Foreword

The Australian Competition and Consumer Commission (ACCC) and the Public Utility Research Centre (PURC) are pleased to be co-publishers of this special volume of papers on the principles and practice of infrastructure regulation. The papers reflect the work of prominent Australian, New Zealand and international authors expert in the field of utility regulation — regulators, academics, consultants and advisers in the utility regulation field, and industry participants. The papers were selected from material presented at a one week utility regulation training program jointly conducted by the ACCC and PURC in Melbourne in November of 1997.

This course was a first for Australia, and was adapted from the successful *International Training Program on Utility Regulation and Strategy* conducted by PURC and the World Bank. The course was notable for the breadth of its coverage of issues involving utility regulation, both economic principles and practice. As such, the sessions were tailored to meet the pressing need to educate and re-educate regulators and policy makers for the rapidly changing regulatory and competitive environment. The success of the course could be judged from the range of its participants, who well represented Australian and New Zealand regulatory and policy-making bodies across all major utilities and jurisdictions.

The quality and breadth of the papers augurs well for infrastructure regulation in the future. The papers cover important fundamental principles, based on recent Australian and international literature and the work of pre-eminent academics and professionals. The discussion of the practice of utility regulation was also invaluable, and similarly draws by way of case studies, on the international environment, especially from the US, UK, New Zealand and Australia. This international flavour not only enhances the dissemination of greater expertise, but also facilitates the drawing of useful comparisons.

The general theme of the course and the papers is the observation that infrastructure regulation is not an end in itself, but rather, is directed at facilitating and augmenting the process of market reform, and contributing to the ongoing vitality, efficiency, and competitiveness of the market economy. The evolving environment of market reform has given rise to new areas of regulation which derive from, and are a response to, competition policy reforms such as legislated access to essential facilities, structural reform, competitive neutrality, and privatisations. The role of infrastructure regulation has correspondingly evolved itself from its traditional purpose of surveillance, detailed price controls, and arbitrary cost allocations. It is well understood that utility regulation is increasingly about maximising the potential for competitive processes to work or at least, helping to simulate market results, and always with an eye to the efficiency implications and resulting trade-offs among objectives. This approach implies a determination and analysis of the prevailing incentives which face industry participants. The regulatory regime introduced will then attempt to create incentives to induce greater efficiency, in addition to blunting the incentives which cause undesirable conduct. This regulatory evolution reflects the broad sweep of competition policy, which has gone beyond a primary reliance on anti-competitive conduct provisions to vigorous pro-competitive initiatives.

The importance of regulatory convergence is also becoming more appreciated. The Commonwealth and States have a role in regulation of infrastructure through the ACCC, NCC and State regulators. It is important that these organisations work together and cooperate so that regulation leads to efficient and competitive markets with good information, consistency of regulatory application, and minimal uncertainty.

We trust that this volume will contribute to the continual development of regulatory theory and practice and promote greater cooperation among regulators both within Australia and internationally. Readers of this publication, will have diverse backgrounds and objectives: you might be a regulator with well-defined mandates, an executive of a regulated firm identifying appropriate business strategies, an official in a government ministry or legislative body developing regulatory procedures, or a scholar studying regulation. Whatever your present role, we trust that upon reading this publication, you will be armed with a greater appreciation of the principles and practice of infrastructure regulation that is designed to help promote the on-going processes of market reform.

Professor Allan Fels
Chairman
ACCC

Professor Sandford Berg
Director
PURC, University of Florida

Acknowledgements

The ACCC and PURC are grateful to the authors who have contributed papers for publication in this volume and wish to express our thanks to them. The papers reflect insightful observations from people eminent in utility regulation or in fields directly related to utility regulation. The authors have been cooperative in reviewing and developing materials originally prepared for the Utility Regulation Training Program for publication in this volume.

The assistance given by Adam Rapoport of the ACCC with the preparation of this volume and in writing the 'Overview' chapter is acknowledged. Charlotte Reynolds assisted with formatting and proofing of the publication. Margaret Arblaster, ACCC, and Mark Jamison, PURC, jointly edited the publication.

The organisations

The Australian Competition and Consumer Commission

The Australian Competition and Consumer Commission was formed on 6November 1995 by the merger of the Trade Practices Commission and the Prices Surveillance Authority. Its formation was an important step in the implementation of the national competition policy reform program agreed by the Council of Australian Governments. An independent statutory authority, the Commission administers the *Trade Practices Act 1974* and the *Prices Surveillance Act 1983* and has additional responsibilities under other legislation.

Under the national competition policy reform program the Trade Practices Act has been amended so that, with State/Territory application legislation, its prohibitions of anti-competitive conduct apply to virtually all businesses in Australia. In broad terms the Act covers anti-competitive and unfair market practices, mergers or acquisitions of companies, product safety/liability and third party access to facilities of national significance.

The Commission is the only national agency dealing generally with competition matters and the only agency with responsibility for enforcement of the Trade Practices Act and the associated State/Territory application legislation.

A legal regime to facilitate third party access to certain essential infrastructure was introduced to the Trade Practices Act (Parts IIIA and XIC) as part of the 1995 competition policy reforms. Access regimes apply only to significant infrastructure facilities with monopoly characteristics and with wide economic influence. Examples of the kinds of facilities which may be covered include gas transmission and distribution pipelines, electricity transmission and distribution wires, railway tracks, airport systems, water pipelines, telecommunication networks and certain sea ports.

The Commission's role under access provisions typically involves assessing the terms and conditions, including prices, on which infrastructure services are provided. This role is undertaken in assessing access undertakings given by service providers or in determining disputes over access.

The Commission's internet website address is: http://www.acccgov.au

The Public Utility Research Centre

PURC is located within the Warrington College of Business Administration at the University of Florida. Founded in 1972, PURC strives to enhance executives', regulators', academics', and students' knowledge of issues confronting public utilities and regulatory agencies through: sponsoring conferences, seminars, and training programs; engaging in research that addresses topics in the energy, telecommunications, and water industries; and preparing students for careers in infrastructure industries.

Funding by energy and telecommunications groups and the Public Service Commission in the State of Florida enables PURC to fulfill its mission. To ensure independence and guarantee impartiality, an Executive committee is composed of representatives from the University, electric and gas utilities, telecommunications firms, the Florida PSC, and the Florida Office of the Public Counsel.

During the past twenty-five years, PURC has grown from a small group hosting one conference per year to an interdisciplinary effort with expanded programs and research whose impact is recognized locally, nationally, and worldwide. To learn more about PURC visit our website at www.cba.ufl.edu/eco/purc.

The International Training Program on Utility Regulation and Strategy, an intensive two-week program held in Gainesville, is a collaboration between PURC and the World Bank. The program is designed to enhance the economic, technical, and policy skills required to design and manage sustainable regulatory systems for infrastructure sectors. The program focuses on a series of cross-sectoral and sector specific topics including trends and drivers of market reform; market structure; financial techniques in utility regulation; methods of incentive regulation; performance standards and enforcement; pricing; and managing stakeholder groups including government, industry and consumers. Over 50 sessions include case studies, practical exercises, and panel discussions with leading experts and international faculty.

About the authors

Mr Graeme Samuel is President of the National Competition Council (NCC). He is also a Company Director. He was a co-founder of Grant Samuel & Associates, corporate advisers. Until 1986, Graeme Samuel was Executive Director of Macquarie Bank Limited (from 1981–1986) in charge of its Victorian operations and a Director of its Corporate Services Division. His career as a Banker was preceded by 12 years as a Partner of leading Melbourne law firm, Phillips Fox & Masel.

Commissioner Rhonda Smith was appointed as a Commissioner with the ACCC in November 1995 for a period of three years. She is an economist and from 1981 was a Senior Lecturer in Economics at the University of Melbourne. From 1988 she was also a consultant economist and was appointed a member of the Copyright Law Review Committee in 1995. Ms Smith has been an expert witness and adviser in major trade practices litigation and acted as a consultant for the Trade Practices Commission, Prices Surveillance Authority and the private sector.

Dr Alan Bollard, Secretary of the New Zealand Treasury, was Chairman of the New Zealand Commerce Commission, the country's anti-trust and fair trading regulatory authority until February 1988. He was appointed as the Government's economic referee for the reviews of the Commerce Act in 1989 and 1992 and was appointed a lay adviser to the High Court. He has published widely in the area of industrial economics and anti-trust. Before becoming Chairman of the NZ Commerce Commission in 1994, Dr Bollard was Director of the New Zealand Institute of Economic Research for seven years, where he was involved in advising the Government and companies on a wide range of applied economic work and forecasting.

Dr Sanford Berg, is Distinguished Service Professor in the Department of Economics at the University of Florida. He is also the Florida Public Utilities Professor and the Director of the Public Utility Research Center (PURC). In addition, he has served as a consultant to various private and public organisations including the Florida Public Service Commission (FPSC), the Florida State Energy Office, the New York Public Congressional Office of Technology Assessment, Southern Company College, the World Bank, the Australian Bureau of Industry Economics, the Tampa Electric Company Citizens' Advisory Task Force for Siting New Generation Capacity (1989–1990), and the Florida Power Plant Licensing Task Force, Technical Advisory Group, 1993. Currently he is responsible for developing and delivering the two-week International Training Program on Utility Regulation and Strategy. Dr Berg has published widely with recent work focusing on public utility pricing policies and transitional regulation in the presence of new technologies.

Professor Stephen King is Professor of Economics at the University of Melbourne. Before this he was Research Fellow in the Economics Program and Director of the Competition and Regulatory Policy Program at RSSS, the Australian National University. Professor King has been consulted by various government bodies including the ACCC, IPART, the Victorian Government, the Industry Commission and AUSTEL. Professor King's recent work on competition policy and infrastructure reform led to the publication in 1996 of a major book: *Unlocking the infrastructure: the reform of public utilities in Australia*, with Professor Rod Maddock.

Dr Denis Lawrence has been a Director of Tasman Asia Pacific, Canberra, the consulting arm of Australia's Tasman Institute, since mid-1996. He has been involved with benchmarking, performance measurement and the analysis of government enterprise issues for the last ten years. Dr Lawrence has worked on a number of infrastructure projects with Tasman. Prior to joining Tasman, he was in charge of the Australian Bureau of Industry Economics program of international benchmarking of Australia's infrastructure.

Mr Alan Booker joined Office of water services (Ofwat), Birmingham, in November 1990, as Deputy Director General of Water Services. He was formerly Managing Director of Biwater Supply Limited and of Bournemouth & West Hampshire Water Companies and Chief Executive of East Worcestershire Water Company. In 1980 Mr Booker became Managing Director of the East Worcester Water Company where he remained until 1988, until he took part in a management buy-out of the company in association with Biwater. Mr Booker is a Civil Engineer by profession.

Professor Henry Ergas is a Senior Research Associate of the Centre for Research for Network Economics and Communications at the University of Auckland. From 1978 to 1993, Professor Ergas held a number of senior positions at the OECD in Paris. During his time at the OECD he was responsible for many studies of regulations and industry policy, including the 1987 Report on Structural Adjustment and Economic Performance. He has taught at Monash University and the Kennedy School of Government at Harvard. Professor Ergas is one of four members of an Expert group advising the Minister for Communications in Australia on regulatory policy.

Mr Mark Jamison is the Director of Telecommunications Studies for the Pubilc Utility Research Centre at the University of Florida. He researches telecommunications competition and strategy issues, and conducts training programs on utility regulation, including assisting in directing the international regulatory training program for the Worl Bank. From February 1993 through June 1996, he was a manager of regulatory policy for Sprint, a telecommunications company. Prior to joining Sprint, he worked for US state commission staffs, including the Iowa Utilities Board where he was Head of Research. He has written numerous papers and testimonies on regulatory costing and pricing issues.

Dr Stewart Joy has enjoyed a diverse career, cutting across both private and public sectors and academia. Currently he is in Consultancy as Principal of Stewart Joy Associates Pty Ltd Economics and Transport Consultants. Other positions have included Deputy Director General of the Ministry of Transport in Victoria, State Manager of National Australia Bank, Chief Economist of British Railways, and Senior Lecturer in Economics at Monash University. Dr Joy has produced an extensive amount of consulting work over many years for a diverse range of organisations including the World Bank, Victorian Government and the US Department of Transportation. His work has covered a broad range of issues, including access principles, transport reforms and development, privatisations, and benefit/cost analysis.

Dr Joshua Gans is Associate Professor and lecturer in economics at the Melbourne Business School (MBS) of the University of Melbourne. Prior to this post at the MBS, Dr Gans was an economics lecturer at the University of New South Wales. Throughout 1996 and 1997, he worked as a Consultant for London Economics in Melbourne. In

this capacity, he undertook a plethora of projects related to infrastructure reform, regulation and privatisation, in areas such as energy, transport and communications reforms.

Professor Philip Williams holds the Foundation Chair of Management (Law and Economics) and is Assistant Director of the Melbourne Business School. The focus of his current research is the economics of law. He has published articles and books on the theory of the firm, trade practices policy, small business and the cost of litigation. He is extensively involved in giving expert advice to leading corporations in Australia and New Zealand in areas of pricing, takeovers and trade practices. This consulting work involves giving opinions and appearing as an expert witness for firms involved in commercial litigation. He is a member of the Trade Practices Committee of the Law Council of Australia.

Professor Kevin Davis is the Colonial Mutual Professor of Finance and Head of the Department of Accounting and Finance of the University of Melbourne. Before this post he was the Director of its Master of Applied Finance Programme. He has been very prolific as an author, mainly specialising in issues of corporate finance, privatisation, treasury management, and capital costs. Professor Davis has also acted as a Consultant for various organisations including Federal Treasury, the ACCC, the ASC, Telstra and Macquarie Bank.

Mr John Handley is currently a Lecturer in Finance at the University of Melbourne where he undertakes research in finance and teaches in various corporate finance and investment subjects at both undergraduate and postgraduate levels. His specific research interests include security valuation and design, capital raising, equity markets and project evaluation. Prior to joining the University of Melbourne in 1993, Mr Handley spent five years with an international investment bank where he provided a range of corporate finance services to both public and private sector clients.

Professor Peter Forsyth is Head of the Department of Economics at Monash University. Prior to this, he taught and researched at the University of New England, Australian National University and the University of NSW, after completing a Doctorate at the University of Oxford. He was editor of, and contributor to, Microeconomic Reform in Australia (Allen and Unwin, 1992). He has undertaken studies on pricing and regulation issues for agencies such as the ACCC and the Industry Commission.

Luke Woodward is General Counsel for the ACCC. He is the former head of the Mergers Branch of the ACCC (and the TPC). Prior to that he was a litigation practitioner with the Attorney-General's Department, where he conducted several major trade practices cases. In these roles, he has come face to face with each aspect of the regulatory information challenge. Mr Woodward has been responsible for using statutory information gathering powers, for the testing of information through formal and informal consultation mechanisms, for the preparation of formal reasons for decision, and for the conduct of judicial and administrative review proceedings.

Overview

The efficient and competitive performance of the industries comprising the infrastructure sector is crucial to the success of competition policy in Australia, and thereby pivotal to the overall effectiveness of microeconomic reform. The latter constitutes extensive efforts to minimise input costs and elevate productivity within the Australian economy. The special role of the infrastructure sector in the achievement of competitiveness and microeconomic goals, apart from standard efficiency reasons, derives primarily from them being such an essential input into all other industries. The nature of most public infrastructure facilities is such that they require regulation of a sort to ensure their efficient operation, especially in the transition to greater deregulation, privatisation and heightened commercial focus. In recognition of the vital importance of improved utility regulation, the ACCC and PURC conducted a Utility Regulation Training Program. This publication is a special volume of selected papers presented at the course, which was held in Melbourne over five days in November 1997, a project considered worthwhile in light of the high quality of the papers and the new regulatory functions continually being developed.

The training program was modelled on the International Training Program on Utility Regulation and Strategy conducted by PURC and the World Bank, held at the University of Florida. The course was a shortened version, which was suitably amended for an Australian and New Zealand regulatory environment, although the valuable lessons contained within the program, and indeed this volume, are widely applicable to regulation globally. The Programs are essentially designed to enhance economic, financial and strategic skills, and problem-solving techniques of senior utility regulators, strategists, and policy-makers, and to promote the exchange of ideas and experiences related to utility regulation. The Australian program featured a wide range of eminent professionals, both local and international, from the academic, regulatory, governmental, consulting and private spheres, as can be gleaned from **'About the Authors'**. It was also equally comprehensive in its coverage of issues relating to utility regulation, including reviews of competition policy, incentive regulation, price/rate structure, financial techniques, and other regulatory issues including information disclosure, and environmental concerns.

This volume is representative of the range and quality of the papers presented. The first section comprises a review of varying aspects of competition law, policy and theory, as they relate to infrastructure regulation, with 'scene-setting' papers by **National Competition Council Chairman Graeme Samuel** and **ACCC Commissioner Rhonda Smith** focusing on Australian developments, and a paper by former **Chairman of the NZ Commerce Commission, Dr Alan Bollard** on the New Zealand experience. The second section covers issues of incentive regulation. Following an 'introduction to the fundamentals' by **Professor Sanford Berg of PURC, Professor Stephen King** focuses in particular on the principles of one common method of incentive regulation, price-capping. **Dr Denis Lawrence** highlights the significance and problems of another method, that of benchmarking. **Deputy Director General of OFWAT Alan Booker** then brings to life these principles in his case study of the price capping incentive regulation of water services in the UK.

Section three deals with issues of price/rate structure and design. **Berg** provides an introduction to the various principles and objectives that should be adhered to in the design of an efficient and practicable price structure. **Professor Henry Ergas** looks at cost allocation from a contestable markets perspective. **Mark Jamison of PURC** follows, engaging in an analysis of price structure as it relates to the problematic issues of interconnection and access, and cross-subsidisation. A case study in access regulation of railway infrastructure by **Dr Stewart Joy** aptly comes after the concepts developed in the above. **Dr Joshua Gans** and **Philip Williams** analyse the usefulness of appropriate access pricing rules as a method of incentive regulation given their influence on the timing of infrastructure investment.

Financial techniques are dealt with in Section Four. **Professor Kevin Davis** and **John Handley** report on issues associated with the measurement of the cost of capital. The final section covers other regulatory issues. **Professor Peter Forsyth** examines some non-price issues of regulation. **ACCC General Counsel Luke Woodward** ends the volume with a paper delving into the important matter of information disclosure.

In his paper, **'Competition Reform and Infrastructure', Samuel** surveys the recent developments in the policy-driven expansion of competition throughout the Australian economy, and specifically, its impact on the utility infrastructure industries. In reading the paper, one appreciates the extent to which competitive forces are being relied upon to generate efficiency and hence better economic performance, reflected in the Competition Principles Agreement initiatives like anti-competitive provisions extension and competitive neutrality. The paper focuses primarily on one of the most significant developments of competition policy for infrastructure — access to essential facilities. This competition initiative recognises that most utility industries have a functional component that has production characterised by natural monopoly technology, such as 'networks' in electricity, gas, rail, etc. The regime understands that in these cases, the facility would be 'uneconomic to duplicate,' an approach exemplifying the philosophy underlying the policy, and enunciated in the legislation, being that competition is merely a *means* to better performance, not an end *per se*. Samuel details the various routes that can be taken to arrive at a fair access solution, an aspect of the policy predicated on the goal of relying on market interaction and minimising regulatory interference and command. However, pre-empting the later discussion incentive regulation, Samuel stresses that imposition of access regulation should be subject to a benefit/cost analysis, as the gains in competitiveness may be offset by regulatory failure and costs, such as investment incentive distortions.

The paper successfully conveys competition policy's impressive endeavours to be market-driven and minimalist. Samuel's concise survey of past access activity clearly indicates this objective, in particular that Part IIIA was constructed vis-a-vis its criteria to ideally capture only those facilities considered to be problematic and worthy of review, i.e. essential facilities of national significance that are uneconomic to duplicate. The Competition Principles Agreement platform of structural reform of government business enterprises facilitates and complements the access drive, as greater 'workable competition' in the vertically separated components of the enterprise requires access to be provided to the remaining monopoly facility. He also outlines the progress on reaching another policy goal, that of regulatory convergence and national markets, and the beneficial effects thereof such as greater certainty, consistency and reduced likelihood of regulatory 'capture'.

Smith delves more into the conceptual framework underlying competitive reform in her paper **'Australian Competition Law and Policy — Theoretical Underpinnings'**. Smith gives perspective and conceptual justification to the new broader wave of competition policy, and contrasts it with the former, more restrictive policy of prohibiting anti-competitive conduct. Economic theory recommends vigorous competition, as it generally best promotes static, (both allocative and productive), and dynamic efficiency. However, the regime shows its alacrity by its allowance for authorisation of certain anti-competitive conduct if it is deemed to be offset by public benefits. This flexibility derives from an embracement of the theory of market failure, whereby competition may not be sufficient to promote good outcomes. Smith illuminates the range of market failures which beset the utility industries, such as economies of scale/scope, and externalities, thus underlining the importance of an 'efficiency defence' via authorisation.

The discussion of the Part IIIA access provisions expands upon the treatment in Samuel. To further validate the strong case for an access law, it is contrasted with the alternative approach to the access problem, which is relying on the s. 46 mis-use of market power provision. However, this alternative approach suffers from various deficiencies, primarily the inability to deal directly with monopoly pricing which is not for a 'proscribed purpose' and associated evidentiary problems.

One other important aspect of the access provisions is the stipulation that competition is to be promoted in a 'separate' market from the facility. This implies the need for an analysis of market definition. The salient point here is that markets are multi-dimensional, and thus, even though the product market is not clearly delineated in utility industries, it is the functional dimension which confers the status of 'distinct' markets. Once again however, regulation should be applied with circumspection, given its problems such as administration, 'capture', and inconsistency. The ACCC's approach to regulation attempts to redress these inherent regulatory concerns. In sum, the case for competition is strengthened, accompanied by the realisation that market failure, as seen in utilities, tempers the axiomatic reliance on competitive forces.

