reservations about DORC for the Dawson Valley Gas Pipeline.
The ACCC (2003) Review of the Draft Statement of Principles for the Regulation of Transmission Revenues identified a number of concerns about DORC similar to those outlined above. In 2004 it made a decision in relation to principles for the regulation of electricity transmission revenues (ACCC 2004b, p.10) that specified a preference to lock in the regulatory asset base (RAB). This approach involves locking in the value of the opening asset base of the prior regulatory period but with adjustments for inflation and depreciation, and capex incurred during the regulatory period on the basis of the capex regulatory arrangements. It also noted that the electricity code provides the discretion to revalue assets and hence, if regulated companies proposed a revaluation, the ACCC would consider the proposal on its merits but that the onus would be on the company to make a case for departing from the preferred principle of locking in the asset base. The ACCC also specified that if it were to consider revaluing the asset base, its preference would be to reopen the entire valuation and consider every element of the asset base.

In making this decision it noted that (ACCC 2004b, pp. 38–9):

- inconsistent treatment and poor accounting information were relevant factors in the adoption of the DORC methodology by all jurisdictional regulators for all first round regulatory caps for the regulation of electricity transmission assets under the NEC;
- periodic valuation using a DORC could result in significant variation in the RAB from one period to the next and, reflecting the scope for a wide asset value range, create considerable uncertainty which might deter investment; and
- locking in the existing RAB values would ensure consistency with the approach it used in regulating gas pipelines.

The AER now has responsibility for the regulation of electricity and gas infrastructure and utilities. For electricity transmission assets, the AER (2005, p.14) has indicated that it will adopt the opening asset valuations set by the ACCC unless the service provider can provide compelling reasons to vary them.

Assessment in terms of preferred evaluation criteria

1. **Supports economic efficiency.** ODV performs poorly in terms of the principle of supporting economic efficiency. As explained above, the underlying economic theory for ODV requires strict assumptions that do not apply when there are significant sunk assets as is the case for network infrastructure. Because there is so much discretion in the estimation of the ODRC component of ODV, there is both great uncertainty as to the actual estimate that will be allowed and great risk of windfall gains or losses being realised. Although uncertainty would discourage investment, in practice the tendency is for ODV to lead to windfall gains given the long life time of infrastructure assets and slow real depreciation. The prospect of windfall gains being realised that were not recognised in regulated income would encourage over-investment with adverse impacts on both productive and allocative efficiencies over time.

2. **Facilitates FCM for prudent investment.** ODV performs poorly in terms of facilitating the NPV=0 principle unless appropriate adjustments are made to regulated income which may be difficult to reach agreement on and implement in practice.
3. **Cost effectiveness.** ODV performs poorly in terms of cost effectiveness because it requires an extensive consultation and complex and expensive calculation of the optimised depreciated replacement cost. The use of ODV is also greatly complicated where mergers and acquisitions occur during the regulatory period.

4. **Consistency and accuracy.** ODV performs poorly in terms of consistency and accuracy given the wide scope for the use of judgement in determining asset values.

5. **Regulatory commitment and predictability.** ODV performs poorly in terms of regulatory commitment and predictability because it is characterised by poor credibility with respect to facilitating economic efficiency and the ex ante NPV=0 rule and there is wide scope for divergences of views and discretion with respect to the asset base.

6. **Transparency.** ODV performs poorly with respect to transparency because of the considerable scope for judgemental assessments in the estimation of ODRC. It is effectively unauditable with meaningful accuracy.

7. **Enables ready conversion of asset values from current to constant prices and vice-versa.** ODRC does enable ready conversion of asset values from current to constant prices and is superior to DHC but not IHC on this principle.

### 5.3 Depreciated historic cost

Under a depreciated historic cost method of asset valuation, the actual written down book value of the assets, defined under standard historic cost accounting conventions, ie the standard accounting book value of the assets adjusted for accumulated depreciation, is used as a basis for determining the regulatory asset base – hence the term depreciated historic cost (DHC). In some jurisdictions the terminology depreciated actual cost (DAC) is used. DHC has tended to be used in the United States to value regulated assets.

**Taking account of inflation**

It is important to recognise the need to adjust for inflation and the relationship between the asset base and the allowed rate of return in the context of inflation. An adjustment for inflation is necessary to ensure that the regulated entity is able to recover the opportunity cost of its investment. Investors need to receive a return on their investment in the regulated asset that compensates for inflation otherwise they will invest in assets with similar risk where they are compensated for inflation. In well functioning markets such adjustments are made continuously and reflected in observable returns and market values of assets.

Adjustment for inflation can be achieved by either using a nominal rate of return and asset values in historic cost terms unadjusted for inflation or using a real rate of return and asset values adjusted for inflation. However, the equivalence of these two approaches depends on the equivalence of actual and expected inflation. This can be explained as follows.

The present value of the income stream obtained by applying the nominal required return to the historic asset base discounted by the nominal required return will be the historic cost of the asset. Note that this approach does not require an estimate of the real rate of return, just an estimate of the nominal rate of return which can be obtained from observable data. However, the present value of the income stream obtained by applying an estimate of a real rate of
Asset Valuation and Productivity–based Regulation

return to the historic cost base, adjusted by inflation and discounted by the nominal required return, will only equal the historic cost of the asset if the expected inflation figure implicit in the required real rate of return is equal to the actual inflation figure used to adjust the asset base.

This can be shown as follows (Patterson 1995, pp.274–6). In mathematical terms, the present value (PV) of an historic cost asset, using the nominal required return is:

\[
P_{H} = \sum_{t=1}^{\infty} H \times n / (1 + n)^t = H \times n / n
\]

The PV of historic cost adjusted for inflation using actual inflation, and applying a real rate of return based on expected inflation is:

\[
P_{I} = \sum_{t=1}^{\infty} H \times (1 + i)^t [(n - i^r)] / (1 + n)^t = H \times (n - i^r) / (n - i)
\]

These two present values are equivalent if:

\[
i^r = i
\]

If expected inflation is less (more) than actual inflation the present value of the asset base will be higher (lower) with the latter method than the former method. This is because the real rate of return is commensurately higher (lower) when expected inflation is less (more) than actual inflation for a given nominal rate of return. The two approaches are more likely to be equivalent in net present value terms when inflation is low and stable as it is then more likely to be predictable.

It is also important to recognise that even where the two approaches mean that investors earn identical returns, the price profiles of the two approaches will differ. In particular an historic cost base combined with a nominal return will imply a constant nominal price but a declining real price while a revalued asset base with a real return will imply a constant real price but a rising nominal price with positive inflation.

Thus, as recognised by the Commerce Commission (2008b, p.48), where DHC is used to value regulated assets, the use of a nominal allowable rate of return (as incorporated into a nominal WACC) provides compensation for expected inflation. If the DHC asset base was also indexed for inflation investors would be doubly compensated for inflation. When a DHC asset base is indexed directly for inflation it is defined as indexed (depreciated) historic cost (IHC) which is reviewed below. When IHC is used the allowable rate of return must be expressed in real terms (as incorporated into a real WACC) as indexing of the asset base will ensure compensation for inflation.

Assessment in Terms of Preferred Evaluation Criteria

Assuming that an appropriate adjustment for inflation is made, the main advantage of an historic cost approach to valuation is the certainty that it creates for the regulated entity and customers. The historic cost approach creates considerable certainty for investors because it is based on readily observable accounting information and it treats each new investment as a long–term contract between the regulated entity and its customers, requiring customers to pay
the original cost of the asset plus a reasonable rate of return, irrespective of changes in circumstances that could affect the value of the asset. This means that the approach does very well in terms of facilitating the NPV=0 rule, i.e., ex ante FCM.

However, the price of achieving this greater certainty for investors is greater risk for customers of the regulated services if capital expenditure is not prudent or efficient. This is because strict application of the historic cost approach would mean that customers bear the risk that investments will not be prudent, that some assets are included that are not being used or some assets are less useful because of technological developments. These problems can be addressed to some extent by developing prudency and used and useful asset tests as in the United States. Note that such tests are a form of asset valuation based on optimising the asset base which, although beneficial to users of the assets at a point in time, may impact adversely on investment incentives. Normally the prudency tests are applied on a forward-looking basis to forecast capital expenditure while the used and useful tests are applied to the existing asset base. The ODV methodology takes the latter approach by applying an ex post optimisation test to the entire asset base but ideally all approaches need to assess the efficiency of forecast capital expenditure over the regulatory period, unless ODV is applied every year.

The historic cost approach also has the advantage that it is based on actual accounting information which greatly reduces the need for the application of judgement in asset valuation. However, the use of prudency and utilisation tests re-introduces the need for some judgement and the potential for variation in asset values. But the scope for judgement to be applied to determine if capital expenditure is prudent and efficient is not likely to be anywhere near as extensive as the range of views that can arise when ODV is applied to the whole existing asset base.

1. **Supports economic efficiency.** DHC provides greater certainty than ODV which helps ensure an appropriate balancing of allocative and dynamic efficiency in terms of the determination of a ‘fair price’ that takes account of the interest of consumers and firms. It provides similar certainty to IHC.

Another intertemporal economic efficiency consideration is that unless offsetting adjustments are made to depreciation provisions (so that the time profile of depreciation differs considerably from standard depreciation provisions while still preserving FCM) DHC will imply more front-loading of capital charges over the lifetime of assets compared to ODV and IHC (for the same dollar value asset base). This will mean higher real prices in the early stages than in the later stages of an asset’s life. Such a price profile would be preferred by investors where they considered there was some probability that regulatory arrangements could change or there was some other perceived threat to the ex ante FCM price path.

However, a higher real price in the early years of an asset’s life could contribute to under-utilisation of the asset which would be inconsistent with allocative efficiency. Furthermore, network assets are typically characterised by economies of scale in construction so that it is optimal to have some excess capacity until demand increases to make better use of that capacity. Thus intertemporal economic efficiency considerations might imply smaller real charges in the early periods of the lifetime of network assets reflecting the low marginal cost of usage and to encourage use of the asset but...
progressively increasing as demand and utilisation of the network increased (see appendix A10.4). Front loading of prices would also imply a greater burden on consumers who make greater use of the asset in the early stages of the asset’s lifetime compared to consumers who make greater use of the asset later in the asset’s lifetime. Assuming similar real incomes, this is likely to be considered inequitable and the effect would be exacerbated when real incomes rise over time, particularly where higher real incomes are associated with diminishing marginal utility of income.

Given the strong commitment of the Commerce Commission to FCM, the likelihood of network assets having some scope to accommodate considerable demand growth and the prospect of rising real incomes over time, it is suggested that a more even time profile of real prices over time would be preferred than one implied by DHC. Thus, based on these considerations DHC would be seen as inferior to ODV and IHC, for the same dollar asset value acquisition cost.

2. **Facilitates FCM for prudent investment.** DHC will help facilitate the NPV=0 principle more readily than ODV, including in circumstances where revaluation gains and losses are reflected in income estimates for the purposes of FCM. This is because revaluation gains and losses do not need to be recognised in order to achieve FCM with DHC, provided an appropriate nominal return on capital is allowed for the regulated firm. As noted, ODV has a high risk that windfall gains and losses will be realised and not properly reflected in income adjustments, thus violating the NPV=0 rule for prudent investment. The ranking on this principle would be similar to IHC except where actual inflation diverged significantly from expected inflation. DHC would be superior in this respect as it relies on nominal returns and does not require an estimate of expected inflation (which is incorporated into the estimate of nominal returns which is based on observable data). However, this difference is not considered to be material in most situations in practice.

3. **Cost effectiveness.** DHC is considered to be more cost effective than ODV because it does not require an extensive and expensive calculation of the optimised replacement cost. This is because historic book values could be used without reconstructing the capital stock to form a real magnitude. But DHC would be significantly less cost effective than IHC, where ready conversion from nominal to real values is required (as per principle 7).

4. **Consistency and accuracy.** DHC is considered to be likely to be more consistent and accurate than ODV as there is less requirement for the use of judgement in determining asset values. The main scope for judgement relates to determining prudent and efficient future capital expenditure which is required under all three methods. Consistency and accuracy would be similar for DHC and IHC except where there was a significant divergence between actual and expected inflation.

5. **Regulatory commitment and predictability.** DHC enables more regulatory commitment and predictability than ODV because it is characterised by greater credibility with respect to facilitating economic efficiency and the ex ante NPV=0 rule and there is less scope for divergences of views and discretion with respect to the asset base. Regulatory commitment and predictability would be the same under DHC and IHC.
Asset Valuation and Productivity–based Regulation

6. **Transparency.** DHC is much easier to understand and document and is more capable of being independently replicated with minimal judgemental assessments than ODV. It would be more difficult to implement and replicate than IHC because of the need to satisfy principle 7.

7. **Enables ready conversion of asset values from current to constant prices and vice-versa.** DHC does not enable ready conversion of asset values from current to constant prices because at each point in time the accounting book value of the asset base is a mix of capital expenditure incurred at different points in time. To convert DHC from a nominal magnitude to a real magnitude would require the conversion of each of the capital expenditure components over the life time of an asset to a real component and then aggregation of those components to form an aggregate real capital measure. This is not seen as practical and would greatly increase the cost of implementing DHC were a constant price asset value series to be used as a proxy for the capital input quantity in total factor productivity measurement.

5.4 **Indexed historic cost**

**Assessment in Terms of Preferred Evaluation Criteria**

The indexed historic cost (IHC) methodology for valuing assets requires the estimation of the asset base in real (inflation adjusted) terms and then the indexing of that asset base by a suitable deflator. In practice, this requires the selection or estimation of an initial asset base and then the estimation of the time profile of that asset base over time by incorporating annual capital expenditure and depreciation. The indexing of the asset base converts it to nominal terms which provides compensation for inflation. An allowable real rate of return and allowable depreciation are then defined to determine allowable capital charges. Note that in order to achieve ex ante FCM the asset base would need to be indexed by the same deflator as used in measuring the allowed expected real return from the investor’s perspective. This would normally be a general deflator such as the consumer price index as this would be most relevant in ensuring capital was maintained in real general purchasing power terms.

This approach has similar ‘certainty’ characteristics to DHC and, like DHC, entails far less need for judgemental assessments than ODV. However, as noted, in practice all three methods require similar judgemental assessments to be made about prudent and efficient future capital expenditure.

1. **Supports economic efficiency.** IHC provides greater certainty than ODV which helps ensure an appropriate balancing of allocative and dynamic efficiency in the determination of a ‘fair price’ that takes account of the interest of consumers and firms. It provides similar certainty to DHC.

As noted above, IHC is considered to be superior to DHC in terms of intertemporal economic efficiency considerations that relate to the time profile of prices. It effectively ‘back–end loads’ the profile of receipts which encourages utilisation of the asset in the early stages of its life while serving to ration use once the asset becomes fully utilised towards the end of its life. This reflects the likelihood of network assets having scope to
accommodate considerable demand growth. This pattern of pricing also comes closest to user pays while recognising the prospect of rising real incomes over time. However, it 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. **Facilitates FCM for prudent investment.** IHC will help facilitate the NPV=0 principle more readily than ODV, including in circumstances where revaluation gains and losses are reflected in income estimates for the purposes of FCM. This is because, as with DHC, revaluation gains and losses do not need to be recognised in order to achieve FCM with IHC, provided an appropriate real return on capital is allowed for the regulated firm. If a nominal rate of return is used with IHC then revaluation gains do need to be recognised to ensure FCM. As noted, ODV has a high risk that windfall gains and losses will be realised and not properly reflected in income adjustments, thus violating the NPV=0 rule for prudent investment. The ranking on this principle would be similar to DHC except where actual inflation diverged significantly from expected inflation. As noted, DHC would be superior in this respect as it relies on nominal returns and does not require an estimate of expected inflation (which is incorporated into the estimate of nominal returns which is based on observable data). However, this difference is not considered to be material in most situations in practice.

3. **Cost effectiveness.** IHC is considered to be more cost effective than ODV because it does not require an extensive and expensive calculation of the optimised replacement cost. As noted, IHC would be less cost effective than DHC, except where ready conversion from nominal to real values is required. However, as this principle (7 below) is required, IHC would be significantly more cost effective than DHC.

4. **Consistency and accuracy.** IHC is considered to be likely to be more consistent and accurate than ODV as there is less need for the use of judgement in determining asset values. The main scope for judgement relates to determining prudent and efficient future capital expenditure which ideally is required under all three methods. Consistency and accuracy would be similar for DHC and IHC except where there was a significant divergence between actual and expected inflation which would tend to favour DHC.

5. **Regulatory commitment and predictability.** IHC enables more regulatory commitment and predictability than ODV because it is characterised by greater credibility with respect to facilitating economic efficiency and the ex ante NPV=0 rule and there is less scope for divergences of views and discretion with respect to the asset base. Regulatory commitment and predictability would be the same under DHC and IHC.

6. **Transparency.** IHC is much easier to understand and document and is more capable of being independently replicated with minimal judgemental assessments than ODV. It would be far less difficult to implement and replicate than DHC because of its ready ability to satisfy principle 7.

7. **Enables ready conversion of asset values from current to constant prices and vice-versa.** IHC enables ready conversion of asset values from current to constant prices and is similar to ODRC in this respect.
5.5 Conclusion

A summary comparison of the performance of each of the asset valuation methods in terms of the preferred criteria is presented in Table 1. Both IHC and DHC are clearly preferred to ODV and of particular importance is that both these methods are seen as superior in terms of economic efficiency, cost effectiveness, consistency and accuracy, and transparency. IHC is clearly preferred to DHC in terms of the criterion for ready conversion of asset values from current to constant prices and vice-versa.

Table 1: Summary assessment of ODV, DHC and IHC against the preferred criteria

<table>
<thead>
<tr>
<th>Principle</th>
<th>ODV</th>
<th>DHC</th>
<th>IHC</th>
<th>Comment</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Supports economic efficiency</td>
<td>x</td>
<td>✓</td>
<td>✓</td>
<td>IHC is superior when constant real prices are required for intertemporal efficiency or front loading of capital charges is considered to be economically inefficient and conventional depreciation is adopted, making it difficult to make offsetting adjustments in defining allowable capital income.</td>
</tr>
<tr>
<td>2. Facilitates NPV=0</td>
<td>x</td>
<td>✓</td>
<td>✓</td>
<td>DHC is superior if there is a significant divergence between actual and expected inflation.</td>
</tr>
<tr>
<td>3. Cost effectiveness</td>
<td>x</td>
<td>✓</td>
<td>✓</td>
<td>IHC is clearly superior if ready conversion from nominal to real magnitudes is required (principle 7).</td>
</tr>
<tr>
<td>4. Consistency and accuracy</td>
<td>x</td>
<td>✓</td>
<td>✓</td>
<td>DHC is superior if there is a significant divergence between actual and expected inflation.</td>
</tr>
<tr>
<td>5. Regulatory commitment and predictability</td>
<td>x</td>
<td>✓</td>
<td>✓</td>
<td>Both DHC and IHC enable more regulatory commitment and predictability than ODV because they are characterised by greater creditability with respect to supporting economic efficiency and the NPV = 0 rule. They also require fewer judgements to be made thus making them more predictable.</td>
</tr>
<tr>
<td>6. Transparency</td>
<td>x</td>
<td>✓</td>
<td>✓</td>
<td>DHC would be more difficult to be replicated than IHC because of the difficulty in converting from nominal to real magnitudes (principle 7). The need for extensive judgements to be made makes ODV less transparent and less replicable.</td>
</tr>
<tr>
<td>7. Conversion of nominal to real</td>
<td>✓</td>
<td>x</td>
<td>✓</td>
<td>DHC performs very poorly on this principle which would be important for total factor productivity measurement if a constant price asset value is used as a proxy for the capital input quantity.</td>
</tr>
</tbody>
</table>

Notes: x = performs poorly. ✓ = performs well. ✓ ✓ = performs very well.

The assessment of the methods clearly supports the use of historic cost rather than replacement cost–based valuations as the preferred valuation method for regulatory asset bases. IHC is the only one of the three methods which satisfies all 7 evaluation criteria and so
is preferred over DHC. However, given the non-commercial nature of the origins of many utilities and the long-lived nature of their assets, in many cases historic cost information does not exist or cannot be recovered. In these cases, the use of the earliest available comprehensive asset valuation – which will usually be a replacement cost-based valuation – can be justified as the starting point. There is then a strong case for ‘locking in’ the starting valuation and rolling forward from that point using data on investment and depreciation under the IHC framework.

Transitional issues are examined further in the following section.
6 TRANSITIONAL ISSUES

Given that true historical or original costs for energy network assets are likely to be unobtainable given the long-lived nature of the assets and changes in organisational structure and record keeping over the decades, all three asset valuation methods would most likely need to use the earliest available reliable ODV as their starting point. In practice then, the difference between the three methods would relate to the method used for roll-forward and whether subsequent new ODV exercises are undertaken and/or recognised.

