DRAFT: 8 November 1998

TSLRIC, TELRIC and Other Forms of Forward-Looking Cost Models in Telecommunications: A Curmudgeon's Guide

Henry Ergas(*)

Centre for Research in Network Economics and Communications
The University of Auckland

A paper prepared for the 1998 EU Competition Workshop at the Robert Schuman Centre of the European University Institute

(*) This paper reflects work and discussions with Alexis Hardin, John Small and Bridger Mitchell, and I am grateful to them, and to my other colleagues for their assistance. It also reflects work and discussions with Paul Paterson at Telstra Corporation, and I am grateful to him and his colleagues for the input they have provided. However, the views expressed are strictly my own, as is responsibility for any errors.

TABLE OF CONTENTS

A Curmudgeon's Guide to TSLRIC, TELRIC and Other Forms of Forward-Looking Cost Models

Regulators in North America, the EU and Australia have mandated the use of forward-looking cost models as the basis for interconnection charging. This paper sets out some of the major issues involved in the construction and use of these models.

The background

Accounting systems structure the collection, analysis and disclosure of cost and revenue information. By so doing, they serve to reduce the transactions costs involved in designing and implementing the explicit and implicit contracts that regulate relations between suppliers and users of resources[1]. Given that this is their purpose, it is not surprising that somewhat differing accounting systems are needed to support differing types of transactions. For example, in most countries, tax accounting differs in important respects from the financial accounting systems used to provide information to the suppliers of firms' financial resources. Equally, it is common for governments to impose special accounting requirements as part of the public procurement process.

In the control of public utilities, *regulatory accounting* serves to support the design and implementation of the regulatory contract – that is, the complex of more or less formalised understandings between regulatory authorities and regulated entities as to the terms and conditions on which services will be provided. Twenty years ago, telecommunications regulation, in most OECD countries, was not separated from service provision; it therefore made little sense to talk of a 'regulatory contract'. Since then, this separation has been effected in virtually all OECD countries. As a result of this separation, the design of regulatory processes, and of the accounting systems needed to support them, has become an issue of obvious importance.

The United States has, of course, long had a clear formal separation between regulatory entities and telecommunications service providers. Issues of regulatory accounting therefore emerged early on in the US, with complex systems being designed to supply the regulatory authorities with cost and revenue information[2]. These systems formed a natural point of reference when similar needs arose elsewhere.

1 See especially Shyam Sunder The Theory of Accounting and Control South Western Publishing: Cincinnati, Ohio 1997.

2 On the early evolution of regulatory accounting for the telephone system see N.R. Danielian AT&T The Vanguard Press: New York 1939, notably 334 and following; and C.L. Weinhaus and A.G. Oettinger Behind the Telephone Debates Ablex: Norwood N.J. 1988.

The systems of regulatory accounting developed in the United States had two salient features[3]. First, they were based on historical costs, which were generally viewed as reflecting (1) the resources investors had devoted to the utilities at issue and hence (2) the claim these investors had on the revenues the systems generated. Second, they relied on exhaustive systems of cost allocation, often referred to as "fully distributed costing", to allocate costs to individual products and services. Both of these features have been heavily criticised.

As regards the historical cost convention, economists[4], as well as economically-minded accountants[5], have long been of the view that the outlays incurred in past periods provide a poor guide to decisions involving resource allocation. In particular, it has been argued that historical costs are merely a collection of variously dated outlays, to which no clear meaning can be attached[6]; that they need not reflect the costs[7] that would be incurred in the current period by an efficient supplier of the service at issue; that they consequently bear no direct relation to the revenue a supplier of such a service would need to obtain so as to maintain intact either its service capability or its financial capital[8]; and that using these costs as the basis for pricing decisions can therefore distort consumption and investment decisions[9].

As for fully distributed costing, scholars of accounting have long argued that the exhaustive allocation of costs relies on judgements that do not reflect cost-causality and hence are essentially arbitrary[10]. Relying on such allocations to determine prices for individual services may not be consistent with preserving the financial viability of the service provider and in any event can seriously distort resource allocation[11].

---

3 A clear formulation of the philosophy underpinning conventional US regulatory accounting, as well as some of the major criticisms levelled against it, can be found in J.C. Bonbright, A.L. Danielsen and D.R. Kamerschen Principles of Public Utility Rates (2nd edition) Public Utilities Reports, Inc: Arlington, Virginia 1988.

4 See for example, Wicksteed's dictum that a man setting his selling price by reference to past costs of production "is either allowing an irrelevant consideration to affect his judgement or else is deliberately taking a commercial risk to gratify a personal feeling". P.H. Wicksteed The Common Sense of Political Economy and Selected Papers and Reviews on Economic Theory Routledge: London 1933 at 386.

5 See K. MacNeal Truth in Accounting University of Pennsylvania: Philadelphia 1939; R.J. Chambers Price Variation and Inflation Accounting McGraw-Hill: Sydney 1983 at 12-22; and G. Whittington Inflation Accounting: An Introduction to the Debate Cambridge University Press 1983 at 39 to 59.

6 R.J. Chambers and P.W. Wolnizer "A 'True and Fair' View of Financial Position" Companies and Securities Law Journal 1990 at 353-368.

7 In the sense of the resources a supplier would need to forego to supply the service at issue.

8 D. Solomons "Criteria for Choosing an Accounting Model" Accounting Horizons IX, 1 1995 at 42-51.

9 R.H. Coase "Business Organisation and the Accountant" in J.M. Buchanan and G.F. Thirlby L.S.E. Essays on Cost Weidenfeld and Nicolson: London 1973 at 97-132.

10 See especially Arthur Thomas' magnum opus: A.L. Thomas The Allocation Problem in Financial Accounting Theory American Accounting Association: Evanston, Illinois 1969, and its successor volume The Allocation Problem: Part Two American Accounting Association: Saratoga, Florida 1974.

11 See W.J. Baumol, M.F. Koehn and R.D. Willig "How Arbitrary is Arbitrary – or, Towards the Deserved Demise of Full Cost Allocation" Public Utilities Fortnightly September 3 1987:120 at 16-21; and R.S. Kaplan and R. Cooper Cost & Effect Harvard Business School Press: Cambridge, Mass. 1998.

Accounting systems based on current costing have been advocated as providing a superior alternative to the approaches criticised above. While the term "current cost accounting" covers a wide range of differing approaches[12], recent interest in telecommunications regulation has focussed on approaches which have two characteristics: they rely on measuring the costs that a hypothetical efficient supplier would incur in the longer term; and they define the relevant costs as those that would be incurred by such a supplier in the provision of a specified increment of output. They thereby combine the optimisation emphasis that characterises the Optimised Deprival Value ("ODV") approach to valuation of the asset base[13] with the marginal approach, and the resulting emphasis on the relevant output increment, characteristic of economic decision analysis[14].