The approach to utility regulation in New Zealand, detailed by **Bollard**, in his paper **'Utility Regulation in New Zealand'**, stands in interesting contrast to that of Australia, as elucidated by Samuel and Smith. Nonetheless, the similarities in history and approach are quite extensive. Public ownership has historically been the 'remedy' to the acute problems of operating public utilities. Bollard attributes the poor performance of utilities to their uncompetitive, uneconomic operation, a view which gave impetus to reforms that mirror the Australian response, such as de-monopolisation, vertical separation, and competitive neutrality.

The paper is excellent not just as reference on the NZ approach, but its concise survey of the foundation of the case for, and problems of, utility regulation. The intrinsic features of utilities are discussed, from economies of scale to high barriers to entry. Some of these problems were accentuated in NZ on account of its small size and correspondingly small, fledgling capital market. But most interestingly, they have diverged from Australia on their approach to access, consistent with their 'light-handed' approach to regulation. There is a reliance on enforcing the relevant (s. 36) mis-use of market power provision, an approach eschewed in Australia in favour of specific access legislation, for the reasons outlined in Smith. As a result, it is evident

that information disclosure plays a correspondingly more significant role in the promotion of competition. However, broadly, a similar philosophy on regulation and competition emerges to that of Australia, with the non-hindrance of competition and economisation of regulation as central. It will be interesting to observe the respective fortunes of the two differing approaches to access — will NZ embrace specific access laws, or will Australia revert to general competition law?

In the beginning of his paper, **'Introduction to the Fundamentals of Incentive Regulation', Berg** makes the fairly obvious, but very poignant point in stating that incentive regulation is about creating incentives for **good** conduct for utilities. Of course regulation of any sort will give rise to incentives, but the challenge is to design and implement a system of rewards and penalties to induce, rather than command, the utility to adopt efficient pricing and patterns of investment. The need for incentive regulation comes about from the inherent informational asymmetries that exist between utility and regulator. As such, the problems of utility regulation form a subset of the set of principal/agent problems. This paper enables us to understand that the *raison d'etre* of incentive regulation is largely an attempt to overcome the distortionary effects which characterise principal/agency scenarios.

The failure of cost-based, or rate-of-return (ROR) regulation can be viewed using this framework. The regulator cannot monitor the potential for the enterprise to engage in opportunistic or slack behaviour. Thus, it is not surprising that regimes of ROR regulation have been characterised by relatively high cost structures and under/overinvestment, due to the guaranteed-return (proportional to cost) nature of the regulation. Berg outlines the alternative regulatory options, such as price-capping and yardstick regulation, which are all essentially methods to help overcome the deficiencies of ROR regulation. The main message here is that, within the general Principal Agency framework, the challenge for utility regulators is to create appropriate 'sharing rules', so that utilities find it in their incentive to act in accordance with the objectives of good regulation. Furthermore, successful regulation will depend largely upon the degree of incorporation of relevant fundamental economic principles in the design of an incentive mechanism for each particular case.

King expounds on the in-principle advantages of price capping over traditional cost-based regulation as a method of incentive regulation in his paper **'Principles of price cap regulation'**. Upon establishing this thesis, the paper focuses on the decisive importance of correct design and implementation in actually capturing these intrinsic benefits. The system not only enables consumers to share in the benefit of productivity, or 'X', improvements, but more significantly, has the power to ameliorate the inherent problem of ROR — incentive to over-invest, in order to take advantage of an overly high allowable rate of return. Price capping, conversely, actually rewards costs-saving, i.e. productive efficiency, since firms want to harness the opportunity to exceed the expected 'X' factor.

King illuminates the delicate balance of incentives which face firms under various price-capping systems, which gives rise to the practical considerations in design and implementation. A key design issue is the setting and re-setting of price caps at review. It is emphasised that information used at review should be beyond the control of the firm. The system would effectively revert to rate-of-return regulation and its associated deficiencies if price caps were based on the firm's realised profit performance. The

regulator must also decide carefully how input prices are to be passed through to consumers, as a lenient approach may blunt the incentives to use the cheapest inputs. Similarly, a harsh or opportunistic excess profit clawback system will soften efficiency incentives. The prime implementation issue concerns regulatory credibility between review periods. The regulator must resist the temptation to raise X or the catch-up factor in the case of a firm exceeding its expected productivity performance. This again will give rise to incentive payoffs which mirror those prevailing under ROR regulation.

The particular design of a price-capping system is fundamentally about an assessment of the relative importance of productive against allocative efficiency, and of the trade-off between cost-cutting incentives and consumer pass-through. Although prices above Marginal Cost (MC) cause a loss of allocative efficiency, the pursuit of its rectification must be cautious due to potential dis-incentives for cost-cutting. The design of a good system is even more complicated when one accounts for multiple-products, an issue also covered in the paper.

As will be discussed below in the case study of UK water, a regulatory tool often used in combination with price capping is performance monitoring. However, as detailed by **Lawrence** in his paper **'Benchmarking Infrastructure Enterprises'**, efficiency measurements constitute a useful method of incentive regulation in their own right. By subjecting enterprises to performance monitoring and subsequent comparison with similar firms, there is a hitherto absent element of competition being promoted, known as yardstick competition. The pressing need for developing specific performance measures for comparison derives from the principal/agency relationship that exists between regulator and firm. In a full information world, a regulator could merely stipulate exact performance goals for each enterprise, along with corresponding rewards and penalties. But, we live in a world of asymmetric information, thus presenting to the agents the opportunity for manipulation of business information. Benchmarking enterprises provides a way of ascertaining the relative performance and identifying abuses of monopoly power and potential room for improvement.

Of the many forms of comparison used, Lawrence argues that external peer comparison provides the best basis, as it can encourage the achievement of world-best-practice. Of the three broad categories of performance measures — accounting, non-financial, and economic — the paper advocates the use of the latter. The paper endorses two types of economic indicators. Total Factor Productivity is the very objective we are trying to maximise, and addresses the incentive problems of using partial productivity measures, which can be prone to giving rise to the incentive to improve performance of the observed unit at the expense of the efficiency of another. The Economic Rate of Return reflects cost recovery based on market values, and hence whether the investment is providing an adequate return to capital. In essence, the point of performance measurement is not merely to provide a 'report card', but further, to actively promote the attainment of higher efficiency and best practice. Therefore, there must be substantial management incentive, by linking rewards to performance.

Booker's paper, **'Incentive Regulation in Water'** is a case study on the implementation of price capping of water services in the U.K. This insight into a real-life process contains valuable lessons for utility regulators. The practical experience of incentive regulation methods greatly reflects what is recommended or can be predicted

using the economic concepts developed in the preceding papers. Regulation in the UK is in the form of setting medium-term (5 years) price caps, a benefit-sharing and cost-cutting incentive mechanism providing a stable investment environment. To ensure real productivity gains, quality service monitoring was introduced. Also included is an element of 'surrogate competition' via comparative performance measurement.

There is much to be learned by regulators from observation of the actual regulatory learning process experienced by water services regulators in the UK. The first periodic review expressly incorporated a quality term in the price capping formula, due to concerns which arose over the first price cap period. The initial caps were allowing substantial excess profits, which were retrieved using a glide path. The importance of an open, transparent and consultative regulatory process was also stressed. The second periodic review will take even more notice of service quality and place more importance on environmental concerns. The review will also address the incentive problems of the clawback system, which encourages companies to delay investment and operating costs at the initial stage of a price cap in order to enjoy out-performance for long periods. In response, it is proposed that companies must actually deliver quality improvements before they benefit from new price limits. In general, the paper places its emphasis on the significant role of the actual process, not just the particular methodology, of price cap implementation. Most importantly, the paper signals a warning to the dangers of over-development of the system and resultant instability.

Berg provides the necessary conceptual underpinnings for any meaningful discussion of the critical subject of pricing principles in his paper, **'Basics of Rate Design: Pricing Principles and Self-Selecting Two-Part Tariffs'.** The essential issue is determining the appropriate method of dealing with pricing in the presence of high fixed costs and relatively low incremental costs, as is the norm in network industries. As a result, efficient pricing must involve an efficient allocation of fixed joint costs throughout their price structure. This paper finds its poignancy in the current environment of greater competition, as the distinction is drawn with monopoly industries which can allocate costs at their discretion. The basic point of the paper is that greater efficiency through optimal pricing can be achieved when one takes an economic, forward-looking view of marginal opportunity costs, rather than arbitrarily use retrospective cost allocation manuals.

Recent years have seen a significant transformation of the infrastructure sectors worldwide. Vertically-integrated public monopolies have become exposed to greater competition through vertical separation, access, and deregulation. The challenge for managers and regulators is to devise rate structures that can efficiently address the recovery of joint costs in a competitive environment. The central problem is how to adopt appropriate pricing schedules and cost recovery when some goods are sold in competitive markets while other goods have no close substitutes. In light of the widespread failings of various approaches, Berg analyses the efficiency advantages of two pricing structures: Two-part pricing and Ramsey pricing.

Price re-balancing that involves a zero-sum game is going to be met with conflict. The feature of multi-part pricing is the opportunity to create win-win options for both firm and customer. Two-part tariffs allow the firm to charge an upfront fee for the service which can cover fixed costs, but the price charged for each unit is exactly marginal cost. This enables the efficient amount of the good to be produced and purchased, as

marginal revenue equals marginal cost. This promotes the twin objectives of optimal pricing — efficient allocation, whilst providing greater facilitation of cost-recovery.

Berg notes that Ramsey Pricing is also a significant contribution of economic theory to the field of utility regulation. The most important feature of this system is that marginal costs play an integral role in design, implying high status being accorded efficiency. The paper explains that taking account of customer groups with differing elasticities opens up a win-win situation. Overall, the paper eruditely conveys the implications for pricing methods of greater competition in the network industries. Whereas in the past particular prices were irrelevant as long as total revenues covered total costs, the prevailing trends now point to the crucial importance of firms identifying incremental costs, and pricing accordingly, taking due account of potential win-win pricing schemes.

Berg's paper, and Jamison's to follow, make significant contributions primarily through their enunciation of rudimentary pricing principles that should be adhered to in the construction of access prices. **Ergas** provides a telling complement to them by delving more closely into the costing and valuation of the infrastructure, which actually comprise the basis of efficient price structures. The paper, entitled, **'Valuation and costing issues in access prices, with specific applications to telecommunications'**, essentially amounts to an injection of economic principles into the traditionally accounting practices of asset valuation, income definition, common cost allocation, and cost of capital determination.

The challenge for regulation, which is addressed here, is how to set access prices to recoup the costs of efficient operation, i.e. those which would prevail in contestable markets. This must be achieved in the presence of asymmetric information, whilst balancing allocative with productive and dynamic efficiency. Ergas adopts a forward-looking economic approach to income flows, so as to maintain the capital base, which has implications for the method of depreciation. The crucial point here is that the system adopted should result in a consistency between the balance sheet and the profit and loss statement given an economic view of assets. The paper advocates the superiority of the replacement cost asset valuation method in this regard over other approaches such as Historical cost, where the income flows would rarely coincide with economic income.

The approach to asset valuation and income flows forms only part of the process of determining access prices. The treatment of joint and common costs is becoming increasingly important due to the proliferation of economies of scope and the presence of CSOs. The recovery of these costs, in the likely absence of direct subsidies, must be made through the access charge. It is important that the regulators allow a mark-up over Total Service Long Run Incremental Cost (TSLRIC) so as to provide adequate incentive for the firm to pursue investment in scale and scope economies. Ergas recommends a multi-part tariff to achieve this goal, as it more accurately reflects cost-causation than does a uniform price structure.

Cost of capital determination is a major issue in capital-intensive industries like telecommunications. The Capital Asset Pricing Model (CAPM) is the traditional model in use, however, according to Ergas, it suffers from serious conceptual flaws which are borne out econometrically. The critique centres on the model's mis-

specification of the major determinant of cost of capital, especially when applied to 'sunk-cost' cases. Under this model, the irreversibility of the investment is not fully recognised in its risk assessment, and therefore its estimated required rate-of-return. Ergas highlights industry trends that may exacerbate this risk understatement.

The recent wave of de-regulation and privatisation of telecommunications markets world-wide has raised new complex regulatory issues. **Jamison** incorporates the pricing and costing principles developed in the preceding papers in analysing this more competitive environment in his paper, **'Regulatory techniques for addressing interconnection, access and cross-subsidy in telecommunications'**. The paper focuses on interconnection and access prices and details their direct effect on profitability and the development of competition. High prices will tend to favour incumbents and inhibit downstream competition, whilst low access prices will encourage entry of new competitors downstream to the network.

Competition reform has prescribed the establishment of rights of access to 'essential' infrastructure industries. Accordingly, in the setting of prices or pricing guidelines, whether in assessing providers' access undertakings or arbitrating commercial access disputes, regulators generally have three basic approaches at their disposal. The Efficient Components Pricing Rule (ECPR) essentially involves setting prices such that incumbent facility owners cover their opportunity costs of providing access, including consequential competition costs. Although this approach implies a preservation somewhat of a monopoly position, it may, under certain conditions, give rise to greater efficiency incentives downstream as entry is only attractive if one can price below the incumbent's downstream firm's incremental costs. Jamison explains however, that the adoption of ECPR, a rare regulatory occurrence, would lead to an inefficient structure given today's telecommunications markets.

The second approach used by regulators is cost-based prices. The Fully Distributed Cost (FDC) method is the common accounting approach used. Jamison's discussion of cost issues sheds light on the confusion surrounding various cost concepts. However, the FDC method applied to interconnection suffers from various deficiencies which derive from the resultant divergences between accounting and economic costs. Given the misleading signals this sends to industry participants and regulators, it is not surprising that the economic -cost approaches, using the more relevant TSLRIC, are the most popular. This approach involves an application of the marginal cost pricing principles advocated by Berg. More specifically, the method prices services at marginal cost, and raises the fixed costs by a fixed contribution. Demand-based prices comprise the third approach to interconnection prices.

The second major focus of the paper is the appropriate methods of dealing with the problematic issues of cross-subsidisation. Abstracting from equity issues, cross-subsidies are considered undesirable due to their anti-competitive nature and source of inefficiencies. However, the challenge for regulators is their detection and remedy, a view of which will depend, in turn, on the particular approach adopted. In doing this, however, the regulator must consider the level of complementarity between interconnection prices techniques adopted and the approach to cross-subsidisation remedies.

Joy applies the principles of pricing, costing, incentive regulation, and interconnection and access pricing in the analysis of rail infrastructure in his paper, **'Regulating access to railway infrastructure'.** The paper surveys the common problems usually experienced in operating rail facilities, and methods to help overcome them. Joy points out that, in principle, vertical integration can probably best achieve optimality in the management of railway capacity. However, as is common with this industry structure, the commercial interests of the major users often interfere with economic concerns of efficiency, resulting in conduct which serves to preserve monopoly interests downstream to the facility. The need for vertical separation turns on the value of economies from vertical integration as against the allocative efficiencies from more vigorous downstream competition. Joy deals with the regulatory approaches to the resolution of the trade-off, in the rail context, between inducing fair, non-prohibitive interconnection to users, and ensuring sufficient revenues to encourage investment in rail infrastructure for providers.

Optimal access pricing regulation is essentially about dealing with the difficulties of cost recovery, and the associated delineation between joint and stranded costs. Users will claim that landlords are covering unrelated costs in their access prices, whilst providers argue the importance of adequate cost recovery for track maintenance. The overriding principle in cost recovery is the need to identify and price based on efficient costs, rather than existing or past costs. The lower bound is of course marginal cost, whilst the upper bound should be stand-alone cost, which refers to the hypothetical cost of serving just the subject's access after rationalising the route's excess capacity. This implies that the onus of managing excess capacity should generally be borne by the facility manager. Costing should account for the presence of adverse incentives, especially in regard to ensuring that present worth, and future economic costs, and not past inefficiencies, are rewarded. The discussion covers familiar pricing principles, in addition to their deficiencies. In sum, cost-minimisation for both owner and provider should be aspired to when designing access regulation, with a view to the ultimate benefit to end-users.

Gans and **Williams** provide a cogent justification of the need and desirability of infrastructure regulation in this environment of market reform in thier paper, **'A primer on access regulation and investment'.** The paper establishes the inextricable relationship between the regulation and investment of infrastructure industries and espouses the potential social benefits of a greater recognition of this nexus by regulators. More specifically, access prices exert an influence on investment incentives by directly affecting the rate-of-return on the provider's investment. A pre-commitment to a regime of access prices by regulators will, accordingly, indirectly impact on investment incentives. The paper's significant contribution is thus the attention it draws to the opportunity for regulators to alleviate the inherent uncertainty and inappropriate signals which prevail under a system of no regulation. Regulation, by contrast, is shown to have the power to vastly improve social welfare insofar as it injects greater certainty and recognition of investment costs into the bargaining environment between access providers and seekers.

As recommended using the pricing principles learnt in the other papers, and assuming full information, efficiency goals entail the use of optimal two-part tariffs. Full information would disclose marginal cost, which would leave to negotiation the fixed access charge which covers all/some of the capital investment costs. The paper thus

illuminates the potential for regulation to bring the sunk costs component into the bargaining situation via this access charge. Previously, under no regulation, capital costs did not directly impinge on decisions post-investment by virtue of them being 'sunk'. Access charges were based on arbitrary notions of income redistribution to the provider, without sufficient regard to efficiency.

In the absence of regulation, providers must balance the conflicting interests of utilising scale economies and limiting downstream competition, whilst recognising that too high an access charge may encourage 'uneconomic' duplication and too low may lead to under-investment. These delicately poised relationships can be largely ameliorated by the increased certainty that regulators can bring by establishing an upfront access regime which can best promote the dynamic aims of the relevant competition legislation and philosophy. In the case where access seekers operate in non-rivalrous markets downstream to the facility, wherein the provider has limited incentive to exploit monopoly power, the benefits from certainty will still be manifested in an optimisation of investment and entry decisions by all market participants.

Gans and Williams separately discuss the case of rivalry in downstream markets, which is more complex, as it involves reconciling the wishes of providers who have an incentive to sub-optimise the access to their facility. Here again, regulation can be shown to provide incentives for efficient and timely investment and competition. Gans and Williams conclude that their analysis suggests the merit in regulators delivering greater clarity in their access pricing guidelines.

The papers thus far have concentrated on the need to incorporate fundamental economic principles in the formulation of various types of infrastructure regulation. The paper, **'The cost of capital and access arrangements'** by **Davis** and **Handley** provides the volume with a discussion of the issues involved in the actual implementation of these economics-based regulatory recommendations. The paper looks at the financial techniques available to regulators in reviewing access prices. In general, access prices are constructed so as to provide a return on the provider's investment. The required rate of return for a company is the market-determined rate of return required by investors to provide capital to the company. In the accounting context, it is referred to as the cost of capital. This paper delves into the issues associated with determining an appropriate measure of the cost of capital which is required to compensate investment costs, whilst also recognising the level of risk involved.

The classic Capital Asset Pricing Model (CAPM) assumes that systematic, or Beta, risk is the only risk remaining in a portfolio, which is rationally compiled to abstract away any unsystematic risk through diversification. In this model, the Beta for a company is a relative risk measure, encapsulating the systematic risk differential between the company and the rest of the market. Beta will be higher, *inter alia*, the higher the company's exposure to debt, the higher the level of fixed costs, and the more sensitive is the company to the vagaries of the business cycle.

The paper covers other issues relevant to cost of capital for utilities, such as the impact of multi-business companies and price regulation on risk estimation. Davis and Handley highlight the issue of the difficulty in assessing risk when the company competes downstream. They also examine the implications for cost of capital

assessment of institutional changes such as dividend imputation and investment income tax reform. The paper, furthermore, begins with an outline of the pro and cons of alternatives to the CAPM.

Forsyth recruits the principles of incentive regulation and access pricing in his analysis of approaches to dealing with the non-price issues of **'Environmental externalities, congestion, and quality under regulation'.** The paper discusses the nature of these issues, and, importantly for infrastructure regulators, their exacerbation under a system of price-capping, now the dominant form of regulation. The analysis involves a recognition of the various adverse incentives which face firms under this system, and the incentive rectification approaches available. The underlying difficulty for regulation in the case of quality is the breaking of the nexus between price and quality. In contrast to the unconstrained monopoly, who enjoys the ability of a price rise in response to increased demands, there is insufficient extra revenue compared with the costs from improving quality, due to the price cap. Congestion, determined by the relationship of highly variable demand to a relatively fixed capacity, can be considered as one specific aspect of quality. Here again we have the problem that the firm has no incentive to reduce congestion, as the benefits are largely appropriated by the users, whilst it bears the cost of investment.

Utilities under access price regulation face similar adverse incentives for supplying low quality services to its users. However, aside from the cost-savings, there is the potentially more adverse incentive of damaging the effectiveness of their downstream competitors. This conduct can restore their full monopoly status, which effectively amounts to circumvention by the provider of the access regime. To combat these incentives, the regulator must then implement appropriate regulation to achieve its quality aims, one example being the incorporation of a quality term in the price cap formula as discussed by Booker.

Externalities pose another challenge for the regulators of price-capped firms. The techniques available for internalisation of the externality, whereby the firm will be induced or forced to produce a socially optimal output level, all have their applicability given certain scenarios. The challenge for regulators is how to incorporate the chosen approach into the price cap. Again, the essential message is to be wary of resultant incentive problems. The paper also stresses the importance of regulatory convergence between economic and environmental regulatory bodies and objectives.