Given that the three methods will most likely be using the same starting point in reality, the transition problems will likely be less severe than if they were starting from ODV on the one hand and true historic cost on the other. Similarly, transition problems are likely to be less severe if the asset value continues to only enter the productivity analysis through the capital user cost rather than also through the capital input quantity. That is, using physical capital quantities to proxy annual capital input quantity is not only likely to better proxy the actual physical depreciation profile but also has the advantage of reducing reliance on asset valuation series which, by their nature, are less robust and, in this case, are likely to be of a hybrid nature.

In preliminary work for the thresholds reset in 2008, the Commerce Commission retained Denis Lawrence to update the Lawrence (2003, 2007) productivity analyses but using IHC asset values rather than the ODV series used in the earlier studies. Although not completed due to the passing of the Commerce Amendment Act, this exercise provided some useful insights into the issues that might be encountered in transitioning from using ODV to using IHC in productivity analysis.

The first obstacle encountered in this exercise was obtaining actual additions and disposals data with which to update and backdate from the starting value. Although additions data are included in the Information Disclosure Data filings up to 2007, these are valued at the relevant ODV unit rates rather than at actual cost. To obtain actual additions and disposals data, the Commerce Commission requested data from the EDBs under Section 98 of the Act. However, the EDBs claimed it was not possible to obtain reliable actual investment and disposals data prior to 2004. It was, therefore, only possible to form an approximate IHC series from 2004 onwards (using the 2004 ODV as a proxy for opening original cost). For the years prior to 2004 it was necessary to splice the year-to-year proportional changes in the ODV series (adjusted to exclude revaluations) onto the series from 2004 onwards.

While not ideal given the need to include the years prior to 2004 using additions calculated using ODV unit rates rather than actual cost, the approach adopted appeared adequate given the TFP specification used. That is, given that the asset value is only used in forming the user cost of capital and not directly in forming the capital input quantity, the spliced series is likely to provide a sufficiently good approximation for use in reliable productivity analysis.

It should also be noted that, in the case of electricity distribution, the Information Disclosure Data requirements for the years from 2008 onwards require explicit disclosure of IHC consistent roll forwards going back to the 2004 ODVs. That is, additions and deletions are shown using actual values and indexation is by means of the CPI. This will make it relatively...
straightforward to construct an IHC series from 2004 onwards and to splice the earlier ODV series as a proxy for IHC for the years prior to 2004.

Transitioning from ODV to DHC would be more problematic because, despite using the same starting point, the DHC series would not be indexing the series to maintain comparable units between the existing stock and subsequent investment. While this may not be a large problem in the space of the first few years (depending on relevant inflation rates), over the medium to longer term the DHC series would likely diverge substantially from the IHC and ODV series leading to the splicing of the ODV series for the years prior to 2004 being less appropriate for productivity analysis in this case.

6.1 Electricity-specific issues

The major issue with the 2004 electricity ODV estimates was the change in the way underground cables were valued. Some of the cables previously allocated overhead lines as the modern equivalent assets have been ‘unoptimised’ in the 2004 valuations and underground unit rates are now used in their valuation. The 2004 ODV was also the first one where disaggregated ODV information was available for all EDBs. This improved information led to significant changes in asset shares used in the productivity analysis for all EDBs and, hence, for the industry level analysis compared to the shares estimated in Lawrence (2003). This raises the question of what constitutes the appropriate starting point for new asset valuation series – is it the older ODV estimates that were less accurate or the 2004 ODVs which rectify some of the significant inaccuracies in the earlier estimates and which provide the required disaggregated information for all EDBs?

For electricity, the 2004 ODVs appear to be the best starting point given that they are the first ones to contain accurate and appropriate treatment of underground cables and sufficiently detailed information to allow asset specific input shares to be used for productivity analysis. This situation differs from the one the Commerce Commission encountered in the gas authorisation inquiry where ODVs undertaken in 2002 and 2003 were thought to be sufficiently accurate and comparable to provide a good approximation to the corresponding historic cost and were, hence, preferred to later 2005 ODVs.

Whereas the Lawrence (2007) productivity update analysis removed the impact of the 2004 revaluations by splicing the new series onto the old series used in Lawrence (2003), in future productivity analysis it would be more appropriate to splice the old series onto the new series from 2004 onward. That is, rather than ‘moving the new series down’ to the level of the old series in the overlap year, it would be more appropriate to ‘move the old series up’ to the level of the new series in the overlap year. This would be consistent with using the 2004 ODV as the starting point for all three asset valuation methods in this instance.

Although the default price paths will be based on industry level productivity growth, it remains important to build the industry level data up from the most accurate information available for each EDB. At this point in time, there do not appear to be any binding restrictions on the Commission’s ability to gather the relevant information required to construct a hybrid IHC series.
Network Regulation and Sunk Costs

APPENDIX A: THE THEORY OF NETWORK REGULATION IN THE PRESENCE OF SUNK COSTS

A1 OVERVIEW

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 appendix 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.

A1.1 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.

A1.2 Modelling summary

In section A2, 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 A3, we explore the comparative statics of the one period model developed in section A2. 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 A4, 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 A3 except that in the continuous time framework, we can deal with technical progress in a reasonably straightforward manner. Not surprisingly, in section A4.2, we find that technical progress in both the regulated and unregulated sectors lead to welfare improvements for the household sector. In section A4.3, 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 A4 relied on the regulated firm’s cost function; ie the total cost of producing the vector of regulated products. Implicitly, section A4 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. 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

---

5 See Jorgenson and Griliches (1967) for an introduction to user costs.
Network Regulation and Sunk Costs

not applicable in this context and it will be necessary to work out a new methodology. In section A6, 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 A5, we prepare for section A6 by reworking the results in section A4 using the opex cost function instead of the regular total cost function. Naturally, we obtain the same results in section A5 as were derived in section A4 but the purpose of section A5 is to develop the opex cost function necessary for subsequent analysis instead of the usual total cost function.

In section A6, 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 A6.1, we set up the intertemporal analogue to the one period regulatory model studied in section A2. In section A6.2, 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 A3 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 A6.4, we generalise the one hoss shay depreciation model to allow for other depreciation profiles. Finally, section A6.5 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
• 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.
Network Regulation and Sunk Costs

In section A7, 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 A8, 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 A7). 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 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 A8. 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 A8, 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 A9, we note that when regulating a group of firms using productivity-based 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.

---

6 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.
Section A10 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.

A2 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 \( y = [y_1, ..., y_N]^T \geq 0_N \);\(^7\)
- J unregulated outputs which are also traded internationally; a vector of unregulated outputs will be denoted by the nonnegative vector \( Y = [Y_1, ..., Y_J]^T \geq 0_J \);
- K primary inputs (not traded internationally); the economy’s endowment of primary inputs will be denoted by the strictly positive vector \( v = [v_1, ..., v_K]^T > 0_K \); the regulated sector’s primary input demand vector will be denoted by the nonnegative vector \( z = [z_1, ..., z_K]^T \geq 0_K \) and the unregulated sector’s primary input demand vector will be denoted by the nonnegative vector \( Z = [Z_1, ..., Z_K]^T \geq 0_K \).

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

\(^7\) 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, \( p^T y = \sum_{n=1}^{N} p_n y_n \) denotes the inner product of the vectors \( p \) and \( y \), \( y \geq 0_N \) means each component of \( y \) is nonnegative, \( y \geq 0_N \) means each component of \( y \) is positive, \( y \geq 0_N \) means \( y \) is nonnegative, and \( (Y, Z) \in S \) means that the point \( (Y, Z) \) belongs to the set \( S \).
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.\(^8\)

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.\(^9\)

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.\(^10\)

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.\(^11\)

Now we are in a position to set up the household’s constrained utility maximisation problem,

\(^8\) 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.

\(^9\) The assumption that \(S\) is a cone means that if \((Y,Z)\in S\) and \(λ \geq 0\), then \((0,λZ)\in S\).

\(^10\) 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 \(0 \leq λ \leq 1\). Then \(U(λx^1+(1−λ)x^2,X^1+(1−λ)X^2) \geq λU(x^1,X^1) + (1−λ)U(x^2,X^2)\). 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.

\(^11\) We are assuming that the country is small and thus the international price vector \(P\) is fixed.
which is (6) below:

\[
\begin{align*}
\max_{u,Y,z,X} & \{u : y - x \geq 0_N ; v - z - Z \geq 0_K ; P^T(Y - X) \geq 0 ; (y,z) \in S ; (Y,Z) \in S ; \nonumber \\}
\end{align*}
\]

\[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.\(^{12}\)

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.\(^{13}\) 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.

\(^{12}\)The nonnegativity constraints \(u \geq 0, y \geq 0_N, y \geq 0_N, z \geq 0_K, Z \geq 0_K, x \geq 0_N, x \geq 0_N, x \geq 0_N, x \geq 0_N\) should also be added to the constraints in (6) but in the interests of brevity, we have omitted these constraints.

\(^{13}\)See Walras (1980; 83), Pigou (1920; 278), Hotelling (1938; 256), Lerner (1944; 182), Meade (1944) and Fleming (1944).
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 $u^* > 0$, $y^* > 0$, $y^* > 0$, $z^* > 0$, $z^* > 0$, $x^* > 0$, and $X^* > 0$ solves (6). Now let $y > 0$ be in a neighbourhood of $y^*$ and consider the following constrained maximisation problem that is conditional on the choice of $y$:

$$\text{(7)} \quad H(y) = \max_{u,y,z,x,K} \left[ u : y - x \geq 0_N \right] : \left( Y, Z \right) \in S \equiv U(\lambda, x, X) - u \geq 0_0, Y \geq 0_N, z \geq 0_K, z \geq 0_K, x \geq 0_K, X \geq 0_K. \right]$$

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 $u$, $Y$, $z$, $x$, and $X$ solutions to (7) are also solutions to the following max-min problem:

$$\text{(8)} \quad H(y) = \max_{u,Y,z,x,X} \left[ \min_{p,w,\lambda} \left[ u + p^T[y - x] + w^T[y - z - X] + \lambda \left( Y - X \right) \right] : \left( y, z \right) \in S \equiv U(\lambda, x, X) - u \geq 0_0, Y \geq 0_N, z \geq 0_K, z \geq 0_K, x \geq 0_K, X \geq 0_K. \right]$$

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 $\lambda$ 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 $y > 0_N$ 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:

$$\text{(9)} \quad c(y,w) \equiv \min_{z} \left( w^Tz : \left( y, z \right) \in S \right).$$

It can be shown that $c(y,w)$ is nonnegative, nondecreasing in $y$, and nondecreasing, (positively) linearly homogeneous$^{16}$ 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:

$$\text{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 $x^0 \geq 0_N$, $X^0 \geq 0_K$, $Z^0 \geq 0_K$, and $Y^0 \geq 0$ such that $y - x^0 >> 0_N, z - z^0 - Z^0 >> 0_K, P^T[Y^0 - X^0] > 0$. This is not a restrictive assumption.

$^{15}$ 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).

$^{16}$ This property means that $c$ satisfies the following equation for all scalars $\lambda > 0$: $c(y,\lambda w) = \lambda c(y,w)$. In what follows, for brevity, we will abbreviate positively linearly homogeneous to linearly homogeneous.

$^{17}$ Notation: $\nabla c(y,w)$ is the column vector of first order partial derivatives of $c$ with respect to the components
Network Regulation and Sunk Costs

(10) $z(y,w) = V_w c(y,w)$. Let $Y \geq 0$, be a vector of output targets and $w >> 0$ be a vector of input prices that the unregulated sector faces. Then the \textit{unregulated sector’s joint cost function}, $C(Y,w)$, is defined as follows:

(11) $C(Y,w) = \min_z \{ w^T z : (Y,z) \in S \}$.

It can be shown that $C(Y,w)$ is nonnegative, nondecreasing, linearly homogeneous and convex in $Y$, 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) $Z(Y,w) = V_w C(Y,w)$.

Let $u$ be a utility target for the household and suppose the household faces the vector of prices $p >> 0$ for regulated outputs and $P >> 0$ for unregulated outputs. Then the household’s \textit{expenditure function}, $e(u,p,P)$, is defined as the solution to the following expenditure minimisation problem:

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

The consumer’s expenditure function will be nondecreasing in all of its variables and linearly homogeneous and concave in prices $(p,P)$. 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) $x(u,p,P) = V_p e(u,p,P)$;

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

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

of $w$, $\partial c(y,w)/\partial w_1, \ldots, \partial c(y,w)/\partial w_K^T$, $\partial^2 c(y,w)/\partial w \partial y$. $\partial c(y,w)/\partial w$ is the $K$ by $K$ matrix of second order partial derivatives of $c$ with respect to the components of $w$, $\partial c(y,w)/\partial y$ 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. 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 $C$.

19 See Diewert (1982; 553-556) for material on these regularity conditions.
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:

\[
\text{max}_{y \geq 0} \left[ \text{max}_{u,Y \geq 0} \left\{ u + w^T v + p^T y - c(y,w) + \lambda P^T Y - C(Y,w) - e(u,p,\lambda P) : u \geq 0 ; Y \geq 0 ; p \geq 0_N ; w \geq 0_K ; \lambda \geq 0 \right\} \right].
\]

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

\[
\text{max}_{y \geq 0} H(y).
\]

We assume that there is a strictly positive solution to the two stage max-min problem defined by (17), say \( u^* > 0 ; y^* >> 0_n ; Y^* >> 0_n ; p^* >> 0_N ; w^* >> 0_K ; \lambda^* > 0 \). 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 \( \lambda \) and setting these partial derivatives equal to zero.

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:

\[
\begin{align*}
(19) & \quad 1 = \frac{\partial e(u^*,p^*,\lambda^* P)}{\partial u} ; \\
(20) & \quad p^* = \nabla_y c(y^*,w^*) ; \\
(21) & \quad \lambda^* P = \nabla_Y C(Y^*,w^*) ; \\
(22) & \quad v = \nabla_u c(y^*,w^*) + \nabla_u C(Y^*,w^*) ; \\
(23) & \quad y^* = \nabla_u e(u^*,p^*,\lambda^* P) ; \\
(24) & \quad 0 = P^T [Y^* - \nabla e(u^*,p^*,\lambda^* P)].
\end{align*}
\]

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 \( \lambda^* = 1 \) in what follows. Thus, with this normalisation, the remaining equations (20)-(24) become the following equations:

\[
\begin{align*}
(25) & \quad p^* = \nabla_y c(y^*,w^*) ; \\
(26) & \quad P = \nabla_Y C(Y^*,w^*) ;
\end{align*}
\]

\[20\] 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.

\[21\] This follows by a generalisation of Samuelson’s (1947; 34) Envelop Theorem to cover the case of max-min problems.
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 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, $\nabla c(y^*,w^*) = \left[ \frac{\partial c(y^*,w^*)}{\partial y_1}, \ldots, \frac{\partial c(y^*,w^*)}{\partial y_N} \right]^T$. 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, $\nabla C(Y^*,w^*) = \left[ \frac{\partial C(Y^*,w^*)}{\partial Y_1}, \ldots, \frac{\partial C(Y^*,w^*)}{\partial Y_N} \right]^T$.

Recalling equations (10) and (12) and Shephard’s Lemma, it can be seen that $z^* = \nabla c(y^*,w^*)$ is the optimal primary input demand vector for the regulated sector and $Z^* = \nabla C(Y^*,w^*)$ 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, $z^* + Z^*$, 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 $\nabla e(u,p,P)$ 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 $P$, $\nabla e(u,p,P)$, it can be seen that equation (29) is equivalent to the following equations:

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

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:

$$ (31) \ \rho(y,w) = \ln c(\lambda y,w) / \partial \lambda_{y,w} = y^T \nabla c(y,w) / c(y,w). $$

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
Network Regulation and Sunk Costs

\( y^T \nabla_c(y,w)/c(y,w) < 1 \) or \( y^T \nabla_c(y,w) < c(y,w). \)

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:

\[ (33) \quad p^T y^* = y^T \nabla_c(y^*,w^*) < c(y^*,w^*) \]

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):

\[ (34) \quad \rho(Y,w) \equiv \partial \ln C(\lambda Y,w)/\partial \lambda|_{\lambda=1} = Y^T \nabla_Y Y(Y,w)/C(Y,w) = 1. \]

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

\[ (35) \quad C(\lambda Y,w) = \lambda C(Y,w) \quad \text{for all} \quad \lambda \geq 0. \]

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

\[ (36) \quad C(Y,w) = Y^T \nabla_Y C(Y,w). \]

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 \( p >> 0 \) and demands that the regulated sector meet all demands \( y \geq 0 \). If the economy’s vector of primary input prices is \( w >> 0 \), 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 \( \nabla_c(y,w) \), 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 \( \nabla_c(y,w) \), we can define the vector of deviations from marginal cost pricing or the vector of margins over marginal cost, \( m \), as follows:

\[ (37) \quad m = p - \nabla_c(y,w). \]

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
Network Regulation and Sunk Costs

(25) by the price equals marginal cost plus margins equations (37). Thus, a general nonoptimal regulatory policy is characterised by the following equations:

\( p = m + \nabla_y c(y,w) \) ;
\( P = \nabla_Y C(Y,w) \) ;
\( v = \nabla_w c(y,w) + \nabla_w C(Y,w) \) ;
\( y = \nabla_p e(u,p,P) \) ;
\( 0 = P^T[Y - \nabla_p e(u,p,P)] \) .

Equations (38)-(42) can be regarded as \( 1 + 2N + J + K \) equations in the \( 1 + 2N + J + K \) endogenous variables \( u, m, y, Y, \text{ and } w \). 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

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

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

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

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

\( e(u,p,P) = p^T y + P^T Y \).

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 \( C \):

\( c(y,w) = w^T \nabla_w c(y,w) \) ;
\( C(Y,w) = w^T \nabla_w C(Y,w) \) .

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

\( w^T v - c(y,w) - C(Y,w) = 0 \).

\( 0 = P^T[Y - \nabla_p e(u,p,P)] \).
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) \( e(u,p,P) = w^Tv + p^T y - c(y,w) + P^T Y - C(Y,w) \).

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

(50) \( P^T Y = C(Y,w) \).

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

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

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 or welfare reacts to small changes in these exogenous variables.