TSLRIC ("total service long run incremental cost") and TELRIC ("total element long run incremental cost") are the main practical forms this approach takes in telecommunications. In essence, these concepts involve three elements: the relevant increment is defined as the total volume of the service at issue; the decision at issue is taken to be whether the increment is supplied over the longer run – so that the capital stock is variable, and hence is included in the cost pool; and the concern is with the resources that would be needed to provide this service with current technology and management practices, as against those that may have been inherited from earlier periods.

A number of strong claims are typically made in support of such an approach. First, it is argued that by focussing on the costs an efficient supplier would incur over the time period stretching out from the present, the approach sets aside the errors or distorted decisions the supplier may have made in the past, and hence reflects the opportunity cost of the services to be provided. Second, these opportunity costs, if they are used to set the regulated supplier's revenue target, should provide an indicator of the income the supplier requires for the maintenance of operating capability (that is, service potential). Third, by focussing on the costs incurred as a result of supplying some increment of output, arbitrary cost allocations are avoided.

From an analytical point of view, these claims are over-stated. As a general matter, there is no meaningful sense in which replacement costs, even of a hypothetical efficient supplier, measure the opportunity cost of using existing assets[15]. Rather, analysis of replacement costs is merely capable of providing the answer to the following thought experiment: what resources would be required to replace the assets that now provide the entirety of the service if for some reason those assets disappeared?

It is not easy to give a strong normative interpretation to this thought experiment. Two related interpretations are nonetheless worth mentioning.

---

12 A particularly useful taxonomy can be found in R. Ma and R. Mathews The Accounting Framework Longman Cheshire: Melbourne 1979 at 478 and follows.

13 See G. Whittington Inflation Accounting: An Introduction to the Debate Cambridge University Press 1983 at 131 to 136.

14 See, for example, W.J. Fabrycky, G.J. Thuesen and D. Verma Economic Decision Analysis (3rd ed.) Prentice Hall: Upper Saddle River, New Jersey 1998.

15 See M. Bromwich "The General Validity of Certain 'Current' Value Asset Valuation Bases" Accounting and Business Research vol. 7, no. 28 1977 at 242-249.

First, the answer to this thought experiment may provide an indicator of the stand-alone cost a consumer would face in opting out of the existing system. In a contestable market, no consumer[16] or coalition of consumers could be charged more than its stand-alone cost. Hence, this indicator can be used to define a ceiling to the revenue that an incumbent supplier could secure in a contestable market.

Second, assume the regulator put out to tender a contract for the supply, on an indefinite basis, of the service at issue. The thought experiment set out above could be structured so as to provide an indicator of the greatest amount the regulator should expect to pay so as to make the incumbent indifferent between entering and not entering into such a contract[17]. Again, this amount is a ceiling price, rather than unambiguously defining a revenue requirement.

The first of these interpretations simulates competition "in the market", while the second simulates competition "for the market". In each case, approaches based on forward-looking costing can be viewed as providing an indicator of the ceiling revenue requirement associated with the provision of a regulated service. However, this requires that these estimates be constructed so as to model the ceiling price of supply.

Seen in this light, estimates of forward-looking costs would seem to be of interest, though they could not be said to be determinative. However, as a practical matter, deriving such estimates involves a significant number of difficulties. As will be seen in the following section, these difficulties have, in many instances, been addressed in a manner that is not consistent with the interpretation of the cost standard as a ceiling test. Moreover, short-cuts and simplifications have been adopted that have little economic justification. The overall may be such as to severely reduce the usefulness of the resulting estimates.

Implementing forward-looking costing approaches

Given a decision to build a forward-looking cost model, nine key issues need to be faced. These are:

(1) determining the relevant increment;

(2) specifying the technology to be used to supply that increment;

(3) determining the time frame in which a network corresponding to that technology will be built;

[16] Including a competitor purchasing the service as an intermediate input.

[17] Note however, that it will not do so if it does not allow the regulated supplier to recover its joint and common costs. As a result, the thought experiment needs to capture the stand-alone costs involved in providing the service, including in these the joint and common costs. Moreover, the costs measured in the thought experiment need to allow the maintenance of the financial capital of the supplier (as against solely providing for the maintenance of physical capital) (see especially R.R. Sterling "The concept of physical capital maintenance" in R.R. Sterling and K.W. Lemke Maintenance of Capital: Physical and Financial Scholars Book Co.: Houston, Texas 1982 at 3-58). Whether or not they do so will depend on the determination of the cost of capital, on the depreciation policy adopted, and on whether costs include the option value associated with deferring investment.

(4) determining the base of existing assets and services to which this new network is to be added;

(5) specifying provisioning rules with respect to the capacity/demand balance;

(6) determining the appropriate level and time path of operating and maintenance outlays;

(7) identifying the level of indirect costs;

(8) determining the treatment of capital charges, including with respect to depreciation, the opportunity cost of capital and the cost of not deferring investment;

(9) validating the model through appropriate sensitivity testing, and determining confidence intervals around the estimates.

Once these issues have been addressed, further questions arise before the results can be used as a basis for charging. In particular, consideration needs to be given to:

(1) determining whether costs will be expressed with respect to a particular year or as a levelised amount – that is, as an annuity factor;

(2) translating these estimates into units useable for regulatory purpose – that is, taking the pool of costs and decomposing it into a cost per line, per peak call attempt, or per call duration.

These issues can, for the sake of simplicity, be roughly grouped under three broad headings: definition and design of the hypothetical network; assessment of the costs of that network; and interpretation of the results for regulatory purposes. Each of these is addressed below.

Definition and design of the hypothetical network

The relevant increment

A first step in constructing a forward looking cost model is to define the outputs being incremented – that is, to specify what it is that is being costed. This involves determining the service, or grouping of services, being incremented and the relevant volume of that increment. At least in an interconnection context, this typically centres on three questions. First, is the relevant increment only the service volume being supplied to one or more competitors, or does it include the volume the incumbent supplies to itself? Second, what base volume should be used in the modelling? And third, when that volume of services is being supplied, are other services being provided as well?

With respect to the first of these questions, it is apparent that an increment defined in terms of (say) the minutes of carriage supplied to one or more competitors will provide a very different answer compared to an increment defined in terms of the minutes of carriage supplied both to competitors and by the incumbent to itself. The “competitor only” approach imputes to the competitor all of the scale economies involved in the service’s provision. Any charge based on such a calculation can only be a floor price, and when the outputs of the incumbent and of competitors are direct substitutes, may well be below any sensible price floor[18]. As a result, it is not apparent that any normative implication can be drawn from analysing an increment so defined. Rather, it seems reasonable that the increment be defined in terms of the *total volume* of the service, including both sales to competitors and the incumbent’s supply to itself.