The importance and benefits of incorporating economic principles into the practice of regulation has been the central theme running through the papers thus far. **Woodward** examines regulation from a different, but equally important perspective in his paper**, 'Information Disclosure in the Australian Regulatory Environment: Legal and Good Practice Aspects,'** which recognises the critical role of information in decision-making. In carrying out their decisions, regulators, as do all decision-makers, require access to information upon which to base those decisions. The paper addresses five questions relating to this information which should be asked by regulators within any given regulatory framework. Underlining the issues is the overriding guiding principle of efficient regulatory implementation — a benefit/cost assessment. In the context of information disclosure, there will be a trade-off between the compliance and administrative costs of obtaining the information and its value.

The paper describes the basic circumstances and relationship between firm and regulator in the information disclosure context. The principal/agent scenario facing regulators imposes upon the regulator the challenge of extracting the right amount of information, with the view that it will usually be incomplete and possibly distorted to serve self-interests. This implies the potential unreliability of the information provided, which leads to the need to *test* the information. The standard of proof required by a test will determine the extent to which the information is tested, although considerations of confidentiality will often impinge on the scope of the testing. In any case, the degree of importance attributed to effective information disclosure is arguably lessened the more effective are the incentive regulation mechanisms in place. Woodward concedes these questions may be obvious to regulators, but their critical importance should warrant greater attention than is being accorded them in practice.

One can observe several common themes interlinking the set of papers. The predominant message to emerge from the papers is the continued elevation of efficiency's status as the central conceptual driving force of infrastructure regulation as we enter new phases of market reform. Greater efficiency can be attained, as the papers concur, by incorporating greater economic analysis into the development of regulation. Market reforms as outlined in Section 1, notably vertical separation and access rights, have heightened the need to develop regulatory techniques that are capable of addressing and complementing these reforms. The papers are bound by their recognition of the inherent limitations of placing a heavy reliance on direct, command-style regulation due to the prominence of information asymmetries. The challenge for regulators is then to design and implement regulation that reshapes prevailing incentive configurations in order to **induce** good, market-oriented performance from firms. In this regard, the papers have also emphasised the importance of an economic approach to costing and pricing issues, and more particularly the merit in devising systems based on marginal cost principles, whilst ensuring a fair rate of return to capital. The themes of minimising regulatory capture and opportunism, and of maximising consistency, certainty, and convergence, whilst retaining flexibility have also continually run through the papers.

Section I. Competition law and policy

Competition reform and infrastructure

Graeme Samuel, National Competition Council, Melbourne, Australia.

Infrastructure sectors such as energy supply, transportation, communications and water supply play a pivotal role in the Australian economy. They generate major business inputs, representing between 7 and 16 per cent of production costs for most Australian industries[1] and also provide essential services to the community.

Any inefficiencies in infrastructure provision directly impact on Australia's growth, competitiveness, and ultimately on living standards. Bringing the cost and efficiency of Australia's infrastructure services at least into line with world best practice is therefore a central focus of micro-economic reform.

The critical role of the utility sectors was recognised by all nine Australian governments in 1995 when they committed to the National Competition Policy (NCP) package, which contained an array of measures to extend competition to previously sheltered parts of the economy.

When adopting the package, governments also established the National Competition Council (NCC). One of the NCC's roles is to assesses State and Territory progress in implementing competition reforms agreed by the Council of Australian Governments (COAG). The NCC must deliver a report to the Federal Treasurer in June of 1997, 1999, and 2001, on progress by each State. The NCC's advice will form the basis of the Treasurer's decision to allocate three tranches of competition payments to each jurisdiction from these dates.

The NCC also administers certain aspects of the reforms, advises on areas where more work is needed, and provides public information on the NCP process generally.

Many of the reforms have direct significance for infrastructure services. In addition, the NCP package incorporated pre-existing reform packages covering four industries — gas, electricity, water and road transport.

What I plan to do today is to address the key NCP reforms of relevance to the infrastructure sectors, the key policy drivers underlying these initiatives and our expectations of the likely benefits.

I will also consider the appropriate role of regulatory processes. The effectiveness of these processes will have a major bearing on the benefits to be reaped by the community from NCP reform.

I will begin with an area central to the NCC's work program — access.

---

[1] Bureau Industry Economics, 1995.

Access

Part IIIA was inserted into the Trade Practices Act (TPA) in 1995 to establish a legal right for businesses to negotiate access to services provided by key infrastructure facilities, with a right to binding arbitration if negotiations failed.

The aim is to promote competition in markets upstream and downstream from a natural monopoly service by allowing firms to use that service on fair and reasonable terms. Access can encourage entry and greater competition in related markets, helping to ensure more efficient use of resources, better service quality and lower prices. For example, access to rail line services could increase the number of train operators and the volume of operations, promoting competition in the market for rail freight haulage and probably in the broader freight market which includes road, sea and air transport.

However, access regulation also has costs. In addition to the direct transaction costs to both governments and business of operating an access regime, arrangements which are poorly designed or operated may deter infrastructure investment and lead to an economic outcome which is less favourable than an unregulated market.

Therefore, getting the right balance of incentives for investment in infrastructure and in related markets is the important challenge for access regulation.

Part IIIA offers three mechanisms to gain access. These are triggered by different parties and follow different routes but their objective is the same — access to the services of key infrastructure on reasonable terms.

The first mechanism is through a certified State or Territory government access regime. The regime is developed by the State or Territory in which the infrastructure is located and requires a framework for the commercial negotiation of access backed by an enforcement process. The regime must also include an independent arbitration process guided by criteria in the Competition Principles Agreement (CPA). It is the NCC's role to assess a State access regime and make a recommendation to the Commonwealth Treasurer on its ability to meet the aims of Part IIIA. Once approved the regime is said to be 'certified as effective' and the service is not open to declaration.

The second mechanism is a supplier undertaking. This is triggered by the infrastructure owner developing a framework for access negotiation and submitting it to the Australian Competition and Consumer Commission (ACCC) for approval. If approval is given, the facility is once again immune from declaration.

If there is no effective State or Territory regime and no ACCC-approved undertaking, the infrastructure is open to the third mechanism — declaration. This is a mechanism triggered by an actual or potential user, who may apply to the NCC to have an infrastructure service 'declared' under Part IIIA. If the service satisfies the declaration criteria, the parties must enter into negotiation, or failing this, legally binding arbitration to determine the terms and conditions of access.

I would now like to offer some perspectives on each of the access mechanisms in which the NCC plays a significant role — certification and declaration.

Certification

Two successful certification applications have been made to the NCC to date — for the services of NSW gas distribution networks and Victorian shipping channels. The National Access Regime for gas pipelines will also follow the certification route.

The NCC's approach to certification has been to apply the CPA principles such that a narrow interpretation is avoided. We believe it is important to apply the tests in such a way that we undertake our statutory responsibilities, while retaining the flexibility to enable access regimes which conform to the policy objectives underlying the CPA to be certified as effective. It is expected that this position will soon be clarified through amendments to the TPA which affirm the CPA principles as having the status of guiding principles rather than binding rules.

While the CPA indicates a preference for the commercial negotiation approach of access (that is, without regulatory intervention), the NCC recognises that limiting commercial negotiation can sometimes promote better policy outcomes by constraining market power, reducing uncertainty and producing more 'workable' outcomes.

However, a regime that imposes considerable constraints on commerical negotiation may require checks and balances to protect the rights of affected parties — for example:

- independent and adequately resourced regulators; and
- appropriate appeals mechanisms.

Declaration

The other access mechanism in which the NCC plays an important role is declaration. Before recommending declaration, the NCC needs to be sure that six criteria specified in Part IIIA are met. The three most critical are:

- firstly, that access must promote competition in an upstream or downstream market. This requires evidence that imposition of the declaration process would result in reduced costs flowing through to a related market — and so act as a catalyst to increase in competition in that market;
- second, it must be uneconomic for another facility to provide that service. This ensures that only entrenched monopoly facilities are declared; and
- third, the facility must be of national significance, judged either by its size, its importance to trade, or its importance to the national economy.

These tests are designed to ensure that access regulation is imposed only when the net benefits are more than trivial.

Interest in declaration has been considerable. Applications have covered a range of services, including electronic payments systems, the use of facilities for offering cargo-related services at airports, and parts of rail networks. There has also been some variety of applicants — including a student union, and small and big businesses.

The NCC has made recommendations on six applications to date — four of which relate to rail services. The NCC recommended declaration in four instances. However, the relevant Minister decided for declaration only once — this was the Commonwealth Treasurer regarding an application by Australian Cargo Terminal Operators Pty Ltd (ACTO) for access to particular airport services at both Melbourne and Sydney international airports. The other decisions — each made by a State Premier — were against declaration. Each decision has been appealed to the Australian Competition Tribunal.

National access initiatives

Looking at access more broadly a welcome development has been the trend towards a single regime to apply nationally in particular industries — this has already occurred in telecommunications, electricity and gas. The benefits of this approach in any industry with a national or interstate market and associated cross-border trade are obvious — especially if a common approach to regulation can be agreed upon.

It is interesting to note that the national access regimes in gas, electricity and telecommunications have taken different paths. While gas is following the certification route via the NCC, access to electricity services has been implemented through the ACCC undertakings route, and to telecommunications services under a specific regime in Part XIC of the TPA. The differences in approach reflects differences in the structure of the industries and the way reform has evolved. However, there are positive signs of regulatory convergence, with the ACCC as the eventual national regulator for major network services — including all interstate services — in each of these industries.

In industries where a specific approach has not been adopted, access activity has fallen back to the generic Part IIIA declaration option. This might explain, for example, the extensive declaration activity in regard to rail services. Recent moves towards regenerating a national rail reform agenda should, if successful, alleviate pressures on Part IIIA here.

Access outcomes

I have talked about the expected benefits of access — but where is the evidence on the ground? While Part IIIA is a relatively recent addition to the TPA, the recent Independent Pricing and Regulatory Tribunal (IPART) determination on access to the AGL gas distribution network in NSW suggests that the potential benefits are substantial. The Undertaking — submitted under the NSW interim gas access regime — provides for average gas transportation tariffs in the contract market to fall by close to 60 per cent in real terms, from $2.26 per GJ in 1995–96 to $1.05 per GJ in 1999–2000. This will deliver major savings to more than 400 industrial and commercial gas users of 20–25 per cent in average delivered gas costs.

Similarly, freight rates for rail freight transport between Melbourne and Perth fell by around 40 per cent following the entry of new freight haulers, including SCT, on that

route in 1995–96.[2] SCT pointed to this drop in freight rates as evidence of potential benefits when it applied for declaration of track services from Sydney to Broken Hill, so that it could on-carry freight to Perth.

Concerns have been raised at various times about the likely implications of NCP reform — and access in particular — for investment. Yet the reform process has accompanied significant ongoing investment in infrastructure in Australia, and the sale prices of affected assets appear to have held up. For example, prospective investment in gas transmission pipelines currently totals upwards of $4.5 billion in anticipation of the national gas access arrangements.[3] And last year, the sale of Victorian electricity generation and distribution entities realised a total of around $18 billion, substantially exceeding expectations.[4]

This suggests that the prospect of access to previously locked-up markets may be supporting investment in infrastructure, and that the regime has not been causing undue uncertainty for infrastructure owners.

Overall, these developments testify to the benefits — both on the ground and in the pipeline — of greater competition and more efficient use of Australia's infrastructure.

While access is central to the NCP reforms affecting infrastructure services, there is much more to competition policy reform. Broadly, NCP is about three things:

- introducing or extending the application of general pro-competitive regulation;
- reforming anti-competitive regulation; and
- reforming public businesses, especially where they compete with the private sector.

I would like to speak briefly on a number of these other reform parameters, with a particular focus on their impact on infrastructure services.

Pro-competitive regulation

The NCP reform package has boosted the role of general pro-competitive regulation in two main areas: the first, the National Access Regime in Part IIIA of the Trade Practices Act, has already been discussed. The second has involved applying the restrictive trade regulation in Part IV of the TPA to all businesses.

---

2 Two companies, SCT and TNT, entered this market to compete against National Rail Corporation (NRC). As well as significant price reductions, improvements in transit time and service quality were recorded.

3 Department of Primary Industry and Energy, figures based on Australian Gas Association, 1997.

4 VDTF, 1997.

Extended reach of the Trade Practices Act

Looking at other reforms, jurisdictions agreed as part of the NCP package to extend the reach of the anti-competitive conduct laws in the TPA to State and local government businesses — which were previously exempt from the Act. This reform has direct implications for bodies still in the hands of State and local government authorities — for example, water and rail in many States.

Reforming anti-competitive regulation

The commitment by governments to review and, where appropriate, reform all laws which restrict competition by the year 2000 has important implications for utility industries. The removal of overly stringent and prescriptive regulations has the potential to promote competition in a number of sectors, with benefits to business and consumers. These benefits are likely to be of particular significance in national utility industries which are currently subject to regulatory frameworks which restrict interstate trade or are inconsistent across the States. For example, South Australia's Cooper Basin Ratification Act — previously identified by the ACCC as a significant barrier to interstate trade in gas — is currently under review, while a number of legislative constraints on gas trade in Victoria are expected to be repealed shortly.

The pace of undertaking the reviews has varied from jurisdiction to jurisdiction. For example, while the NSW and Victorian governments have scheduled a large number of important reviews early in the review period, other jurisdictions have a significant task ahead.

The NCC expects jurisdictions to apply independent reviews which are conducted in an open and rigorous manner. While there should be a bona fide eye towards reform, some reviews have recommended retaining anti-competitive provisions for public interest reasons. For example, a South Australian review found that while aspects of the Water Resources Act are restrictive, they generate net benefits by mitigating the risk of environmental degradation and disputes over water usage. It therefore recommended that they be retained.

GBE reform

The NCP reform package also included a number of specific measures for Government Business Enterprises. These included measures aimed to ensure:

- competitive neutrality between public and private businesses; and
- the appropriate structure for public monopolies.

Competitive neutrality

The competitive neutrality reforms have also commenced, although the scope and pace of implementation has again varied across the States.

In essence, competitive neutrality is the application to public businesses of the same taxes, incentives and regulations as those facing private businesses. Corporatisation and full cost pricing are two ways competitive neutrality can be introduced.

Competitive neutrality reform has significance for the utility sector where a private operator wishes to compete against a government-owned incumbent enjoying competitive advantages by virtue of its ownership. Reforms have now been applied — or are in the process of being applied — to most contestable government utility businesses through privatisation or corporatisation, taxation and pricing reform.

A vital element in competitive neutrality reform is a complaints mechanism to provide a forum for allegations of non-compliance. For example, views expressed in the media by Victorian electricity producers — that NSW electricity suppliers enjoy net competitive advantages by virtue of public ownership — could now be formally directed to the NSW complaints mechanism.

Structural reform of government monopolies

Governments have agreed to consider the structural reform of government monopoly businesses to facilitate competition. Two important aspects of structural reform are:

- Separating industry regulation from business activities. Vesting this dual role in a single body creates a potential conflict of interest between the commercial interests of an enterprise and wider public interests through the exercise of regulatory powers. As an example, technical regulation of the telecommunications industry was transferred from Telstra to an independent regulator, AUSTEL, in 1989.

- Separating natural monopoly elements from contestable business activities in vertically integrated industries. Where the natural monopoly elements of an industry are integrated with potentially competitive elements, there may be two kinds of competition policy concerns:

    - monopoly rents in the monopoly market may be used to cross-subsidise prices in the competitive end of the market, driving out potential competitors.

    - control over access to the monopoly element may be misused to stifle competition in contestable sectors. For example, a vertically integrated electricity business might use its control over the transmission grid to stifle competition in generation and distribution.

Considerable progress with structural reform has occurred in several utility sectors, including telecommunications, gas and electricity. This has contributed to new firms entering the contestable segments of these industries.

Industry reform packages

In conjunction with the reforms I have just outlined, the NCP package also incorporates pre-existing reform packages in the electricity, gas, water and road transport industries. These packages in part reiterate generic competition principles such as access, regulatory reform and structural reform — but also address industry specific matters. I will refer to the state of play in each industry shortly.

Projected outcomes

While the application of reform is more advanced in some sectors than others — for example, the reform process began earlier in telecommunications and airlines than in the energy sector — there is mounting evidence of the potential benefits which competition can bring.

I have already commented on evidence of substantial benefits arising from the introduction of third party access arrangements. To quote two more examples of the benefits of competition reform:

- real average airfares were around 22 per cent lower in September 1996 than their pre-deregulation level;[5] and
- prices of government trading enterprises fell on average by around 15 per cent in real terms over the five years to 1995–96.[6]

It is important to note, however, that while prices should generally decline, in some cases the NCP reforms could result in higher prices. This is because the introduction of competition in some industries involves removing subsidies. For example, part of the water reforms involves a shift to cost-reflective pricing. At present, many water users do not pay the full costs of storing, treating and transporting water. They therefore have incentives to use more water than is either environmentally or economically sound. Consequently, the application of reform is likely to result in higher prices for some users as subsidies are reduced.

This highlights that the fundamental aim of NCP is not to reduce prices, but to promote more efficient use of resources. Prices which are too low can be just as detrimental in terms of providing incentives for people to use resources efficiently as prices that are too high. Increased efficiency will in itself exert downward pressure on prices — but this may sometimes be offset for particular customer groups by the phasing out of subsidies.

I would now like to comment on competition reform more specifically with regard to some of the industries in which the NCC has played — or continues to play — a role.

---

5 ACCC, 1997.

6 Steering Committee National Perform ance Measurement Government Trading Enterprises, 1997.

Electricity

The NCC's role in electricity reform is to assess progress for the purpose of the competition payments. The NCC has identified electricity reform as a 'big ticket' item for the purpose of the assessments.

In June 1993, six governments — the Commonwealth, NSW, Victoria, Queensland, South Australia and the ACT — committed to undertake reforms necessary to allow a competitive electricity market to commence from July 1995 in sourthern and eastern Australia. They agreed to establish an interstate electricity transmission network involving those States already inter-connected, together with Queensland. They also agreed to separate the transmission elements of their existing electricity utilities from the generation elements, and turn them into stand-alone corporations.[7]

At the April 1995 COAG meeting, these reforms were extended and brought within the NCP process — with payments to the States and Territories depending partly on adequate progress in implementing the reforms.

Progress has been slower than expected. A competitive national market was originally envisaged to commence from July 1995, but an interim market — comprising NSW, ACT and Victoria — did not commence until May 1997, and full implementation has been rescheduled to the end of March 1998. Queensland recently commenced its interim electricity market, and will join once the National Electricity Market (NEM) commences, although full integration with the NEM will not occur until physical interconnection with NSW (around 2001). By 2001 all customers will also have the freedom to choose their electricity supplier.

Most jurisdictions have undertaken the required reforms necessary to create an effective competitive electricity market. Some jurisdictions, realising the benefits of competitive forces, have gone beyond the minimum requirements by splitting electricity generation and distribution/retail utilities into multiple independent and competing businesses. In addition, the Victorian Government has undertaken a comprehensive privatisation programme, selling all five distribution businesses, four generation utilities, and its transmission company.

As an indicator of the potential gains, a 1996 survey of Victorian electricity customers in the contestable part of the market found that 78 per cent of firms claimed they were better off, with close to one-third of firms reporting an improvement in retailer service.[8]

---

[7] At the time, South Australia stated that it would consider a subsidiary structure pending the resolution of cost issues associated with separating transmission from its vertically integrated authority. Resolution of those issues would enable the adoption of the model.

[8] Australian Chamber of Manufacturers, 1996.

Gas reform

On current trends, gas shortages are likely in eastern Australia within the next decade. Investment in new infrastructure will be needed to address this problem. While a number of new projects are currently being considered, many are awaiting much needed competition reform in the industry before receiving the green light to proceed.

To address these concerns, COAG agreed in February 1994 to remove impediments to free and fair trade in natural gas. One of the NCC's many roles in gas reform is to assess State and Territory progress in implementing these reforms for the purposes of the competition payments.

A central plank in the reform process has been the development of a national access regime for gas pipelines. This will allow persons to negotiate access to natural gas transmission and distribution pipeline services on reasonable terms and conditions approved by an independent regulator — with a right to binding arbitration to resolve disputes. This will promote competition between producers and retailers and fuel market development, creating the necessary conditions for investment. The construction of new infrastructure to link major producers with markets will set the state for a truly national competitive market for gas.

I am delighted to be able to report that the National Code was signed off by heads of government at the COAG meeting on 7November 1997. I would like to congratulate all parties for reaching this important juncture following a long and sometimes difficult process. What is most pleasing to me about this agreement is the remarkable commitment shown by all parties to the goals of gas reform, and the spirit of co-operation shared by all.

The stage is now set for legislative implementation in all States and Territories by 30 June 1998. South Australia — the lead legislator — has now passed its legislation, which is expected to be proclaimed once the relevant Commonwealth legislation is passed — expected in March-April 1998. Each jurisdiction will apply to the NCC in 1998 for certification of their access regimes established under the National Code.

The COAG gas agreements recognised two other essential ingredients to achieve free and fair trade:

- Firstly, there has been extensive structural reform of gas utilities since 1994. Contestable services such as gas retailing are being separated from gas transmission and distribution — normally regarded as natural monopoly activities[9] — through privatisation, corporatisation or ring fencing.
- Second, the removal of legislative barriers to trade provides the regulatory framework to open up markets. For example, reform of the exclusive franchise arrangements between Esso/BHP and Gascor are expected to be implemented shortly in Victoria as part of a legislation package to accompany privatisation.

---

[9] However, market growth and technological change may result in certain transmission pipelines becoming economically viable to duplicate — there are indications the Dampier-Bunbury pipeline may fall into this category in the future.