## A3 THE COMPARATIVE STATICS PROPERTIES OF THE ONE PERIOD REGULATORY MODEL

### A3.1 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; i.e., 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
Network Regulation and Sunk Costs

level, say $u^*$. Thus, we have:

\begin{equation}
Y^* = e(u^*, p, P) = u^*
\end{equation}

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 $\lambda Y^*$ where $\lambda$ is a positive proportional scaling of the initial income, the new budget constraint for the consumer will be:

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

Thus, if $\lambda$ is greater than 1, the initial budget set expands upwards in a parallel fashion, whereas if it is $\lambda$ 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 $\lambda Y^*$. 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:

\begin{equation}
A = \begin{bmatrix}
1 & -m_r^T & 0_j^T & 0_k^T \\
\n & -\nabla_{yw}^2 c & I_N & 0_{N \times J} \\
0_j & 0_{J \times J} & \nabla_{x_j}^2 C & \nabla_{y_i}^2 C \\
0_k & \nabla_{w_j}^2 C & \nabla_{x_j}^2 C + \nabla_{w_j}^2 C
\end{bmatrix}
\end{equation}

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., $\nabla_{yj}^2 C = \nabla_{yj}^2 C(Y, w)$ (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., $\nabla_{wy}^2 c$ is $\nabla_{wy}^2 c(y, w)$ (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 $\nabla_{pw}^2 e = \nabla_{pw}^2 e(u, p, P)$ 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 $b$:
Network Regulation and Sunk Costs

(57) \[ \mathbf{b} = \nabla \mu \gamma \mathbf{c}(u, p, P) \]
\[ = \partial \mathbf{x}(u, p, P)/\partial u \]

where the second line in (57) follows from Shephard’s Lemma, (14); i.e., the consumer’s vector of Hicksian demands for regulated outputs, \( \mathbf{x}(u, p, P) \), is equal to the vector of first order partial derivatives of the consumer’s expenditure function \( \mathbf{e}(u, p, P) \) with respect to \( p \), \( \nabla \mathbf{e}(u, p, P) \). Thus, \( \mathbf{b} \) 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 \( \mathbf{b} \) vector will be positive.\(^{22}\)

A3.2 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 \( \mathbf{v} \), regarding the components of \( u, y, Y, w \) and \( m \) as functions of \( \mathbf{v} \), so that these functions are \( u(\mathbf{v}), y(\mathbf{v}), Y(\mathbf{v}), w(\mathbf{v}) \) and \( m(\mathbf{v}) \) and \( \nabla_u u(\mathbf{v}), \nabla_y y(\mathbf{v}), \nabla_Y Y(\mathbf{v}), \nabla_w w(\mathbf{v}) \) and \( \nabla_m m(\mathbf{v}) \) 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 \( \mathbf{v} \). After differentiating (49) with respect to the components of \( \mathbf{v} \) and using equations (38)-(41) and (52), we obtain the following equation:

\[ \nabla_v u(\mathbf{v}) = w(\mathbf{v}) + [\nabla_y y(\mathbf{v})]^T m(\mathbf{v}). \]

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, \( \nabla_y \mathbf{c}(y, w) \), then the vector of markups over marginal costs, \( m(\mathbf{v}) \), will equal \( \mathbf{0} \) and equations (58) collapse down to

\[ \nabla_v u(\mathbf{v}) = w(\mathbf{v}) \gg 0_K. \]

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 \( \nabla_v u(\mathbf{v}) \) will generally be close to \( w(\mathbf{v}) \) 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, \( \nabla_v y(\mathbf{v}) \), 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 \( \mathbf{v} \). Thus, rewrite equation (41) as follows:

\[ y(\mathbf{v}) = \nabla_p \mathbf{e}(u(\mathbf{v}), p, P). \]

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

\(^{22}\) If the \( n \)th component of \( \mathbf{b} \), \( b_n \), is negative, then the consumption of the \( n \)th regulated product will decrease as real income increases; i.e., the \( n \)th regulated product is an inferior good.
Network Regulation and Sunk Costs

$\nabla_y v (v) = \nabla y u (v)^T$

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

Thus, if

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

then $\nabla u (v)$ equals $(1 - m^T b)^{-1} w >> 0$, 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) $a \equiv \nabla u^T e(u, p, P) = \partial X(u, p, P) / \partial u.$

Thus, (63) is a counterpart to our earlier definition of $b$ equal to $\nabla y c(y, w)^T$. 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 $\lambda$ and then setting $\lambda$ equal to 1 leads to the following equation:

(65) $\partial e(u, p, P)/\partial u = \nabla y e(u, p, P) p + \nabla y e(u, p, P) P$

$= b^T p + a^T P$

using definitions (57) and (63)

$= 1$

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, $\nabla y c(y, w) \geq 0$:

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

Thus

(67) $1 - m^T b = 1 - [p - \nabla y c(y, w)]^T b$

using (66)

$= p^T a + \nabla y c(y, w)^T b$

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):

\[(68) \nabla_u(v) = [1 - m^T b]^{-1} w >> 0_k.\]

**A3.3 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 A3.2.

Differentiate equation (49) with respect to the components of the vector of international prices \( P \), regarding the components of \( u, y, w \) and \( m \) as functions of \( P \), so that these functions are \( u(P), y(P), w(P) \) and \( m(P) \) 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:

\[(69) \nabla_P u(P) = Y(P) - X(u(P),p,P) + [\nabla_P y(P)]^T m(P)\]

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, \( \nabla_c(y,w) \), 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; i.e. if \( Y_j - X_j \) is positive, then at the initial equilibrium, the \( j \)th unregulated commodity is exported while if \( Y_j - X_j \) is negative, then at the initial equilibrium, the \( j \)th 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 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 a 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:

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

using definition (57).

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

\[
(73) \left[ 1 - m^T b \right] \nabla_P u(P) = Y - X + \nabla_P^2 e(u, p, P) m
\]

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 \( \nabla_P^2 e(u, p, P) m \). Now the J by N consumer substitution matrix \( \nabla_P^2 e(u, p, P) \) 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.\(^{23}\) 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 equilibrium.\(^{24}\)

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

### A3.4 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 \( u, y, w \) and \( m \) as functions of \( p \), so that these functions are \( u(p), y(p), w(p) \) and \( m(p) \) and \( \nabla_P u(p), \nabla_P y(p), \nabla_P w(p) \) 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:

\[
(74) \nabla_P^2 u(p) = [\nabla_P y(p)]^{T} m(p)
\]

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, \( \nabla_P c(y, w) \), then the vector of markups over marginal costs, \( m(p) \), will equal \( 0_N \) and equations (74) collapse down to:

\(^{23}\) Hicks (1946; 311-312) noticed this tendency and proved it must hold if \( N = 1 \) and \( J = 1 \).

\(^{24}\) 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.
Network Regulation and Sunk Costs

(75) \( \nabla_u u(p) = 0 \).

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 \( 0 \), then the vector of regulated prices \( p \) is equal to the corresponding vector of marginal costs of producing these regulated products \( \nabla_c y(w)^{-1} \) and in this case of first best optimal regulation, we attain the first best equilibrium that was discussed in section A2.

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 \( \nabla_u u(p) \) will generally be close to 0 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 matrix equation:

(77) \[ \nabla p y(p) = \nabla p u(p)^{-1} \nabla p e(u(p),p,P) \]

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

(78) \[ (1 - m^T b) \nabla_u u(p) = \nabla p^2 e(u,p,P)m \]

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, \( \nabla_u u(p) \), will be equal to the signs of the vector on the right hand side of (78). This right hand side vector is equal to \( \nabla p^2 e(u,p,P)m \).

Now the \( N \) by \( N \) consumer substitution matrix \( \nabla p^2 e(u,p,P) \) 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.\(^{26}\)

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

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

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

\(^{25}\) Recall that equations (38) hold in our regulatory model. These equations are \( p = m + \nabla c y(w) \).

\(^{26}\) Again, Hicks (1946; 311-312) noticed this tendency and proved it must hold if \( N = 2 \).
Network Regulation and Sunk Costs

\( e_{11} = \frac{\partial^2 e(u,p_1,p_2,P)}{\partial p_1 \partial p_1} \) and \( e_{22} = \frac{\partial^2 e(u,p_1,p_2,P)}{\partial p_2 \partial p_2} \) 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 \( e_{12} = \frac{\partial^2 e(u,p_1,p_2,P)}{\partial p_1 \partial p_2} = \partial x_1(u,p_1,p_2,P)/\partial p_2 > 0 \) and the demand for regulated product 2 will increase as the price of the first regulated product \( p_1 \) increases (this means \( e_{21} = \frac{\partial^2 e(u,p_1,p_2,P)}{\partial p_2 \partial p_1} = \partial x_2(u,p_1,p_2,P)/\partial p_1 > 0 \). 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:

\[(82) \ \partial u(p_1,p_2)/\partial p_1 = [1 - m_1 b]^{-1}[m_1 e_{11} + m_2 e_{12}] = [+] [(-)(-)] > 0 ;\]
\[(83) \ \partial u(p_1,p_2)/\partial p_2 = [1 - m_1 b]^{-1}[m_1 e_{21} + m_2 e_{22}] = [+] [(-)(+) < 0 .\]

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

\[(84) \ \partial u(p_1,p_2)/\partial p_1 = [1 - m_1 b]^{-1}[m_1 e_{11} + m_2 e_{12}] = [+] [(+)(-) < 0 ;\]
\[(85) \ \partial u(p_1,p_2)/\partial p_2 = [1 - m_1 b]^{-1}[m_1 e_{21} + m_2 e_{22}] = [+] [(+)(-) > 0 .\]

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.\(^{27}\)

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.

\(^{27}\) Unfortunately, it will not be easy for the regulator to determine this in most cases.
Network Regulation and Sunk Costs

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):

$$ (86) \quad p^Ty = c(y,w). $$

Equation (86) simply sets the regulated firm’s revenues, $p^Ty$, 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 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:

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

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

$$ (88) \quad \nabla_p G(p) = y + [\nabla_y y(p)]^T p - [\nabla_y y(p)]^T [\nabla_y c(y(p),w(p))] - [\nabla_y w(p)]^T [c(y(p),w(p))] - [\nabla_w w(p)]^T z(p) $$

Using (38) and (10)

$$ = y + [\nabla_y y(p)]^T m - [\nabla_w w(p)]^T z. $$

The matrices of partial derivatives $\nabla_y y(p)$ and $\nabla_w w(p)$ 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):

$$ (89) \quad [\nabla_p G(p)]^T dp = 0 $$
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.\(^{28}\)

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

\[(90) \quad c(y,w) = y^T \nabla_c c(y,w).\]

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

\[(91) \quad p^T y - c(y,w) = m^T y.\]

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 A3 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:

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

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

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, then using (75), we have:

\( \nabla p u(p^*) = 0_N. \)

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:

\[
\begin{align*}
(94) \quad u(p^*) - u(p) & \approx \frac{1}{2} \left[ \nabla p u(p^*) + \nabla p u(p) \right]^T [p^* - p] \\
& \approx - \frac{1}{2} \left[ (1 - m^T b)^{-1} m^T \nabla_{pp^2} e(u,p,P)m \right] \quad \text{using (92) and (93)} \\
& \geq 0
\end{align*}
\]

where the inequality follows from assumption (62) and from the negative semidefiniteness of the \( N \times N \) 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.

**A4 A CONTINUOUS TIME APPROACH TO THE DETERMINANTS OF WELFARE CHANGE**

**A4.1 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.

**A4.2 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.\(^{29}\) 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

\(^{29}\) The first line in (94) follows using Diewert’s (1976; 118) Quadratic Identity.

\(^{30}\) This continuous time approach to economic measurement was pioneered by Divisia (1926).
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; that is in the case of technical progress in the two sectors, the partial derivatives of these cost functions with respect to time will be negative:

\[
\frac{\partial C(Y,w,t)}{\partial t} < 0; \quad \frac{\partial c(y,w,t)}{\partial t} < 0.
\]

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:

\[
(96) \quad u'(t) = w(t)'v'(t) + [Y(t) − X(t)]'P'(t) + m(t)'y'(t) − \frac{\partial c(y,w,t)}{\partial t} − \frac{\partial C(Y,w,t)}{\partial t}
\]

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 \( w(t)'v'(t) \), 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 \( 0 \), so that prices equal the corresponding marginal costs in the regulated sector, then this term vanishes. The fourth term is \( −\frac{\partial c(y,w,t)}{\partial t} \) and it can be seen that this term is the rate of technical progress in the regulated sector and the fifth term is \( −\frac{\partial C(Y,w,t)}{\partial t} \), 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:

---

31 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).
Network Regulation and Sunk Costs

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

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

\[
(98) b(t) = \nabla_u^2 e(u(t),p(t),P(t)) = \partial x(u(t),p(t),P(t)) / \partial u.
\]

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

\[
(99) y'(t) = b(t)u'(t) + \nabla_p^2 e(u(t),p(t),P(t))p'(t) + \nabla_p P^2 e(u(t),p(t),P(t))P'(t)
\]

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

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

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:

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

Recall our comparative statics results for the change in utility due to changes in \( v \), equation (68) in section A3.2 above, for the changes in utility due to changes in \( P \), equations (73) in section A3.3 above and for the changes in utility due to changes in \( p \), equations (78) in section A3.4 above. For convenience, we will rewrite these equations below as equations (103)-(105).

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

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

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

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 A3.2.

However, our new equation (102) represents a generalisation of the results in section A3 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...
Network Regulation and Sunk Costs

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:

\[
(106) \quad p'(t) = -\alpha m(t)
\]

where \( \alpha \) 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 \( m(t)^T[\nabla_{pp^2}e]p'(t) \) in (102) to:

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

where the inequality in (107) follows from the negative semidefiniteness of \( \nabla_{pp^2}e \). Thus, the pricing rule (106) will generally lead to an improvement in welfare.33

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

---

33 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.
approximations to the technical change parameters, \( \frac{\partial c(y,w,t)}{\partial t} \) and \( \frac{\partial C(Y,w,t)}{\partial t} \).

### A4.3 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:

\[
Z(t) = \nabla_w C(Y(t),w(t),t).
\]

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:

\[
P(t) = \nabla_Y C(Y(t),w(t),t).
\]

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

\[(110) \quad C(Y(t),w(t),t) = w(t)^T \nabla_w C(Y(t),w(t),t) = Y(t)^T \nabla_Y C(Y(t),w(t),t) = w(t)^T Z(t) \quad \text{using (108)} \]

\( = P(t)^T Y(t) \quad \text{using (109)}. \]

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:

\[(111) \quad w(t)^T Z'(t) + w'(t)^T Z(t) = \nabla_Y C(Y(t),w(t),t)^T Y'(t) + \nabla_w C(Y(t),w(t),t)^T w'(t) + \frac{\partial C(Y,w,t)}{\partial t} \quad \text{using (108) and (109)}. \]

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

\[(112) \quad \frac{\partial C(Y,w,t)}{\partial t}/C(Y,w,t) = Y_D'(t) - Z_D'(t); \]

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:

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

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

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 \( t \):

\[(115) \quad -\frac{\partial C(Y,w,t)}{\partial t}/C(Y,w,t) = Y_D'(t) - Z_D'(t); \]

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
Network Regulation and Sunk Costs

one output.\textsuperscript{34}

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.\textsuperscript{35}

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:

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

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 $\nabla_y c(y(t),w(t),t)$:

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

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

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

$$c(y(t),w(t),t) = w(t)^T \nabla_w c(y(t),w(t),t) = w(t)^T z(t)$$

using (116).

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

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

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

using (116).

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

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

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

\textsuperscript{34}The one output case was generalised to many outputs by Jorgenson and Griliches (1967).

\textsuperscript{35}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).
observable output prices for the unregulated sector, \( P(t) \).\(^{36}\)

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, \( \rho \), in the regulated sector. In the present context, this definition becomes the following one:

\[
(121) \quad \rho(t) = y(t)^{\top} \nabla_c y(t) / c(y(t),w(t),t) = \rho(t) \mu(t)/c(y(t),w(t),t).
\]

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

\[
(122) \quad \mu(t)^{\top} y(t) = \rho(t)c(y(t),w(t),t).
\]

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

\[
(123) \quad y_D'(t) = \sum_{n=1}^{N} \left[ \mu_n(t) y_n'(t) / \mu(t)^{\top} y(t) \right] \left[ y_n'(t) / y_n(t) \right] ;
\]

\[
(124) \quad z_D'(t) = \sum_{k=1}^{K} \left[ w_k(t) z_k'(t) / w(t)^{\top} z(t) \right] \left[ z_k'(t) / z_k(t) \right].
\]

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 progress in the regulated sector at time \( t \):\(^{37}\)

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

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, \( z_D'(t) \).\(^{38}\)

\section*{A5 OPTIMAL ONE PERIOD REGULATION WITH FREELY VARIABLE CAPITAL SERVICES}

\subsection*{A5.1 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:

\[^{36}\] 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 \textit{not} the vector of output prices \( p(t) \); ie marginal cost weights must be used in (120) rather than revenue weights.

\[^{37}\] The decomposition (125) is equivalent to a decomposition obtained by Denny, Fuss and Waverman (1981; 1996).

\[^{38}\] 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).
• 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 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 A2 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 ≥ 0 to the regulated sector as an additional input. The remaining three classes are the N regulated outputs, y ≥ 0_N, the J internationally traded outputs produced by the unregulated sector, Y ≥ 0_J, and the K primary inputs with the 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.

As in section A2, 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

---


40 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.

41 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.
assume that $S$ satisfies assumption (2); i.e. $S$ is a closed convex cone. Our assumptions on the regulated sector technology set are as follows:

(126) For any $y \geq 0$, 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 A2, we assume that the household has preferences defined over nonnegative consumption vectors of regulated products, $x \geq 0$, and nonnegative vectors of unregulated products, $X \geq 0$. 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 A2, 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) $P^T[Y - X] - P_kk \geq 0$.

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

where $u$ is the household’s utility level and all other decision variables, $y,Y,Z,x,X,k$ 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 $y \geq 0$ be a vector of feasible output targets given this $k$ and let $w >> 0$ be a vector of (noncapital) input prices that the regulated sector faces. Then the regulated sector’s variable cost function or operating cost function or opex function, $c(y,w,k)$, is defined as follows:

(131) $c(y,w,k) = \min_{z} \{w^Tz : (y,z,k) \in S\}$.

It can be shown 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

---

42 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 $(y,z,k) \in S$; i.e. 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.

43 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).
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:

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

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

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

It can be verified that (133) is a concave programming problem and hence the Karlin (1959) regularity conditions, that correspond to the balance of trade constraint (129).

Using 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:

(135) $H(y) = \max_{u,Y} \min_{p,w,k} \{ u + w^T v + p^T y - c(y,w,k) + \lambda P^T Y - C(Y,w) - \lambda P_k k \}
- e(u,p,\lambda P) : u \geq 0; Y \geq 0; p \geq 0; w \geq 0 k; \lambda \geq 0 \}$.

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, we can see 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) $\max_{y,0} \min_{u,Y} \{ u + w^T v + p^T y - c(y,w) + \lambda P^T Y - C(Y,w) - \lambda P_k k \}
- e(u,p,\lambda P) : u \geq 0; Y \geq 0; p \geq 0; w \geq 0 k; \lambda \geq 0 \}$.

We also require that $c(y,w,k)$ be finite in a neighborhood around the point $(y,w,k)$.

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 $x^* \geq 0$, $X^* \geq 0$, $z^* \geq 0$, $Z^* \geq 0$, and $Y^* \geq 0$, such that $y - x^* > 0$; $v - z^* - Z^* > 0$; $p^T[Y^* - X^*] > 0$ and $k^* > 0$. This is not a restrictive assumption.
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 \(u^* > 0, y^* > 0, p^* > 0, w^* > 0, \lambda^* > 0\). 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 \(\lambda\) and setting these partial derivatives equal to zero.

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:

\[
\begin{align*}
(138) \quad 1 &= \partial e(u^*, p^*, \lambda^* P)/\partial u; \\
(139) \quad p^* &= \nabla_y c(y^*, w^*, k^*); \\
(140) \quad \lambda^* P &= \nabla_Y C(Y^*, w^*); \\
(141) \quad v &= \nabla_u c(y^*, w^*, k^*) + \nabla_u C(Y^*, w^*); \\
(142) \quad y^* &= \nabla_p e(u^*, p^*, \lambda^* P); \\
(143) \quad P_k &= -\partial c(y^*, w^*, k^*)/\partial k \\
(144) \quad 0 &= P^T[Y^* - \nabla_p e(u^*, p^*, \lambda^* P)] - P_k k^*. 
\end{align*}
\]

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 A2; ie we will set \(\lambda^*\) equal to unity in what follows. Thus, with this normalisation, the remaining equations (139)-144) become the following equations:

\[
\begin{align*}
(145) \quad p^* &= \nabla_y c(y^*, w^*, k^*); \\
(146) \quad P &= \nabla_Y C(Y^*, w^*); \\
(147) \quad v &= \nabla_u c(y^*, w^*, k^*) + \nabla_u C(Y^*, w^*); \\
(148) \quad y^* &= \nabla_p e(u^*, p^*, \lambda^* P); \\
(149) \quad P_k &= -\partial c(y^*, w^*, k^*)/\partial k \\
(150) \quad 0 &= P^T[Y^* - \nabla_p e(u^*, p^*, \lambda^* P)] - P_k k^*. 
\end{align*}
\]

Equations (145)-(148) and (150) are the counterparts to equations (25)-(29) in section A2 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

---

46 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.

47 This follows by a generalisation of Samuelson’s (1947; 34) Envelop Theorem to cover the case of max-min problems.
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 A2 where regulated prices were set equal to \textit{marginal costs} rather than \textit{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).

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:

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

Thus, the value of net exports of goods and services, \( P^T Y^* \), must equal the value of capital imports, \( P_k^* \).

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 \( w \). \textit{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 \( p >> 0 \) and demands that the regulated sector meet all demands \( y \geq 0 \). Given the vector of selling prices \( p \) and the vector of opex marginal costs \( \nabla_y c(y,w,k) \), we can define the \textit{vector of deviations from opex marginal cost pricing} or the \textit{vector of margins over opex marginal cost}, \( m \), as follows:

\[
(152) m = p - \nabla_y c(y,w,k).
\]

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 \textit{general regulatory policy} is characterised by the following equations:

\[
(153) p = m + \nabla_y c(y,w,k);
(154) P = \nabla_Y C(Y,w);
\]

\[48\] Let the full cost function equal \( c'(y,w,P_k) \). In equilibrium, we will have \( c'(y^*,w^*,P_k) = \min_{y,w} [c(y^*,w^*,k) + P_k k^*] = \nabla_y c(y^*,w^*,k) + P_k k^* \). Equation (149) and the Envelop Theorem will imply that \( \nabla_y c'(y^*,w^*,P_k) = \nabla_y c(y^*,w^*,k) \).
Network Regulation and Sunk Costs

(155) \(v = \nabla_v c(y,w,k) + \nabla_v C(Y,w)\)

(156) \(y = \nabla_y e(u,p,P)\)

(157) \(P_k = -\frac{\partial c(y,w,k)}{\partial k}\)

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

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 A2, 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 A2, 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 A2.