This total volume is, however, itself uncertain. In principle, a regulator seeking to replicate the outcomes associated with a contestable market should seek to cost the output level that would prevail in such a market (as against using the output level that happens to prevail in the ‘real world’). Where prices are marked up above the contestable market benchmark, then actual output may be lower than the output level that should be costed. Using this lower level as the relevant increment will over-state costs whenever there are unexploited scale economies.

In practice, the difference between actual and hypothetical output will be small, at least in most mature networks. In these networks, access and local calls account for the largest share of the output to be modelled, and demand for these is extremely inelastic[19]. As a result, in the typical mature network, moving prices to the contestable market level will have little effect on the aggregate load, and hence on costs. It is therefore reasonable to use actual loads, and their projected growth, as the basis for determining the volume change being modelled.

[18] When the outputs are direct substitutes, the marginal cost to the incumbent of supplying interconnection includes the foregone revenue in the downstream market. If the downstream services are not priced at cost (because of the need to fund joint and common costs), (1) the incumbent would never use marginal resource costs as a floor price for the sale of interconnection services and (2) a regulator which required these to be used for that purpose would threaten the supplier’s financial viability.

[19] For example, it is reasonable to suppose that the price elasticity of demand in Australia is in the order of -0.05 for residential access, -0.06 for business access, and -0.06 for local calls (which are untimed). The elasticities may be slightly higher for corporate customers, as these face a broader range of substitution possibilities.

Assuming the volume change is determined in this manner, there is still the issue of whether the increment only encompasses the service directly at issue. For example, in an interconnection context, the primary concern will typically be with originating and terminating PSTN access[20]. The relevant increment would then be the PSTN Customer Access Network, as well as the switching and transmission networks used to carry PSTN traffic to and from points of interconnection[21].

However, it might be argued that the facilities used by the PSTN are also used by other services, such as leased lines, and by other networks, such as the ISDN and the IN[22]. The services provided by these other networks and facilities, it is then argued, should also be included in the relevant increment so as to capture the cost sharing that occurs in the network. Reflecting this argument, the model built by NERA for the Australian Consumer and Competition Commission (the industry regulator) defines the relevant increment as originating and terminating PSTN access *and* originating and terminating ISDN access *and* leased line services.

It is difficult to know what to make of this approach. To begin with, it seems fairly arbitrary. Thus, the relevant increment, as NERA defined it in Australia, includes some services which share significant facilities with the PSTN but excludes others. At the same time, since costs now need to be distributed among the services sharing the facilities at issue, it re-introduces the allocation difficulties that an incremental cost standard was supposed to avoid. Unless the services are perfect substitutes, resolving these difficulties by apportioning costs on the basis of usage factors (as NERA does) is not economically defensible[23]. The resulting estimates cannot be given an economic interpretation either as a price ceiling or as a price floor.

Rather, if the outcome of the exercise is to be interpreted as a price ceiling, then the relevant increment should be PSTN originating and terminating access, and it alone. Unfortunately, none of the major TSLRIC models takes this approach[24].

[20] The PSTN is the Public Switched Telecommunications Network. “Originating and terminating PSTN access” refers to the provision of the facility of transporting calls between customer premises and the points of presence of the incumbent and/or competitors, as well as the actual conveyance of those calls between customer premises and these points of presence.

[21] The Customer Access Network (CAN) links customers’ premises to points of traffic concentration, which may be either local exchanges or sub-units such as Remote Integrated Multiplexors and Remote Switching Units. The nodes at which the CAN terminates, together with the transport facilities linking these nodes to each other and to higher level exchanges (generally, trunk network exchanges), form the Inter-Exchange Network (IEN).

[22] The ISDN is the Integrated Services Digital Network, which is an end-to-end digital circuit-switched network that (in ETSI systems) provides capacity in bearer units of 64kbit/s. The IN is the Intelligent Network, which is a network that uses the signalling network to supply enhanced control capabilities to local and trunk exchanges.

[23] This is because the services will face different demand elasticities, and hence should bear a differing share of any joint and common costs: “When economies or diseconomies of scale are present, both the state of demand and the structure of costs must be taken into account in the setting of efficient [access] prices.” G. Sidak and W. BaumolTowards Competition in Local Telephony, The MIT Press, Cambridge: Mass. 1994 at page 50.

[24] The major US and UK models include leased lines, but exclude ISDN and data networks. The Australian model is unique in including ISDN.

Specifying the technology to be used

Given the increment being modelled, the next task is to specify the technology by which it will be supplied. “Technology” here is to be interpreted broadly, as comprising the facilities to be used and the manner in which they will be maintained and operated.

The major issue that arises in this respect is that of the extent and nature of optimisation. In practice, telecommunications networks are extremely complex systems linking many millions of terminal and intermediate nodes. Moreover, they are multi-purpose systems, comprised of inter-dependent but distinct sub-systems. Re-optimising an entire existing network would be an extremely demanding task – indeed, it is questionable whether the task could even be carried out to any acceptable level of accuracy. So as to make this complexity tractable, forward-looking cost models typically adopt three important simplifications.

First, they keep certain key features of the existing network constant. In particular, they usually assume that the location of local exchanges is not changed – that is, that the broad geography of local switching nodes is kept intact. For this reason, they are often referred to as “scorched node” models (that is, models in which everything other than the location of the nodes is treated as scorched) as against “scorched earth” models (in which the location of nodes is also treated as variable).

Second, rather than using the “best available technology”, they generally seek to embody the “best technology in widespread use”. This reflects two considerations:

- To the extent to which the cost estimate is used to set a revenue cap for a carrier, modelling “best available technology” would penalise carriers for not constantly adopting the most recent breakthrough. Such a standard seems unreasonable, and might well have a range of undesirable consequences[25].

- No less importantly, the cost properties of recently developed technologies are only poorly known, and the optimisation tools available for modelling their application are often experimental. As a result, cost estimates for networks embodying such technologies would be highly sensitive to the precise assumptions made and would have very wide confidence intervals[26].

[25] For example, a carrier which did develop and adopt a breakthrough would, under such a standard, not earn any innovator’s rents.

[26] See, for example, L.A. Ims “Economics of Residential Broadband Access” IEEE Network, vol.11 no. 1 1997 at 51-57.

Third, while the modelling is based on the assets corresponding to the “best technology in widespread use”, substantial parts of the rules according to which these assets are deployed and operated are not re-optimised. For example, it is usual to scale switching capacity by using Erlang processors that correspond to the conventional distribution of voice telephony. However, it is well known that these do not do a good job of representing the traffic patterns and switching needs now emerging from the growing use of the Internet or more generally, of mixed voice/data networks[27].