Water

Several factors have focussed attention on the need to improve efficiency of water delivery services. These include the high costs of developing new water supplies, environmental considerations and concerns as to the continued availability of quality water in some parts of the country. In addition, prices charged for water in most parts of Australia do not cover the costs of supply.

In February 1994, COAG agreed to develop a 'Strategic Framework' for the efficient and sustainable reform of the Australian water industry. It entails, among other things:

- pricing reform based on the principles of consumption-based pricing, full-cost recovery, and removal of cross-subsidies, with remaining subsidies made transparent;
- tradeable water allocations or entitlements, including allocations for the environment, to allow water to flow to higher value uses;
- structural separation of the roles of service provision from water resource management, standard setting and regulatory enforcement; and
- future investment being undertaken only after appraisal indicates economic viability and ecological sustainability.

In April 1995, governments agreed to bring the water reform agenda within the ambit of the NCP process. The pace of reform has been slow to date, with most effort focussed on formulating policies and considering technical matters in relation to the proposed reforms.

In particular, the intergovernmental task force[10] established to coordinate the reform program has developed generic national 'milestones' as advice to jurisdictions in implementing the Strategic Framework requirements.

The water reforms are important and extend beyond competition policy matters to embrace social policy issues such as recognising the environment as a legitimate user of water. If fully implemented, they could have a greater impact on community welfare (broadly defined) than any other single measure in the NCP agenda. The NCC will therefore give high priority in the second and third tranche assessments to the timely implementation of agreed water reforms.

Rail

I have already discussed the considerable activity in access declaration applications in rail. Without a comprehensive national rail reform agreement in place, the business community, in its attempts to obtain improved service quality and lower prices, has had to rely on the general provisions of Part IIIA.

---

[10] The Standing Committee on Agriculture and Resource Management (SCARM) Task Force, under the auspices of Agricultural and Resource Management Council of Australian and New Zealand (ARMCANZ).

While the performance of Australian rail freight services has traditionally lagged behind world best practice, two recent developments are encouraging:

- the proposed privatisation of NRC is likely to boost competition in rail freight services. The new owner is expected to compete aggressively in several interstate and intrastate markets, particularly with the state-owned rail freight companies.

- the National Rail Summit in September 1997 provided a first step towards co-ordinated reform when Transport Ministers recognised the need for urgent reform and reached agreement on parameters. This was followed by an Intergovernmental Agreement between Transport Ministers in November 1997 to establish the Australian Rail Track Corporation, which will aim to develop a more commercially viable system of interstate rail infrastructure and access arrangements.

While these are positive developments, it is likely that the declaration route will continue to drive rail reform until effective access regimes for rail services have been developed. The NCC notes that the November 1997 Agreement does not address intrastate rail services — which have also been the subject of considerable declaration activity.

While Part IIIA can help drive rail reform, this process is more difficult than if a national rail reform program was in place. There are at least two reasons for this:

- without a coordinated process, it becomes difficult to establish consistent access principles between the jurisdictions, and to establish efficient principles for 'first mover' jurisdictions;

- to significantly improve the quality and competitiveness of rail freight services in Australia, access reform must encompass a package of interrelated issues covering the efficiency of rail track and related infrastructure, safety accreditation, timepath, pricing principles and dispute resolution.

The electricity and gas reform agreements both dealt with these matters, and Australia is close to achieving a national market in these sectors.

But rail services are a long way behind, even though:

- existing rail infrastructure is more developed nationally than both electricity and gas infrastructure; and

- rail is publicly owned — a characteristic which made electricity reform simpler when compared to gas industry reform.

Airport and airline services reform

Partial deregulation of airline services in 1990 led to an immediate decrease in fares and an increase in flight frequency. The new era of contestability has provided a discipline on the incumbents, despite the failure of some new competitors to win market share.

Structural reform of airports has continued with the recent long term leasing of Melbourne, Brisbane and Perth Airports to private operators. It is the Government's intention that all airports will be operated privately under long term leases, with a further 15 airports to be privatised next year.

The *Airports Act 1996*, is part of the new regulatory regime which sees the ACCC responsible for administering a CPI-X price cap for aeronautical services at the major leased airports as well as prices oversight and performance monitoring. The Airports Act also provides for the automatic triggering of the access provisions of Part IIIA of the Trade Practices Act in respect of services provided by significant airport facilities.

The NCC has already had cause to consider access issues in respect of airport services related to international air freight at Sydney and Melbourne airports. The Commonwealth Treasurer accepted the NCC's recommendation earlier this year to declare services relating to the loading and unloading of freight from international aircraft at both Sydney and Melbourne airports. In making this decision the Treasurer agreed with the NCC that facilities such as hard stands and freight and passenger aprons must be co-located with other airport infrastructure such as runways, if they are to fulfil their functions. In order to duplicate these facilities, it would be necessary to duplicate the airports.

Road transport

In October 1992, Australian Transport Ministers agreed on a national approach to road transport reform. The aim was to improve transport efficiency, increase road safety and reduce the administrative and compliance costs of regulation. This was to be achieved through, among other things, uniform national arrangements for vehicle roadworthiness and driver licensing, and vehicle charges which reflect the full cost of providing road transport services.

In April 1995, the road transport reforms were brought within the ambit of the NCP process.

Progress has been slower than anticipated, with only one of the six reform modules being implemented to date. A revised timeframe and agenda was agreed by Transport Ministers in February 1997. All jurisdictions accpeted that this new agreement and timetable will be the basis on which the NCC assesses competition payments for road transport reform.

In November 1997, the Ministerial Council of Road Transport (MCRT) revised the timetable and delayed implementation dates for many of the reforms. The Council will consider the implications for competition payments early in 1998.

Conclusion

The Industry Commission predicted that some of the biggest gains in the application of the NCP agenda can be reaped in the area of infrastructure services — with electricity and gas alone set to yield more than 25 per cent of total forecast benefits.

While progress has been slower than anticipated, new reform timetables have been, or are expected to be, agreed in relation to the gas, electricity, water and road transport sectors. Significant reforms for all four sectors are scheduled to be implemented over the next two years.

Considerable progress has also been made in the reform of the telecommunications and air transport sectors. The rail sector remains a focus for considerable concern. While not covered by an NCP appended industry agreement, the extent of declaration activity in this sector gives us one indication of the value in developing a national approach to infrastructure reform. The current approach to rail access — and rail reform more generally — is difficult, time-consuming and continues to shelter arrangements which restrict competition.

In the broader context, a significant development in the reform agenda has been the shift towards independent regulators in the States and Territories — with generic rather than industry specific responsibilities. While IPART and the Office of the Regulator General (ORG) have been in place for some time in New South Wales and Victoria respectively, independent regulators are now being established in Queensland, South Australia and Western Australia. This welcome development has been partly driven by the gas reform process, a critical linchpin of which is the requirement for regulators to be independent of industry players and governments.

The next step towards efficient regulatory outcomes may be towards regulatory convergence, with a single national regulatory process for particular industries, and ultimately, across all infrastructure services. A common approach can promote certainty among industry players and confidence in regulatory processes generally, and encourages efficient patterns in infrastructure investment. However, there are two other drivers of crucial significance:

- Firstly, the emergence of national markets for many utility services once largely confined within particular jurisdictions — already this is a reality in gas and electricity. A common approach to regulation of interstate infrastructure is essential for effective competition to emerge.

- Second, the trend towards multi-utility bodies. For example, Telstra applied in August 1997 for retail electricity licenses in NSW and Victoria — and United Energy, the Victorian power utility, has been awarded a telecommunications carrier license. The next step may be multi-utilities selling packages of electricity, gas, water, telecommunications and other services. As the distinction between traditional utility services blurs, the need for a common approach to regulation becomes obvious.

The development of national regulatory regimes for gas, electricity, telecommunications and airport services are important steps in this direction. So too are the regulatory forums between national and State-based regulators now occurring on a regular basis, the appointment of the heads of State regulatory agencies in Victoria and NSW as associate Commissioners of the ACCC, and the establishment of a new Energy Division within the ACCC, with formal involvement of State regulators. The NCC would envisage that other State regulators would undertake similar roles once the appropriate regulatory agencies are established in these jurisdictions.

Ultimately, the NCC would argue for a single national regulator for all infrastructure services, and I believe that the jurisdictions are starting to think in the same direction. Already, there are moves in this direction, with the ACCC now acting as the regulator for airport issues, telecommunications and gas transmission services. It will also regulate electricity transmission services from 2001. As this process continues, the NCC would envisage that regulatory regimes will naturally converge towards the approach which is demonstrated as the most effective.

Bibliography

Australian Competition and Consumer Commission, 1997, 'Movements in Average Airfares: June and September 1996 quarters', *ACCC Journal* (forthcoming).

Australian Chamber of Manufactures, 1996, *Customer Feedback on Victoria's Competitive Electricity Market*, November.

Australian Gas Association, 1997, *Gas Supply and Demand Study 1997*, AGA, 1997.

Bureau of Industry Economics, 1995, *International Performance Indicators Overview*, AGPS, February.

Industry Commission, 1994, *Improving the Efficiency of GBEs*, May.

Standing Committee on National Performance Monitoring of Government Trading Enterprises, 1997, *Government Trading Enterprises: Performance Indicators 1991–92 to 1995–96*, Report, May.

Victorian Department of Treasury and Finance, 1997, *Victoria's Electricity Supply Industry: Towards 2000*, June.

Competition law and policy — theoretical underpinnings

Rhonda L. Smith, Australian Competition and Consumer Commission.

Introduction

Traditionally competition policy has been perceived fairly narrowly, with legal prohibitions on anti-competitive agreements, misuse of market power, and anti-competitive mergers. However, views have altered and competition policy is now perceived to have a wider role. Extension of competition law to government business activity and unincorporated businesses and its relationship to micro-economic reform will be integral to this discussion.

Reasons for competition policy

In most countries unrestricted competition is not a goal in itself. Competition generally promotes efficient allocation of resources and ultimately economic growth which benefits all participants in the economic process. However, the aim of competition policy is not exclusively related to efficiency. Competition should be thought of as a dynamic process of rivalry for sales between market participants and potential market participants, who invest capital in the production and development of goods and services. In addition, competition policy may encompass a broader set of policy objectives including consumer welfare, more equitable income distribution and encouragement of small business. However, there is a presumption in favour of competition unless it can be shown that efficiency or some other public policy goal overrides it.

Why is competition preferred?[1]

Central to competition policy is the assumption that there is something undesirable about an environment in which there is less competition compared with one in which there is more competition. Under normal circumstances, a competitive market structure will allocate resources in such a way as to produce the goods and services which consumers value most highly and are prepared to pay for, and it does so at the lowest possible cost in terms of resource use. Such a market is held to be efficient. In a highly competitive market, the individual firm is so small in terms of total market supply that it will have no impact on market price, irrespective of whether it chooses to produce a very large output or a very small output. Assuming that the firm sets out to maximise its profits, it will choose to produce the output which results in the lowest average cost of production. In this way, production efficiency is achieved. Similarly, as there are no

---

[1] For a more comprehensive coverage see: F.M. Scherer and David Ross, 1990, *Industrial Market Structure and* Economic *Reform*, Houghton Mifflin Company, Boston, Chapter 2 or Dennis W. Carlton and Jeffrey M. Perloff, 1994, *Modern Industrial Organisation*, Harper Collins College Publications, Chapters 4 & 5.

barriers to entry/exit to a competitive market, resources will move into/out of the market in response to price changes which reflect the value consumers place on these products, thereby also achieving allocative efficiency.

On the other hand, if there is only one seller in a market, the monopolist will restrict output below the competitive level in order to raise prices. This means that, for corresponding technology, higher average costs will be incurred and so the industry is technically inefficient. Neither does the monopolist achieve allocative efficiency as too few resources will be allocated to production. X-inefficiency is commonly associated with monopoly.

Thus, competition policy is based on the belief that a competitive market will result in economic efficiency and increased social welfare. Ideally competition policy would be non-interventionist and non-regulatory. It would leave market forces to operate. In practice, however, this is not possible. Not all markets are competitive and even those that are competitive initially may change as a result of market conduct.

A difficulty encountered in competition policy/law is the possibility that conduct may impair static efficiency in the short run but may increase dynamic efficiency in the long run. This is especially difficult in a legal environment where a judge wants to see evidence of the outcome claimed. Yet, as it is dynamic efficiency which drives growth, on public interest grounds, this would surely be a priority. It is an issue which may arise in relation to privatised and/or deregulating industries.

Conduct of concern from a competition policy perspective

It follows then that there is concern amongst policy makers about conduct which results in an increase in market power or conversely which results in a substantial lessening of competition. Market power may be defined as the ability to 'give less and charge more'. It refers to a situation where a firm (or group of firms acting jointly) has discretion in its decision making because it is free from constraints imposed by competition[2]. The type of conduct which may be of concern for this reason can be categorised as:

- contracts, arrangements and understandings between competitors (horizontal) — such as market sharing, price setting;
- misuse of existing market power — where firms already have significant market power, most governments would not seek to make such power illegal. Rather they would seek to curb its abuse (but not seek to prevent the maximisation of profits);
- exclusive supply arrangements (vertical restraints) and other vertical relationships (such as resale price maintenance) — the problem here is that these may in some circumstances appear to be efficiency enhancing but there are usually other ways of achieving the same results without the associated reduction in competition;

---

[2] Trade Practices Commission, 1990, *Misuse of Market Power*, Section 46 of the *Trade Practices Act 1974*, pp 16ff.

- mergers and acquisitions — these are different in that they are the only conduct which has the potential to directly alter market structure (although other conduct may change structure indirectly).

The economic analysis of competition issues

Obviously not all (or even most) conduct of this sort would raise the concerns of the competition authorities. A test is needed to identify conduct which does give rise to concern. The test generally applied is:

- does the conduct result in a substantial lessening of competition in the relevant market or markets; or

- in the case of a firm with substantial market power, is there a misuse of that power. This might be more usefully expressed in the form: is the firm behaving in a manner which would be unlikely if it operated in a competitive market?

Economic analysis of competition — structural or strategic[3]

The conceptual framework for the analysis of competition issues raised by the types of conduct referred to above, is derived from the structure-conduct-performance paradigm which had its origins in the writings of E. S. Mason in the 1930s. A causal relationship between the three elements is assumed, such that structure determines conduct and this in turn determines performance, although backward linkages are recognised. The relevant market is defined and then the conduct at issue is assessed according to whether the structural features of the market indicate that as a result of the conduct there will be an increase in the firm's market power and consequently a substantial lessening of competition. The structural features of the market are the basis on which this conclusion is drawn. Little or no attention is paid to the dynamic interplay of rivalrous interaction between current sellers, as well as between these incumbents and potential entrants. Dynamic non-price rivalry is given scant attention. Few conclusions can be drawn about competition simply by looking at the structural characteristics of markets.

Corporate behaviour has become much more sophisticated and strategic as oligopolistic structures have increased in importance in the economy. Under oligopolistic conditions firms have an incentive to alter their relative position in the market through strategic behaviour. Strategic behaviour is not inherently anti-competitive. Generally, it is pro-competitive.

To date competition authorities have had difficulty dealing with most forms of strategic behaviour perhaps because they have sought comparatively simple rules. The primary difficulty is that a particular piece of conduct is often viewed in isolation where it is really part of a package of actions (e.g. innovation and learning by doing) and because generally such conduct is open to various interpretations (e.g. a price cut may be predatory or it may

---

[3] Rhonda L. Smith and David K. Round; 'Strategic Behaviour and Dynamic Efficiency in Competition Analysis: A New Approach for Australia', 1997, Trade Practices Workshop, Glenelg; subsequently revised and published as A Strategic Behaviour Approach to Evaluating Competitive Conduct, 1998, *Agenda*, vol. 5, no. 1.

be intended to secure market share in order to achieve economies of scale or scope). It may be that the only way to assess strategic behaviour in relation to competition policy is to seek to identify whether it is likely to result in increased efficiency in the long run or whether it is likely to result in foreclosure of markets.

There can be little doubt that in most public utility businesses there is the potential for strategic behaviour. Of particular concern has been the role of strategic ownership and the ability to use this to influence prices. The solutions to date appear to be restrictions on cross ownership and ring fencing arrangements.

Market failure — authorisation[4]

The theory of market failure points to situations in which preserving or increasing competition may not be enough to promote good economic performance. An increase in competition will only result in an increase in efficiency in the absence of market failure i.e. anything which causes a divergence between the opportunity cost of the resources used in production and the value which consumers place on the good or service. A variety of factors can give rise to market failure including economies of scale and/or scope; externalities/spillover effects; public goods; transactions costs, including those associated with information asymmetries; and a non-competitive market structure.

Where economies of scale are such that unit costs continue to decline for all levels of output, market failure arises because of the non-existence of a competitive equilibrium. The consequence is 'natural monopoly'. Under such conditions, even if price were set equal to marginal cost to promote allocative efficiency, the monopolist would be operating at a loss. Under such conditions, price discrimination may be a second best solution. Natural monopoly is obviously a key characteristic of at least part of the supply chain of most public utilities.

Mergers may allow the merged firm to achieve economies of scale and/or of scope through pooling production and distribution facilities and these savings may encourage investment in research and development which reduce costs and represent dynamic efficiencies. However, to the extent that mergers reduce competition they may adversely affect at least short run static efficiency.

Where some costs or benefits associated with production or consumption are not paid for, i.e. where there are externalities, producers or consumers are able to free ride on others. Private costs and benefits diverge from social costs and benefits resulting in a misallocation of resources. An example would be where a firm invests in innovation only to have other firms copy the results without compensation to the original innovator. There are frequently externalities associated with traditional public sector activities — for example street lighting not only results in greater safety, it also increases personal security.

Transaction costs are another source of market failure. Market transactions involve costs of search, advertising and promotion, contract specification and monitoring. The transactions costs problem is essentially one of insufficient markets and/or monopoly

---

[4] For a fuller discussion see Australian Competition and Consumer Commission (ACCC), 1995, *Guide to authorisations and notifications*, AGPS, Canberra, November.

power over information. These problems tend to be greatest in situations of asset specificity, creating the possibility of opportunistic behaviour by purchasers, where transactions are frequent and/or where there is uncertainty. Asset specificity is again a feature of many essential facilities. Such transaction costs tend to encourage vertical integration or other vertical relationships. Where market power exists, vertical restraints or vertical integration may allow an extension of market power to another stage in the supply chain.

In the presence of market failure, although proposed conduct may result in a substantial lessening of competition, it may also give rise to various public benefits. These may include fostering business efficiency; industry rationalisation resulting in more efficient resource allocation and lower costs; increased employment or reduction or avoidance of increased unemployment; promotion of industry cost saving resulting in lower prices; promotion of competition; promotion of equitable dealings in the market; growth of exports; development of import replacement; economic development by encouraging R & D, mineral exploration and development etc; assistance to efficient small business which promotes competition; industrial harmony; improved quality and safety of goods and services and expansion of consumer choice; and supply of better information to consumers.[5]

Rather than lose these benefits a means must be found of determining whether they are sufficient to compensate for the loss of competition associated with the conduct. One means of dealing with this problem is to make increased efficiency a defence against a breach of the antitrust or trade practices provisions. Another is to authorise conduct which results in a net benefit to the community and thereby exempt it from the relevant provisions of the Trade Practices Act.

Changes to competition policy

The late 1980s and the early 1990s saw changes which had important implications for competition policy in Australia and especially for the treatment of public utilities. Difficult and increasing economic problems forced a reassessment of the performance of Australian businesses and also of government businesses. It was accepted that ways had to be found to make the former more internationally competitive. Both as part of this, but also as part of reducing the cost of government, it was also accepted that government business activity in many cases needed to be more market oriented. Solutions varied but governments generally embraced corporatisation, privatisation and deregulation.

Privatisation

A reassessment of why governments were operating particular businesses revealed opportunities for exit. Often in the past, identification of what was regarded as a natural monopoly component within an industry resulted in the government of the day concluding that it was in the best interests of the public if the whole operation was

---

5 Australian Competition and Consumer Commission, 1995, *Guide to authorisations and notifications*, p. 18.

government run. This view has been substantially reassessed as is evident from the vertical separation introduced into the Victorian electricity industry for example.

A number of competition issues may arise in the context of privatising government owned assets. There is a temptation to introduce anti-competitive arrangements for these businesses prior to privatisation to increase the sale value of the assets. An important concern from a competition policy perspective is that government monopolies do not simply become private sector monopolies. Section 4G is relevant in this context. To avoid this, it is preferable where possible to divide up the assets to maximise the number of market participants. For example, ports may be able to be sold as separate docks. The competition authority also needs to be alert to re-aggregation of assets via subsequent mergers.

Where competition cannot be injected through disaggregation a monopoly situation continues it may be appropriate to provide for:

- access arrangements;
- undertakings; and/or
- prices oversight.

Hilmer reforms

As part of government policy to encourage international competitiveness, tariff reform was commenced in 1988 with across the board tariff cuts. It was also accepted that vigorous competition was an effective means of increasing business efficiency and so the government established the Hilmer review of competition policy. The key recommendations of this review were incorporated into the Competition Policy Reform Act. This effectively:

- made changes to Part IV of the Trade Practices Act to the effect that its coverage was extended to all unincorporated businesses and the shield of the crown protection of government business activity was removed; and
- Part IIIA, an access provision to the services of infrastructure facilities of national significance, was introduced.

As a consequence of these changes, government businesses became subject to the provisions of the TPA in exactly the same way as private sector activity. This put them at risk for conduct such as agreements between competitors if such agreements were likely to substantially lessen competition. As a consequence, the electricity industry sought authorisation for its code of conduct; the gas industry is following a similar course.