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

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.

A5.2 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 A3, it can be shown that we obtain the same comparative statics results as were obtained in section A3 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) \(\nabla_P u(P) = [1 - m^T b]^{-1} \{Y + \nabla_P^2 e(u,p,P) m\}\)

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

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, \( \nabla_e c(y, w) \), then the vector of markups over marginal costs, \( m(P_k) \), will equal 0 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) \[ y'(P_k) = \nabla_{pp} e(u(P_k), p, P) u'(P_k) = b u'(P_k) \]

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.

In the following section, we rework the analysis that was presented in section A4.2 above; ie we look at welfare growth in a model where the opex cost function replaces the total cost function.

**A5.3 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 A4.2, we assume that all variables are differentiable functions of time. Thus, equation (159) is now rewritten as follows:

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

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:

\( u'(t) = w(t)^T v'(t) + [Y(t) - X(t)]^T P'(t) - P'_k(t)k(t) + m(t)^T y'(t) - \frac{\partial c(y,w,k,t)}{\partial t} - \frac{\partial C(Y,w,t)}{\partial t} \)

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 \( -\frac{\partial c(y,w,k,t)}{\partial t} \) 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 \( -\frac{\partial C(Y,w,t)}{\partial t} \), 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:

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

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:

\( u'(t) = \left[1 - m(t)^T b(t)\right]^{-1} \left\{ w(t)^T v'(t) + [Y(t) - X(t)]^T P'(t) + m(t)^T [\nabla_p^2 e] P'(t) - P'_k(t)k(t) + m(t)^T [\nabla_p^2 e] P'(t) - \frac{\partial c(y,w,k,t)}{\partial t} - \frac{\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
Network Regulation and Sunk Costs

counterpart to our previous expression for the rate of growth of welfare, (102), that we obtained in section A4.2 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 A4.3 (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 A5.4 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 A4.2, 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, \(\nabla_{pp} e(u(t),p(t),P(t))\), then the regulator can attempt to determine a vector of price changes for the regulated products, \(p'(t)\), such that \(m(t)'[\nabla_{pp}^2 e]p'(t)\) 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 A4.3, 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, \([- \partial c(y,w,t)/\partial t]/c(y,w,t)\).

### A5.4 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, \([\partial c(y,w,k,t)/\partial t]/c(y,w,k,t)\).

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:

\[
(172) \ z(t) = \nabla_w c(y(t),w(t),k(t),t).
\]

\[50\] This measure of technical progress can also include changes in variables that reflect exogenous operating conditions for the regulated firm.
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:

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

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

$$(175) \ \ w(t)^Tz'(t) + w'(t)^Tz(t) = \nabla_y c(y(t),w(t),k(t),t)^T \dot{y}(t) + \nabla_w c(y(t),w(t),k(t),t)^T \dot{w}(t) + [\partial c(y,w,k,t)/\partial k]k'(t) + \partial c(y,w,t)/\partial t \ \text{using (172) and (174)}.$$  

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

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

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, $\sum_{k=1}^K \omega_k(t)z_k'(t)$, 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 A4.3, 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).

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:

$$(177) \ \rho(y,w,k,t) = \partial \ln c(\lambda y,w,k,t)/\partial k|_{\lambda=1}$$  

$$= [y^T \nabla_y c(y,w,k,t) + \dot{k}c(y,w,k,t)/\partial k]/c(y,w,k,t).$$  

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 \( c(\lambda y, w, \lambda k, t) = \lambda c(y, w, k, t) \) for all \( \lambda \) 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:

\[
\rho(t) = \rho(y(t), w(t), k(t), t) = \frac{\nabla_y c(y(t), w(t), k(t), t) + k(t) \partial c(y(t), w(t), k(t), t)}{c(y(t), w(t), k(t), t)}
\]

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:

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

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:

\[
y_D'(t) = \sum_{n=1}^N \left[ \mu_n(t) y_n(t) / \mu(t)^T y(t) \right] \left[ y_n'(t) / y_n(t) \right],
\]

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

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 \( t \):

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

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, \( y_D'(t) \), 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, \( k'(t)/k(t) \).

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.

---

51 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).

52 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).
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.

A6 MULTIPLE PERIOD REGULATION WITH SUNK COSTS

A6.1 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 $t = 1,2$.

As in section A5.1, 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

---

53 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.
Network Regulation and Sunk Costs

follows:

\[ c_t(y,w,k) = \min \{ w^Tz : (y,z,k) \in \mathbb{S}_t \} ; \quad t = 1,2. \]

As in section A5.1, 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:\(^{54}\)

\[ z_t(y,w,k) = \nabla_w c_t(y,w,k) ; \quad t = 1,2. \]

Let \( Y \geq 0 \) be a vector of output targets and \( w >> 0 \) 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:

\[ C_t(Y,w) = \min \{ w^TZ : (Y,Z) \in \mathbb{S}_t \} ; \quad t = 1,2. \]

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:

\[ Z_t(Y,w) = \nabla_w C_t(Y,w) ; \quad t = 1,2. \]

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, \( x \geq 0 \), and nonnegative vectors of unregulated products, \( X \geq 0 \). 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 \( p_1 >> 0 \) for regulated outputs and \( P_1 >> 0 \) 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:

\[ e(u,p_1,P_1,p_2,P_2) = \min \{ p_1^Tx + P_1^TX + p_2^Tx + P_2^TX : U(x_1,X_1,x_2,X_2) \geq u ; \]

\[ x_1 \geq 0_N ; X_1 \geq 0_J ; x_2 \geq 0_N ; X_2 \geq 0_J \}. \]

The consumer’s expenditure function will be nondecreasing in all of its variables and linearly homogeneous and concave in the prices \((p_1^1,p_1^2,P_2^1,P_2^2)\). If \( e(u,p_1,P_1,p_2,P_2) \) is differentiable with respect to the components of the commodity prices \( p_1^1 \) and \( P_1^1 \), then Shephard’s Lemma implies

\[^{54}\] We also require that \( c(y,w,k) \) be finite in a neighbourhood around the point \((y,w,k)\) in order to derive this result.
that the consumer’s period t system of Hicksian demand functions for regulated commodities, \( x_t(u, p^1, p^2, P^2) \), is equal to the vector of first order partial derivatives of \( e(u, p^1, p^2, P^2) \) with respect to the components of \( p^1 \) and the consumer’s system of Hicksian demand functions for unregulated commodities in period t, \( X_t(u, p^1, p^2, P^2) \), is equal to the vector of first order partial derivatives of \( e(u, p^1, P^2) \) with respect to the components of \( P^2 \) for \( t = 1, 2 \); ie in the differentiable case, we have:

\[
\begin{align*}
(188) \quad & x_t(u, p^1, p^2, P^2) = \nabla_{p^1} e(u, p^1, p^2, P^2) ; \\
(189) \quad & X_t(u, p^1, p^2, P^2) = \nabla_{p^1} e(u, p^1, p^2, P^2) ;
\end{align*}
\]

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

\[
\begin{align*}
(190) \quad & y_t \geq x_t ; \\
(191) \quad & v_t \geq z_t + Z_t ; \\
(192) \quad & P^1T(Y^1 - X^1) + P^2T(Y^2 - X^2) - P_k k \geq 0.
\end{align*}
\]

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) and \( k \geq 0 \) is the number of units of capital purchased at the beginning of period 1. Note that \( P \) 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:

\[
\begin{align*}
(193) \quad & U(x_1, X_1, x_2, X_2) - u \geq 0 ; \\
(194) \quad & (y_t, z_t, k) \in S_t ; \\
(195) \quad & (Y_t, Z_t) \in S_t ;
\end{align*}
\]

What is fundamentally different about the new intertemporal utility maximisation problem as opposed to the one period utility maximisation problem considered in section A5.1 above is the treatment of capital in the constraints (194); ie note that the capital variable \( k \) is

---

\(^{55}\) 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.
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, 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 \( u^* > 0 \), \( y^* > 0 \), \( Y^* > 0 \), \( z^* > 0 \), \( k^* > 0 \), \( Z^* > 0 \), \( x^* > 0 \) and \( X^* > 0 \) for \( t = 1, 2 \) solves the household’s constrained intertemporal utility maximisation problem. Now let \( y^* \geq 0 \) be in a neighbourhood of \( y^* \) and consider the following constrained maximisation problem that is conditional on the choice of \( y^* \) and \( y^2* \):

\[
H(y^1, y^2) = \max_{u_1, y_1, x_1, z_1, s_1, x_1, s_1} \left\{ u : y^1 - x^1 \geq 0, t = 1, 2 ; y^2 - z^2 \geq 0, t = 1, 2 ; P^T[Y^1 - X^1] + p^T_1[y^1 - x^1] + p^T_2[y^2 - z^2] - p_k \geq 0 ; (y^1, y^2, k) \in S^1 ; (Y^1, Z^2) \in S^2, t = 1, 2 ; U(x^1, x^2, X^2) - u \geq 0 ; u \geq 0, Y^1 \geq 0, z^1 \geq 0, k^* \geq 0, x^1 \geq 0, x^2 \geq 0, X^2 \geq 0, t = 1, 2}\right\}
\]

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 \( u, k, Y^1, z^1, Z^1, x^1 \), and \( X^1 \) solutions to (196) are also solutions to the following max-min problem:\(^{56}\)

\[
H(y^1, y^2) = \max_{u_1, y_1, x_1, z_1, s_1, x_1, s_1} \min_{p^1, p^2, \lambda} \left\{ u + p^1T[y^1 - x^1] + p^2T[y^2 - z^2] + \lambda_1[P^1T[Y^1 - X^1] + P^2T[Y^2 - X^2] - p_k] : (y^1, y^2, k) \in S^1, t = 1, 2 ; (Y^1, Z^2) \in S^2, t = 1, 2 ; U(x^1, x^2, X^2) - u \geq 0 ; u \geq 0, \lambda \geq 0, Y^1 \geq 0, z^1 \geq 0, k^* \geq 0, x^1 \geq 0, X^1 \geq 0, p^1 \geq 0, w^1 \geq 0, t = 1, 2}\right\}
\]

where \( p^1 \) can be interpreted as a vector of prices for the regulated products in period \( t \), \( w^1 \) can be interpreted as a vector of (noncapital) primary input prices for period \( t \), \( t = 1, 2 \), and \( \lambda \) 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 \( z^1, Z^1, x^1 \) and \( X^1 \). Thus, (197) is equivalent to the following max-min problem which now has no constraints other than nonnegativity constraints:

\[
H(y^1, y^2) = \max_{u_1, y_1, x_1, z_1, s_1, x_1, s_1} \min_{p^1, p^2, \lambda} \left\{ u + w^1T[y^1 + p^1T] - c^1(y^1, w^1, k) + \lambda_1 p^1T[y^1] - c^1(y^1, w^1) + w^2T[y^2 + p^2T] - c^2(y^2, w^2, k) + \lambda_2 p^2T[y^2] - c^2(y^2, w^2) - \lambda p_k \right\} : u \geq 0, \lambda \geq 0, Y^1 \geq 0, p^1 \geq 0, w^1 \geq 0, t = 1, 2\right\}
\]

\(^{56}\) In order to obtain the equality of (196) and (197), we need to assume that a constraint qualification condition holds.
Note that the period t opex cost functions for the regulated sector, \( c(Y,t) \), and the period t cost functions for the unregulated sector, \( C(Y,w) \), have made their appearance in (198). It can be seen that solutions to the household’s initial constrained utility maximisation problem 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:

\[
\text{(199) max}_{y_s} \left[ \max_{u,v,k} \min_{p,w/k} \{ u + w^1Y^1 + p^1Y^1 - c^1(y^1,w^1,k) + \lambda P^1Y^1 - C^1(Y^1,w^1,k) \} \right]
\]

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

\[
\text{(200) max}_{y_s} H(y^1,y^2).
\]

We assume that there is a strictly positive solution to the two stage max-min problem defined by (199), say \( u^*, v^*, k^* \), \( Y^r \), \( z^r \), \( k^r \), \( Z^r \), \( x^r \), \( K^r \), \( w^r \), \( P^r \) for \( t = 1,2 \). We also assume that the four cost functions, \( c^1 \) and \( C^1 \) for \( t = 1,2 \) and the consumer’s expenditure function \( e \) are finite and differentiable in a neighbourhood of this equilibrium point. 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 \( y^1, k, u, Y^1, p^1, w^1 \) and \( \lambda \), and setting these partial derivatives equal to zero. Thus, under our assumptions, we find that the optimal solution to the consumer’s constrained utility maximisation problem satisfies the following first order conditions:

\[
\text{(201) } 1 = \partial e(u^*,v^*,k^*,p^1,p^2,\lambda^*)/\partial u ;
\]

\[
\text{(202) } p^r = \nabla_v,c(y^r,w^r,k^r) ; \quad t = 1,2 ;
\]

\[
\text{(203) } \lambda^*p^r = \nabla_x,C(y^r,w^r) ; \quad t = 1,2 ;
\]

\[
\text{(204) } v^r = \nabla_u,c(y^r,w^r,k^r) + \nabla_u,C(y^r,w^r) ; \quad t = 1,2 ;
\]

\[
\text{(205) } y^r = \nabla_y,e(u^*,v^*,k^*,p^1,p^2,\lambda^*,\lambda^*) ; \quad t = 1,2 ;
\]

\[
\text{(206) } P_t = - \partial e(u^*,v^*,k^*,p^1,p^2,\lambda^*)/\partial k ;
\]

\[
\text{(207) } 0 = P^1[Y^2 - \nabla_y,e(u^*,v^*,k^*,p^1,p^2,\lambda^*,\lambda^*)] + P^2[Y^2 - \nabla_y,e(u^*,v^*,k^*,p^1,p^2,\lambda^*,\lambda^*)] - P_t k^* .
\]

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 A2; ie we will set \( \lambda^* = 1 \) equal to unity in what follows. Thus, with this normalisation, the remaining equations (202)-(207) become the

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

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

\[\text{Again, this follows by a generalisation of Samuelson’s (1947; 34) Envelop Theorem to cover the case of max-min problems.}\]
Network Regulation and Sunk Costs

following equations:

\[(208) \quad p^*=\nabla_c(y^*,w^*,k) ; \quad t = 1,2 ;\]

\[(209) \quad P^t = \nabla_x C(Y^t,w^t) ; \quad t = 1,2 ;\]

\[(210) \quad v^t = \nabla_a c(y^t,w^t,k) + \nabla_a C(Y^t,w^t) ; \quad t = 1,2 ;\]

\[(211) \quad y^t = \nabla_u e(u^t,p^t,a^t,P^t) ; \quad t = 1,2 ;\]

\[(212) \quad P_k = -\frac{\partial c(y^t,w^t,k)}{\partial k} - \frac{\partial c(y^t,w^t,k)}{\partial k} ; \quad t = 1,2 ;\]

\[(213) \quad 0 = P^1T[Y^t - \nabla_y e(u^t,p^t,a^t,P^t)] + P^2T[Y^t - \nabla_y e(u^t,p^t,a^t,P^t)] - P_k k^* .\]

Equations (208)-(211) and (213) are the counterparts to equations (145)-(148) and (150) in section A5.1 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 \(\partial c(y^t,w^t,k)/\partial k\) plus \(\partial c(y^t,w^t,k)/\partial k\) of the period 1 and 2 opex cost functions with respect to the (fixed) capital input variable, \(k\). In section A5.1, when capital was freely variable from period to period, the optimality condition was that \(-\partial c(y^t,w^t,k)/\partial k\) 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 A6.3.

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.
price of regulated outputs in period t at \( p^1_t \gg 0 \) and demands that the regulated sector meet all period t demands \( y^1 \geq 0 \) for \( t = 1, 2 \). Given the vector of selling prices \( p^1 \) and the vector of opex marginal costs \( \nabla_p c^1 (y^1, w^1, k) \), we can define the period t vector of deviations from opex marginal cost pricing or the period t vector of margins over opex marginal cost, \( m^1 \), as follows:

\[
(215) \quad m^1 = p^1 - \nabla_p c^1 (y^1, w^1, k) ; \quad t = 1, 2.
\]

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:

\[
(216) \quad P^t = \nabla Y_c^t (y^t, w^t) ; \quad t = 1, 2 ;
\]

\[
(217) \quad v^t = \nabla Y_e^t (y^t, w^t, k) + \nabla V_c (Y^t, w^t) ; \quad t = 1, 2 ;
\]

\[
(218) \quad y^t = \nabla Y_e^t (-e(u, p^t, p^2, P^t)) ; \quad t = 1, 2 ;
\]

\[
(219) \quad P_k = -\partial c^t (y^t, w^t, k) / \partial k - \partial c^2 (y^2, w^2, k) / \partial k ;
\]

\[
(220) \quad 0 = P^1 [Y^1 - \nabla Y_e^1 (-e(u, p^1, p^2, P^1)) + P^2 [Y^2 - \nabla Y_e^2 (-e(u, p^1, p^2, P^2))] - P_k k .
\]

These equations can be regarded as 2 + 4N + 2J + 2K equations in the 2 + 4N + 2J + 2K endogenous variables \( u, k, m^1, y^1, Y^1 \) and \( w^1 \) 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 A2, 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.\(^{61}\)

Using the same techniques that were used in section A2, 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 A5.1.

\[
(221) \quad e(u, p^1, p^2, P^2) = w^1 T v^1 + w^2 T v^2 + P^1 T y^1 - c^1 (y^1, w^1, k) + p^2 T y^2 - c^2 (y^2, w^2, k) - P_k k.
\]

\(^{61}\) 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).
It should be mentioned that the model defined by equations (215)-(220) does not assume that 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 A6.2, we will briefly look at the comparative statics properties of the present model, using the same techniques that were used in section A5.2 above. In section A6.3, 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 A5 above.

A6.2 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 A5 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:

\[
e(u,p^1_p^1, P^1_p^1, p^2, P^2) = u \text{ for all } u \geq 0.
\]

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

\[
\partial e(u,p^1_p^1, P^1_p^1, p^2, P^2)/\partial u = 1.
\]

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 \(u, k, m^1, y^1, w^1\) 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:

\[
\nabla_p u(p^1_p^1) = \left[\nabla_p y^1(p^1)\right]^T m^1(p^1) + \left[\nabla_p y^2(p^1)\right]^T m^2(p^1).
\]

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 \(0\) (so that we have marginal cost pricing in each period), then

\[
\nabla_p u(p^1_p^1) = 0;\]

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, \(x^t(u,p^1_p^1, P^1_p^1, p^2, P^2)\), in terms of partial derivatives of the consumer’s expenditure function, \(e(u,p^1_p^1, P^1_p^1, p^2, P^2)\). Differentiating both sides of equations (188) with respect to \(u\) leads to the following equations (and definitions):
(226) $\partial x(u,p^1,p^1,p^2,p^2)/\partial u = \nabla^2_{p^1} e(u,p^1,p^1,p^2,p^2) = b^1$; \hspace{1cm} 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) $\nabla_{p^1} y^1(p^1) = \nabla_{p^1}^2 e(u,p^1,p^1,p^2,p^2) \nabla_{p^1} u(p^1) + \nabla_{p^1}^2 e(u,p^1,p^1,p^2,p^2)$;

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

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, $\Sigma$:

(229) $\Sigma = \begin{bmatrix} \nabla_{p^1} y^1 & \nabla_{p^1} y^2 \\ \nabla_{p^1} y^1 & \nabla_{p^1} y^2 \end{bmatrix}$

Since $\Sigma$ is a negative semidefinite symmetric matrix, so are the submatrices, $\Sigma_{11}$ and $\Sigma_{22}$. Moreover, the symmetry of $\Sigma$ implies $\Sigma_{21}$ equals $\Sigma_{12}$. 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) $(1 - m^1 b^1 - m^2 b^2) \nabla_{p^1} u(p^1) = \Sigma_{11} m^1 + \Sigma_{12} m^2$.

Typically, $m^1 b^1 + m^2 b^2$ will be much smaller than 1 so that the vector of derivatives $\nabla_{p^1} u(p^1)$ of utility with respect to the components of the vector of first period vector of regulated prices will be equal to the vector $\Sigma_{11} m^1 + \Sigma_{12} m^2$ times the positive number $(1 - m^1 b^1 - m^2 b^2)^{-1}$. 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, $\Sigma_{11}$ and $\Sigma_{12}$, 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) $(1 - m^1 b^1 - m^2 b^2) \nabla_{p^2} u(p^2) = \Sigma_{21} m^1 + \Sigma_{22} m^2$.

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

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

Note that the intertemporal comparative statics results derived in this section are very similar to the results derived in section A3.4 above which worked out expressions for the effects on household welfare of changes in the price of regulated commodities. Our advice in section A3.4 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
Network Regulation and Sunk Costs

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, \( \Sigma \).