Combined, these assumptions mean that the networks embodied in “forward looking” cost models are not terribly forward looking. They do not, in fact, reflect the network that would be built today, were the assets used to provide existing services to disappear. Rather, they involve a limited optimisation in which clearly obsolete vintages are scrapped in favour of the vintages more recently placed into service. This makes it difficult to interpret the results as reflecting the outcomes that would prevail under contestability.

The difficulties associated with either building the model or interpreting the results are compounded when the service increment is defined too broadly. For example, the NERA model in Australia involves some optimisation of the PSTN, notably in terms of greater use of distributed multiplexing and switching. However, although it includes leased lines and ISDN in the service increment, no optimisation has been done of these services. The result is an uneasy amalgam, in which a hypothetically re-engineered PSTN shares capacity with the inherited leased line and ISDN services.

The time frame for network construction

Having specified the network facilities that need to be constructed, the next task is to define the manner and timing of their construction. From a practical point of view, there are very substantial cost differences between building a network “in one bang” and building it over a long period of time[28]. The Australian experience in telecommunications is that construction costs are highest for very fast and very slow roll-out speeds, reflecting a mix of factors that includes the availability of skilled personnel, the cost of stock-piling materials and the extent of volume discounts for aggregated purchases.

In practice, forward looking cost models are generally based on build costs that reflect the recent experience of network service providers. This implies that they embody a relatively slow roll-out speed, characteristic of the pace of replacement and expansion investment in mature networks. However, this means, as a logical matter, that costs should include a substantial amount for capitalised interest during construction. This is almost universally over-looked, adding to the difficulty involved in interpreting the results.

[27] See, for example, D. Minoli Broadband Network Analysis and Design Artech House: Boston 1993 at 99 and following.

[28] This point, and its implications for cost modelling, were emphasized by A. Alchian in his masterly essay on “Costs and Outputs”; see A.A. Alchian Economic Forces at Work: Selected Works by Armen A. Alchian Liberty Press: Indianapolis 1977 at 273-300.

The base of existing assets and services

Whatever the time frame adopted, the cost of building capacity to handle the relevant increment will clearly depend on the assumptions made about the base to which the increment is to be added. If other telecommunications services are being provided, or other assets which can be used already exist, costs will be lower than they would be for construction from scratch.

As a practical matter, it is unlikely that other telecommunications services would be provided in the absence of originating and terminating PSTN access. Moreover, even if they were, the asset base and resulting cost structure associated with their provision is difficult to model with any accuracy[29]. As a result, the best assumption that can be made is that the relevant increment is the first service to be provided.

Even so, there are still choices to be made as between modelling on a “greenfield” versus “brownfield” basis. In brownfield modelling, conventionally used for systems such as railway networks, it is assumed that perpetual structures, such as embankments, landfills, tunnels and other corridor formation assets, do not need to be reproduced. In contrast, in greenfield models, all structures are modelled as needing replacement or reproduction. The cost differences between these modelling approaches can be very substantial.

The choice between these boils down to the treatment of sunk costs. As a matter of theory, forward looking cost models are intended to act “as if” sunk costs did not exist. As a result, it seems inconsistent with the purpose to assume that some sunk costs (say, those associated with trenching) should be treated as sunk, while others (say, those associated with cabling) are not. Moreover, the line drawn between these would seem to be arbitrary, and would hence reduce the significance of the results[30].

Consequently, it seems best to consistently adopt a greenfield approach. However, this has not always been the approach adopted. For example, the NERA model developed for the ACCC treats the existing stock of connections (that is, the cost associated with actually linking premises to the distribution network and activating service) as pre-existing, as well as assuming that all the links from the Network Termination Point to the Property Entry Point do not need to be costed. It is not apparent for what reason these network elements have been treated as sunk, while others are not.

Specifying provisioning rules and the capacity/demand balance

In practice, networks are built to handle demand that varies over time according to a pattern that has both stationary and non-stationary elements. The former typically reflect the pattern of traffic over the time of day/day of week/week of year, while the latter reflect long run changes in the structure and level of demand.

---

29 For example, it is conceivable that leased lines would be provided in the absence of originating and terminating PSTN access. However, it is unlikely that they would be ducted, except in CBDs. Rather, more extensive use would be made of fixed radio technologies.

30 Additionally, a regulator who mandates that certain costs are to be treated as sunk signals a risk that a similar writing-off of costs may be carried out in the future, perhaps with a wider coverage. This increases regulatory risk and hence will increase the required rate of return.

Given this variability, the efficient provision of network facilities involves building capacity that both handles peak demand at some reference date (say, in the year being modelled) and caters for continuing growth in demand. *The margin over average load provided to handle the peakiness of demand* should reflect a grade of service target (for example, in terms of how long customers may have to wait before securing a second line, or how many calls are blocked at peak). *The margin provided for growth* should reflect optimisation for a given degree of lumpiness of investment – as a general matter, it is more costly to make frequent additions to capacity than to provision somewhat ahead of demand[31].

Forward looking cost models can generally cater for the first of these needs. Thus, information on traffic distributions is used to dimension traffic-sensitive facilities such as multiplexers, exchanges and transmission paths to meet a specified grade of service targets. However, greater difficulties arise in providing for the continuing growth of demand.

Thus, none of the major models makes any attempt to directly model growth over time and on that basis minimise costs on an explicitly inter-temporal basis. Rather, growth is typically handled by using the standard provisioning rules employed by major carriers and by adopting fill factors for equipment that reflect a certain degree of sparing (that is, provision in excess of instantaneous demand). However, the standard provisioning rules and fill factors are generally designed to optimise the costs of incrementally expanding an existing network[32]; using them to capture the efficient structure of inter-temporal provision in a greenfields network is inappropriate.

Determining the costs of that network

The level of Operating and Maintenance (O&M) outlays

Given the design of the hypothetical replacement network, the cost of operating and maintaining it needs to be determined. This will generally differ from that of the existing network, as the asset base has changed, work practices may be subject to some degree of optimisation and input prices may also be varied.

In principle, the O&M outlays associated with the new network should be modelled explicitly. This can be done by examining O&M outlays in the carrier being modelled for the "best in widespread use" assets, and then grossing these outlays up to reflect the prevalence of these assets in the hypothetical network. However, this assumes that these outlays reflect efficient work practices and input prices.

---

31 For example, in designing a new network, it is generally cheaper to provide 2 pairs per anticipated year 1 service in operation than to have to add small pair gain systems so as to cope with increased demand some years down the track.

32 See, for example, K.J. Carroll Economic Studies, Telecom Australia, 1977 and J. Freidenfelds Capacity Expansion, North-Holland, New York 1981.