In this context Fisse and Simpson, 1995,[6] point to the constraint imposed on operations such as (former) public utilities and telecommunications companies via legislative

[6] Brent Fisse and Andrew F. Simpson, 1995, 'Compelled to Compete? Assessing Market Power in Regulated Markets', *Competition and Consumer law review*, vol. 3, no. 2 pp 113–143.

provisions, and query whether this diminishes in any way their market power in a Trade Practices context.

Changes to the regulatory system can:

- restrict the decisions made by market participants;
- create competition where none existed before e.g. telecommunications and entry of new players;
- alter markets by bringing products which were previously not in close competition into competition e.g. removal of regulations which required that all wheat had to be carried by rail rather than by truck; access arrangements.

Other changes of significance which flowed from the Hilmer review were:

- legislative review
- requirement for competitive neutrality

Part IIIA of the TPA

A question which merits some consideration is the need for a separate access provision. One might ask: why, as under the Essential Facilities doctrine in the United States, could access not be adequately dealt with under Section 46 of the TPA?

The Hilmer report defines an 'essential facility' as a natural monopoly which cannot be economically duplicated and which occupies a strategic position in the industry which permits it to reduce output and/or service and to charge monopoly prices to the detriment of users (p. 239). Where the owner of the essential facility also operates in upstream or downstream markets in competition with those seeking access, there is obviously the possibility that the market power arising from ownership will be misused.

There are a variety of means by which this problem can be addressed and one of them is via the use of s. 46 of the Trade Practices Act. Both Queensland Wire Industries (QIW)[7] and Pont Data[8] could be interpreted as such an application. For s. 46 to be applicable it must be shown that:

- the facility owner has market power;
- there is a misuse of that market power; and
- the purpose is a proscribed purpose.

It has been argued that the essential nature of the facility establishes market power and that refusal to deal in this environment would be interpreted as taking advantage for a

---

7 Queensland Wire Industries Pty Ltd v The Broken Hill Pty Ltd Co and Amcor (1989), 167 CLR 177.

8 Pont Data Australia Pty Ltd v ASX Operators Pty Ltd (1991) ATPR 41–007.

proscribed purpose because in a competitive environment refusal would be unlikely i.e. the competition test of QWI would be met. However, s. 46 is a particularly difficult section of the Act to deal with. In QWI, it was argued that BHP's control of the steel and steel products market should be regarded as control of an essential facility and so, following the US essential facilities doctrine, BHP was required to supply Y-bar to QIW to allow it to compete in the rural fencing market. The Full Federal Court rejected this. On appeal, the High Court failed to address the issue.

Hilmer[9] concluded that s. 46 could not adequately address the access problem because:

- the inability of s. 46 to deal directly with monopoly pricing that is not for a proscribed purpose;
- the evidentiary problems of proving in court that refusal to supply on reasonable terms is for a proscribed purpose;
- the cost, time and risk involved in obtaining a court resolution of a commercial dispute;
- doubts about the ability of the courts to determine optimal pricing, and terms and conditions of access to essential facilities.

The last of these is of particular concern. The decision of Wilcox J. when he ordered marginal cost pricing in relation to Pont Data clearly illustrates the lack of desire or expertise of judges to set prices.

An alternative to relying on s. 46 is to create special access rights. An example is the *Telecommunications Act 1991* which established the right of any carrier to connect its facilities to the network of any other carrier and have its calls carried and completed over that network. However, this still begs the question of how the access price is to be determined.

Part IIIA of the Trade Practices Act is in some respects an intermediate position. It establishes a right of access with certain specified conditions. The conditions which the National Competition Council (NCC) must find in order to recommend Declaration by the relevant minister under Part IIIA are set out in s. 44G(2) of the TPA:

> (a) that access (or increased access) to the service would promote competition in at least one market (whether or not in Australia), other than the market for the service;
>
> (b) that it would be uneconomical for anyone to develop another facility to provide the service;
>
> (c) that the facility is of national significance having regard to:
>
> (i) the size of the facility; or

---

[9] Report by the Independent Committee of Inquiry, 1993, *National Competition Policy*, AGPS, Canberra, August, Chap 11, pp 239–268.

(ii) the importance of the facility to constitutional trade or commerce; or

(iii) the importance of the facility to the national economy;

(a) that access to the service can be provided without undue risk to human health and safety;

(b) that access to the service is not already the subject of an effective access regime;

(c) that access (or increased access) to the service would not be contrary to the public interest.

Part IIIA provides for access either via Declaration, with pricing and other conditions to be negotiated between the parties or failing agreement provision is made for arbitration. Alternatively, the facility owner can provide an undertaking to the ACCC in relation to the terms and conditions on which access will be provided.

There will undoubtedly be a variety of problems/issues in relation to Part IIIA but at this stage it is too early to make much comment. Of some significance is the role of the NCC which recommends Declaration or not to the relevant minister but the minister may decide to reject the recommendation for various reasons which are unrelated to competition issues. In relation to the Declaration process, as undertaken by the NCC, difficulties may arise due to the requirement to establish an increase in competition in 'a separate market'. [10]

Direct regulation

Regulation has been defined as the use of a government's power to coerce 'for the purpose of restricting the decisions of economic agents'.[11] Direct regulation may be appropriate where the government is desirous of achieving some social objective or where competition is absent or weak and is unlikely, at least for some time, to develop. The latter may be because activities are being deregulated and/or privatisation of formerly government owned assets is occurring. In Australia, competition policy can normally be described as general regulation, that is the framework is provided but specific requirements or directives in relation to particular decision variables is avoided. There are, however, exceptions. Regulation in many areas can be seen to be in conflict with the aims of competition policy, e.g. when it results in price fixing, market allocation. The aim of the review to be carried out by all governments of their legislation as it affects competition, it to remove these anti-competition effects subject to certain conditions. There are also other well documented problems associated with direct regulation and these are briefly outlined below.

Costs

---

[10] Rhonda L. Smith and Jill E. Walker, 'Part IIIA, Efficiency and Functional Markets', Competition and Consumer Law Journal (forthcoming).

[11] W. Kip Viscusi, John M. Vernon and Joseph E. Harrington Jr, 1996, Economics of Regulation and Antitrust, MIT Press, p. 307.

It is generally accepted that in relation to any regulation the benefits resulting from it must exceed the costs associated with it. The costs of regulation include compliance, administrative and enforcement costs. In addition, direct regulation is likely to result in distortions as illustrated by the Sites Act in the petrol industry.

General v industry specific regulator

When a decision is taken to impose direct regulation on industry, a secondary issue is whether oversight of the regulation is via a general body such as the competition authority or via an industry specific regulator. A general regulator provides greater consistency in relation to competition policy, probably requires less resources to operate, and is less susceptible to industry-capture and empire building. However, a general regulator lacks the detailed technical expertise to understand the operation of the industry. As a consequence expensive additional staff may be needed and even this may not adequately address the issue. Consultancies are a possible solution. In Australia (in contrast to the UK) the various governments have in most cases deliberately rejected specialist industry regulatory bodies. As a consequence the non-technical aspects of Austel, the specialist telecommunications were subsumed into the ACCC.

National v state regulators

Another issue to be decided is whether regulation should be at a national level or state level. Most of the industries, such as the utilities, that are of concern in this respect are either national, are moving toward becoming national or are involved with at least some other states. Problems associated with state based regulation include:

- the risk of inconsistent decision-making where there are a number of separate state regulators.
- it is likely to be more costly, less efficient.
- there may be difficulty finding sufficient suitably qualified people to fill the regulatory roles in these offices.

In reality, Australia has a mix of national and state regulation and of general and industry specific regulation. Specific regulation has been chosen in some areas such as telecommunications but the economic/competition aspects are the responsibility of the general competition regulator.

Direct price controls

The task of competition policy should not be to replace the market, but rather to create an environment in which it can function better. Nevertheless, where markets are not competitive or where the product is one with significant implications for the cost structure of the economy, some form of prices oversight may be appropriate. If lower prices are seen as a worthwhile policy aim, then one element of competition policy may be direct price controls. Especially where monopoly is prevalent, or at least where market power is widespread, there may be a temptation to think that the simplest way

to overcome the problem is to regulate prices. This is likely to be the case where public utilities are involved. In many of the public utilities, there is a significant degree of market power, whether they are in public or private ownership.

A major problem with this approach is that it leaves the regulators to determine the price. The track record of price setters is poor. Usually prices are set on the basis of cost plus a return so price increases are justified by and related to cost increases. This provides little incentive to increase productivity and efficiency. Linking prices to the inflation rate less a discount factor to encourage productivity improvements, is one way of addressing this problem. Alternatively, instead of directly setting prices, the authorities might set a maximum rate of return which can be earned: anything above this would incur a tax surcharge of 100 per cent. Each of these approaches has problems and they will be addressed in later sessions of this conference.

Conclusion

In conclusion, competition policy and competition law are not about removing or outlawing monopolies. However, they are concerned with preventing the abuse of existing market power and about maintaining a competitive market structure. The basis for this is the acceptance that generally such a market structure results in economic efficiency and encourages economic growth. Nevertheless, an effective competition policy should include provisions to deal with market failure where the promotion of competition will not necessarily result in an efficient outcome. The traditional approach to competition policy was structuralist, but increasingly it is recognised that to allow adequate consideration of dynamic efficiencies strategic behaviour may need to be analysed.

Utility regulation in New Zealand[1]

Alan E. Bollard, Chairman, New Zealand Commerce Commission.

Reform of the utilities sector

The term 'utility' is difficult to define precisely, but generally it is applied to the communications, energy, transport, and 'public amenities' sectors, including telecommunications, broadcasting, electricity, gas, railways, sewerage and water (Kahn, 1995). These industries usually exhibit two characteristics which may be shared by the ports, airports and roads. Firstly, utilities provide a distribution, transmission, or transport service through a network of cables, pipes or other facilities which tend to enjoy such large scale economies as to become natural monopolies. Secondly, since the service they provide is often regarded as an 'essential' input to other industries, the efficiency of utilities has a widespread impact on the efficiency of other firms. A further characteristic in New Zealand, and often elsewhere, is that historically the social importance of such industries, and doubts about their ability to function in competitive markets, have resulted in a history of public ownership.

Until the mid-1980s, utilities in New Zealand were generally statutory monopolies under state ownership, and run by Government trading departments or divisions. In the early 1980s the Treasury estimated that public sector enterprises in New Zealand, which covered but were not limited to utilities, accounted for about 12 per cent of GDP, and for some 20 per cent of investment in the economy (Treasury, 1984).[2] Since they often produced 'essential' inputs used by firms in the private sector, their efficiency, price-setting and investment behaviour had a major impact on the competitiveness of the economy as a whole. However, their performance was judged by critics to have been poor:

- Over the twenty years to 1985–86 the government invested $5000 million (in 1986 dollars) of taxpayers' money in the departmental trading activities of the Airways System, the Lands and Survey Department and Forest Service, the Post Office, the State Coal Mines, and the Electricity Division of the Ministry of Energy. In 1985–86 these organisations managed assets valued at over $20 billion but returned no net after tax returns to taxpayers.[3]

---

1 This section has been drawn from A. Bollard and M. Pickford 'The New Zealand Solution: An Appraisal', in: M E Beesley (ed), Regulating Utilities: Broadening the Debate, IEA Monograph, London: IEA, 1997, chapter 4.

2 In 1995 the three sectors accounted for the following percentages of GDP: 'Electricity, Gas and Water', 3.4 per cent; Transport, 6.8 per cent; and 'Communications', 6.0 per cent.

3 S. Jennings and R. Cameron, 'State-Owned Enterprise Reform in New Zealand', in: A. Bollard and R. Buckle, 1987.

Several reasons have been adduced for the inadequate performance: the conflict they faced between various commercial and social objectives; an operating environment in which competition was usually lacking; access to funding from government sources at favourable rates of interest; lack of accountability to, and inadequate monitoring of performance by, government; and political interference.[4]

As part of wider economic liberalisation policies, these industries were progressively reformed in the years following 1985. The most important steps were:

- the removal of nearly all statutory monopoly rights so as to expose utilities to competition;
- the corporatisation (and in some cases, privatisation) of numerous state trading departments so as to place them in a company form of organisation, with commercial objectives;
- their restructuring to isolate the natural monopoly elements from the more contestable parts of the industries;
- the abolition of social service obligations or their explicit funding by government rather than, as previously, by cross-subsidy with profits earned by the business in incontestable markets (Duncan and Bollard, 1992).[5]

Under the *State-Owned Enterprises Act 1986* the major government trading departments were corporatised on 1st April 1987, from which time they were required to operate as profitable and successful businesses, comparable with their private sector counterparts. Goods and services were to be marketed on a user pays basis, unless an explicit subsidy was provided to finance non-commercial activities. Regulatory barriers were dismantled, thereby exposing the new corporations to private sector competition, and other forms of special assistance — such as subsidised government loans — were removed. Taxes and dividends had to be paid to the government. In short, 'competitive neutrality' was to apply.

Monitoring of performance was enhanced by the establishment of measurable targets based on profitability, although asset values were hard to assess. Departmental organisation was replaced by a company (limited liability) structure with the government as sole shareholder. Managers were given greater independence in decision-making, but were accountable to boards of directors appointed from the private sector, and ultimately to Parliament through the Minister of Finance and the responsible minister. The intention was to remove decision-making in State-Owned Enterprises (SOEs) from direct political interference.

While these reforms marked a major step forward, it was argued that certain problems remained with the SOE model (New Zealand Business Roundtable, 1988). Firstly, since the ownership rights in SOEs are diverse and cannot be transferred (ownership being vested in the Crown), managers lack the incentives to perform normally provided

---

[4] See previous footnote; also Farrar and McCabe, 1995, for an agency cost approach.

[5] A social service obligation is an activity undertaken by a government enterprise which meets a social objective required by the government, but which makes a financial loss at current prices.

through the share market. They face no threat of take-over, and the monitoring of performance by shareholders and investment analysts is attenuated (the free rider problem). Secondly, the incentives provided by the possibility of bankruptcy are regarded as minimal because of an implicit government guarantee. By reducing risk, this may distort the cost of capital in a downwards direction. Finally, some claim that SOE decision-making is subject to residual government interference, since the directors are political appointees, and an annual 'statement of corporate intent' has to be approved by the government. Moreover, interest groups may pressure the government to hold inquiries into particular management decisions, as happened with the pricing of the Electricity Corporation (ECNZ) and New Zealand Post's rural mail charges in the early 1990s.

Such considerations have been argued to justify privatisation of the SOEs, as has happened in many cases, including Telecom and New Zealand Rail.[6] Competition and other concerns have hindered moves to privatise other major utilities such as Electricity Corporation, Coal Corporation, and Television New Zealand.

The problem of utility regulation

In most Western countries the regulation of utilities poses more complex problems than the regulation either of non-utility SOEs, or (generally) of private sector firms through competition policy. This is partly because utilities — particularly those in the energy sector — have intrinsic features which, by leading to small numbers of industry participants and by raising significant barriers to entry and exit, serve to attenuate competition.

These features include: substantial economies of scale, sometimes to the point of natural monopoly (e.g. high voltage electricity transmission lines); economies of scope (e.g. in the provision of different telecommunication services); and large, lumpy, immobile investments in sunk assets (e.g. natural gas production facilities and distribution networks; railway networks).

Further regulatory problems are raised by networks and plants (e.g. hydro-electric dams) typically having low marginal costs of expanding output up to full capacity, but high fixed costs associated with that capacity; by the potential for substantial externalities, especially environmental (e.g. coal mining, power stations); and in some cases, by inelastic demand curves (e.g. for electricity because of appliance ownership), which raise the gains from the exercise of market power.

The utility problem is most acute in those industries which provide a basic service for consumers and an essential input for other firms, and where production involves the use of non-contestable or non-economically reproducible facility services owned by

[6] However, SOEs typically are both large and complex organisations, which would tend to make them more resistant to take-over (and bankruptcy), and more difficult for the capital market to monitor because of information asymmetries and possible managerial opportunism. Thus privatisation may not be able to solve all of the incentive problems posed by large, complex SOEs.

incumbents, to which access is required by entrants in order to compete with the same incumbent in upstream or downstream markets.

Typical examples are in local electricity supply, where independent retailers need access to the vertically integrated incumbent's local distribution lines; or in the case of a geographically isolated port, where an independent stevedoring company needs access to the wharves to be able to compete with the horizontally integrated port company's stevedoring service. It is in these sorts of areas that regulators encounter most difficulties, both in the technical, legal and economic aspects, and also frequently in trying to deal with asymmetric information, conflicting objectives, aggressive and sophisticated incumbents, politically astute entrants, and technological change.

In terms of Western economies, these problems of utility regulation are probably at their most acute in New Zealand, given the small size of the economy and the difficulty of gaining the economies of scale needed for the efficient operation of many utilities. In addition, the country's geographical isolation removes any possibility of international trade in utility services with adjoining countries. Thirdly, the state has throughout New Zealand's colonial history traditionally played a major role in the utilities sector, initially because the fledgling capital market was not prepared to bear the risk in major developmental projects, such as the building of the main trunk railway network. By the 1970s almost every utility was occupied by a large state trading organisation, usually operating with statutory monopoly rights, and suffering from political interference in terms of price-setting, investment planning, and other operating conditions. Deregulation was required as a first step, but this tended to generate problems for the regulator in that the industry was thrust into an initial state of disequilibrium, with incumbent and new firms both struggling to come to terms with an unfamiliar competitive environment, and sometimes with individual markets being difficult to discern after years of statutory, vertically integrated monopoly.

Left to themselves, such deregulated utilities may generate significant resource misallocation and inefficiency from their monopoly pricing, and from exclusionary behaviour aimed at deterring competition from new entrants. A totally 'hands off' approach is not considered an option. At the other extreme, direct control imposed by an industry-specific regulator will likely generate its own inefficiencies. These are likely to include: the costs of operating the regulatory body; the 'paper burden' or information supply costs imposed on the regulated firm; the scope for it to engage in 'opportunistic' behaviour; the compliance costs arising from the distortions caused by imperfect regulation; the losses associated with the possible corruption of the system through 'regulatory capture'; and dynamic losses associated with the control by regulators of industry structure and conduct, which may inhibit new entry, competition, investment, and innovation.

The design of a regulatory regime must therefore weigh up the potential costs and benefits involved. Success in this area is a relative concept, and has to be judged against the background of a realistic counterfactual, rather than some sort of ex-post, socially-optimal nirvana. All forms of regulation, including non-regulation, produce adverse outcomes; the relevant question is which form produces the least adverse result when assessed over a range of criteria.

Light-handed regulation in New Zealand

Utility regulation has been approached in New Zealand in a characteristically straightforward manner, using the same set of competition principles used to regulate other industries, together with some additional light-handed elements. The concept of 'light-handed' regulation emerged from the policy debate concerning the deregulation and privatisation of telecommunications during the period 1987–90. Given the market power which a deregulated Telecom would wield, alternative regulatory regimes were evaluated. Conventional 'heavy-handed' forms of regulation were rejected in favour of a novel approach, support for which was reiterated in a policy statement by the Government in December 1991:[7]

> The Government sees competition as the best regulator of telecommunications markets. Accordingly, there will continue to be no statutory or regulatory barriers to competitive entry into (the) . . . market . . . To maintain the conditions of effective competition, the Government places primary reliance upon the operations of the *Commerce Act 1986* . . . The following supplementary measures will continue to apply: (a) the Telecommunications (Disclosure) Regulations 1990; and (b) the Telecommunications (Internal Services) Regulations 1989. If it proves to be necessary, the Government will consider the introduction of other statutory measures or regulation.

The overall thrust of the policy, in common with that in other Western countries, is to encourage competition where markets are potentially contestable, and to focus regulation on the non-contestable markets controlled by incumbent utilities.

This regulation takes a number of forms:

- There is a reliance on general competition law, as expressed in the *Commerce Act 1986*, under s. 36 of which dominant firms must not (in general terms) behave abusively towards actual or potential competitors. Contravention of s. 36 exposes dominant firms to court action by private parties or by the Commerce Commission.

- Information disclosure regulations, especially for natural gas and electricity, and to a lesser degree for telecommunications, require the disclosure of accounting and other information annually by incumbent operators of 'essential facilities' with market power. This is intended to encourage self-regulation through the market and to underpin the effectiveness of the Commerce Act. The regulations are administered by the Ministry of Commerce, which publishes key pricing and performance data.

- In addition to being required not to misuse their dominant position under the Commerce Act, utilities also face the standard legal requirement that they should not lessen competition through their trade practices.

---

[7] Quoted in Ministry of Commerce/The Treasury, 1995, pp 21–22.

- Utilities involved in business acquisitions which lead to the acquisition or strengthening of a dominant position breach the Commerce Act, although such acquisitions may be sanctioned under the standard authorisation procedure.

- Regulatory barriers to entry have been minimised through the termination of statutory monopoly rights for almost all utility areas (including, shortly, postal services), so as to encourage the entry of new competitors. Entry has occurred subsequently in the case of telecommunications, television and radio broadcasting, and domestic airlines. These industries, all formerly state-owned statutory monopolies, are thus now treated like any other industry.

- In the regulatory reform of some utilities, there has been separation of contestable and non-contestable elements of the utility's activities, either by the core non-contestable business being established as a separate company (e.g. the separation of Trans Power, the high voltage electricity transmission system, from the electricity generator), or by the accounting ring-fencing of the two businesses (as required of local electricity distributors who are also electricity retailers).

- In cases where there has been no such separation,[8] incumbents who refuse to provide access to natural monopoly facilities on reasonable terms may breach s. 36 of the Commerce Act, which can result in punitive damages being imposed.

- A further possibility is to break up the state-owned utility into competitors; this has happened in the case of state-owned electricity generation. A wholesale electricity market has recently been established to facilitate competition between the two generators, while allowing technical co-ordination of generation and transmission from the company's major hydro and thermal power plants.