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.

### A6.3 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 A5.4 above; ie recall equation (157) in section A5.4. However, in section A5.4, 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:

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

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

\[
(234) \quad P_k = P_k^1 + P_k^2
\]

where \( P_k^t > 0 \) 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:

\[
(235) \quad \min_{k>0} \{ c^1(y^1,w^1,k) + c^2(y^2,w^2,k) + P_k k \}.
\]
The first order necessary condition for an interior solution to (235) is:

\[ \partial c^1(y^1,w^1,k)/\partial k + \partial c^2(y^2,w^2,k)/\partial k + p_k = 0. \]

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.\(^{62}\)

Recall equation (176) in section A5.4 which allowed us to obtain an estimator for the rate of technical progress in the regulated sector. We repeat this equation for convenience:

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

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.

### A6.4 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\(^{63}\) 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 \( \delta \) 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

---

\(^{62}\) 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.

\(^{63}\) 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.
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 in period 3, the asset is retired. Under these conditions, we can still define the opex cost function for the regulated sector in period \(t\), \(c'(y,w,k)\), 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 \(c'(y,w,k')\) where \(k' = k\) and the period 2 opex cost function is \(c'(y,w,k')\) where \(k' = (1-\delta)k\). Under these conditions, the max-min problem defined by (197) becomes the following problem:

\[
\begin{align*}
\text{(238)} \quad H(y^1,y^2) &= \max_{u,Y^1,T\in S} \left\{ u + w^T v^1 + w^T v^2 - c'(y^1,w^1,k) + \lambda P_1^{1T} Y^1 - c'(y^2,w^2,(1-\delta)k) + \lambda P_2^{1T} Y^2 - C^1(Y^1,w^1) - \lambda P_1\right\} \quad \lambda \geq 0, \quad u \geq 0, \quad Y^1 \geq 0, \quad Y^2 \geq 0, \quad \lambda \geq 0, \quad Y^1 \geq 0, \quad Y^2 \geq 0, \quad u \geq 0, \quad \lambda \geq 0, \quad \lambda \geq 0.
\end{align*}
\]

where as before, \(p^1\) can be interpreted as a vector of prices for the regulated products in period \(t\), \(w^1\) can be interpreted as a vector of (noncapital) primary input prices for period \(t\), \(t = 1,2\) and \(\lambda\) is a Lagrange multiplier that corresponds to the balance of trade constraint (192). Using the same arguments as were made in section A6.1, it can be shown that (238) is equivalent to the following max-min problem which now has no constraints other than nonnegativity constraints:

\[
\begin{align*}
\text{(239)} \quad H(y^1,y^2) &= \max_{u,Y^1,T\in S} \left\{ u + w^T v^1 + w^T v^2 - c'(y^1,w^1,k) + \lambda P_1^{1T} Y^1 \right\} \quad \lambda \geq 0, \quad u \geq 0, \quad Y^1 \geq 0, \quad Y^2 \geq 0, \quad \lambda \geq 0, \quad \lambda \geq 0.
\end{align*}
\]

We can now follow the rest of the analysis in section A6.1 and derive counterparts to equations (215)-(220), which characterised a regulatory equilibrium. The new equations are:

\[
\begin{align*}
(240) \quad m^1 &= \lambda k - \nabla_u c'(y^1,w^1,(1-\delta)k); \\
(241) \quad p^1 &= \nabla_w c'(y^1,w^1); \\
(242) \quad v^1 &= \nabla_w c'(y^1,w^1,(1-\delta)k) + \nabla_w C^1(Y^1,w^1); \\
(243) \quad v^2 &= \nabla_w e(u,p^1,p^2,p^3); \\
(244) \quad P_0 &= -\delta c'(y^1,w^1,k)\delta k - (1-\delta)\delta c'(y^2,w^2,(1-\delta)k)\delta k; \\
(245) \quad 0 &= P_1^{1T} [Y^1 - \nabla_p e(u,p^1,p^1,p^2)] + P_1^{2T} [Y^2 - \nabla_p e(u,p^1,p^1,p^2)] - P_1 k.
\end{align*}
\]

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 A6.1, equation (219), which we later rewrote as equation (234) in section A6.3. We will now obtain counterparts to equations (233) and (234) in section A6.3. 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

\[\text{The new equations are exactly the same as the old ones if the geometric depreciation rate } \delta \text{ equals 0.}\]
and there is geometric depreciation, then we have the following relationships between \( k, k^1 \) and \( k^2 \):

\[
(246) \quad k^1 = k ; \quad k^2 = (1-\delta)k.
\]

Define the user benefit price for the use of one (new) unit of the fixed capital input in period \( t \), \( P_k^1 \), as follows:

\[
(247) \quad P_k^1 = -\partial c_1^1(y^1,w^1,k^1)/\partial k ; \quad P_k^2 = -\partial c_2^2(y^2,w^2,k^2)/\partial k.
\]

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:

\[
(248) \quad P_k = P_k^1 + (1-\delta)P_k^2
\]

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 \( P_k \) 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 \( P_k \) into intertemporal components:

\[
(249) \quad P_k = P_k^1k^1 + P_k^2k^2.
\]

Thus, the beginning of period 1 cost of the sunk cost asset, \( P_k \), is equal to the sum of the period by period user benefit prices, \( P_k^1 \), times period \( t \) capital input in constant efficiency units, \( k^t \), for \( t = 1,2 \). Note that (249) is equivalent to (244).65

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.

### A6.5 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:

\[
(250) \quad -P_k + (1+r_1)^{-1}P_1^1[T_1 Y_1 - X_1] + (1+r_1)^{-1}(1+r_2)^{-1}P_2^2[T_2 Y_2 - X_2] \geq 0.
\]

\[65 \] 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.
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 value, $P_1k$, 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 so that the period 1 value of net exports (excluding the imported sunk capital), $P^1[Y^1 - X^1]$, 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 $P^1$ 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, $P^2[Y^2 - X^2]$, discounted to the beginning of period 1 is $P^2[Y^2 - X^2]/(1+r_1)(1+r_2)$. 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 $p^1$ and $w^1$, 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:

$$
H(y^1, y^2) = \max_{u, y, w, z, x_1, x_2, k} \min_{p^1, w^1, k} \{ u + (1+r_1)^{-1}p^1[y^1 - x_1] \\
+ (1+r_1)^{-1}w^1[y^2 - z^2] + (1+r_1)^{-1}(1+r_2)^{-1}p^2(y^2 - x_2) + (1+r_1)^{-1}(1+r_2)^{-1}w^2[y^2 - z^2] \\
+ \lambda (1 - P^1 k + (1+r_1)^{-1}p^1[Y^1 - X^1] + (1+r_1)^{-1}(1+r_2)^{-1}p^2[Y^2 - X^2]): (y^1, y^2, z^2, x_1, x_2) \in \mathbb{S}^1, t = 1, 2; \\
(y^1, z^2) \in \mathbb{S}^2, t = 1, 2; u(x^1, x^2, x^3) - u \geq 0; u \geq 0, \lambda \geq 0, y^1 \geq 0, z \geq 0, k, Z^2 \geq 0, k, \\
x^1 \geq 0, x \geq 0, p \geq 0, w \geq 0, t = 1, 2).$$

In (197), the vectors $p^1$ and $w^1$ were interpreted as discounted future expected price vectors for regulated products and for primary inputs in period $t$ whereas in (251), $p^1$ and $w^1$ 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 A6.1 proceeds in much the same manner. The counterparts to equations (215)-(220) which characterise a regulatory equilibrium using explicit discounting are the following equations:

$$
(252) m^1 = p^1 - \nabla V_C(y^1, w^1, k); \quad t = 1, 2; \\
(253) P^1 = \nabla V_C(Y^1, w^1); \quad t = 1, 2; \\
(254) v^1 = \nabla v_C(y^1, w^1, k) + \nabla v_C(Y^1, w^1); \quad t = 1, 2; \\
(255) y^1 = \nabla \rho C(1+r_1)^{-1}p^2, (1+r_1)^{-1}p^2, (1+r_1)^{-1}p^2, (1+r_2)^{-1}w^2; \quad t = 1, 2; \\
$$

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).
(256) $p_k = -\partial c^2(y^t,(1+r_t)^{-t}w^t,k)/\partial k - \partial c^2(y^t,(1+r_t)^{-t}w^t,k)/\partial k$;
(257) $0 = (1+r_t)\partial P^1\partial [Y^1 - X^t] + (1+r_t)\partial P^2\partial [Y^2 - X^t] - P_k k$
where $X^t = \nabla \mu (u, (1+r_t)^{-t}p^1,(1+r_t)^{1-t}P^1,(1+r_t)^{1-t}P^2,(1+r_t)^{-t}(1+r_t)^{-t}p^2) \text{ for } t = 1,2$. 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) $p_k = - (1+r_t)^{-t}\partial c^2(y^t,w^t,k)/\partial k - (1+r_t)^{-t}\partial c^2(y^t,w^t,k)/\partial k$.

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 A6.3 as follows:

(259) $p_k^1 = -\partial c^2(y^t,w^t,k)/\partial k$; $p_k^2 = -\partial c^2(y^t,w^t,k)/\partial k$.

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

(260) $p_k = (1+r_t)^{-t}p_k^1 + (1+r_t)^{-t}p_k^2$.

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) $\min_{k \geq 0} \{(1+r_t)^{-t}c^1(y^t,w^t,k) + (1+r_t)^{-t}c^2(y^t,w^t,k) + p_k k\}$.

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 Appendix.

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

(262) $p_k = (1+r_t)^{-t}p_k^1 + (1+r_t)^{-t}p_k^2 k^2$
where the $p_k^1$ are defined by (259) and the $k^2$ are defined by (246); ie $k^1 = k$ and $k^2 = (1-\delta)k$. 107
Network Regulation and Sunk Costs

A7 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. This determines the price cap for the following period.

We use the same notation that was developed in section A5.4 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 \( y(t), w(t) \) and \( k(t) \) 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:

\[
C_z(t) = c(y(t),w(t),k(t),t) \tag{263}
\]

The vector of time \( t \) marginal costs, \( \mu(t) \), is defined as in section A5:

\[
\mu(t) = \nabla_y c(y(t),w(t),k(t),t) \tag{264}
\]

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

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

\[
(265) \quad \tau(t) = -\frac{\partial \log c(y(t),w(t),k(t),t)}{\partial t} = \frac{\partial c(y(t),w(t),k(t),t)}{c(y(t),w(t),k(t),t)} \geq 0
\]

Recall equation (176) in section A5.4 which explained how an estimator for \( \tau(t) \) could be obtained. Using the notation in the present section, the counterpart to (176) is:

\[
(266) \quad \tau(t)C_z(t) = \mu(t)^T y'(t) - w(t)^T z'(t) - \pi(t)k'(t).
\]

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 A5.4, the vector of opex input demands at time \( t \) is \( z(t) \) defined via Shephard’s Lemma as usual:

\[
(268) \quad z(t) = V_w c(y(t),w(t),k(t),t).
\]

Differentiating period \( t \) opex cost \( C_z(t) \) defined by (265) with respect to time \( t \) and using (264)-(268), we obtain the following expression for \( C'_z(t) \):

\[
(269) \quad C'_z(t) = \mu(t)^T y'(t) + z(t)^T w'(t) - \pi(t)k'(t) - \tau(t)C_z(t).
\]

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, \( \pi(t)k'(t) \), 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 \( y_n(t) \):

\[
(270) \quad R(t) = p(t)^T y(t).
\]

Thus, the rate of change of time \( t \) revenues is:

\[
(271) \quad R'(t) = p'(t)^T y(t) + p(t)^T y'(t).
\]

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 A10, 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 \( t \) 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:

\[(272) \quad C_k(t) = \sum_{m=1}^{M} P_{km}(t)k_m(t) = P_k(t) \cdot k(t)\]

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

\[(273) \quad C_k'(t) = P_k'(t) \cdot k(t) + P_k(t) \cdot k'(t).\]

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:

\[(274) \quad \Pi(t) = R(t) - C_k(t) - C_d(t).\]

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 $\Pi'(t)$:

\[(275) \quad \Pi'(t) = p'(t) \cdot y(t) - w'(t) \cdot z(t) - P_k'(t) \cdot k(t) + \tau(t)C_d(t)\]

\[+ [p(t) - \mu(t)]y'(t) + [P_k(t) - \pi(t)]k'(t).\]

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, $\tau(t)C_d(t)$, 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, $\mu(t)$, plus 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:

\[(276) \quad p'(t) = \alpha'(t)p(t)\]

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

\[(277) \quad \Pi'(t) = \beta R(t).\]

Substituting (275) and (276) into (277) gives us the following solution for the allowable rate

---

69 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.
of increase in regulated prices $\alpha'(t)$:

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

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. 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 $\beta$ equal to 0 and (287) would simplify to:

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

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

Define the Divisia index of output price growth at time $t$, $w'_D(t)$, and the Divisia index of amortisation prices for sunk capital stocks at time $t$, $P'_D(t)$, as follows:

$$w'_D(t) = \sum_{k=1}^{K} w_k(t)z_k(t)/w(t) \sum_{k=1}^{K} z_k(t)$$

$$P'_D(t) = \sum_{m=1}^{M} P_m(t)k_m(t)/P(t) \sum_{m=1}^{M} k_m(t)$$

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):

$$\alpha'(t) = [C_k(t)/R(t)]w'_D(t) + [C_k(t)/R(t)]P'_D(t) - [C_k(t)/R(t)]\tau(t)$$

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 output costs to revenues, $C_k(t)/R(t)$, times the Divisia rate of increase in output 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, $P'_D(t)$, less the ratio of output costs to revenues, $C_k(t)/R(t)$, 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, $\tau(t)$.

---

70 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.”

71 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 $-\{p(t) - \mu(t)\}y'(t)$ 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.

72 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 $\tau(t)$ 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.
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) 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.

A8 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. 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 \( y_0(t) \), of variable input growth \( z_0(t) \) and of (sunk cost) capital services growth, \( k_0(t) \) at time \( t \):

\[
(283) \quad y_0'(t) = \frac{p(t) \cdot y'(t)}{p(t) \cdot y(t)} = \frac{p(t) \cdot y'(t)}{R(t)} ;
\]

\[
(284) \quad z_0'(t) = \frac{w(t) \cdot z'(t)}{w(t) \cdot z(t)} = \frac{w(t) \cdot z'(t)}{C_z(t)} ;
\]

\[
(285) \quad k_0'(t) = \frac{P_k(t) \cdot k'(t)}{P_k(t) \cdot k(t)} = \frac{P_k(t) \cdot k'(t)}{C_k(t)}.
\]

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:

\[
(286) \quad T'(t) = y_0'(t) - s(t)z_0'(t) - s(t)k_0'(t)
\]

73 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 \( w'(t) \).

74 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 Dievert (2006).
where the time t input cost share weights $s_d(t)$ and $s_u(t)$ and total cost $C(t)$ are defined as follows:

\[(287) \quad s_d(t) = C_d(t)/C(t) \quad ; \quad s_u(t) = C_u(t)/C(t) \quad ; \quad C(t) = C_d(t) + C_u(t).\]

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 $C_d(t)$ plus $C_u(t)$ equal to $w(t)\cdot z(t)$ plus $P_d(t)\cdot k(t)$. Time t pure profits are $\Pi(t)$ equal to $R(t)$ less $C_d(t)$ less $C_u(t)$. Thus, the time t derivative of profits is equal to the following expression:

\[(288) \quad \Pi'(t) = p'(t)\cdot y(t) - w'(t)\cdot z(t) - P_d'(t)\cdot k(t) + p(t)\cdot y'(t) - w(t)\cdot z'(t) - P_d(t)\cdot k'(t).\]

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

\[(289) \quad \tau(t)C(t) + [p(t) - \mu(t)]\cdot y'(t) + [P_d(t) - \pi(t)]\cdot k'(t) = p(t)\cdot y'(t) - w(t)\cdot z'(t) - P_d(t)\cdot k'(t)\]

\[= R(t)\cdot y'(t) - C_d(t)\cdot z'(t) - C_u(t)\cdot k'(t) \quad \text{using (283)-(285)}\]

\[= C(t)\left[\left[R(t)/C(t)\right]y'(t) - s_d(t)z'(t) - s_u(t)k'(t)\right]\]

\[= C(t)\left[\left[\Pi(t)/C(t)\right]y'(t) + T'(t)\right] \quad \text{using (286).}\]

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

\[(290) \quad p'(t)\cdot y(t)/C(t) - s_d(t)\cdot w_d'(t) - s_u(t)\cdot P_{dD}'(t) = T'(t) + [\Pi'(t)/C(t)] - [\Pi(t)/C(t)]\cdot y_d'(t)\]

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

\[(291) \quad w_d'(t) = w'(t)\cdot z(t)/w(t)\cdot z(t) = w'(t)\cdot z(t)/C_d(t) ; \]

\[(292) \quad P_{dD}'(t) = P_d'(t)\cdot k(t)/P_d(t)\cdot k(t) = P_d(t)\cdot k'(t)/C_d(t).\]

Equation (290) is the sunk cost generalisation of equation (2.5) in Bernstein and Sappington (1999; 9). 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, $\alpha'(t)$:

\[(293) \quad \alpha'(t) = \left[\left[R(t)/C(t)\right]s_d(t)w_d'(t) + s_u(t)P_{dD}'(t) - T'(t)\right] + [\Pi'(t)/R(t)] - [\Pi(t)/R(t)]\cdot y_d'(t)\]

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:

\[(294) \quad \alpha'(t) = s_d(t)w_d'(t) + s_u(t)P_{dD}'(t) - T'(t).\]

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, $s_d(t)w_d'(t) + s_u(t)P_{dD}'(t)$, less the anticipated rate of

---

75 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).
growth of TFP, $\tau(t)$.\footnote{Note that if $\tau(t)$ is positive, then this \textit{direct method of price cap regulation} is definitely not equivalent to old style rate of return regulation (which would use formula (294) with $\tau(t)$ 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).}

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:

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

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

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 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. 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 $\tau(t)$:

$$T'(t) = \tau(t)s_z(t)$$

(296)

where $s_z(t)$ is the share of operating costs in total costs at time $t$, $C(t)$. 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.79

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. 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

78 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.

79 See Balk (2003) for a fuller explanation of this phenomenon. Balk called the factor $s_z(t)$ in (296) the “Domar factor”.

80 We will call the resulting formula (304) the indirect method for implementing the price cap formula given by (293), which is the direct method.
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) $R_E(t) = P(t)Y(t) = W(t)Z(t) = C_E(t)$.

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, $W_D'(t)$:

(298) $Y_D'(t) \equiv P(t)Y'(t)/P(t)Y(t) = P(t)Y'(t)/R_E(t)$;
(299) $Z_D'(t) \equiv W(t)Z'(t)/W(t)Z(t) = W(t)Z'(t)/C_E(t)$;
(300) $P_D'(t) \equiv P'(t)Y(t)/P(t)Y(t) = P'(t)Y(t)/R_E(t)$;
(301) $W_D'(t) \equiv W'(t)Z(t)/W(t)Z(t) = W'(t)Z(t)/C_E(t)$.

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) $T_E'(t) = Y_D'(t) - Z_D'(t)$.

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:

(304) $\alpha'(t) = P_D'(t) + \left\{ \frac{C(t)}{R(t)}[s_d(t)w_D'(t) + s(t)P_{LD}'(t)] - W_D'(t) \right\} - \left\{ \frac{C(t)}{R(t)}[T'(t) - T_E'(t)] + \left[ \frac{\Pi(t)}{R(t)} \right] y_D'(t) \right\} = P_D'(t) - X(t)$

where the $X$ factor at time $t$, $X(t)$ is defined as follows:

(305) $X(t) = \left\{ \frac{C(t)}{R(t)}[T'(t) - T_E'(t)] + \left[ \frac{\Pi(t)}{R(t)} \right] y_D'(t) \right\} + \left[ \frac{\Pi(t)}{R(t)} \right] y_D'(t) - \frac{W_D'(t) - \left[ \frac{C(t)}{R(t)}[s_d(t)w_D'(t) + s(t)P_{LD}'(t)] \right]}{T'(t)}$.

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, \( W_D(t) \), less \( C(t)/R(t) \) times a share weighted rate of growth of non sunk cost input prices for the regulated firm, \( w_0(t) \), and the rate of growth of allowable amortisation charges for sunk cost capital inputs, \( p_{0D}(t) \). 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, \( s_z(t) = \frac{C_z(t)}{C(t)} \) and the period \( t \) ratio of allowable amortisation costs to total cost, \( s_k(t) = \frac{C_k(t)}{C(t)} \) 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, \( W_D(t) \).

---

81 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.

82 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.”
Network Regulation and Sunk Costs

• 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.  