Whether this is viewed as a drawback depends on an assessment of the appropriate degree of optimisation. “Inefficient” work practices or input prices above those paid elsewhere may simply reflect differing factor endowments and policy distortions across jurisdictions – for example, many materials costs are higher in Australia than in the United States because of differences in the structure and level of taxes. In other instances, they may reflect the extraction of monopoly rents by input suppliers, with consequent income transfers. In either case, it is by no means apparent that they are under the control of the network service provider. As a result, merely making the market in which that network provider operates contestable would not necessarily alter the level of these costs[33]. To the extent to which this is the case, they should not be changed in the calculation of costs for the hypothetical replacement network. Rather, the best option seems that proposed above – that is, to set O&M charges on the basis of the best in use in the carrier being modelled, taking account of likely productivity improvements genuinely within the operator’s control.

What is done in practice can differ greatly from this recommendation. For example, the NERA model developed for the ACCC calculates O&M charges as a percentage of the total capital cost, based on estimates provided by various operators in Australia and NERA’s experience elsewhere. However, this approach raises a number of problems, most of which were identified by NERA themselves in their assessment of the UK bottom-up model of BT’s costs.

First, operating to capital cost ratios are, by way construction, specific to a particular mix of capital and operating expenditure and will therefore vary significantly among different operators[34]. Unless capital costs estimates and operating to capital cost ratios are estimated from the same source, as recommended above, then the results produced by applying operating to capital cost ratios will be misleading.

Even with the same level of efficiency, operating to capital cost ratios may differ between operators. At least to some extent, this reflects the trade-off that exists for operators between capital and O&M expenses. NERA identified a number of other reasons why the ratios may differ significantly. These include the different ways in which operators categorise their operating costs, different assumptions used by operators in their estimates of incremental costs of network components and different methods of splitting costs into network and retail components. Therefore, using an average ratio estimated across a number of operators will not be consistent with the choice of capital expenditure adopted in the model.

Second, the operating to capital cost ratios used by NERA are calculated for existing operators. In practice, operators’ networks are not entirely new and the operating cost information that they submit will reflect the maintenance requirements of both new and old assets. This will mean that the operating cost information is not consistent with the annual capital cost which is estimated on a forward-looking, best-in-use basis.

[33] Though it would arguably increase the pressure on the network service provider to achieve any efficiencies that were within its control.

[34] In their review of the UK bottom-up model of BT’s costs, NERA found a large degree of difference between BT and other operators for the estimated ratio of aggregate operating to capital cost for exchanges.

The level of indirect costs

The forward looking cost models being surveyed are essentially engineering models and do not pretend to measure the level of many indirect costs – going from those associated with personnel functions through to those of corporate governance. However, it is apparent that these costs need to be factored into any estimate of the price ceiling associated with stand-alone supply.

Where indirect costs are causally related to the service increment, they should be modelled explicitly. For example, vehicle fleet costs should be related to the output being provided within the framework of an activity based costing model, and the extent of these costs required to provide the service increment determined. Where costs are not attributable, such as the costs associated with corporate governance, the actual costs incurred by the operator should be used (at least in the absence of tangible proof of inefficiency).

As a general matter, the costs being modelled in an interconnection setting are those of a wholesale service provider – that is, of an operator supplying PSTN terminating and originating access to competing retail service providers (including its own such operation) which then market services to end-customers. The indirect costs associated with the retailing function (such as advertising and promotion, end-user billing and end-user revenue collection) should therefore generally be excluded. However, this does not justify excluding *all* retail costs; rather, the only retail costs that should be excluded are those that a hypothetical wholesale service provider would avoid.

For example, the models universally assume that customers have been supplied with access lines, and hence can use these lines to place and receive traffic. This implies that the hypothetical wholesale service provider is carrying out the functions involved in the retail supply of access lines. The costs associated with these retail functions ought therefore to be included.

In short, it would seem desirable to (1) explicitly model indirect costs; and (2) include in these the retail costs the network service provider would not avoid when it supplies service to downstream competitors.

Practice often departs in significant respects from these recommendations. For example there are a number of models in which indirect costs are determined by using ratios relative to direct costs. For example, in the model it built for the ACCC, NERA estimated indirect capital and operating costs as a percentage of total direct network capital costs and total direct network operating costs, respectively. Data from the US as well as data for BT were used to construct these ratios to which two "adjustments" were then made. First the ratios were adjusted to take into account any differences in the Australian environment (though, as noted below, in a manner that seems arbitrary). The relevance of the cost item to interconnection charges was then estimated and the figures were again "adjusted" to ensure only relevant costs were included. While NERA noted that these adjustments involved a degree of judgement, no indication was provided on how those judgements were made.

Reliance on these ratios is flawed for a number of reasons.

First, the valuation base according to which the ratios are calculated may, and often does, differ from that being used in the model. Thus, in the NERA model referred to, the ratios for indirect capital costs were calculated using historical costs in the numerator; the assumption that this provides any indication of the "right" level of outlays relative to assets valued at replacement cost seems entirely arbitrary.

Second, the ratios are typically calculated for each line item over a sample of operators – for example, by looking at the ratio of vehicle fleet expenses to network expenses for a range of US local exchange carriers. Some average of the values of this ratio in the sample is then taken as reflecting the appropriate level of that item of outlay. However, whether this is reasonable depends on the assumptions being made about the relation between the various items. If they are substitutes— that is, if greater expenditure in one category is associated, for an efficient operator, with lower expenditure in another — then taking averages across disparate carriers may be inappropriate as it would not capture these substitution effects.

Rather, the conventional approach in economics would be to use Stochastic Frontier Models or the non-parametric Data Envelopment Analysis method to estimate a production function, and then derive a reasonable configuration of the relevant costs from the function so estimated.

Last but by no means least, there is no obvious rationale for assuming that the ratios derived from one country (say the United States) are applicable to another (say Australia), even when these ratios are correctly measured. Rather, it is obvious that adjustments need to be made – but there is no simple or reliable way in which to do this. As a result, adjustments, when they are made, tend to be highly arbitrary, and undermine the credibility of the overall results.

It has, for example, been noted above that the NERA model developed for the ACCC makes an adjustment to US ratios for legal expenses, presumably on the basis that Australia is less litigious than the United States. The amount of the adjustment is not explained. However, NERA does not increase the estimate of vehicle fleet costs, despite the far stronger argument for doing so (Australian tariffs on vehicles being relatively high, increasing Australian prices for vehicles relative to those in the US).

The level and treatment of capital costs

Three complex issues are involved in determining the level of capital costs: the required rate of return on capital; the role of option values; and the treatment of depreciation.