- Competition from substitute goods, or from new technologies has been encouraged. Industries which historically have been the preserve of regulated monopolists, such as long distance transport (New Zealand Rail) and parcel post (New Zealand Post), have experienced intense competition from road freight and courier services respectively. In other industries, like telecommunications, technological advances in cellular, radio and satellite-based technologies are undermining the advantages of network ownership.

- In addition to these generally light-handed forms of regulation, there is a more heavy-handed measure of last resort: the use of price control as provided for in Part V of the Commerce Act (or potentially some other form of statutory regulation which the government may wish to introduce). The Commerce Commission can recommend to the government that it impose price control. In addition there have been specific warnings of direct regulation for particular industries. However, there are no goods currently subject to price control, and the government has historically shown no desire to re-introduce such controls into the newly deregulated environment.

---

[8] Examples include the major gas and telecommunication utilities, which were sold complete with networks. While there may be efficiency reasons for this integration, it has complicated the entry conditions for potential new competitors.

- In some cases the Government has retained a 'Kiwi share' to prevent changes to the Articles of Association (e.g. to maintain a 49 per cent domestic ownership of Air New Zealand to allow the government to negotiate international bilateral access agreements). Government trading enterprises have statements of corporate intent which may also limit the scope of their operations. For example, they may be prevented from diversifying their investments.

In addition, there are other forms of regulation in New Zealand which are of a social or safety nature, rather than being a form of light-handed regulation. Some privatised enterprises are required to meet particular service obligations (e.g. the universal service obligation on Telecom), or social obligations agreed with the government (e.g. New Zealand Post's obligations include six day delivery and a universal service requirement).

While the emphasis is to admit and encourage competition wherever possible, New Zealand's policy of light-handed regulation clearly does not mean zero regulation, as has sometimes been asserted.

The role of competition law and information disclosure

As indicated above, the critical elements of the policy of light-handed regulation are competition law and information disclosure. These two elements are now examined briefly in turn.

Under New Zealand's light-handed policy, owners of natural monopoly networks are likely to be in a dominant position in a market, and thus to be subject to the s. 36 prohibition of the Commerce Act that they must not use that position for the purpose of limiting competition. Specifically, firms which have a dominant position in a market must not:

> . . . use that position for the purpose of —

- restricting the entry of any person into that or any other market; or
- preventing or deterring any person from engaging in competitive conduct in that or in any other market; or
- eliminating any person from that or any other market.

While s. 36 prohibits dominant firms from using a dominant position for the purpose of restricting competition, it does not prohibit the dominant firm from using its market power for other than anti-competitive purposes. For instance, the charging of a 'monopoly' price in itself is not prohibited, although the inference is that monopoly profits should be competed away where entry is possible.

There has been considerable discussion in the courts about the meaning of a 'dominant position in a market', how the presence or absence of dominance may be determined, and what constitutes 'use' and 'purpose'.[9]

In the utilities area, access by Clear Communications (the new entrant) to the local telephone network of Telecom (the incumbent) has been the subject of extensive litigation, which has potentially important ramifications for other network industries. Telecom put forward the Baumol-Willig rule (Baumol and Sidak, 1994) as the way of setting the interconnection price; it states that monopolists are entitled to provide services to competitors at the same price they implicitly charge themselves, including monopoly profits. This rule was accepted by the New Zealand High Court, rejected by the Court of Appeal, and finally in 1994 sanctioned by the Privy Council, New Zealand's highest court. Critics of the rule have argued that it has undermined the effectiveness of s. 36 as a mainstay of light-handed regulation.

In 1995 Government officials undertook an 'analysis of the experience of telecommunications interconnection negotiations in New Zealand', produced a detailed report, and invited submissions (Ministry of Commerce/The Treasury, 1995). After lengthy consideration, the Government issued a media release on 26th June 1996 in which it pledged 'to continue, for the time being, with the present regulatory regime based upon the Commerce Act.' (Minister of Commerce, 1996). In addition, it stated that:

> . . . developments since the Privy Council decision demonstrate that the major telecommunications industry players do not support the Baumol-Willig rule as a satisfactory basis for interconnection pricing. The Government considers that the . . . rule has the potential to lessen competition, thereby limiting the rate of introduction of new products and services and lessening the benefits to users. The Government would be concerned to see the . . . rule being applied in future.

Note, however, that this statement has not yet been enshrined in statute.

Information disclosure has been developed to its greatest extent in the electricity industry. This reflects the relative lack of competition and of substitute forms of energy in electricity, and the relative slowness of technical change which might otherwise serve to undercut market power.

The Electricity (Information Disclosure) Regulations 1994 introduced the requirement to disclose information, along with the accounting separation of distribution and retailing activities. Three years information is now available, including line charges to all consumers, prices, and other key conditions of contract; costs and revenues by load group; separate financial statements from line owners on generation, line and retailing activities; and financial and other performance measures (with assets valued at optimised deprival values). By facilitating comparisons of the prices and performance between different power companies, this information is intended to discourage monopoly pricing, excessive cross-subsidisation and uneconomic generation, and to promote competition by facilitating access to distribution lines. Any cross-

---

[9] An excellent review is given by McGechan J, in: Commerce Commission *v* Port Nelson Ltd., CP NO 12/92 (NN), June 1995, pp 39–53.

subsidisation, which occurs between commercial and domestic consumers, and between urban and rural areas, will be vulnerable to price undercutting by a competitor.

Where a business in either electricity or gas combines activities which have natural monopoly characteristics and others which are potentially competitive, separate audited statements are required, in order to detect instances where market power in the latter is enhanced by the use of monopoly leverage in the former. Consumers' bills must be unbundled so as to reveal the separate line and energy components, undermining the scope for price discrimination or cross-subsidy. However, comparisons between companies will have to make allowance for the scope in the regulations for companies both to define their businesses and to allocate assets and costs between them, as well as standardising for regional variations in network density, geology, and customer profile.

So far the Commerce Commission has had to resolve three main competition issues involving utilities: business acquisitions; the price and other terms of access to network facilities; and complaints of monopoly pricing. As an example, in the local markets for the supply and distribution of electricity, the Commerce Commission has received a number of complaints alleging inability by one company to gain access to another's network, usually because of onerous terms and conditions in the 'use of system' agreements. To date the Commerce Commission has taken the view that such difficulties can often be resolved by negotiation, and by further refinement to access agreements, during the 'shakedown' period of industry adjustment to developing competition; but it also acknowledges that persistently anti-competitive behaviour (whether by a network owner towards an entrant seeking access, or by arrangements between incumbent suppliers and downstream buyers) may have to be taken to court. The Commission monitors industry developments continuously, and is currently investigating some more serious cases.

Bibliography

Baumol, W. J. and J. G. Sidak, 1994, 'The Pricing of Inputs Sold to Competitors,' *Yale Journal on Regulation* 11, pp 171–202.

Bollard, A. E. & R. Buckle (eds.), 1987, *Economic Liberalisation in New Zealand*, Allen & Unwin, Wellington.

Bollard, A. E. and M. Pickford, 1995, 'New Zealand's "Light-Handed" Approach to Utility Regulation,' *Agenda*, vol. 2, forthcoming.

Duncan, I. and A. E. Bollard, 1992, *Corporatisation and Privatisation: Lessons from New Zealand*, Oxford University Press, Auckland.

Farrar, J. and B. McCabe, 1995, 'Corporatisation, Corporate Governance and the Deregulation of the Public Sector Economy,' *Public Law Review*, vol. 6, no. 1, pp 24–43.

Kahn, A. E., 1995, The Economics of Regulation: Principles and Institutions, Cambridge, Mass., MIT Press.

Minister of Commerce, 1996, 'Government Signals Future Directions for Regulation of Telecommunications, Electricity and Gas,' Wellington, Office of the Minister of Commerce, 26th June.

Ministry of Commerce — The Treasury, 1995, 'Regulation of Access to Vertically-Integrated Natural Monopolies: A Discussion Paper', Wellington, Ministry of Commerce.

New Zealand Business Roundtable, 1988, *State Owned Enterprises: Issues of Ownership and Regulation*, NZBR, Wellington.

The Treasury, 1984, *Economic Management,* The Treasury, Wellington.

Section II. Incentive regulation

Introduction to the fundamentals of incentive regulation

Sanford V. Berg, Public Utility Research Center, University of Florida, United States.

Some might argue that the term 'incentive' in incentive regulation is redundant. Government intervention **by definition** establishes a system of rewards and penalties for private decision-makers. The resulting incentives can be dysfunctional, but regulation cannot help but create incentives of some type. Clearly, regulation affects behaviour, as evidenced by traditional and more recent varieties of regulation. The key question is how to make the intervention productive, promoting the achievement of economic objectives at minimum cost. The purpose of this training program is to identify lessons that emerge from worldwide experience and to develop policies that build on fundamental economic principles.

Fundamentals of incentive regulation

Over the past two decades, economists have emphasised regulation as a response to information problems:

- problem of monitoring performance, and
- problem of specifying performance targets.

Incentive regulation can partly overcome information problems. Lewis and Garmon, 1997, define the process as follows:

> Incentive regulation is the use of rewards and penalties to induce the utility to achieve desired goals where the utility is afforded some discretion in achieving goals.

They note that there are three important elements of this definition:

- use of rewards and penalties provide inducements to motivate the utility to perform. This system replaces a command and control form of regulation.
- utility assists in setting goals or performance targets. Goals are not unilaterally dictated by the regulator. The firm's information on complex performance interdependencies (as between output and quality) may need to be taken into account.
- utility decides how to achieve goals. Specific actions are not prescribed by the regulator, which allow the utility to utilise its internal information and to establish internal incentives appropriate for improved performance.

In comparing incentive regulation with control and command regulation, he notes that under the latter, the regulator provides detailed instructions of duties to be performed by the utility. For example, 'the utility might be instructed to construct a particular

type of power plant, to adopt select particular fuels to burn, to maintain a specified work force, and to follow specific operating procedures.' Control and command sets performance goals for the utility to achieve. Command and control regulators monitor personnel to insure goals are achieved using specified procedures.

The preference for incentive regulation or control and command will depend on a number of factors:

- regulator's knowledge of utility operations;
- regulator's ability to monitor utility;
- administrative costs of regulation;
- motives of the utility;
- political environment;
- capital market discipline; and
- underlying market structure.

The following generalisations emerge from the literature.

- Cost-of-service regulation (including return on rate base regulation) provides an opportunity to cover costs. It also provides companies with an incentive to over/under invest in plant, inflate costs, and cross-subsidise. Regulators generally try to remedy these perverse incentives through regulatory lag, sliding scales, and efficiency audits/reviews.
- Price cap regulation provides companies with incentives to cut costs. It also dampens the effects of cost information asymmetries between companies and regulators. Service quality and infrastructure development may suffer. However, incentives to over-invest in capital and to cross-subsidise are less than with cost-of-service regulation.
- Yardstick regulation provides companies with incentives to cut costs. It also dampens the effects of information asymmetries between companies and regulators. However, developing appropriate yardsticks is resource intensive and may not be possible in some situations.
- Performance-Based Regulation (PBR) is another term applied to situations in which performance measures are used to incent the utility. Good performance measures should be accurately observed and verifiable, should reflect the utilities' efforts, and should be structured to reduce the impact of random variation. In addition, performance measures should be adjusted according to how specific or diverse are the utility's performance areas.

- ✍ Franchise Regulation provides incentives for cost containment through competitive bidding for the right to provide service in a geographic area. The firm bidding the lowest price (subject to meeting quality requirements) becomes the supplier. However, re-contracting issues arise at the end of the period (e.g. capital maintenance and 'lock-in' issues).

Note that hybrids of these can also be utilised.

Incentives for cost containment

Case studies, economic theory, and empirical evidence suggest that cost minimisation is problematic under traditional U.S.-style cost-based regulation. The principal-agent/incentive literature shows the implications of firms and regulators having different information, capabilities, and objectives. Agents (corporate executives) engage in opportunistic behaviour that promotes **their** objectives rather than those of the regulator. The terms organisational slack, 'shirking behaviour,' and 'engaging in abuse' are synonyms for the resource-utilising actions of firms that cannot be monitored by regulators. Although regulators can collect information on firm expenditures, these reported accounting numbers consist of a combination of necessary production costs and other outlays that provide some benefit to managers (and perhaps stockholders).

The fundamental point of the principal-agent literature is that actual managerial abuse (or 'reduced managerial effort') cannot be easily monitored (Sappington, 1994). 'Reduced effort' is a slightly more value-neutral term, and reflects the realistic view that corporations are capable of different levels of efficiency. The achievement of best-practice can be quite costly from the standpoint of managerial effort. Whatever the behaviour is labelled, it is clear that necessary costs and other outlays cannot be easily disentangled. It should be noted that in practice, regulators punish detected inefficient behaviour through cost disallowances and other regulatory penalties. In addition, yardstick regulation represents one way to benchmark firms — rewarding good performance and penalising poor performance. As noted earlier, care must be shown in developing appropriate comparisons. The traditional regulatory emphasis on procedural fairness in the U.S. can be shown to induce inefficiencies: when regulators cannot count what really counts, they are more comfortable following precedents!

In correspondence, Stephan King notes that nations instituting reforms face additional problems:

> A lot of regulators in Australia at present are dealing with rules that are basically new and have little if any court precedent. A significant problem at present with those reforms is that the courts have not interpreted (for example) the rules on declaration of a service. There are currently a group of appeals to the Australian Competition Tribunal over successful and unsuccessful declarations. Until the courts give some precedent and interpret the relevant statutes, the regulators are trying to operate in a rather fuzzy legislative environment. In my opinion, this tends to lead to a bias towards cost-based regulation unless otherwise explicitly stated in the legislation. Price cap rules have been put in place for airports but the legislation on access and for telecommunications just refers to regulatory criteria such as '... the legitimate interest of the owners ....' In the latter case, it is easiest to interpret 'legitimate interest' by using rate-of-return rules.

Thus, lack of precedents can lead to delays and/or regulatory inertia.

In addition, regulatory commissioners can have their own agendas. Regulators can also engage in 'opportunistic behaviour' — related to future income opportunities or to political aspirations beyond current responsibilities. Just as firms have access to unique information, agencies obtain confidential data from a wide range of entities. Such information can be misused or deliberately misinterpreted to the personal advantage of the regulator. The 'public interest' theory of regulation takes a different view of the process: policy makers and politicians are basically benevolent designers of government institutions which correct market failures and reduce market imperfections. Economic theories of regulation introduce a greater role for special interests, including the interests of regulators themselves.

Incentives issues arise whichever view of regulation one takes. Traditional cost-based regulation in the U.S. involved a bottom-up approach: costs were aggregated and service prices reflected the results of complex (and arbitrary) cost allocations. Postage-stamp pricing resulted within a utility's territory with customers in high cost rural areas and low cost urban areas paying the same prices. The political advantages of such regulatory treatment made for a stable situation. The beneficiaries of the rules were well aware of their gains, and the costs were spread over a larger population. Politicians also argued that income distributional concerns or universal service obligations justified the higher prices borne by some customers. For example, in telecommunications, since local telephone rates were held down due to transfers from long distance customers, the cross-subsidies came at the expense of 'others'. The political economy of rate design raises numerous issues. For example, initial asset valuation becomes a key issues at the time of price cap review — since it affects allowed profits. For now, we focus on the recent movement away from cost-based regulation to price-based regulation, generally involving some form of price caps, often with profit-sharing between stockholders and customers. The U.K. pioneered in the introduction of price caps (Beesley and Littlechild, 1989; Weyman-Jones, 1990).

The purpose of these regulatory policy changes has been to avoid the inefficiencies associated with cost of service regulation (Braeutigam and Panzar, 1989). These inefficiencies have been identified as the AJ effect (over capitalisation which leads to production and allocative inefficiencies), cross-subsidisation (entering competitive markets and recovering costs from core customers), excessive or inadequate service quality, and employee/managerial slack. The extent of these inefficiencies is an empirical question, although studies of the impacts of competitive pressures and 'incentive' regulation suggest that the inefficiencies associated with traditional regulation were significant. However, it is hard to distinguish between the competition effect and the regulation effect, since reductions in entry barriers and incentive regulation often occur simultaneously. Perhaps competitive pressures should be given more credit for the productivity advances and new service introductions arising in recent years.

As noted earlier, information asymmetries create problems: firms have information and opportunities which enable them to benefit at the expense of ratepayers. Agencies incur administrative costs to deal with these issues. In 1988, the direct cost of

operating regulatory agencies in the U.S. amounted to about $.50 a month per household — a trivial sum for oversight activity in comparison to the amount consumers spent on electricity, telecommunications, and other infrastructure services. In the U.S., some argue that additional regulatory monitoring activity and longer regulatory lags between rate reviews can promote efficient production; others see opportunities for reductions in regulatory oversight (as agencies move toward more 'light-handed' intervention).

New types of regulation may still fall short of potential gains from improved incentive mechanisms. Within the U.S., the state telecommunications plans adopted in California, Illinois, Michigan, and elsewhere tend to have a range over which cost reductions do not lead to price reductions, so profits are earned on a dollar for dollar basis. Usually, sharing sets in at some point (say, at a particular realised rate of return), and beyond some return, all of the savings are passed on to customers (in the form of lower prices or rebates in a future period). The Washington plan has one especially interesting feature: the firm captures an increasing share of the profits — so the disincentives associated with a profit cap are avoided. However, according to Blackmon, 1994, other features of that particular incentive plan are problematic: excessive rewards to small increases in efficiency and inadequate rewards for larger improvements, distorted risk-taking (depending on whether one is at the top or bottom of the sharing scheme), and intemporal manipulation of outlays — so the bunching of expenses can increase profits.

As will be seen, both giving firms an option regarding price caps and instituting sharing rules have some desirable features. Firms can have input on setting performance objectives, but regulators should recognize that firms have an incentive to understate their abilities to reduce costs. With a policy of creating options, the regulator establishes several plans, with different productivity ('X') factors. Low performance targets (prices that fall more slowly) are linked to lower rewards, with high performance targets having sharing rules that yield higher (possibly unlimited) profit potential for the firm. Such optional schemes induce firms with substantial potential for cost containment to self-select into the appropriate plan. The FCC price caps applied to local exchange carrier access charges had this feature. This scheme has some excellent incentive properties and builds upon insights from the principal-agent literature.

Incentive plans in infrastructure industries are quite complicated. Academic researchers often use highly stylised characterisations of regulation to allow us to highlight the strengths and limitations of alternative regimes and sharing rules. However, the buy-ins (initial prices, plant modernisation mandates, and plan durations) accompanying the actual incentive plans are important aspects of transitional deregulation. Nevertheless, it is easy to demonstrate the deficiencies of the bottom-up cost-plus approach. Regulatory micro-management cannot induce efficient production. A top-down 'price caps and sharing rules' approach represents an improvement over cost-based regulations — though problems still exist (Weisman, 1994).

There is some evidence that the transition to more competitive markets has often involved new types of regulation rather than less regulation in the evolution to new industry structures. Numerous contentious issues are still being fought in hearing rooms in the U.S. and debated around the world: funding universal service, maintaining

network interoperability, ensuring service quality, developing number portability, continuing supplier of last resort obligations, determining the appropriate extent of unbundling, establishing interconnection charges, and designing efficient prices for network components. In telecommunications, the convergence of voice, data, information services and video markets raise numerous issues regarding entry, service quality and interconnection. Spectrum, fibre optics, and other technologies make it difficult to predict optimal configurations of firms in an industry. These issues warrant attention as they illustrate the complex problems facing policy-makers. In particular, entry policy represents a potential substitute for price and profit regulation. The competitive route to disciplining market power is being followed in many countries — with respect to electricity generation, natural gas, and telecommunications.

Guidelines and principles

There is no simple and comprehensive roadmap for policymakers in this decade of dramatic change. Rough maps of relatively unexplored territory are bound to contain errors and omissions. Mistakes will be made — some turns to the left or right will lead to dead-ends. Then politicians will have to retrace their steps or strike out over uncharted territory. The regulatory lessons from the U.S. suggest that the politics and economics of infrastructure industries are complex. The best decisions are those which are based on reality. However, incumbents, potential entrants, and consumers have different views of technological realities. From the standpoint of public policy, that which seems familiar is not necessarily appropriate in new territory.

The central message of this overview has been the need to design institutions which promote efficiency. The U.S. is slowly abandoning rate of return on rate base regulation, though it sometimes seems like two steps forward and one step back. The competitive forces unleashed by new technologies, court rulings, and new legislation can be channelled but not totally diverted. Entrants become stakeholders, and while they too attempt to manipulate the political system to their advantage, incumbent suppliers will not depart the field of battle without a fight. If the incentives are such that least cost suppliers win markets and those who introduce valued new services obtain profits, then the economy as a whole is the winner.

We have learned some principles that can be useful in new situations. However, in some instances, the art of policy development requires compromise. Pricing of access to essential facilities is probably the toughest issue confronting regulators in the U.S. and elsewhere. Arrangements have emerged in some state jurisdictions, based on a blend of economic principles or political compromise. Other principles have been enunciated for the regulatory transition in energy and telecommunications. In his survey of incentive regulation, Sappington, 1994, identified ten guidelines for designing incentive regulation plans:

- use incentive regulation to better employ the firm's superior information;
- prioritise regulatory goals and design incentive regulation to achieve stated goals;
- link the firm's compensation to sensitive measures of its unobserved activities;

- avoid basing the firm's compensation on performance measures with excessive variability;
- limit the firm's financial responsibility for factors beyond its control;
- adopt broad-based performance measures where possible, unless their variability is excessive;
- choose exogenous performance benchmarks;
- allow the firm to choose among regulatory options, while recognising the interdependencies among the regulatory options that are offered to the firm;
- promise only what can be delivered, and deliver whatever is promised;
- plan for the rare, unforeseen event, but minimise after-the-fact adjustments to the announced regulatory policy.