• 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 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

---

83 Different treatments of the user cost formula can give different TFP growth estimates; see Harper, Berndt and Wood (1989).
Network Regulation and Sunk Costs

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.

A9 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
Network Regulation and Sunk Costs

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:
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 \( \text{anticipated peak load demand} \ V_P \) 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;
Network Regulation and Sunk Costs

- 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.

A10 FINANCIAL CAPITAL MAINTENANCE AND SUNK COSTS

In section A7 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:

\[
(308) \quad P_1 = (1+r_1)^{-1}C_1 + (1+r_2)^{-1}(1+r_1)^{-1}C_2 + ... + (1+r_1)^{-1}... (1+r_T)^{-1}C_T.
\]

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 \( P_k \) defined by (259) times \( k \). Thus, (308) is a counterpart to equation (260), which assumed one hoss shay depreciation.

\[\text{footnote}{\text{We have neglected any scrap value that the asset may have at the end of its useful life and we have neglected any disposal or decommissioning costs associated with disposing of the asset.}}\]
Network Regulation and Sunk Costs

If the modified one hoss shay model is more applicable, then equation (262) related the initial purchase charge, $P_k$, to the period $t$ user charges, $P_{kt}$. Equation (262) can be put into the form defined by (308) if we set $P_1$ equal to $P_k$ and the period $t$ user charges $C_t$ equal to $P_{kt}$ for $t = 1,2,...,T$ where the $P_k$ were defined by (259) and the $k'$ were defined by (246); ie $k' = (1-\delta)^{-1} k$.

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 addressed in due course is: how exactly should the regulator choose the $C_t$?

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:

$$P_2 = (1+r_2)^{-1}C_2 + (1+r_2)^{-1}(1+r_3)^{-1}C_3 + \ldots + (1+r_T)^{-1}C_T.$$

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:

$$D_1 = P_1 - P_2.$$

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 $C_1$:

$$(312) C_1 = (1+r_1)P_1 - P_2 = (1+r_1)P_1 - [P_1 - D_1] \quad \text{using} \quad (310)$$

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 $D_1$.

---

85 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.

86 See Schmalensee (1989; 294) and Johnstone (2003; 4) for a similar formula.
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:

\[
(313) \quad P_t = (1+r)^t C_1 + (1+r)^{t-1} C_{t+1} + \ldots + (1+r)^{T-1} C_T; \quad t = 2, 3, \ldots, T;
\]

\[
(314) \quad D_t = P_t - P_{t+1}; \quad t = 2, 3, \ldots, T;
\]

where $P_{T+1}$ is defined to be 0. The period $t$ counterpart to equation (311) is:

\[
(315) \quad P_t = (1+r)^t C_1 + (1+r)^{t-1} P_{t+1}; \quad t = 2, 3, \ldots, T.
\]

Equation (315) can now be rearranged to give us the following user cost like formula for the period $t$ user charge $C_t$:

\[
(316) \quad C_t = (1+r)^t P_t - P_{t+1} = (1+r)^t P_t - [P_t - D_t] = r_t P_t + D_t = r_t [P_1 - D_1 - D_2 - \ldots - D_{t-1}] + D_t
\]

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 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.\(^7\)

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:

\[
(317) \quad (1+r_1) P_1 = C_1 + P_2.
\]

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,

\(^7\) 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 $r_t$.\(\)
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:

\[(318) \quad (1+r_2)P_2 = C_2 + P_3.\]

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 $P_2$ 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 $P_0$), creates a problem for the regulator and possibly for the regulated firm.\(^8^9\) 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

\(^8^8\) 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 $P_t$ and $P_{t+1}$ and $C_t$ is Peasnell’s counterpart to our $C_t$. 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).

\(^8^9\) 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), $P_t - iA_{t-1}$, 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 ($t = 2, \ldots, N-1$) 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 ($RAB_{t+1} - RAB_t$), PV is constant regardless of the time pattern of period depreciation charges. It makes no difference over what interval $0 \leq t \leq T$ assets are written down, or how aggregate depreciation expense is distributed within this interval.” David Johnstone (2003; 6)
Network Regulation and Sunk Costs

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\(^9\). 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.

A10.1 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 \( C \):

\[
(319) P_1 = C \left( \frac{1}{1+r_1} + \frac{1}{1+r_1} \frac{1}{1+r_2} + \ldots + \frac{1}{1+r_1} \frac{1}{1+r_2} \frac{1}{1+r_T} \right)
\]

where the discount factors \( \delta_t \) are defined as follows:

\[
(320) \delta_t = \frac{1}{1+r_t} ; \quad t = 1,2,...,T.
\]

Solving equation (319) for the constant user charge \( C \), we have

\[
(321) C = C^* = P_1 / \left\{ \delta_1 + \delta_1 \delta_2 + \ldots + \delta_1 \delta_2 \ldots \delta_T \right\} ; \quad t = 1,2,...,T.
\]

Once the constant \( C \) 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 \( P_t \):

\[
(322) P_2 = C^* \delta_2 + \delta_2 \delta_3 + \ldots + \delta_2 \ldots \delta_T ;
\]

\[
(322) P_3 = C^* \delta_3 + \delta_3 \delta_4 + \ldots + \delta_3 \ldots \delta_T ;
\]

\[
(322) ... \]

\[
(322) P_T = C^* \delta_T.
\]

Now equations (314), \( D_t = P_t - P_{t+1} \), 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

\[
(323) r_t = r \neq 0 ; \quad t = 1,2,...,T.
\]

Under assumptions (323), the discount rates \( \delta_t \) 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
\[(324) \, r = 0 \text{ and hence all } \delta_t = 1/(1+r) = 1 ; \quad t = 1,2,\ldots,T.\]
Under these conditions, we obtain the following solution for the constant period \(t\) charges, \(C_t\), regulatory asset values \(P_t\) and depreciation amounts \(D_t\):
\[
\begin{align*}
(325) \quad C_t &= C^* = P_t/T ; \\
(326) \quad P_t &= C^* \{T+1-t\} ; \\
(327) \quad D_t &= C^* ;
\end{align*}
\]
Thus, in this case where the cost of capital \(r\) 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
\[(328) \, r > 0 \text{ and hence all } \delta_t = 1/(1+r) \neq \delta ; \quad t = 1,2,\ldots,T,\]
where \(0 < \delta < 1\). In this case, equation (319) becomes:
\[
\begin{align*}
(329) \quad P_1 &= C \{\delta_1 + \delta_2 + \ldots + \delta_{T} \} \\
&= C \{\delta + \delta^2 + \ldots + \delta^T \} \\
&= \delta C \{1 - \delta^T\}/(1-\delta). \\
\end{align*}
\]
Solving equation (329) for the constant user charge \(C\) leads to the following equations for \(C_t\):
\[
\begin{align*}
(330) \quad C_t &= C^* = P_t/(1-\delta) \{1 - \delta^T\} ; \\
&= 1,2,\ldots,T.
\end{align*}
\]
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:
\[
\begin{align*}
(331) \quad P_t &= \delta C^* \{1 - \delta^{T-1}\}/(1-\delta) ; \\
&= 2,3,\ldots,T.
\end{align*}
\]
Using (329) and (331) for \(t = 2\), we obtain the following expression for period 1 allowable depreciation, \(D_1\):
\[
\begin{align*}
(332) \quad D_1 &= P_1 - P_2 = \delta C^* \{1 - \delta^T\}/(1-\delta) - \delta C^* \{1 - \delta^{T-1}\}/(1-\delta) \\
&= \delta \cdot 1. \\
\end{align*}
\]
Note that if \(r\) equals 0 so that \(\delta\) 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 \(\delta\) is less than one so that allowable first period depreciation in this case will be less than straight line depreciation since \(\delta^T\) 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:
Network Regulation and Sunk Costs

(333) $D_t = P_t - P_{t-1}$

$t = 1, 2, ..., T$

$= [\delta C^* (1 - \delta^{T-L})/(1-\delta)] - [\delta C^* (1 - \delta^{T-L-1})/(1-\delta)]$

$= \delta^{T-L} C^*.$

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).

A10.2 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 $\rho_t$ 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) $C_t = C(1+\rho_1) ... (1+\rho_t); \quad t = 1, 2, ..., T.$

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 $C$:

(335) $P_1 = (1+r_1^*)^{-1} C(1+\rho_1) + (1+r_2^*)^{-1}(1+r_2)(1+\rho_2)

+ ... + (1+r_T^*)^{-1}(1+r_T)^{-1}C(1+\rho_1)...(1+\rho_T)

= C[(1+r_1^*)^{-1} + (1+r_2^*)^{-1}(1+r_2^*)^{-1} + ... + (1+r_T^*)^{-1}(1+r_T^*)^{-1}]

= C[\delta_1^* + \delta_1^* \delta_2^* + ... + \delta_1^* \delta_2^* ... \delta_T^*]$

where the period t real interest rates $r_t^*$ are defined by equations (336) below and the period t real discount factors $\delta_t^*$ are defined by (337):

(336) $(1+r_t^*) = (1+\rho_t)/(1+\rho_t); \quad t = 1, 2, ..., T$

(337) $\delta_t^* = 1/(1+r_t^*); \quad t = 1, 2, ..., T.$

Solving equation (335) for the constant $C$, we have

(338) $C^* = P_t/\{\delta_1^* + \delta_1^* \delta_2^* + ... + \delta_1^* \delta_2^* ... \delta_T^*\}.$

Once $C^*$ has been determined, the sequence of nominal period t charges is determined by
Network Regulation and Sunk Costs

equations (334) which we repeat here for convenience:

\[ C_t = C^{**}(1+\rho_1)\ldots(1+\rho_t) ; \quad t = 1,2, \ldots, T. \]

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 \( t = 2, \ldots, T \) 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 \( P_t \):

\[ P_2 \equiv (1+\rho_1)C^{**}\{\delta_2^* + \delta_2^*\delta_3^* + \ldots + \delta_2^*\delta_T^*\} ; \]
\[ P_3 \equiv (1+\rho_1)(1+\rho_2)C^{**}\{\delta_3^* + \delta_3^*\delta_4^* + \ldots + \delta_3^*\delta_T^*\} ; \]
\[ \ldots \]
\[ P_T \equiv (1+\rho_1)\ldots(1+\rho_{T-1})C^{**}\delta_T^* \]

where \( C^{**} \) is defined by (338) and the real discount factors \( \delta_t^* \) 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 \textit{real} weighted average cost of capital in period \( t, r_t^* \), is a positive constant \( r^* \) for all periods. Thus, we assume that:

\[ r_t^* = r^* > 0 \quad \text{and hence all} \quad \delta_t^* = 1/(1+r^*) = \delta^* ; \quad t = 1,2,\ldots,T \]

where \( 0 < \delta^* < 1 \). In this case, equation (335) becomes:

\[ P_1 = C\{\delta^* + \delta^*\delta^* + \ldots + \delta^*\delta^T\} \]
\[ = \delta^* C\{1 - \delta^*T\}/(1 - \delta^*). \]

Solving equation (342) for \( C \) leads to the following solution:

\[ C^{**} = P_1(1-\delta^*)/(\delta^*(1 - \delta^T)). \]

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:

\[ P_t = (1+\rho_1)\ldots(1+\rho_{t-1})\delta^T C^{**}\{1 - \delta^Tt-1\}/(1 - \delta^*); \quad t = 2,3,\ldots,T \]

where \( C^{**} \) is defined by (343) and \( \delta^* \) 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:

\[ D_t = P_t - P_{t+1} \quad t = 1,2,\ldots,T \]

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 \( \delta \) will be less than \( \delta^* \). Hence the \( C^{**} \) which satisfies (342) and (343) will be smaller than the corresponding \( C^* \) which satisfies (329) and (330).

---

\[^{91} \text{It will generally be the case that the constant nominal interest rate } r \text{ can be assumed to be greater than the corresponding real } r^* \text{ and hence } \delta \text{ will be less than } \delta^* \text{. Hence the } C^{**} \text{ which satisfies (342) and (343) will be smaller than the corresponding } C^* \text{ which satisfies (329) and (330).} \]
\[
(1 + r_1) \times \cdots \times (1 + r_T) \frac{\delta^* C}{(1 - \delta^*)} \]

\[
= (1 + r_1) \times \cdots \times (1 + r_T) \frac{\delta^* C}{(1 - \delta^*)}
\]

where the last approximation will be satisfactory if \( \rho_t \) 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.

A10.3 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 \( t = 1, \ldots, T \). Let the user charge in period \( t \) be:

\[
(346) \quad C_t = C(1+i_1) \times \cdots \times (1+i_t) ; \quad t = 1,2, \ldots, T.
\]

Now substitute (346) into (308) and get the following equation for \( C \):

\[
(347) \quad P_1 = (1+r_1)^{-1} C(1+i_1) + (1+r_2)^{-1}(1+r_2)^{-1} C(1+i_1)(1+i_2) + \cdots + (1+r_T)^{-1}(1+r_T)^{-1} C(1+i_1)(1+i_2) \cdots (1+i_T).
\]

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.

A10.4 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. 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 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

---

\[ P_k^* = -\partial c_t^* (y^t, w^t, k^t)/\partial k \] 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.

---

\[ \text{In our intertemporal theoretical model of regulation described in section 6, recall that equations (259) defined our optimal sequence of user charges } P_k^* = -\partial c(t) y, w, k^t)/\partial k \] 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.
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.
A11 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.
APPENDIX B: THE RATIONALE FOR AND HISTORY OF THE ‘BUILDING BLOCKS’ METHODOLOGY

B1 BACKGROUND

This appendix provides an overview of the rationale and historic development of the ‘building blocks’ methodology for regulating prices in network industries. The ‘building blocks’ methodology reviewed here refers to an approach where prices or revenues are regulated by calculating forward looking, allowable cost components and summing those cost components to define allowable revenue. The cost components are described as cost building blocks and the allowable revenue is sometimes described as ‘building blocks allowable revenue’. Allowable revenue can then be defined as the target regulatory variable (after making additional adjustments based on other objectives) or allowable revenue can be converted to an average price as the target regulatory variable.

It is important to recognise that the methodology is forward looking and that normally costs are both forward looking and defined to reflect prudent expenditure and realistically achievable operational efficiencies. In addition, adjustments are normally made to remove asset revaluation gains and losses (that are not related to efforts to achieve efficiency). The methodology is thus designed so that on an ex-ante basis investors can expect that funds prudently invested in regulated assets will be fully recouped in net present value terms (based on a discount rate that reflects the opportunity cost of the investment) provided actual costs are expected to be comparable to allowable efficient costs. This latter condition is generally referred to as ex-ante financial capital maintenance (FCM) which means that there is an expectation that the value of invested capital will be maintained in real terms over the life of the investment.

Note that ex-ante FCM is intended to be achieved as opposed to ex-post FCM and investors are allowed to retain realised returns in excess of those required to achieve ex-ante FCM and required to bear the costs of realised returns lower than expected over a defined regulatory period. The rationale for adopting ex-ante FCM as a regulatory principle is that it is consistent with ensuring efficient investment occurs.

Baumol (1971, section 4.1.5) was an early regulatory analyst who drew attention to the importance of the investor recovering the full ‘opportunity cost’ of their investment. Baumol noted that, “from the point of the investor, if no more than replacement cost is returned, the entire asset purchase can turn out to be a mistake. That is, the investment decision will have been worth his while only if at the end he receives back his initial purchasing power plus compensation for the use of funds”. Baumol also noted that, from the point of view of society, if consumers of the services produced with the aid of that investment are unwilling to pay the opportunity cost (in real terms) of obtaining the asset in question, then construction of that asset represents a wasteful use of resources. Consequently, payments to capital should “return funds whose discounted value, after correction for changes of the price level, is equivalent to the cost of the investment. This may or may not be equal to the replacement

---

93 Eunica Aure of the Commerce Commission assisted with research for and drafting of this appendix.
cost of the asset”. This is effectively an early exposition of the FCM concept.

The FCM concept is central to the application of the building blocks methodology as applied in Australia, New Zealand and the United Kingdom. However, there are differences in how each jurisdiction implements the FCM concept and the building blocks methodology, particularly with respect to the determination of allowable efficient costs.

Understanding the rationale and evolution of the building blocks methodology and the incorporation of the FCM criterion provides information relevant to determining the appropriate method of asset valuation and the design of methods to determine efficient costs. This appendix focuses on the regulatory approaches in Australia, New Zealand and the United Kingdom to identify when, why and how the building blocks methodology has been used in certain network industries to help ensure that the application of FCM in the New Zealand context will be as effective as possible.

Although there is a focus on these three countries some references are made to experience in other European countries and the United States as well. Selected academic articles were also surveyed to help confirm the theoretical foundations for the development of the building blocks methodology.

B2 THE ORIGINS OF FCM AND CPI–X AS REGULATORY CONCEPTS

B2.1 Rate–of–return regulation, FCM and CPI–X regulation

The origins of the FCM criterion in the regulation of prices for network industries really lie in the adoption of rate–of–return regulation, as this approach is intended to ensure that investors receive a “fair rate of return” on their capital invested after allowing for all costs incurred. Costs need to include appropriate allowances for depreciation so that investors receive a return of capital as well as a fair return on their capital and, to the extent that both are achieved, then financial capital will be maintained in real terms. However, the advantage of the FCM criterion is in how tightly it is defined so that there is no doubt that when applied on an ex–ante basis financial capital is expected to be maintained in real (inflation–adjusted and risk–adjusted) terms.

The essential differences between rate–of–return regulation and CPI–X price cap regulation are outlined below for context before focusing on the more specific methodology for determining price caps known as ‘building blocks’ incorporating FCM.

As rate–of–return regulation was the precursor to more explicit adoption of the FCM concept and CPI–X regulation, it is important to understand the weaknesses of rate–of–return regulation. The main flaws in rate of return regulation were that if actual costs were used there would be no incentive to achieve efficiencies and (with a cap to the rate of return) there would be an incentive to over invest in the capital base known as the Averch–Johnson (1962) effect. Rate–of–return regulation without indexing of costs also became problematic when inflation was significant.

Over time, rate–of–return regulation was modified so that estimates of efficient forward
Looking costs could be used rather than actual costs (Thompson 1991, p.201) and this together with lags between the setting of prices, with scope for excess profits to be retained in the interim period, improves incentives for efficiency. Although various adjustments can ameliorate the adverse efficiency incentives associated with rate–of–return regulation, this approach can still essentially be seen as providing a ‘low powered’ incentive mechanism (Baldwin and Cave 1999, p.225) to the extent that revenues are not decoupled from actual costs. The consensus view is that there is less scope for the decoupling of revenues from costs to occur under rate–of–return regulation compared to CPI–X regulation.

To the extent that price capping by a forward looking CPI–X mechanism makes it easier to decouple revenues from costs it can be described as a ‘high powered’ incentive mechanism to achieve efficiencies. The decoupling of revenues from costs arises in the CPI–X system because a forward looking price cap requires an ex ante assessment of efficient opex and capex (Beesley and Littlechild 1989, p.461 and Dassler, et al 2006, p.167) and the time period between price reviews is usually of several years’ fixed duration. These characteristics constituted the main economic efficiency rationale for the CPI–X approach to setting prices. Other advantages included greater flexibility in practice for the regulated firm to set individual prices under a total price cap and greater discretion for the regulator in setting X than under the US tradition in determining cost components (Beesley and Littlechild 1989, p.461). The main disadvantage is that regulated costs may diverge too much from true efficient costs and compromise an objective of maintaining ex-ante financial capital maintenance.

Although there are a few US precursors, CPI–X regulation was first applied on a large scale in the UK to British Telecom in 1984 and then extended to other UK utilities as they were privatised (Baldwin and Cave 1999, p.226). Littlechild (1983) authored an influential report proposing a CPI–X approach for British Telecom. The approach was concerned to avoid the pitfalls of the US style rate–of–return regulation.

However, as subsequently explained by Littlechild, this simple formulation did not emerge elegantly from the draft version of his paper but was motivated more by considering the economic and political constraints on the privatisation of BT. The specific idea was based on the so–called “Buzby Bond” in the context of a privatisation option, where it referred to an RPI–2 per cent cap on BT’s tariffs but the bond never transpired.

Littlechild highlights the role of Professor Alan Walters, the UK Prime Minister’s economic advisor at the time of the introduction of CPI–X for British Telecom, and in particular the argument–clinching quote that Walters drew on from his earlier writing in a key ministerial meeting (Littlechild in Bartle 2003, p.37):

“The imposition of a maximum rate of return has many of the characteristics of a tax rate which is fairly low until the maximum rate is achieved, then it becomes a hundred per cent. We all know the consequences of that sort of tax system on cost control and enterprise.”