For most purposes, it can be assumed that the required rate of return will be modelled using the framework of conventional finance theory. In this approach, a weighted average cost of capital is calculated, based on the estimated costs of debt, the cost of equity and the firm's optimal capital structure. Typically, the most contentious element in this process, the cost of equity, is derived through application of the Capital Asset Pricing Model (CAPM).

Without reviewing the merits and demerits of the CAPM, it is clear that it can only be used if the cash flows being discounted correspond to the model's assumptions. In particular, the expected future cash flows being discounted must come from a stationary normal distribution. Where there is some net present value to the option of delaying investment[35], then the resulting expected cash flows need not be so distributed[36]; the problems this creates are especially acute when the CAPM is being applied in a context where regulation makes the distribution of risks one-sided (the firm can incur losses on some investments, while being prevented from earning supra-normal returns on others[37]).

In these circumstances, the expected cash flows need to be transformed to be consistent with the CAPM. This is best done by including, as an outlay in the cash flows being modelled, the actuarial value corresponding to the risk of stranding. In practice, this is equivalent to assuming that the firm needs to self-insure against those risks whose cash flow consequences are not normally distributed, and that the costs of this self-insurance need to be included in the relevant cost base[38]. Equivalently, the required grossing-up of outlays can be expressed as a mark-up over the weighted average cost of capital[39].

Given corrected cash flows, the CAPM-based required rate of return can be used to measure the return on capital for a given value of assets. To obtain the total capital charge for the period, the depreciation charge must be added to this amount.

[35] In practice, an investment in new assets can almost always be delayed. Furthermore, under some conditions, an investor will prefer to delay the project while demand evolves, because immediate investment incurs a risk that the project will fail to earn the cost of capital. If the expected value of delaying investment is positive, the ability to delay has the same advantages as a call option in finance. Because it is defined over a real asset, however, it is referred to as a real option. The value of a real option is simply the expected additional profit from delaying the investment. Equivalently, it is the expected value of the profit that would be lost by investing immediately. The source of this loss is the risk that the asset will be incapable of earning the cost of the capital employed. In other words, an appropriately calculated real option value is also the expected cost of the asset becoming stranded. See H. Ergas and J. Small "One Way Bets on Regulated Monopolies: Theory and Evidence" CRNEC, University of Auckland, 1998.

[36] For option values to be positive, the earnings series must display first-order stochastic dominance. The series may be stationary but autocorrelated. The inconsistency with the underlying CAPM assumptions arise when the series is not stationary. See for example M.C. Ehrhardt The Search for Value: Measuring the Company's Cost of Capital Harvard Business School Press: Cambridge, Mass. 1994 at 212 and follows.

[37] The use of TSLRIC to set interconnection prices may itself have this effect. Thus, the regulated firm will take a loss whenever best-in-use equipment costs fall more rapidly than expected, but (if regulatory reviews are sufficiently frequent) will not earn supranormal profits when costs fall less rapidly than expected.

[38] For example, under a TSLRIC rule, firms would need to insure against stranded asset risk, since (in a TSLRIC world) this is not a risk that investors can diversify. However, there is no market for third-party provided insurance against asset stranding, and any instruments which attempted to provide such insurance would be vulnerable to moral hazard. As a result of these transactions costs considerations, efficiency requires that firms self-insure.

[39] It is sometimes claimed that the same effect can be achieved by accelerating the depreciation schedule. Even in the circumstances in which this can be done (and it depends on accelerated depreciation increasing the net present value of the firm – which it does not always do), the correct amount of the acceleration needs to be determined by calculating the charge referred to in the text. In other words, the cost associated with the option value (the value of the ability to defer investment) needs to be computed explicitly.

Economic, rather than accounting, depreciation should be used in this calculation. In essence, accounting depreciation allocates the historical cost of assets to the time periods in which those assets are used[40]. In contrast, economic depreciates measures, in each period, the holding cost associated with using that asset in that period, this cost being assessed as the change in the value of the asset in that period[41].

Viewed in an *ex ante* sense, economic depreciation can be regarded as a contract between future periods that is established when a non-current asset is acquired[42]. The terms of this contract ensure that, if capacity can be adjusted in an accounting period, that period bears a charge equal to the ex ante (anticipated) change in replacement cost during that period[43], with the total charge across such periods 'adding up' to the capacity's original cost. If capacity cannot be adjusted within accounting periods, then replacement costs are not relevant and the depreciation charge, while continuing to 'add up', is imputed to each period based on the asset's value-in-use[44].

As a general matter, the pattern of charges generated by accounting depreciation need not be sustainable in a contestable market, even when changes in asset values are correctly anticipated[45]. As a result, it the *ex ante* concept of economic depreciation that should be used in forward-looking cost models[46].

In short, the return on capital should be calculated correcting the cash flows so as to make them consistent with the CAPM (if the cost of equity is being calculated on a CAPM basis); while the return of capital should be based on anticipated economic depreciation.

The practice of forward looking cost models departs substantially from these recommendations.

Thus, none of the major models explicitly corrects the level of cash flows before applying a CAPM-based weighted average cost of capital. As a result, they mis-state, and likely understate, the return an investor would require to finance the facilities at issue.

---

40 See, for example, G. Peirson and A. Ramsey Depreciation of Non-Current Assets Australian Accounting Research Foundation: Melbourne, Victoria 1994.

41 That is, the amount that must be added to the balance sheet in order to keep wealth intact. (See C.R. Hulten and F.C. Wykoff "Issues in the measurement of economic depreciation" Economic Inquiry, vol. 34 1996 at 10-23.)

42 This view of economic depreciation as an implicit contract can be traced back to Ladelle in 1890. See P. Brief "A late Nineteenth Century contribution to the theory of depreciation" Journal of Accounting Research 1967 at 27-38.

43 Note that the replacement concept must itself be interpreted in an *ex ante* sense: that is, replacement cost in the future as anticipated in the firm's optimal plan.

44 See A.A.Atkinson and W.R. Scott "Current Cost Depreciation" Journal of Business Finance and Accounting vol. 9 1982 at 19-42.

45 See E.E Zajac Political Economy of Fairness The MIT Press: Cambridge Mass. 1995 at 240-242 and especially J. Panzar "A Critique of the WIK Cost Model" CRNEC, University of Auckland, 1998.