As Sappington points out, ' ... the design of sound, effective regulation in particular settings will require careful attention to the idiosyncratic features of the environment. The best incentive regulation plan in any given setting will vary according to regulatory goals, institutional and technological factors, the nature of the information asymmetry between regulatory and firm, and the commitment abilities of the regulator.' (p. 269).

Performance-based incentive mechanisms are emerging in state regulatory jurisdictions; these new initiatives include price caps, revenue caps, yardstick regulation, and profit sharing. Some of these alternative rate plans have emerged as a result of a collaborative process involving the participation of major stakeholders. The design issues (such as those noted earlier in conjunction with price caps) require the resolution of a set of interrelated problems. Attention to fundamental economic principles has strengthened new regulatory initiatives.

Some read the history of regulation, and conclude that new initiatives are **not** called for. Shepherd, 1992, states, 'The 1980s search for a mechanical, automatic method of 'incentive regulation' was largely illusory. In complex situations, there is no easy substitute for sophisticated, effective regulation.' (p. 71). In contrast to Shepherd, Strasser and Kohler, 1989, describe the overlapping command and control mechanisms comprising cost of service regulation as tools which are ' ... at best blunt and crude, preventing the worst abuses, but not sharp enough to encourage anything better. An incentive approach promises more.' (p. 137). Later, they state, 'Controls can keep managers from doing specific things, but they cannot command managers to use management processes energetically and creatively to tackle the problem of more efficient operation, although improved processes are essential to improved performance.' (p. 169). Movements away from cost-of-service regulation are illustrated by profit sharing via banded returns and various forms of incentive regulation. Generalised incentive regulation could be characterised as decoupling prices from costs via new regimes, such as yardstick regulation or price caps. As regulators move away from command and control micro-management, they are lowering entry barriers and utilising incentive regulation in those markets with residual market power.

In summary, note that making firms more profit-driven (reducing incentives for 'abusive behaviour') can be a win-win situation, although some customers might be worse off under price caps and sharing rules. For example, firms may be less tolerant

of non-paying (generally, low income) customers and their willingness to promote environmental investments may be reduced. Electricity demand-side management programs also suffer to the extent that they tend to decrease net case flows to the firm. Many observers would conclude that these perceived negatives do not outweigh the efficiency gains from adopting carefully-designed incentive regulation. Going beyond existing plans, some industry researchers (such as Blackmon) would provide the utility with 100 per cent of changes in profits (on the margin) so that incentives are not capped (as with many state plans). To address the regulatory commitment problem, he proposes that plans be established for a fixed time period, with regulators deciding halfway through a plan whether it should be rolled over for another term. Whether the accompanying rate freeze sufficiently promotes allocative efficiency is another question. Finally, when price cap regimes have been implemented, the review processes resemble rate of return regulation — strong efficiency incentives end up being balanced against consumer calls for sharing in the cost savings.

Bibliography

Beesley, M. E. and S. C. Littlechild, 1989, 'The Regulation of Privatised Monopolies in the United Kingdom,' *Rand Journal of Economics*, 29, pp 454–472.

Blackmon, Glenn, 1994, *Incentive Regulation and the Regulation of Incentives*, Kluwer Academic Publishers, Boston, pp x + 133.

Braeutigam, R. and J. Panzar, 1989, 'Diversification Incentives under "Price-Based" and "Cost-Based" Regulation,' *Rand Journal of Economics*, 20, pp 373–391.

Lewis, Tracy and Chris Garmon, 'Fundamentals of Incentive Regulation.' PURC/World Bank International Training Program on Utility Regulation and Strategy, June 1997.

Sappington, David E. M., 1994, 'Designing Incentive Regulation,' *Review of Industrial Organisation*, 9, pp 245–272.

_____ and Dennis L. Weisman, 1994, 'Designing Superior Incentive Regulation: Accounting for All of the Incentives All of the Time' and 'Modifying Plans to Preclude Recontracting and Promote Performance,' *Public Utilities Fortnightly*, 15 February and 1 March.

Shephard, William G., *Regulation and Efficiency: a Re-appraisal of Research and Policies*, National Regulatory Research Institute (NRRI) 92–14, July 1992.

Strasser, Kurt A. and Mark F. Kohler, *Regulating Utilities with Management Incentives: A Strategy for Improve Performances*, Quarum Books, New York, 1989, ix–197.

Weisman, Dennis L., 1994, 'Why Less May Be More Under Price-Cap Regulation,' *Journal of Regulatory Economics*, 6, pp 339–361.

Weyman-Jones, T. G., 1990, 'RPI-X Price Cap Regulation: The Price Controls Used in UK Electricity,' *Utilities Policy*, 1, pp 65–77.

Principles of price cap regulation

Stephen P. King, Department of Economics, The University of Melbourne, Australia.

The background to price cap regulation

Price cap regulation, as an alternative to traditional rate-of-return regulation, developed as a practical regulatory tool in the early 1980s in Britain. While both the undesirable features of rate-of-return regulation and the ideas underlying price caps were understood well before this time, it was in the UK in the 1980s that price caps were first widely used as an alternative regulatory rule for limiting abuse of market power[1].

Following a Mergers and Monopolies Commission (MMC) report in 1982, price cap regulation was used in Britain to limit abuse of market power by a dominant supplier of contraceptive sheaths (Armstrong, Cowan and Vickers, 1994). The MMC argued against traditional rate-of-return regulation on the grounds that (1) it weakened incentives for cost efficiency, (2) for a multi-product firm where only some products were regulated, rate-of-return regulation involved arbitrary allocations of cost and assets and (3) it was difficult to set an appropriate rate-of-return for the regulated company.

The newly privatised British Telecom (BT) was regulated by price caps after the recommendations of a report by Stephen Littlechild in 1983. In his report, Littlechild argued that price cap regulation would give BT desirable incentives to achieve and improve productive efficiency, while reducing the information burden of regulation. Unlike rate-of-return regulation, price caps do not require imprecise and often arbitrary measures of a rate base or return on capital, and eliminate the need to allocate costs when only some parts of a firm are regulated. He also argued that the simplicity of price caps would reduce the likelihood of regulatory capture.

Since the early 1980s, price cap regulation has been adopted in a wide range of countries. In Australia, price caps have been used in the telecommunications, energy and transport industries. For example, the newly privatised airports in Australia are subject to price cap restrictions on some of their services.

What is a price cap?

In its simplest form a price cap simply sets a maximum allowed inter-temporal path for the price of a specific product. The rules for the path are set in advance and only depend on factors that are beyond the control of the regulated firm. For example, the price of a specific product in any given year may be capped at a level which alters over time in response to a price index that is exogenous to the regulated firm and a factor set in advance by the relevant regulator. For example, a standard price cap in the UK (an RPI-X cap) or Australia (a CPI-X cap) involves the regulator setting an initial

---

[1] For example, the possibility of overcapitalisation under rate-of-return regulation was formalised in the economics literature in the early 1960s. See Averch and Johnson, 1962.

maximum price. This maximum price then rises in line with the main index of retail prices (the retail price index in the UK and consumer price index in Australia), but falls at a rate X set in advance by the regulator.

The value of X is meant to reflect potential cost savings by the firm due to either increased efficiency or technological progress. The X factor enables these cost savings to be shared with consumers without adversely effecting the incentives of the firm to minimise and achieve these savings, between review periods. The value of X and the absolute level of the price cap index are reviewed at set intervals. For example the review period for BT was originally set at five years.[2]

In some cases it may be considered that the utility should be allowed to increase its maximum allowed price at a rate faster than the rate of increase of the relevant price index. For example, privatised UK water companies are regulated by an RPI+K system that allows prices to rise at a rate faster than the retail price index.

In practice, price caps tend to be more complex than simply a set price path on a single product. Many regulated firms produce multiple products and these products may be bundled together in the price cap. The price cap may automatically adjust for exogenous changes in specific prices that have strong implications for the profitability of the regulated firm. For example, a price cap on a gas supplier may allow that firm to increase its price if there is an increase in the purchase price of the gas that it buys or in an index of gas prices. Price regulation may have associated regulation covering service quality. These additional features need to be designed to retain the basic desirable incentive properties created by a price cap.

Price cap regulation and efficiency

Why does a price cap promote efficient production? Suppose a regulated firm produces one product. For example, consider a gas retailer/distributor that only sells gas to residential consumers. We will consider two potential regulatory schemes. First, the firm may operate under a regulatory regime that limits profits to a set level, say no greater than $1 million. Alternatively, the firm may operate under price cap regulation. It can set the price of the gas subject to a cap of 5 cents per MJ. At this price it can sell 100 million MJ resulting in $5 million. Firm profits are given by revenue less the costs of production. These costs include billing and servicing customers, routine and emergency maintenance of the distribution system and, most importantly, the cost of wholesale gas. If the costs of retailing and distributing 100 million MJ of gas are (initially) $4 million then the firm will make $1 million in profit under the price cap. In other words, the firm initially makes identical profits under either regulatory regime with identical levels of output and costs of supply.

Suppose the owners and managers of the private gas firm can undertake a variety of activities that effect costs. For example, they can investigate and implement improved

---

[2] Armstrong, Cowan and Vickers, 1994, refer to the length of time between reviews as the period of regulatory lag. Note also that the value of X need not be constant between reviews. Rather a simple price cap requires the time path of X to be fixed in advance by the regulator at each review.

systems of billing, monitor the performance of maintenance teams and seek out cheaper sources of gas. The incentive for owners and managers to pursue activities that reduce costs will depend on the rewards that they receive from any cost reduction. If they receive the full benefit of any cost reduction, then they will have strong (and socially desirable) incentives to pursue activities that reduce costs. In contrast, if they receive little benefit from activities that reduce costs, they will have no incentive to engage in these activities.

Consider profit regulation. If the owners and managers reduce costs (and increase profits) under this regulatory regime, then this will have to be offset by a reduction in profits, for example, by lowering the price of gas. Any benefits of cost reductions are passed immediately on to gas consumers, but this will destroy the incentives for owners and managers to seek out ways to minimise costs. Consequently, there is little incentive for the firm to operate efficiently and engage in minimum cost production under profit regulation.

In contrast, under price cap regulation, any cost reduction achieved by the firm's owners and managers will be retained by the firm. If the managers and owners are able to reduce the costs of supplying 100 MJ of gas from $4 million to $3 million, then the entire $1 million reduction in costs will be translated into increased profits. Under price cap regulation, the firm can still sell gas at 5 cents per MJ so the 25 per cent reduction in costs will result in a 100 per cent rise in profits. Clearly, under price cap regulation there are strong incentives for the firm to operate efficiently and seek out cost minimising production techniques.[3]

Because the firm retains the benefits of cost reductions under price cap regulation, this regulatory regime provides powerful efficiency incentives. In reality, these incentives may be reduced by a number of factors. If the wholesale price of gas makes up a substantial part of the regulated firm's costs, then it may be desirable to allow the firm to pass through any changes in this wholesale price to customers. In the absence of such pass-through, the gas distributor could be driven into bankruptcy by a sudden rise in the price of gas that pushes its costs above the price cap. However, if gas costs can simply be passed through to the consumer, then the firm will have reduced incentives to seek out the cheapest sources of gas. In general, any allowed cost pass through under the price cap will reduce incentives to minimise costs.

Price caps are never set 'forever'. In some industries, price caps may be used as a one shot regulatory tool while competition emerges. More generally, price caps are reviewed and reset at regular intervals.[4] When caps are reset, attention is inevitably paid to the profits of the regulated firm and cost savings gained between reviews, at least partially, are passed on to consumers. Any 'allowed profit' aspect to price cap reviews will tend to reduce the incentives of a regulated firm to reduce costs between review. Further, to the degree that the regulator uses past performance to set the future

---

[3] For a more formal, technical presentation of this material, see Sappington, 1996.

[4] In Britain, where price caps were initially set for BT as a transitional procedure, the caps have been both reviewed and extended over time as the expected level of competition has not emerged. See Armstrong, Cowan and Vickers, 1994.

price cap, all the administrative issues associated with traditional rate-of-return regulation, such as cost allocation rules, re-emerge.

Even given the caveats above, price cap regulation will provide more incentives for efficient production than traditional profit or cost based regulation. At the same time, the design of the price cap regime will effect these incentives and the regulatory design of price caps must carefully consider the trade-off between providing incentives to reduce costs and the desire by regulators to pass on any cost savings to customers.

CPI-X

The regulated price cap is designed to adjust automatically between regulatory reviews. This occurs in two ways. First, the regulator sets the X-factor which determines the rate of decrease of the cap. The X-factor provides a way for the regulator to allow consumers to benefit from cost reductions and improvements in productive efficiency under price cap regulation without diminishing the incentives for managers and owners to undertake activities that create these efficiencies.

The regulator sets the X-factor to reflect (i) expected firm productivity improvements in excess of those expected for the general economy and (ii) expected changes in input prices for the regulated firm that differ from the general economy-wide rate of price change. When setting the value of X, the regulator often has to rely on historic information about performance of the regulated firm or of other firms in similar industries. The regulator may gain information about productivity improvements from profit studies based on the past rate-of-return achieved by the firm. This raises a number of important issues.

- If the firm has previously been regulated by profit or cost based regulation, or has been inefficiently operated under public ownership, then future productivity improvements may significantly exceed historic rates. It may be desirable for the regulator to set a relatively high value of X, particularly in the first few years of the price cap regime.

- If productivity improvements achieved by the firm exceed the X factor by a substantial amount then the firm will make significant profits and there may be pressure on the regulator to adjust either the value of X upwards or the base level of the price cap downwards between review periods. Such regulatory opportunism generally should be avoided. The reputation of the regulator in abiding by the regime between reviews will be an important input to the success of the regulations. If the firm believes that any successful reduction in costs below the amount allowed for in the X factor will simply lead to a reduction in the price cap, then the managers and owners will not seek to achieve these extra cost savings.

- If rate-of-return or other profit measures are used to reset the base value of the price cap or the future path of X when these values are reviewed, then this will feed into the incentives facing owners and managers in the period leading up to the review. In particular, the incentives for efficient production will tend to fall as the review approaches if managers and owners know that these cost savings will be taken from them in the review process. The regulator can use a number of

simple devices to help reduce this problem. For example, the regulator can base the price cap review on industry wide performance rather than firm specific performance. This may involve explicit benchmarking between similar regulated firms. The regulator may also commit in advance to using a 'glide path' to adjust the price cap at a review. This means that rather than using a one-shot revaluation to set the new price index at the review, the regulator sets a path by which the base must move down to its new level over time. Thus the price cap will be reduced by the X factor (reflecting future expected productivity gains) and the glide path factor (reflecting past gains above the X factor) between reviews.[5]

The regulator will also need information about input price changes that are idiosyncratic to the regulated industry or firm. Again, this information may be gained from historic data, subject to similar caveats as presented above.

The approach to resetting X will depend both on how the regulator evaluates profits compared to consumer benefits, and the relative importance placed on allocative and productive efficiency. The higher the relative weight placed on consumer benefits relative to profits, the more the regulator will wish to claw back excess profits as they emerge. At the same time, if the regulator seizes profits created through productivity gains in an opportunistic way, then this will simply reduce the incentive for these gains to be made in the future, making both the firm and the consumers worse off in the longer term.

Even if the regulator evaluates profits and consumer benefits equally, it is still desirable to have a positive X factor and to adjust the cap at regular reviews so that excess profits are eventually returned to the consumer. Pricing above true marginal cost (or above a minimum sustainable price) will lead to a loss of allocative efficiency by eliminating potentially mutually beneficial trade. When considering how quickly excess profits should be removed by tightening the price cap, the regulator must weigh up these allocative benefits with the potential reduction in productive efficiency due to weakened incentives for firm owners and managers.

An alternative to a fixed X factor between reviews is a scheme of earnings sharing. This form of price cap, which is more popular in the US than in either Britain or Australia, requires the regulated firm to share any 'excess' earnings with consumers. However, such a sharing scheme reduces efficiency relative to pure price caps.[6] Furthermore, it may raise the costs of regulation by requiring continual monitoring of firm returns.

---

5 For a summary on the incentive aspects of different methods to 'claw back' excess profits, see Cave, 1997.

6 Of course, to the degree that regulators implicitly use sharing schemes when reviewing price caps, there may be little practical difference between the incentives under a scheme of explicit earnings sharing and under an actual price cap regime. In fact, incentives may be improved by making the form of sharing explicit rather than implicit. I thank Mark Jamison for pointing this out to me.

The price cap also adjusts for increases in general input prices over time. Both the adjustment for general and idiosyncratic increases in input prices must be beyond the control of the firm to avoid reducing the incentives to search out the cheapest input prices or creating incentives for managers to manipulate the input price index. It is desirable to use a general inflation measure to adjust the price cap. In practice, it is standard to use a measure of consumer prices to adjust the price cap (the CPI). Such an index is clearly beyond the control of almost any regulated firm. Any differences between the rate of increase in the consumer price index and the rate of increase in the prices of the inputs for the regulated firm should be included in setting the value of X.

Even if the rate of general inflation is used to reflect general economy wide effects on input prices, and idiosyncratic increases in input prices are built into the setting of X, there may still be deviations between the actual change in input prices and those allowed for in the price cap. To prevent firm distress caused by increases in specific input prices that are beyond their control, it may also be desirable to adjust the price cap to allow for the pass through of specific input prices. This is particularly the case when some input prices make up the bulk of the regulated firms costs. For example, 95 per cent of the costs of electricity suppliers in Britain are derived from the costs of generation, transmission and distribution and these costs are passed on directly to the consumers through the price cap (Armstrong, Cowan and Vickers 1994).

Where pass through is allowed, it is desirable to base it on a price index rather than the price paid by the regulated firm, to create appropriate incentives. For example British Gas can pass through an index of wholesale gas prices rather than its particular purchase costs.

Other factors may also be used to adjust the price cap between reviews. For example, where the price cap embodies specific assumptions that can be confirmed at a later date, then the cap may adjust if the initial assumptions are proved false (Cave, 1997).

Multiple products

A variety of forms of price cap are used when a regulated firm produces multiple products. Some of these are presented below.[7]

Suppose that the firm produces two products. Say that the price cap is set at the end of year zero and is to be reviewed at the end of year five. For the present, we ignore both the CPI and X-factor adjustments. The firm sets prices for its products in each year *t* subject to the price cap. Let the prices set by the firm in year *t* (where *t* is 1, 2, 3, 4 or 5) be denoted by p_1^t and p_2^t. The firm sells the amount of each product demanded at

[7] For example, Taylor and Weisman, 1996, present a slight variant on the current quantity average revenue regulation formula presented below where *p* refers explicitly to average revenue rather than a uniform price and the fixed price entering the right hand side of the formula is set each year at the previous years average revenue for each product.

these prices, $q_1^t\left(p^t\right)$ and $q_2^t\left(p^t\right)$ (where p^t represents the two prices p_1^t and p_2^t). The firm's revenue in year *t* will be $p_1^t q_1^t + p_2^t q_2^t$.

Fixed weight price cap. The firm can set prices p_1^t and p_2^t in year *t* so long as $p_1^t \overline{q}_1 + p_2^t \overline{q}_2 \le \overline{p}_1 \overline{q}_1 + \overline{p}_2 \overline{q}_2$. The values of $\overline{p}_1$ and $\overline{p}_2$ may be set at their *t* = 0 levels with $\overline{q}_1 = q_1^0\left(p_1^0\right)$ and $\overline{q}_2 = q_2^0\left(p_2^0\right)$.

Average revenue regulation (current quantities). The firm can set prices so long as $p_1^t q_1^t + p_2^t q_2^t \le \overline{p}_1 q_1^t + \overline{p}_2 q_2^t$.

Average revenue regulation (lagged quantities). The firm can set prices so long as $p_1^t q_1^{t-1} + p_2^t q_2^{t-1} \le \overline{p}\, q_1^{t-1} + \overline{p}_2\ q_2^{t-1}$.

Tariff basket regulation. The firm can set prices so long as $p_1^t q_1^{t-1} + p_2^t q_2^{t-1} \le p_1^{t-1} q_1^{t-1} + p_2^{t-1} q_2^{t-1}$.

Under each of these formulae the firm retains some ability to rebalance prices. In other words, the firm may raise the price of one product and reduce the price of another product without violating the cap. The fixed weight price cap is administratively simpler than the other caps and limits any possibility for the firm to manipulate the cap by setting the quantity weights on the left hand side of the formula and the allowed revenue on the right hand side of the formula at a fixed level that is invariant between reviews. However, this is also the least flexible cap and may prevent the firm from designing price changes that are both profitable and in the consumers' interest.

Tariff basket regulation allows the firm to both alter prices in response to demand changes and to rebalance prices in a way beneficial to both the firm and the customer. To see this, note that under tariff basket regulation, consumers in any year *t* can always buy the same bundle as they did in the previous year (*t* – 1) for the same total cost. However, to the degree that prices have changed between year *t* – 1 and year *t* the consumers may prefer to buy a different bundle of goods. As they can still buy the original bundle at the same cost but may prefer to buy a different bundle, consumers cannot be made worse off by the rebalancing.[8] If the firm finds it profitable to rebalance prices then this reflects a mutual gain — both the firm and consumers are at least as well off under rebalanced prices as under original prices.

[8] Of course, if consumers' income alters between years, consumers may be better or worse off regardless of any price rebalancing by the firm.

Tariff basket regulation has been used for BT and the British water companies. Sappington (1997) presents a variant of tariff basket regulation designed to avoid penalising a firm for pricing below authorised levels in any year.