The initial formulation of the CPI–X approach did not incorporate the forward looking consideration of efficient costs that is a feature of the ‘building blocks’ approach. But in applying the CPI–X approach, attention naturally turns to estimating future efficient cost levels in order to determine an appropriate X factor.
However, it is emphasised that there is obvious convergence of the rate–of–return and CPI–X approaches when the time period between regulatory reviews is shortened and the methodology for estimating costs to determine starting point prices between the two approaches is similar. In the limiting case where there are annual reviews and the methodology for estimating costs is the same, both CPI–X and rate–of–return regulation based on efficient costs collapse to being simple cost–plus pricing. However, in practice rate–of–return regulation has been increasingly displaced by CPI–X regulation including in the United States (at least for telecommunications, although not for energy) (Littlechild in Bartle 2003, p.40).

B2.2 The capital maintenance concept

Capital maintenance is a longstanding financial accounting concept. It is closely related to the definition of capital that one seeks to explain and, hence, must be consistent with the valuation method in use. Financial capital refers to equity or net assets, while physical capital is the productive or operating capacity of the assets.

Financial capital maintenance (FCM) is the maintenance of the expected income earning power of the shareholders’ investments (or investments’ purchasing power). As long as the net present value (NPV) based on the appropriate opportunity cost discount rate is greater than or equal to zero, financial capital maintenance will be achieved.

Operational capability maintenance (OCM) treats physical assets, instead of the shareholders’ funds, as its main interest. Profit is only recognised after the specific operating capacity of assets has been maintained or when the operating capacity of the enterprise at the end of the period exceeds the operating capacity at the beginning of the period, after excluding any distributions to and contributions from owners during the period. OCM essentially determines asset prices and depreciation charges based on the cost of replacing assets in order to maintain operational capability at a defined level.

For reference, note that for a defined level of operational capability FCM differs from OCM by recognising capital gains and losses as well as the standard OCM charge.

From the reports and other documents reviewed in this study, ex–ante FCM is much more commonly used than OCM in overseas jurisdictions particularly in the UK, Australia and in the member countries of the European Union. Ofgas and British Gas used OCM for a time but switched to FCM after the 1997 Price Control Review.

B2.3 Capital maintenance in the UK

Byatt et al (1986) in an influential two volume report discussed the appropriate choice of accounting rules in the measurement of economic costs including in the context of regulating state owned enterprises. The reports advocated FCM as a common standard for comparing returns and as the ex–ante standard relevant for defining an investor’s expectation with respect to recovering the opportunity cost of capital for a specific investment (Byatt et 1986, Volume I paragraphs 18–19 and Volume II paragraph 3.54).

FCM appears to have been in use by UK regulators since the early 1990s. According to
Whittington (1998), the Regulatory Asset Base in the UK is considered to represent shareholder financial investments rather than the physical assets or operating capacity of the firm. This is reflected by the referencing of the initial financial capital base for the electricity and water sectors to the market value of equity in the immediate post–privatisation period.

The Regulatory Accounting Guidelines (RAG) issued by Ofwat (1992, 2003 and 2007) have been explicit in discussing the concept of capital maintenance as a measure of a company’s profits. These guidelines were formulated to ensure that the accounting statements published by companies are consistent with the economic framework in which they are regulated. The following excerpt from the 1992 guidelines (pp. 3–4) explains the concepts and position of Ofwat (whose director was then Ian Byatt):

“1.4 Profit measurement

1.4.1 The ASC Handbook on 'Accounting for the effects of changing prices' discusses two alternative measures of a company's profits which can be summarised as follows:

Real Financial Capital Maintenance ('FCM') is concerned with maintaining the real financial capital of a company and with its ability to continue financing its functions. Under real FCM, profit is measured after provision has been made to maintain the purchasing power of opening financial capital. This involves the use of a general inflation index such as the RPI. Real FCM therefore addresses the principal concerns of the shareholders of a company. In the absence of general inflation real FCM is equivalent to conventional HCA, with the exception of the treatment of unrealized holding gains (paragraph 1.7.8).

Operating Capability Maintenance ('OCM') is concerned with maintaining the physical output capability of the assets of a company. Under OCM, profit is measured after provision had been made for replacing the output capability of a company's physical assets which involves the use of specific inflation indices such as the Baxter index. This will typically be a major concern for the management of a company and was the approach used in Statement of Standard Accounting Practice ('SSAP') 16.

1.4.2 The Director has a duty to ensure that companies can finance the proper carrying out of their functions. In this regard he has a responsibility to customers to ensure that the return earned by the providers of capital to efficiently operated water companies is sufficient, but no more than sufficient, to induce them to hold shares and to make loans. The Director has therefore decided, following discussions with the Working Group on Accounting Issues for Regulation ('WGAR'), that the regulatory current cost accounts should be prepared on a real FCM basis since this will provide a measure of profit that is well suited to achieving a balance between the providers of capital and customers.

1.4.3 The Director also has a duty to ensure that the companies maintain the required level of physical operating capability. The July Returns to the Director, on the level of service and capital expenditure, are however specifically designed to monitor operating capability plans against required service standards, and the Director has concluded that there is no need to reflect OCM concepts in the current cost accounts.”

Note that the UK water regulator considered using both concepts of capital maintenance at the outset following the duty placed on the Director to guarantee that companies can both
finance the proper carrying out of their functions and, at the same time, maintain the required level of physical operating capability. In the end, Ofwat decided to apply real FCM only for the following reasons:

• The report on the level of service and capital expenditure contained in the July Returns to the Director was specifically designed to monitor operating capability plans against required service standards. Hence, there was no need to reflect OCM concepts in the current cost accounts anymore.

• Preparing the regulatory current cost accounts on a real FCM basis would provide a measure of profit that was well suited to achieving a balance between the providers of capital and customers. This would ensure that the return earned by the providers of capital to efficiently operated water companies was sufficient, but no more than sufficient, to induce them to hold shares and to make loans.

• It is usual for the accounts in a normal competitive environment to focus on the returns to shareholders.

• The use of RPI as a measure of the change in the purchasing power of the unit of account, the relevant index in measuring real financial performance, was already built into the price control formula as a measure of general inflation. Because of this, the value was readily available and the estimates were stable.

The earliest regulatory report that appeared to apply the contents of these Guidelines in the UK is the first volume of Ofwat’s Cost of Capital Consultation Paper (1991). The report reiterated the duty of the regulator to ensure that regulated businesses can finance their functions. It also stated that maintaining the financial capability was something broadly mirrored by the legislation for other privatised utilities already. Their question at that point was how to incorporate the rather novel reference to “reasonable returns on capital” with the initial setting of the value of K, a parameter relevant to setting the price level over time.

According to Ian Byatt, the Director of Ofwat, the accounts prepared on a real FCM basis better reflect the impact of the financial performance of companies than accounts based on the calculation of profit using maintenance of operating capability accounting. However, Ofwat’s Cost of Capital Consultation Paper (1991) also considered allowable rates of return in the initial price setting period following privatisation had been too generous and proposed a significant downward revision to the allowable return on capital. Although this is not necessarily inconsistent with achieving FCM, an important assumption used in the Ofwat 1994 periodic review is inconsistent with FCM as a principle. The Ofwat (1994a) paper that sets out the framework and approach to the 1994 periodic review contains an assumption that is inconsistent with ex–ante FCM when it says (p.5):

“In the absence of persuasive arguments to the contrary, the Director will assume that companies can in future deliver at least current levels of service at prices which are, in real terms, no more than those charged at present. That is what would be expected from companies in competitive markets.”

The problem with this principle, as stated above, is that it is not a universal theoretical principle but rather a principle based on average outcomes for the economy and it does not guarantee ex–ante FCM for a reasonable rate of return since it sets the standard for regulated...
prices as no increase in real prices. On average there is not a real increase in prices in an economy since the general price deflator is an average price deflator and by definition there will be no real increase when the same price deflator is used to convert a general price change to a real price change. However, if ex–ante FCM is to be maintained it is not logically possible to set a cap on prices so that there is no real increase in regulated prices. Forward looking real prices may or may not increase to achieve ex–ante FCM – it will depend on what is required to finance forward looking capital and operating expenditures which, among other things, depends on the movements in input prices the utility faces.

Prior to the 1997 Price Control Review British Gas used the OCM version of current cost accounting in which assets are revalued to current replacement cost and where any holding gains or losses are taken to reserves and not through the profit and loss account (Ofgas 1996). OfGas subsequently recommended that assets be revalued for regulatory purposes in line with changes in the Retail Price Index, rather than in line with changes in the cost of asset replacement for two reasons. This was essentially a shift from OCM to FCM. The reasons for this move were:

a) There is a need for consistency between the basis for estimating TransCo’s cost of capital and the basis for estimating the regulatory value of its assets.

b) Current cost revaluations involve an element of subjectivity which complicates the regulator’s task. Linking revaluation to a general inflation index removes the problem of companies having an incentive to exaggerate or to understate changes in the cost of asset replacement.

The ACCC (2004, p.24) noted that British Telecom purportedly uses the FCM convention in accordance with the principles set out in the handbook “Accounting for the effects of changing prices” published in 1986 by the Accounting Standards Committee. Under this convention current cost profit is normally arrived at by adjusting the historical cost profit to take account of changes in asset values and the erosion in the purchasing power of shareholders’ equity during the year due to general inflation. However, the Commission also noted that: the approach to FCM, as implemented in the UK context, can produce hybrid accounting systems, in which enterprises could combine a looser capital maintenance concept with one of a number of asset measurement bases (irrespective of the degree of conceptual and practical compatibility); and the ability to adopt such combinations allows greater flexibility in the process by which the profit of the enterprise is determined. Note, however, that an issue is whether greater flexibility in the measurement of profit also means that FCM is effectively compromised.

**B2.4 Capital maintenance in the European Union**

A detailed discussion of FCM and OCM can be found in a report on the implementation of cost accounting methodologies and accounting separation by telecommunication operators with significant market power prepared by Andersen Business Consulting (2002) on behalf of the European Commission DG Information Society. This study provided the backbone to the Commission’s explicit rejection of OCM for application to telecommunications regulation. The objective of this study was to assess the different practices and initiatives implemented in
member states to ensure compliance with the Directives and Recommendations on cost accounting and accounting separation issued by the European Commission. The study also assessed the effectiveness of the Commission’s recommendations on accounting separation and cost accounting.

The Andersen Business Consulting report (2002, p.15) argued that FCM is the superior capital maintenance concept as follows:

“The use of the OCM concept may systematically incorporate insufficient or excess returns into the level of allowed revenue (depending, respectively, on whether asset–specific inflation was expected to be lower than or higher than general inflation). This is not a desirable feature of any regulatory regime, as it would not provide appropriate investment incentives. Under FCM, however, the returns to the providers of capital would equal the required return (as measured by the cost of capital) irrespective of whether replacement costs were rising or falling relative to general prices. Hence, if current cost accounting information is used as the basis to determine interconnection charges, FCM is the preferred capital maintenance concept.”

A recent NERA (2008) report provides a useful review of the European Regulator’s Group for Electricity and Gas (ERGEG) consultation paper on “Principles for Calculating Tariffs for Access to Gas Transmission Networks”. NERA (2008, p. 12) note that with respect to asset valuation and accounting standards for regulation the most serious omission in the ERGEG consultation paper is the lack of any general regulatory principles to guide the choice of valuation method or the associated rate of return.

The NERA report (2008, p.15) also notes the following examples of regulatory systems that do not meet the standard of FCM in Europe:

- Germany’s method of regulating energy sector assets still applies OCM standards, in which asset values are inflated by a different (asset–specific) price index without any offsetting compensation for rising/declining real values.
- A recent “draft method decision” from the Dutch energy regulator proposed a combination of real WACC and non–revalued RAB for gas distribution networks. That combination is also a mistake, since it deprives investors of any compensation for inflation, and so exposes them to a steady decline in the real value of their assets.
- The situation in Finland is hard for me to determine with precision (a description of the latest decisions is available only in Finnish), but I understand that some regulatory decisions apply an estimate of the nominal rate of return to a revalued asset base. The combination would offer compensation for inflation twice over – were it not for the fact that the estimated nominal rate of return seems to be extremely low.

**B2.5 Capital maintenance in Australia**

Documents in Australia discussing FCM and OCM largely draw from the Andersen Business Consulting report to the European Commission, the Network Economics Consulting Group...
History of Building Blocks Regulation


NECG, before the Telecommunications Act came into force, suggested that the Act require FCM for investments in regulated assets that were prudent at the time. It was one of the supplements to the legislated access pricing principles recommended by NECG, alongside compensation for regulatory risk and recognition of the impact of social obligations. NECG (2001b p. 3) noted:

“Financial Capital Maintenance (“FCM”) ensures that funds prudently invested in regulated assets will be recouped. No regulatory arrangement can be sustainable if investors in regulated assets cannot reasonably expect the regulatory contract to ensure FCM. NECG recommends that FCM be used as a guiding access pricing principle.”

Ergas (2003) discussed ex-ante FCM in his commentary on the Western Australia Supreme Court’s decision on Epic explaining it in the context of determining efficient costs or efficient investments. The case focused on the decision by the regulator, Offgar, to value the Dampier to Bunbury Natural Gas Pipeline at about half the cost paid for it in a competitive tender for a publicly owned asset. The Epic case involved judicial review of the regulator’s decision. In granting relief to Epic, the Court required the regulator to reconsider its decision and that in effect the price paid by Epic for the pipeline was a matter which the regulator had to consider in the context of his decision about access charges for the use of the pipeline.

The Court did not explicitly recommend the FCM approach but noted that recovery of the cost of the investment, even if it reflected an expectation of monopoly prices, was not contrary to a legitimate business in accordance with requirements of the National Third Party Access Code for Natural Gas Pipeline Systems. The Court further noted that taking account of the actual investment cost was consistent with another Code objective of not distorting investment decisions but also noted that accepting any cost could also lead to a distortion of investment decisions.

The Court noted that the Code also required that access prices should seek to ensure that revenue is sufficient to recover efficient costs and there was support for the view that only forward looking costs should be considered. However, the Court did not attempt to resolve any inconsistency in these objectives under the Code and in effect left it to the regulator. Ergas (2003, p.11) argued that rather than widening the factors for consideration to non-economic matters, the Court could have taken a more economic perspective and in particular adopted the concept of ex-ante financial maintenance so long as prudent or efficient expenditures were made.

Offgar issued a revised decision that increased the regulatory asset base from $A1.234 billion to $A1.55 billion compared to the price of $A2.407 billion paid by Epic. Ergas (2003, p.13) noted that Offgar concluded that Epic did not undertake a prudent or objective assessment of a future regulator’s position on the rate of return and that there was a need to take account of the interest of users and the public interest. Ergas argued the asset base was estimated to meet what was considered an expected tariff level which was also sufficient for Epic to fund its debt commitments. Ergas (p.15) also noted that although the regulator’s approach provided very detailed calculations, it was weak in terms of setting out economic concepts to guide access charging.
It is clear from this decision that an ex-ante FCM concept for prudent expenditure was far from the guiding principle used by the regulator. Instead, in the end the criterion used was an asset value that avoided bankruptcy for Epic and, as noted by Ergas (p.15), highlighted weaknesses in the Gas Code.

A similar issue arose following the privatisation of the electricity distribution sector in Victoria where American investors paid $A 8.3 billion for assets for which the regulatory asset base was subsequently set at $A 3.8 billion (Fearon 2006, p.15).

Turning to telecommunications, in January 2004 the ACCC (2004) released a framework document outlining the current cost asset valuation and capital maintenance methodologies to apply in the longer term to Telstra in relation to accounting separation of its retail and wholesale operations. The document specified the use of modern equivalent asset (MEA) valuation (based on replacement cost) and the use of FCM as the basis of this reporting. It argued that the valuation of assets was a separate issue to the measurement of profit and capital maintenance within the current cost accounting framework. Furthermore, it noted that its approach to the valuation of assets was also adopted in the UK by Oftel and recommended by the EU (ACCC 2004a, p.15). However, it is important to recognise that if both the FCM and MEA concepts are concurrently adopted then it is inconsistent to use the FCM concept in recovering the asset value represented by MEA. In some cases, however, the historic cost of past investments may not be available, particularly for long-lived assets, and the earliest available depreciated replacement cost estimate may be used as the best available substitute for historic cost.

FCM is widely used and applied by Australian regulators in the measurement of profit, particularly in the regulation of the telecommunications industry. It is also used by all energy regulators in setting building block allowable revenues. In the latest Current Cost Accounting Report relating to the Accounting Separation of Telstra, ACCC (2008, s.2.1) explains:

"Approach to financial capital maintenance

In determining the level of profit in the current-cost profit and loss statements, the concept of financial capital maintenance (FCM) has been employed.

FCM is concerned with maintaining the real financial capital of the company so that it can continue financing its functions. Capital is maintained if shareholders’ funds at the end of the period are maintained in real terms at the same level as at the beginning of the period. Under FCM, profit is therefore only measured after provision has been made to maintain the purchasing power of opening-period financial capital.

The FCM basis of capital maintenance requires adjustments to be made to the current-cost profit and loss statement to reflect holding gains or losses arising from changes in the value of the assets over the relevant period, depreciation differences between historical cost and current cost accounting, and the effects of inflation on the resources invested in the enterprise."
B3 THE DEVELOPMENT OF THE BUILDING BLOCKS METHODOLOGY

There has previously been little attention directed to the history, development and rationale for the building blocks methodology as applied to economic regulation in Australia, New Zealand and the United Kingdom. Related literature usually comprises descriptive accounts of this type of methodology or what its strengths and weaknesses are relative to other approaches. They are either especially prepared for or prepared by the regulators (see Farrier Swier Consulting 2002 and Productivity Commission 2001). A few articles made mention of building blocks analysis but descriptively only (e.g. a method used in Australia and the UK, a method of building up costs faced by regulated businesses) and did not go beyond that level of discussion (Carrington, Coelli and Groom 2002, Cowan 2006, Fearon 2006).

In the documents reviewed that relate to the UK experience, the process of cost build-up in the determination of a price cap or maximum allowable revenue was rarely referred to as a ‘building blocks’ approach. However, the most recent Electricity Distribution Price Control Review does use the term “building block framework”, regarding the individual building blocks to be used by distribution network operators when presenting their forecasts for cost components for the price control review for the period 2010–2015 (Ofgem 2008, pp. 63–64 and Appendix 8). In addition Ofgem (2009, p.23) in a recent review of the History of Energy Network Regulation in the UK does describe the approach to setting price controls in the energy sector as based on a building block approach as the term is understood in Australia and New Zealand.

The key elements of the “building blocks” methodology, including the incorporation of ex-ante FCM and forward looking estimates of costs were adopted in the UK with the work of both Byatt (Byatt et al 1986 and Ofwat 1991) and Littlechild being highly influential (Littlechild 1986, Littlechild in Bartle 2003, Littlechild 2007 and 2008).

B3.1 The first steps in the UK – electricity and water

CPI–X regulation (referred to as RPI–X in the UK) was adopted at the time various utilities in the UK were privatised. The UK government typically set the initial values of X at the time of flotation of each utility company for a period of 5 years but did not explain how the original decisions on the X values were reached (Littlechild 2007, p.5). However, in the water industry X factors known as K factors (as they determined the rate of increase of real prices) were derived in order to ensure the net present value of each firm’s future cash flows was equal to the value implied from assuming the cash flows that firms might have been expected to earn on existing assets if the previous regulatory regime had continued (Armstrong, et al 1995, p.345). This is equivalent to ex-ante FCM for both existing and new investment at the time of privatisation.

Price paths were initially set for electricity transmission, supply and distribution from 31 March 1990 for three, four and five years, respectively (Armstrong, et al 1995, p.299). Price paths were initially set for water and sewerage companies in 1989 for a ten year period which was subsequently changed to five years in 1991 (Ofwat 1994, p.5).

However, the development of a more rational, systematic and transparent approach occurred
when the water and electricity regulators came to reset the initial price controls. The resetting of prices occurred for both electricity distribution and water and sewerage companies in 1994.

According to Littlechild, at the time there was a need for a method that could explicitly show that no individual company had been favoured or disfavoured relative to its peers. Until then the emphasis had tended to be on regulators proposing the toughest value of X they could and investors had been applying substantial pressures on regulators to explain their calculations (Littlechild 2007, p.6 and Littlechild in Bartle 2003, p.47).

In developing the approach a key consideration of Littlechild was to ensure that “the level of X offered a rate of return at least comparable to what investors could get elsewhere (for a comparable risk and requirement to be efficient and innovate, etc)” (Littlechild in Bartle 2003, p.46). Littlechild considered numerous approaches but focused on two. One was developed by Michael Beesley and one by his regulation and business affairs director, Geoff Horton (Littlechild in Bartle 2003, p.46). Both Beesley’s and Horton’s methods involved assessing the efficient levels of operating costs, capital expenditure and cost of capital for the forthcoming period and beyond, and projecting cash flows (Littlechild 2007, p.7). This feature together with the concern to ensure an adequate rate of return on capital are key aspects of the ‘building blocks’ approach.