46 See more generally A. Hardin Accounting and Economic Depreciation in Forward Looking Cost Models at www.necg.com.au

At the same time, they do not use economic depreciation to calculate the return of capital. The main US models default to straight-line depreciation, as do the WIK model of the German network and the NERA model developed for the ACCC. A more complex approach has been adopted in the UK. In BT's top-down model, straight-line depreciation was used plus a factor to account for price changes which resulted in a slight forward tilting of the depreciation profile. In the initial version of the UK bottom-up model a tilted annuity method was used to allocate capital costs. In their assessment of these models NERA correctly concluded that the depreciation methods used would not provide a good approximation of economic depreciation. Subsequent to NERA's review, model developers estimated economic depreciation profiles for each of the main categories of assets in the bottom-up model so the depreciation charge in each period would reflect economic depreciation. However, no change was made to the treatment of depreciation in the top-down model and the 1994/95 version of this model still uses straight-line rather than economic depreciation.

Model testing and validation

Forward looking cost models are inevitably based on important simplifications. Typically, even in the most detailed models, crucial parameters are based on summary statistics derived from samples of complex distributions. For example, the highly detailed model developed by Bellcore of the costs of providing service in high-cost areas in Australia relies on a 1 per cent stratified sample to estimate summary statistics for loop length distributions[47]. Further uncertainties are introduced by the process of analysis, including in terms of the assumptions made about such matters as the extent of trench sharing, the degree of sparing, the level of indirect costs and the pace and pattern of depreciation.

Given these uncertainties, the resulting estimates are no more than that – estimates. Before they can be relied on, some sense is needed of the likely margin of error.

In principle, the risk of error should be assessed using formal tools for model validation. In essence, this requires making explicit the confidence intervals associated with the sampling used to derive input data. It also requires testing the sensitivity of the outcomes to changes first, in the input data and second, in the key functional relations.

[47] According to the documentation, the goal was to divide PNLAs into sets based on cost drivers and then to measure the variability of the sample by selecting several ESAs from each set. Costs for each of the selected ESAs were to be estimated so the variance of the distribution of costs within each set and hence the required number of sample exchanges needed to achieve the desired level of precision could be calculated. Bellcore recommended a 95% confidence interval at 10% variance. The 1% sample actually used is intended as an approximation to the recommended Bellcore method. .

What is done in practice falls well short of this. For example, NERA, in its Report to the ACCC, carried out sensitivity tests by varying some individual parameters; however, it did so holding everything else in the model constant. This can only be justified when the parameters being varied are independent of those being held constant. While this is true for some parameters, a properly built model should have quite significant interdependencies. When this is the case, model validation requires that these linkages be accounted for in the sensitivity analysis[48]. This can be done through a Monte Carlo analysis that traces out the response plane of the estimate to changes in sets of input values and in functional relationships.

When such an analysis is not carried out, the best that can be done is to gross up the estimate so as to account for the risk of error. Thus, in engineering economics, it is standard to use contingency factors which vary with the degree of project completion. For a typical major project – say, construction of a new port facility – it is assumed that the likely error range in the initial cost estimate (that is, a cost estimate made before 2 per cent of the project has been completed) is plus or minus 40 per cent. Investors funding such a project would generally require a contingency allowance of some 30 per cent to be added to costs – that is, costs would be grossed up to 130 per cent of the original estimate to insure against the risk that that estimate was unsustainably low.

Making use of the results

Levelisation and the choice of a reference year

The result of the steps outlined above should be a stream of annual costs, consisting of year-on-year capital charges and expenses arising from operations and maintenance. This stream will start in year 1 (the first period in which the network is in operation) and stretch out over the network's useful life[49].

As a general matter, the costs associated with each year will vary – that is, the values associated with year 1 will not be same as those in year 2, which themselves will differ from those in year 3 (and so on). As a result, the choice of which year is taken as a base will matter – that is, the revenue ceiling arising from the estimate will be different depending on whether it is assumed that we are in year 1 (the network has just been built) or in year 10 (the network is now 10 years old).

---

48 See for example G.J. Thuesen and W.J. Fabryky Engineering Economy (8th edition) Prentice-Hall: Englewood Cliffs New Jersey 1993 at 503 and follows.

49 If the stream is truncated before that time, there will be a terminal value capitalising the stream of costs to be incurred subsequent to the end-date.

This proposition – that the choice of reference year matters – is at times disputed on the basis of the following argument. Suppose that the regulator requires the capital charge to be calculated on a year 1 basis, that is, at the end of the first year. If we are in 1998 then the network is built in 1998. The capital charge will be the sum of economic depreciation over 1998 and the cost of capital on the average written down value of the assets in that year. Now assume that the regulator specifies that the capital charge should be calculated on a year 10 basis. Then, still in 1998, this means the network was built in 1988, and hence built at 1988 prices. With economic depreciation being applied, the annual capital charge today on a network built ten years ago should be no different from the annual charge today on a network built one year ago. As a result, the choice of reference year does not matter.

This argument is flawed. To see this, assume that the cost K of constructing a network falls at the constant rate α per annum. The relationship between the costs of constructing a network in 1988 and 1998 is therefore given by $K_{98} = K_{88}(1-\alpha)^{10}$. Given a construction date, the connection between written down values in various years is simply $w_t = w_0(1-d)^t$ and assuming that the rate[50] of depreciation is also constant at d we can therefore see that the 1998 capital charge for a network constructed in 1988 is:

$$C_{98,88} = (d + r)K_{88}(1-d)^{10}$$

whereas the corresponding charge for a network constructed in 1998 is:

$$C_{98,98} = (d + r)K_{98}(1-d)$$

$$= (d + r)K_{88}(1-a)^9\,(1-d)$$

Thus, unless the rate of decline in construction costs α is exactly equal to the depreciation rate d, the capital charge will be different for networks of different ages.

For the economic depreciation rate and the rate of decline in construction costs to be identical economic depreciation would therefore have to reflect technological obsolescence *only*. However, in practice, assets are replaced for a range of different reasons, including capacity constraints, physical deterioration and technological obsolescence. A depreciation method that is consistent with economically efficient decision making will take account of these replacement characteristics. Therefore, in addition to technological obsolescence, economic depreciation reflects physical deterioration and capacity constraints in which case today's annual capital charge on a network built ten years ago is likely to be substantially different from today's annual capital charge on a network built one year ago[51].

---

50 Note that the implied connection between d_t and d is simply $d_t = dw_t$.

51 A year 1 charge may also involve double counting. This is because the price of new capital goods usually include an element of pre-paid maintenance – that is, that they can be viewed as "steady state" assets which have just had a comprehensive overhaul. If the maintenance costs typical for a steady state asset are added to the annualised cost of an entirely new asset, actual year 1 costs will be over-estimated.

An alternative to calculating the capital charge for a particular year is to levelise the capital charge such that its value is equal in each year of the asset's life. Adopting this approach, the annual capital charge would be calculated as:

$$C_l = \frac{r}{1-(1+r)^{-n}} \cdot \sum_{t=1}^{n} \frac{(\overline{w}_t \times r + d_t)}{(1+r)^t}$$

where C_l is the levelised capital charge;

r is the WACC;

n is the useful life of the asset;

w is the written down value of the asset; and

d is economic depreciation.

Unlike the year 1 approach, a levelised capital charge is consistent with the contestable market standard. In a contestable market in which long-term contracting could occur, a firm could only charge the year 1 price in year 1 if it could credibly commit to charging the year 2 price in year 2 and so on. Since any higher price would induce entry (by firms offering a levelised charge on a long-term contract basis), the firm would have to accept an expectation of price defined by the levelised charge.

The levelised charge therefore corresponds most closely to the thought experiment discussed at the outset – that is, how much revenue would the regulator need to commit to allowing the service provider to earn to make that firm indifferent as to entering into an indefinite term contract for the supply of the services at issue? It consequently seems the most reasonable way of translating the estimated stream of costs into annual terms[52].

Here too, however, practice often departs from the approach theory suggests. In the UK, the bottom-up model annualises BT's estimated capital costs by simply taking the first year capital charge, even though substantial efforts were made to estimate economic depreciation profiles over the entire useful lives of each asset group. Similarly, in the model developed by NERA for the ACCC, the annual capital charge was calculated as the sum of depreciation and the cost of capital in the first year of the network's life.

[52] The claim is sometimes made that use of the year 1 charge is desirable because (like accelerated depreciation) it provides some insurance against the regulator acting opportunistically. This argument is unconvincing. To begin with, with very long lived assets, use of the year 1 charge would only have a slight effect in terms of insuring against regulatory opportunism. As a result, the year 1 approach is a poor instrument for dealing with the problem. Rather, to the extent to which there is such a risk, it ought to be reflected in the weighted average cost of capital. If it is not, then it is the cost of capital that should be adjusted.

Unitisation

The main output of a TSLRIC model is an amount of costs – that is, a dollar total. This total amount can be viewed as the ceiling revenue requirement described above. In practice, this total is almost always re-expressed in more disaggregated terms, both as a series of sub-totals and ultimately as a cost per physical unit.

A typical first step in this respect is to decompose total direct costs into "line related" and "traffic related" costs, where the line-related costs are those that do not vary with traffic. There are two problems with what is done in this respect.

First, the purported allocations may bear little relation to cost sensitivities within the model at issue. For example, using the Hatfield model for seven US States, the percentage of the dominant local exchange carrier's total costs that is traffic sensitive can be calculated as ranging from 15.0% to 21.1% (see **Table 1**)[53].

Table 1: Share of Traffic Sensitive Costs in Total Costs

Hatfield Model 5.0a	
State	**Traffic Sensitive Costs**
California	19.4%
Florida	17.1
Montana	18.4
New York	21.1
Georgia	15.0
Missouri	15.4
Maryland	17.9

[53] The percentage of total costs in the Hatfield model that is usage sensitive is estimated by assuming that all loop and switch port costs are NTS, and all remaining switch costs, all signalling, and all transport costs are TS. Note that the model computes the costs of the UNEs inclusive of attributed common and overhead costs, including the costs that are shared with the private line services provided by the network. However, dedicated transport has been excluded from the calculations reported.

However, increasing traffic by 30 per cent above the default level specified in the model increases total costs in each of these States by less than 1 per cent. The split between traffic and non-traffic sensitive costs therefore maps very poorly into incremental, and even less so marginal, costs. As a result, using this split to set unit charges, 'as if' the split reflected cost causality, could be seriously distorting.

Secondly, decomposition into sub-total creates issues about the allocation of common costs. Thus, in practice, costs do not fall neatly into "traffic" and "non-traffic" sensitive categories as line elements share many costs with traffic systems. As a result, there are substantial common costs between these component parts.

The standard practice is to distribute these common costs through a proportionate mark-up on the attributable costs. However, there is no obvious economic justification for this practice. References are at times made to Shapley values as being consistent with such an equi-proportionate mark-up[54]. However, it is not apparent why the Shapley value is the appropriate standard for this purpose. As with all cost-allocation rules based purely on costs, the Shapley value will not yield an economically efficient mark-up. No less importantly, there is no *a priori* reason to assume that the Shapley value is sustainable – that is, that the regulated firm could actually secure the revenue patterns and levels determined by the Shapley rule[55]. When by-pass is a possibility, the use of equi-proportionate mark-ups may result in revenue targets for each sub-component that cannot be achieved. As with Fully Distributed Costs, the use of this allocation rule may therefore undermine the network providers' financial viability.

These issues of cost allocation are exacerbated by the allocation to sub-pools of indirect costs. Here too, equi-proportionate mark-ups are generally used, raising the same issues as those discussed above.

Rather than such essentially arbitrary unitisations, it may be preferable to simply view the cost estimate generated by the models as an aggregate revenue cap. It would then be up to the network service provider to ensure that the relevant revenues did not exceed this cap, subject only to the constraint that the charges through which this is achieved are broadly competitively neutral[56].

[54] The Shapley value is a solution concept in cooperative games. Formulated loosely, it is the expected value of the pay-off to each player when the game is played an infinite number of times, so that each possible sequence of play is fairly represented.

[55] The Shapley value need not, in other words, be in the core of the cost allocation game. This is because it takes no account of the opt-out possibilities open to sub-coalitions within the game.

[56] The welfare properties of such an approach are discussed in E. Ralph "Regulating an Input Monopolist With Minimal Information" which can be found at http://www.necg.com.au

Conclusions

Properly constructed forward-looking cost models can be viewed as providing an indication of the revenue requirement associated with the provision of a regulated service. When the service at issue is modelled as fully including the joint and common costs of service provision, this amount defines a ceiling to the revenues a regulator would need to allow the network provider to earn. However, absent special assumptions, replacement cost modelling will not provide more information than that.

Whether it even provides that information depends on the reliability of the estimates generated. The practice of forward-looking costing departs in many respects from the theory: as has been shown above, the models reviewed are littered with short-cuts and assumptions that limit their validity. The authors of the models often justify these assumptions by asserting that each assumption's consequences are small. However, what matters is the cumulative error range the sequence of assumptions creates. With minor exceptions, the models at issue do not provide estimates of this range and in some cases cannot be used to do so.

It is well-known that attempts to introduce forward-looking costing into statutory financial accounts have generally been unsuccessful because investors have had little confidence in the accuracy of the resulting estimates[57]. Regulators are of course better placed than individual investors to audit the information they require. Even so, there is a clear need for realism about the results of models such as those reviewed above, and for the results of these models to be compared with those of other approaches to determining regulatory revenue requirements.

---

[57] See for example W.R. Scott Financial Accounting Theory Prentice Hall: Upper Saddle River, New Jersey 1997, at 21 and follows, 152 and follows, and 205 and follows.