Beesley’s method was based on future cash flows, share prices and other financial indicators. The method entailed projecting the dividend streams and borrowing necessary to finance the sustained operation of the company, but implied an explicit role for the regulator in determining or influencing share prices that Littlechild found difficult to accept. Geoff Horton, on the other hand sought to reconcile the forward-looking thinking embodied in Beesley’s approach with the more conventional approach of incorporating a return on existing capital (Littlechild in Bartle 2003, p.47). Littlechild decided on the Horton approach as he considered it would be easier to explain, implement and defend (Littlechild 2007, p. 7). Over time this approach evolved into what became known as the ‘building blocks’ approach or methodology which has underpinned the explicit calculations that accompany the resetting of X in the utilities sector.

The approach was used in setting the Regional Electricity Companies’ (RECs) distribution price controls in the August 1994 proposals document and in all subsequent work. The model underlying it was circulated to RECs in 1994. It was more explicitly expounded in numerical detail in the Monopolies and Mergers Commission Report on Scottish Hydro in 1995. Details of the history of the approach and its current application in the energy sector in the UK are set out in Ofgem (2009).

Under the current regime Electricity Distribution Network Operators (DNOs) have to submit their cost forecasts in the defined building blocks, each with clearly identifiable assumptions, costs and outputs. They also have to provide information on other options considered, sensitivities to changes in assumptions and required outputs including the impact of any stakeholder engagement.

The objectives of this detailed methodology are:

1. To allow the regulator to assess DNOs’ forecasts taking into account the baseline
expenditure from the modelling.

2. To provide forecasts that the regulator can compare across DNOs without compromising flexibility on the part of DNOs in the submission of their cost forecasts.

The building block approach was also developed by Ian Byatt in his role as the water regulator at around the same time. However, he considered the K factors that were decided for the initial regulatory period were too generous (Ofwat 1991a, Ofwat 1991b and Armstrong, et al 1995, p.346) and proposed and decided on a lower rate of return for new investment. It is important to recognise that even if ex-ante FCM is adopted there will still be the issue of what is the appropriate cost of capital as well as what is appropriate for operating costs. However, as noted earlier it is notable that the Ofwat (1994a, p.5) paper that sets out the framework and approach to the 1994 periodic review contains an assumption that is inconsistent with ex-ante FCM.

The building blocks approach was used in the 1994, 1999 and 2002 price reviews for water and sewerage companies (Ofwat 1994b, 1999 and 2002). Based on the price review documents covered, there seems to be no significant changes on the rationales behind the adoption of the Building Blocks approach in the water sector. Table B1 provides a summary of these rationales.

Table B1: Ofwat rationale for building blocks analysis

<table>
<thead>
<tr>
<th>Year</th>
<th>Rationale</th>
</tr>
</thead>
<tbody>
<tr>
<td>1994</td>
<td></td>
</tr>
<tr>
<td>1. Consistency with the law: the Water Industry Act requires the regulator to exercise his powers in the manner that he considers is best calculated in order to ensure that companies are able to finance proper carrying out of their functions</td>
<td></td>
</tr>
<tr>
<td>2. Incentivise companies to reward shareholders from greater efficiency as well as to deliver better services to their customers</td>
<td></td>
</tr>
<tr>
<td>3. Ensure that profit is (just) sufficient to attract and retain capital in the business</td>
<td></td>
</tr>
<tr>
<td>4. Nature of the water industry (universalisity of the essential service, difficulty of differentiating quality, indefinite life of assets and the appropriate capital maintenance, low growth in demand and limited opportunities to enhance market share)</td>
<td></td>
</tr>
<tr>
<td>1999</td>
<td></td>
</tr>
<tr>
<td>1. Efficiency and incentives</td>
<td></td>
</tr>
<tr>
<td>2. Maintaining service to customers</td>
<td></td>
</tr>
<tr>
<td>3. Quality program</td>
<td></td>
</tr>
<tr>
<td>4. Maintaining the balance between supply and demand</td>
<td></td>
</tr>
<tr>
<td>5. Financial Issues</td>
<td></td>
</tr>
<tr>
<td>2004</td>
<td></td>
</tr>
<tr>
<td>1. Allow businesses to meet all existing obligations, and make sufficient revenue to finance their operating expenditure and capital investment programs</td>
<td></td>
</tr>
<tr>
<td>2. Maintain assets for current and future customers</td>
<td></td>
</tr>
<tr>
<td>3. Ensure a sufficient balance between supply and demand for the water and sewerage services</td>
<td></td>
</tr>
<tr>
<td>4. Future efficiency gains</td>
<td></td>
</tr>
<tr>
<td>5. Incentive allowance for outperformance</td>
<td></td>
</tr>
<tr>
<td>6. Efficient and transparent approach to the review and a framework that</td>
<td></td>
</tr>
</tbody>
</table>
B3.2 Application by other regulators in the UK

Gas and telecommunications industries in the UK were not faced with the same pressure to present a formal model from regulated companies as Offer was. While there was only one telecommunications or gas company, there were more than a dozen electricity companies and over 40 water companies on a comparable basis. Detailed calculations underlying the resetting of the price formulae for British Telecommunications and British Gas have not been published. On a general note, the regulators of these two industries favoured a financial approach to regulation paying more attention to profit forecasts and other related information. It is understood that typically, in the early stages of regulation following privatisation, the FCM principle was effectively adopted in terms of a goal of ensuring prices were sufficient to cover the cost of capital.

In the regulated gas sector the initial price control period following privatisation operated from 1987 to 1992. In that period there was a three part formula for the setting of maximum allowable average revenue known as an RPI–X+Y cap where the Y denoted a cost pass through term (Armstrong, et al 1995, p.256). The formula allowed for the full pass through of the cost of gas, specified that the non–gas cost component was to grow by no more than RPI–X and also included a correction factor to allow for any under– or over–charging in one year to be corrected in the following year to recognise that outturn can never be precisely the same as forecast. The justification for full pass through of gas costs was that gas was supplied under long term contracts that could not be changed. The X factor was focused on providing discipline on non–gas costs including the cost of operating and maintaining transmission, distribution and storage facilities and the costs of marketing and metering gas sales.

Some deficiencies that have been noted about this approach are as follows. The formula allows the average cost of all gas purchases including those for the non–tariff (non–regulated) market to be passed through to tariff customers which can be higher if marginal costs exceed average costs and marginal expansion occurs in the non–tariff market. This could lead to underpricing of other cost components in selling to the non–tariff market in order to raise average costs that can be recovered in higher tariffs in the tariff market. Complete pass through of gas costs also removed the incentives to purchase efficiently. Finally, as the cap was a constraint on average revenue there was an incentive to set prices that were not close to Ramsey prices. See Armstrong et al (1995, pp.256–7).

A comprehensive accounting exercise, called the Cost Apportionment Program, provided OfGas with more detail about how BG apportions both gas and non–gas costs between its major market segments. It commenced in 1987 and was carried out on a confidential basis in an effort to find a more satisfactory basis for the price formula.

In the 1990–91 review a new formula was defined as RPI–X+GPI–Z+E (Armstrong, et al 1995, p.259). The X factor was increased from 2 to 5. The GPI–Z term refers to a gas price index less an efficiency factor which was defined to be 1 per cent. The E factor was an energy efficiency factor that allowed for reasonable expenditures that act to reduce demand to...
be passed on to consumers.

It was not until the 1997 Price Control Review that Ofgas explicitly used building blocks analysis to calculate allowed revenues of Transco (Ofgas 1996a). The approach was based on the methodology first formalised by MMC in the 1993 Gas and British Gas Reports under the Fair Trading Act (Monopolies and Mergers Commission 1993a, b, c and d). However, there was a difference in how pre–privatisation assets were valued compared to the approach used by the water and electricity regulators. This reflected differences in how to best incorporate into the valuation the reasonable expectations of shareholders at the time of privatisation. However, Ofgas (1997c, p.81) summarised the common thinking of the water, electricity and gas regulators as follows:

“Rather the intention in setting price controls for those industries where companies were initially privatised at substantial discount to net book values has been to reconcile the need for new investment to be remunerated at the companies’ respective cost of capital with the desire not to give shareholders windfall gains through allowing equivalent returns on the current cost net book values of pre–privatisation assets. The debate has, been about how to value pre–privatisation assets and how to incorporate into that valuation the reasonable expectations of shareholders at the time of privatization.”

Oftel did not use the building blocks approach in either the first review of prices in 1988 or the second review in 1992, although it did use a RPI ± X approach (Oftel 1988 and 1992b and c). In 1992 X was reported to have been set at a level which provided BT with an expectation of covering the cost and risk of capital, while providing demanding targets for improvements in customer service and increased efficiency (Armstrong, et al 1995, p.227).

B3.3 Developments in the UK

Refinements to the building blocks approach used by gas and electricity regulators in the UK are as follows (Armstrong, et al 1995, Littlechild 2007, Ofgem 2006 and 2009):

- The form of the control and whether to set a price cap or a revenue cap. The initial controls typically set a price per unit cap which was subsequently replaced for electricity transmission by a total revenue cap because of concerns about risk for the company. However, this increased the risk to consumers as reflected in fluctuating prices if output fluctuated unexpectedly. A subsequent price control on distribution companies embodied a 50–50 weighting on actual and expected output thereby sharing output risk between the company and its customers.

- Adjusting the regulatory asset base to reflect actual rather than assumed capital expenditure, seeking greater agreement on future capital expenditure plans, providing incentives to forecast accurately and making capital expenditure more conditional on the growth of demand. In particular, a sliding scale mechanism was introduced to provide for a more flexible approach to capital expenditure, without disadvantaging those companies that have provided more reasonable forecasts.

- Incentive mechanisms to ensure specified levels of quality and to reduce loss factors in distribution companies. An information quality incentive mechanism was introduced to provide efficiency rewards to licensees who manage to deliver savings against the most
History of Building Blocks Regulation

• Smoothing of allowable operating and investment expenditure to avoid gaming of the system.
• Adjustments of cost pass through terms to encourage cost efficiency.
• Use of correction mechanisms to adjust the price control for any previous over- or under-recovery against allowed revenues. The mechanisms could apply to revenue or cost parameters.
• Re-opener mechanisms to enable a price review for specific events or circumstances.
• Use of benchmarking to help determine allowable costs in implementing the building blocks approach.

B3.4 Application by Australian regulators

The building blocks approach has been the “dominant method” of determining the \( P_0 \) and \( X \) factors in Australia. The Productivity Commission (2001, p.341) claims that regulators have adopted the approach because it is seen to be objective, transparent, and results in prices which closely track individual facilities’ costs. The ACCC (2004, p.21) described the approach in a Statement of Regulatory Principles for the Regulation of Electricity Transmission Revenues as follows:

“The building block approach is used to ensure that the expenditure of each TNSP is appropriately amortised over time to ensure that each TNSP, given efficient expenditure practices and decisions, is adequately compensated for the cost of providing the transmission services to customers in the long run.

The building block model consists of two equations which are known as the revenue equation and the asset base roll forward equation. These two equations are used to determine an allowed stream of revenues for each TNSP for as long as it remains regulated. Ignoring any incentive rewards or penalties, these equations together ensure that the present value of the allowed revenue stream is equal to the present value of the expenditure stream of the regulated firm.”

Reports from Australian regulators, particularly from the Australian Competition and Consumer Commission (ACCC), Australian Energy Regulator (AER), the Independent Pricing and Regulatory Tribunal of NSW (IPART), Queensland Competition Authority (QCA) and Essential Services Commission of South Australia (ESCOSA) confirm that building block analysis continues to be widely used in Australia. Victoria has been at the forefront of advocating a shift from the building blocks approach to a TFP based approach to CPI–X price setting where the X factor relies more on external (to the regulated company) industry TFP benchmark trends and less on information specific to regulated firms (Fearon 2008). The Australian Energy Market Commission (AEMC) is currently conducting a review into whether TFP based regulation should be allowed as an alternative to the building blocks approach.
In general the building blocks approach adopted in each Australian jurisdiction consists of three components: efficient operating and maintenance costs, an allowance for the return on capital and an allowance for the return of capital (depreciation). However, there are some differences in cost components and the methods for their determination. These include: efficiency carry over mechanisms, unders and overs adjustments (to take account of any under or over recovery of allowable revenue), trigger mechanisms, asset valuation, the cost of capital and the time profile of X factor adjustments.

In New South Wales the approach is called the Cost Building Block Revenue Model and is designed to be used in conjunction with the Weighted Average Price Cap model. It provides the option of using the rolled forward Regulated Asset Value or entering a new Regulated Asset Base using straight line depreciation. A return on working capital is also allowed and an unders and overs adjustment has been allowed in the past.

In Victoria, an important feature incorporated in the 2001–05 price review for electricity distribution network service businesses was an efficiency carry over mechanism to reflect efficiency savings relative to forecasts, effectively allowing the full value of an efficiency gain to be retained by the regulated company for a 5 year period. However, following concerns about its operation, for the 2006–10 regulatory period, the mechanism only applies in relation to operating and maintenance expenditure.

In Queensland where a total revenue cap applies for electricity distribution network service businesses, there is a demand trigger mechanism based on maximum demand and customer numbers. An unders and overs account also exits.

It is worth noting that an issue of difference for Australian regulators and across sectors is the form of price control and in particular whether to use a weighted average change in prices (sometimes referred to as a pure price cap), a total revenue cap or a revenue yield (average revenue) price cap. A pure revenue cap reduces risk to the firm relative to a pure price cap and discourages innovations to achieve growth while an average revenue cap can lead to profits higher than allowed for in setting regulatory parameters. This contrasts with a pure price cap where revenue is allowed to move in line with the specific tariff applying to marginal consumption (Office of the Regulator General, Victoria 1998, pp. 46–47).

For electricity distribution New South Wales and Victoria adopted a weighted average price cap while Queensland adopted a total revenue cap and South Australia adopted an average revenue cap. In transmission New South Wales and Victoria adopted a total revenue cap. All these regulatory functions have either now passed or are in the process of passing to the AER.

**B3.5 Developments in Australia**

It is important to recognise that the application of the building blocks methodology in Australia to date has emphasised firm-specific costs with a regulatory lag of generally five years. The derivation of the X factor depends on a judgement of the extent to which reasonable efficiencies can be achieved based on the specific circumstances of the firm. This contrasts with an approach where the X factor is based more on industry trends in total factor productivity.

A report prepared by Farrier Swier Consulting (2002) for the Utility Regulators Forum (URF)
undertook a comparison of the building blocks and indexed approaches”. The report concluded that a TFP based approach to price regulation was likely to create superior economic efficiency incentives (p.84) provided the approach tended to operate (p.72) “mechanistically” without triggering excessive reviews; earnings sharing mechanisms are either not incorporated or have wide bands; and the approach was implemented within an appropriate and robust decision-making framework.

As noted above, the TFP based approach is currently the subject of a review by the Australian Energy Market Commission.

Other problems identities with the building blocks approach in Australia are as follows (most of these are summarised by Fearon 2006 and 2008):

• tensions in a privatised industry with monopoly characteristics between the firms seeking to maximise returns and the expectations and objectives of customers
• the clear information asymmetry exacerbated by reliance on the information provided by the utility with incentives to “talk up” costs and “talk down” future sales
• underestimation, in hindsight, of the challenges in relying on reported costs
• restructuring of EDBs including arrangement with entities with common ownership, but which are not directly covered by the regulatory regime, and the possibility that such arrangements may not be at arm’s length, with the potential to inflate or obscure reported costs
• the challenges generally of obtaining transparent cost data and unravelling complex and changing cost allocations making comparisons and forecasts difficult
• the considerable difficulty obtaining information per se, with delays in some cases and others where information was withheld entirely
• asset valuation based on depreciated optimised replacement costs provides considerable scope for judgement and a wide divergence of views
• divergence in regulatory decisions about cost of capital parameters.

B4 CONCLUSIONS

The key findings with respect to the origins and evolution of FCM and the building blocks approach in the regulation of prices for utilities are as follows:

• The FCM concept has its origins in accounting literature but its relevance in providing incentives for efficient investment, as an ex–ante application, is well recognised by the regulatory authorities in Australia, New Zealand, the United States and the United Kingdom.

• The origin of the FCM concept in economic regulation lies in the adoption of rate–of–return regulation to the extent that such an approach is concerned to ensure investors receive a “fair rate of return”. However, its explicit use as an ex–ante concept in economic regulation followed the initial privatisation of utilities in the UK in the late 1980s.
• The ex–ante FCM concept has been applied by most utility regulators in the UK and continues to be widely used.

• Byatt, prior to becoming the first Director of Ofwat, was the joint author of an influential report in 1986 that advocated FCM for measuring returns and as the ex–ante standard for an investor.

• Ofwat’s Regulatory Accounting Guidelines first issued in 1992 appeared to be the first document to discuss the concepts of FCM and OCM in relation to the regulation of a utility in the UK. Other regulators from the UK did not have the power to formulate similar accounting guidelines at the time.

• An important divergence from the application of the FCM concept was by Ofwat in its 1994 price review where it specified a principle it was adopting (based on a competitive markets analogy) was that prices for current service levels could not increase in real terms. However, this position seems to be related to a view that rates of return were too generous which is not an unreasonable position and can be addressed while still adopting FCM. However, the assumption as specified conflicts with FCM for future investment.

• Ofgas used OCM until the 1997 Price Control Review.

• FCM has tended to be preferred over OCM in the UK, Australia and member countries of the European Union.

• However, there are exceptions: Germany applies OCM rather than FCM in regulating the energy sector while the Netherlands and Finland also do not adopt the FCM criterion.

• In addition, a recent consultation paper by the European Regulator’s Group for Electricity and Gas is notable for the lack of any general regulatory principles to guide the choice of valuation method or the associated rate of return.

• The ex–ante FCM concept is the dominant approach to calculating regulatory allowable revenue in the electricity sector in Australia although in many cases the historic cost of past investments has not been available and the earliest available depreciated replacement cost estimate has been used as the best available substitute for historic cost.

• However, in telecommunications the ACCC has drawn a distinction between the use of FCM for calculating profits (which it considers is appropriate) and its use in the valuation of assets where it prefers a modern equivalent asset (replacement) approach.

• Stephen Littlechild, Michael Beesley and Geoff Horton developed the building blocks approach in the UK in the early 1990s for Offer’s first reset of the value of X in the electricity sector.

• Ian Byatt of Ofwat used a model similar to Offer’s building blocks approach almost at the same time as its application in the electricity sector.

• There was relatively little material provided on the rationale for the building blocks approach in the UK, however its origins relate to the adoption of CPI–X price regulation and the need to adopt a forward looking perspective on efficient costs in order to determine an appropriate X factor.

• The rationales and assumptions underlying the use of the approach have not changed
much through time.

• However, various refinements have been made to encourage efficiencies and avoid gaming problems.

• Over time the approach has become more transparent.

• Challenges in obtaining appropriate estimates of efficient costs in a cost effective way are an increasing issue.

• Regulators from the UK and Australia are now considering other approaches to determining the value of X in price or revenue cap regulation. They are trying to find a more light-handed approach that could better incorporate incentives for efficient investment.
REFERENCES


Australian Competition and Consumer Commission (ACCC) (2004b), Statement of Principles for the Regulation of Electricity Transmission Revenues, Background Paper, Decision, 8 December.


Australian Energy Regulator (AER) (2005), Compendium of Electricity Transmission Regulatory Guidelines, August.


Group, Energy and Resources Division, Ministry of Commerce, for briefing the Electricity Supply Association of New Zealand (ESANZ), August, Wellington.


Essential Services Commission of South Australia (ESCOSA) (2004), Statement of Principles for the Regulation of Electricity Transmission Revenues, Adelaide.

Farrier Swier Consulting (2002), Comparison of Building Blocks and Indexed Base Approaches, Report prepared for the Utility Regulator’s Forum


References


Lerner, A.P. (1944), The Economics of Control, New York: Macmillan.

Littlechild, S.C. (1983), Regulation of British Telecommunications’ Profitability, Report to the Secretary of State, Department of Industry, London.


Littlechild, S. (2008), Correspondence with D. Newberry.


Office of the Regulator General, Victoria (ORG) (1998), Weighted Average Cost of Capital for Revenue Determination: Gas Distribution, Staff Paper, Number 1, May.


Ofgas (1996c), *Price Control review Supply at or below 2,500 therms a year– British Gas Trading The Director General’s initial proposals*. London, Office of Gas and Electricity Markets.


Queensland Competition Authority (QCA) (2003), Form of Regulation of Electricity Distribution to Commence from 1 July 2005, Brisbane.