Average revenue regulation with current quantities requires the regulator to forecast demand in advance and to establish procedures to compensate either the firm or consumers if these forecasts are in error. Using lagged quantities avoids this problem and lagged quantity average revenue regulation was used in the US to regulate AT&T. Sappington and Sibley (1992) show that if two-part tariffs are incorporated into this scheme (interpreting one of the products as 'access') then rebalancing may lead to a reduction in both consumer surplus and total welfare over time.[9]

The above formulae ignore the adjustment factors for input prices and productivity improvements. These need to be added into the formulae. For example, tariff basket regulation is $p_1^t q_1^{t-1} + p_2^t q_2^{t-1} \leq [1 + CPI - X]\left(p_1^{t-1} q_1^{t-1} + p_2^{t-1} q_2^{t-1}\right)$.

Care must be taken under price cap regulation to avoid rebalancing, for example, to prevent or delay entry into one part of the market. In part, this issue overlaps with the question of which prices should be included in the price cap. If competition is expected to emerge for one product but not for another, then usually only the product that is not going to be subject to competition should be included in the price cap.

Price caps often incorporate additional restrictions on particular prices. This may be required for equity reasons, for example, to avoid the rapid dismantling of historic cross-subsidies to certain consumer groups. The individual products may be constrained by a separate specific price cap or may be allowed to move within a certain price band. The band, involving both a ceiling and a floor on the relevant price, may be designed to limit the rate of change of a specific product price.

Conclusion

Price cap regulation offers substantial benefits for regulators, consumers and regulated firms if applied with care. However, if price caps are poorly designed and are subject to arbitrary re-evaluation based on firm performance, then they open the scope for potential abuse by the regulated firm and may be worse than traditional profit and cost based regulation.

This paper has presented a brief overview of the theory of price caps and the practical problems that need to be addressed when establishing price cap regulation. Three points need to be emphasised. First, when designing a price cap, the regulator must carefully consider the bundle of goods and services to be covered by the cap. If the bundle is poorly designed, then the regulation may be subject to potential anti-competitive abuse. However, if the price cap provides too little flexibility to firms,

---

[9] See also Armstrong, Cowan and Vickers, 1994, on the potential for rebalancing that harms consumers under average revenue regulation.

then opportunities to rebalance prices for both consumer and firm gain will be limited. Second, the regulator needs to carefully design the review process. Where possible, information used in a price cap review needs to be beyond the control of the regulated firm. If reviews are based on realised profits, then the price cap may degenerate to standard rate-of-return regulation. Thirdly, regulatory credibility is crucial to the success of the price cap. If the regulator sets an X-factor that is too low, then the regulated firm will make large profits. However, if the regulator uses current profit information to arbitrarily revise up the X-factor between reviews, the regulated firm will have little incentive in the future to pursue the cost savings that lead to those profits.

Bibliography

Armstrong, M., Cowan, S. and Vickers, J., 1994, *Regulatory reform: economic analysis and the British experience*, MIT Press, Cambridge.

Averch, H. and Johnson, L., 1962, 'Behaviour of the firm under a regulatory constraint', *American Economic Review*, 52, pp 1052–69.

Cave, M., 1997, 'Aspects of the revision of price caps in the electricity supply industry', Report to the Energy Projects Division, Department of Treasury and Finance, Victoria.

Sappington, D., 1997, 'Basic elements of price cap regulation', mimeo, Department of Economics, The University of Florida, Gainesville.

Sappington, D. and Sibley, D., 1992, 'Strategic nonlinear pricing under price-cap regulation' *RAND Journal of Economics*, 23, pp 1–19.

Taylor, L. and Weisman, D., 1996, 'A note on price cap regulation and competition', *Review of Industrial Organisation*, 11, pp 459–71.

Benchmarking infrastructure enterprises

Denis Lawrence, Tasman Asia Pacific, Canberra, Australia.

The development of international performance measures offers a way of indirectly introducing competitive pressures to infrastructure industries by comparing actual performance to international benchmarks. Efficiency measures for infrastructure services can be used by enterprise managers, government, regulators and business users to identify areas where there is potential to improve performance. They are an ideal tool for promoting so-called 'yardstick competition'. By providing a means of comparing performance between similar infrastructure enterprises in different regions, states, or in different countries, efficiency measurement can promote attainment of best practice.

Before we can assess the merits of different performance measures, it is necessary to first consider why we bother to measure performance and what the purpose of management and business information systems is. These systems are usually put in place to overcome problems associated with so-called 'principal/agent' relationships. However, the nature of these principal/agent relationships varies across different types of organisations.

The role of performance measurement

In most business enterprises a good information system is required to enable those who own the enterprise, the principals, to control the actions of those they employ to run the enterprise, their agents. This permits those who supply capital to the enterprise to judge its performance and take action to correct poor performance.

In the private sector this line of accountability is relatively clear-cut, as is the overall objective of the enterprise — to maximise the net present value of profits. Investors have an incentive to continuously assess whether to buy, sell or retain equity in the enterprise and have an effective and immediate sanction to use against poor management in the form of a withdrawal of funds. Competitive pressures ensure that poor performers do not survive in the long run.

Relatively simple financial ratios and, more recently, discounted cash flow analysis have formed the principal monitoring devices in the private sector. The ratio of current assets to current liabilities has long been used as an indicator of potential financial distress. As more sophisticated techniques have become available we have seen the emergence of specialist intermediaries whom investors can employ to advise them on financial performance. The emphasis is now for enterprises to disclose comprehensive information in their financial statements and for investors and their advisers to form their own measures of profitability.

In the public sector the situation is considerably more complex. There are several dimensions to the principal/agent relationship for government enterprises, lines of accountability are much less clear cut and there is often an absence of competition. Furthermore, the relationship between the organisation and its customers is more

complex in the case of government enterprises. Whereas in the private sector the customer is normally assumed to be kept informed of output quality and options by advertising and competition, in the public sector the customer typically does not have access to alternative suppliers and comparative data may form the only basis for judging output quality.

Performance for any enterprise will be multi-faceted. However, the complexity of this for public enterprises providing infrastructure services is likely to be much greater leading to the need for more complex performance measurement information systems than those typically found in the private sector. As a result, rather than simply focusing on measures of profitability and potential financial distress, performance measurement of government enterprises has tended to focus on measures associated with inputs, processes, outputs and outcomes.

Performance measurement is playing an increasingly important role in the regulation of infrastructure enterprises. They provide the regulator with a means of assessing whether the enterprise is operating as efficiently as possible, whether cost savings are being passed on to users of the service and whether the service provider is exploiting any monopoly power they may have.

Before considering the various measures available, we will first review some of the desirable characteristics of a set of performance measures.

Considerations in choosing performance measures

The principal criteria for choosing a set of performance measures are that they be comprehensible, comprehensive, useable and timely. The main requirement for the set of measures to be comprehensible is that they be relatively few in number. It is no use having hundreds or even dozens of measures presented, as no-one will be able to distil the essential information on the enterprise's overall performance. This reinforces the need for and importance of comprehensive summary measures. While particular detailed technical measures may provide important information on specific aspects of operation, it is important to obtain an overall indication of how the enterprise is performing relative to its peers or comparable organisations in other industries. A large number of technical or, for that matter, easily calculated partial productivity measures may in total provide an accurate picture but will typically be difficult to comprehend readily and may create confusion if some indicators move in opposite directions.

Timeliness is of the essence in any performance measurement exercise to provide feedback as close to the event as possible. 'Historical' studies of performance will be of interest but are less helpful to management than current information. Current information also makes it more difficult for managers to dismiss poor performance observed some time ago with the line that the problem has now been remedied, even if it in fact still persists. Government enterprise information systems are generally being upgraded to provide both better quality and more timely data.

These considerations point to the need to have a small number of comprehensive indicators of overall performance in the first instance. These should be followed up

with some technical or partial measures to give details of performance in particular areas and hence an explanation of why the overall result may be either good or poor. These supplementary measures may also be able to provide more up-to-date information than the comprehensive indicators, which may only be available on an annual basis.

Equally important, however, is the consideration that the measures be useable. This means that the intended audience must understand them and how they can be influenced by obvious actions. This points to the need for relatively simple measures which are easy to interpret and for which the audience will understand what actions would lead to observed improvements.

One of the problems in getting ownership of a set of performance measures, particularly ones used internally within an enterprise, is that individual managers may feel that they have little influence over the result because of factors beyond their control. Within the enterprise this does not reduce the need for comprehensive summary measures of overall business unit performance but it does point to the need to perhaps supplement these summary measures by specific measures concentrating on factors under the unit's control. There is, thus, a need to tailor the set of measures to the purpose or audience at hand.

When making comparisons between enterprises at the aggregate enterprise level there is also a need to make adjustments for factors beyond management's control such as differing climatic, geographic or demographic influences. This highlights the importance of comparing like with like. This could be achieved by comparing units of organisations that conduct largely similar functions under similar conditions. Alternatively, performance measures can be adjusted for 'operating environment' characteristics beyond the control of management.

What should we compare performance with?

The forms of comparison most commonly used are targets, time-series, between business units and external comparisons. The setting of targets to be achieved or reached in various performance measures has often been advocated. However, target setting often encounters ownership problems as there may be a feeling that the target was imposed from outside by people who do not understand the organisation and it is hence unrealistic.

Time-series comparisons of performance within the one enterprise provide a more useful indication of changes in performance but still fail to put performance in appropriate context relative to other enterprises. For instance, while it is always desirable to have rapid improvements in performance through time, less comfort can be taken from this if the enterprise is starting from a low base relative to its peers. It may also be difficult to judge whether abnormal results in one year are because of actions affecting only that enterprise or because of more general considerations. Technological change may also make it difficult to interpret results for one enterprise in isolation through time.

Comparisons between comparable business units within the enterprise form an important basis of so-called 'yardstick competition'. Provided the measures are comprehensive and readily communicated to the shop floor level then a useful incentive mechanism can be introduced where incentives or competition may have previously been lacking.

External comparison of an organisation's performance with that of its peers, both domestically and internationally, provides the best basis for comparison. This highlights where the organisation stands in regard to international best practice and places the focus on the need to improve the performance of the organisation in question. Combined time-series, cross-section or 'panel' data provide the fullest picture possible by showing not only how the organisation compares to others at the moment but also how performance has changed through time. Normally we would expect those organisations starting from the lowest bases to have the most rapid rates of improvement as there are more options available to them to increase performance and catch up to industry leaders. Industry leaders, on the other hand, may have quite low rates of improvement because they have already exhausted available avenues for efficiency improvement.

Finally, it is important to bear in mind that a set of performance measures will inevitably evolve through time. As experience is gained in forming and using the measures the enterprise will refine the set to meet its requirements. Enterprises may start off with relatively simple measures initially and progress to more sophisticated measures as experience is gained and better quality data becomes available. Comprehensive summary measures will always remain important but the focus of supplementary partial or technical indicators may change through time as different performance problem areas receive attention.

With these considerations in mind we are now ready to examine various types of performance measures. For infrastructure services, these can generally be divided into three broad categories: accounting, non-financial and economic indicators.

Accounting measures

Accounting ratios are typically classified as: operating ratios, leverage and liquidity ratios, and market ratios. Market ratios are, of course, not relevant for government enterprises where stocks are not traded on the sharemarket. Operating ratios are concerned with performance in terms of assets employed, sales and expenses while leverage and liquidity ratios reflect the risks inherent in the particular financing policies of the enterprise. Accounting measures commonly used for infrastructure enterprises include the following:

- return on assets;
- sales to total assets;
- profit to equity or shareholders' funds;
- dividend payment to profit;
- debt to equity;
- current assets to current liabilities; and

- profit to interest.

The main advantage of accounting measures is that they can usually be calculated from information readily available in infrastructure enterprise annual reports and financial statements. However, caution needs to be exercised in their interpretation as not all enterprises use the same accounting conventions and a range of valuation methods are typically used for determining the value of assets. Furthermore, as outlined in the first section of this paper, performance monitoring of government enterprises is not as straight forward as for the private sector due to less clear lines of accountability and market imperfections or the absence of markets in some instances. This means that private sector monitoring mechanisms such as the use of accounting ratios may not be appropriate or may need to be supplemented.

Most infrastructure enterprises are moving towards accrual accounting and the use of current cost valuation methods. However, a number of government enterprises still use cash-based accounting concepts and the use of historic cost as the basis of valuing assets remains widespread. Historic cost valuations do not allow for the effects of inflation or technological obsolescence and may bear little resemblance to current asset values. While the use of current cost accounting methods is highly desirable, the economic rate of return (discussed later) should be used wherever possible as the return on assets measure since it more accurately reflects the enterprise's true performance.

Accounting ratio measures also assume that there are constant returns to scale for enterprises and suffer from the difficulties mentioned in the previous section of determining what the target value of the ratio should be. Using the industry average as a reference point instead will place less pressure on the enterprise to attain international best practice.

Care also needs to be exercised that accounting ratio results are not being improved at the expense of other aspects of performance. For instance, the return on assets could be being improved by an unsustainable running down of the capital stock or deterioration in service quality. For many government infrastructure enterprises the absence of effective competition or the threat of takeover may mean that no pressures are applied to correct this situation. Consequently, accounting measures typically have to be examined in conjunction with measures that provide information on other aspects of the enterprise's performance such as service quality.

Non-financial measures

Non-financial measures are used to supplement simple accounting measures and comprehensive summary measures. The composition of non-financial measures will usually need to be tailored to the enterprise and its industry but will typically cover three broad areas:

- efficiency;
- effectiveness; and
- service quality.

Non-financial efficiency measures usually consist of either simple partial productivity ratios such as labour productivity or technical, engineering measures of system performance. Examples of technical efficiency indicators include reserve plant margins, load factors and capacity factors in the case of electricity supply. Other efficiency measures include working time lost due to sickness, industrial accidents or industrial disputes, time required for maintenance and labour turnover.

The main advantage of these measures is that they provide additional detail to isolate areas of poor performance observed from overall summary measures of performance. However, they can prove misleading when used in isolation and should not generally be used for purposes of target setting as perverse incentives may be created with adverse effects on overall performance.

Effectiveness indicators are designed to measure aspects of the enterprise's performance set by often loosely defined policy or social objectives. They include environmental considerations, the uniformity of and access to services provided, indices of real prices and the number of accidents. The advantage of effectiveness measures is that they capture aspects of the complex requirements placed on government enterprises not reflected in more market-oriented measures. However, they should only be used in a supplementary capacity or perverse management incentives may again be created.

As outlined above, given the absence of competitive pressures facing many infrastructure enterprises, measures of service quality form an important adjunct to other performance measures. Service quality measures cover such matters as reliability, response time, accessibility and security. Examples include response times to enquiries and provision of service, rate of service difficulties and equipment failures. Surveys of customer satisfaction are likely to play an increasingly important role in determining the quality of service provision.

Economic indicators

Economic indicators provide the best comprehensive summary of an enterprise's overall performance. The two major economic performance indicators commonly used are Total Factor Productivity (TFP) and the economic rate of return (ERR). They can provide information on how the infrastructure enterprise is performing through time and how well it is performing relative to its peer enterprises. This information can be used to determine areas requiring improvement, as well as helping to determine appropriate pricing and investment policies.

Productivity measures

Productivity is a measure of the physical output produced from the use of a given quantity of inputs. All enterprises use a range of inputs including labour, capital, land, fuel, materials and services. If the enterprise is not using its inputs as efficiently as possible then there is scope to lower costs and increase profitability through productivity improvements. This may come about through the use of better quality inputs including a better trained workforce, adoption of technological advances, removal of restrictive work practices and other forms of waste, and better management through a more efficient organisational and institutional structure.

In practice, productivity is measured by expressing output as a ratio of inputs used. There are two types of productivity measures: TFP and Partial Factor Productivity (PFP). TFP measures total output relative to all inputs used. Output can be increased by using more inputs, making better use of the current level of inputs and by exploiting economies of scale. The TFP index measures the impact of all the factors effecting growth in output other than changes in input levels. PFP measures one or more outputs relative to one particular input (e.g. labour productivity is the ratio of output to labour input).

Partial productivity measures are widely used as they are simple to calculate. However, as noted in the preceding section, PFP measures should be interpreted with caution.

By concentrating on the productivity of one particular input a misleading impression of overall performance may result. Take, for example, a railway whose productivity is being measured by the number of passenger journeys per employee (i.e. the partial productivity of labour). A ten-year comparison of the results may indicate substantial productivity gains during the last four years. The source of these gains may be due to increased output or a decrease in labour input or a combination of both. Assuming that output increased by investing in faster and bigger trains (i.e. more frequent services and more seating capacity during peak hours) while keeping the number of employees constant, the observed improvement in labour productivity has most likely been achieved at the expense of a deterioration in the partial productivity of capital. To assess whether the railway has become more efficient overall we need to measure output relative to both labour and capital inputs. Only by using a measure of total factor productivity can the true picture of performance be obtained.

Similarly, we need to be wary of setting performance targets in terms of partial productivity measures. Managers can often meet the specified target by simply using more of the inputs not included in the measure (e.g. by substituting capital for labour if a labour productivity target was set). The net result overall may be a worsening of performance.

TFP indices calculated for one enterprise through time provide information on how its TFP has grown. However, information on TFP growth rates gives no indication of how close the enterprise is to being as technically efficient as possible or on how technically efficient it is relative to its peers. To obtain this information we need to know the enterprise's TFP levels as well as growth rates. This information can be obtained by extending measurement to look at a number of infrastructure enterprises providing similar services through time and calculating multi-lateral TFP indices which give information on both TFP levels and growth rates (Lawrence, Swan and Zeitsch 1991). The multilateral TFP technique provides a ready means of benchmarking enterprises both domestically and relative to international best practice (Swan Consultants (Canberra) 1991, 1992b; Zeitsch and Lawrence 1996).

Another measure of productive efficiency sometimes used is Data Envelopment Analysis (DEA) which calculates a synthetic efficiency frontier in a number of dimensions for comparable organisations with a common number of inputs and outputs. To provide maximum information DEA requires a large number of observations making it less likely to be a practical tool for an individual enterprise or small group of enterprises. TFP, on the other hand, can provide valuable information with only a few

observations. Because of its linear programming basis, DEA efficiency scores will be particularly sensitive to the number and accuracy of observations and efficiency scores can be increased by adding more categories of outputs and inputs.

DEA does not use prices as the basis for aggregating outputs and inputs. This may be desirable when looking at non-market activities such as health, education and local councils (the areas where DEA has mostly been applied) where it is difficult or impossible to attach prices to outputs but may lead to overly generous treatment of market sector enterprises with inefficient combinations of outputs or inputs.

However, all measures have their disadvantages and a number of useful insights can be obtained from periodically using DEA in large scale comparison exercises in conjunction with the ongoing use of TFP. Rather than being viewed as substitutes, the two measures are better used to complement each other.

The economic rate of return

The ERR is defined as the ratio of economic income to the opening market value of the enterprise's assets. Economic income is defined as the enterprise's cash flow (earnings before interest and tax plus accounting depreciation) plus the change in the market value of the enterprise's assets during the year. This measure overcomes the problems of accounting indicators based on historic cost and accounting rates of depreciation. Many studies have shown that the ERR can strongly differ from the accounting rate of return (Shinnar, *et al* 1989). The ERR is completely consistent with discounted cash flow analysis which provides the most rigorous means of determining economic or commercial value (Brealey and Myers 1984).

While measures such as TFP and DEA provide information on an enterprise's productive efficiency, the ERR indicates whether costs are being adequately recovered and hence provides a means of assessing whether the community is getting a reasonable return on the assets it has tied up in government infrastructure enterprises. The ERR also provides the best means of comparing government enterprise performance with that of the private sector.

Another useful economic measure of infrastructure enterprise performance is the enterprise's terms of trade or ratio of prices received to prices paid. If an enterprise faces a declining terms of trade — its input prices are increasing faster than the prices it receives or is allowed to charge for its output — then it must offset this by an equivalent rate of productivity growth if it is to maintain its returns to costs ratio. This concept lies behind the application of so-called 'CPI – X' rules.

Who uses performance measures?

Efficiency measures for infrastructure services can be used by a range of interested parties to identify areas where there is potential to improve performance. The main categories of users include:

- enterprise managers, who can use efficiency measures such as DEA to identify where there are gaps between actual and best practice levels of performance.

These measures can form a useful part of enterprise or business unit benchmarking initiatives;

- governments, which can use efficiency measures to identify and reward those service providers who are meeting their objectives most cost effectively. It also allows policy-makers to assess the impact of reforms, such as output-based funding or contracting-out of service provision;

- regulators, who can use efficiency measures to assess the scope for further efficiency improvements and whether infrastructure enterprises are exploiting monopoly power; and,

- the community and business users, who are able to use public information on the performance of different service providers to keep governments accountable and make decisions on their preferred supplier.

What's in it for enterprise managers?

Just as TFP measurement provides governments as regulators and owners of infrastructure enterprises with a ready means of assessing the overall performance of those enterprises and whether reforms are in fact being successful, it also provides infrastructure enterprise managers with a useful management tool. Efficiency measurement is an ideal tool for promoting so-called 'yardstick competition'. By providing a means of comparing performance between similar infrastructure enterprises in different states, or in different countries, efficiency measurement can promote attainment of best practice.

Comparing efficiency levels between similar enterprises will provide information on who is the best performer. This then prompts management in the other infrastructure enterprises to ask why the best performer is outperforming them. To find the answer to this question, infrastructure enterprises may have to engage in more detailed benchmarking exercises. This will enable them to identify the techniques being used by other enterprises and assess whether those techniques could usefully be adopted in their own operations to improve performance. A first step towards finding out why efficiency levels differ between similar enterprises is to compare the partial productivity levels of the various inputs being used. This will provide a focus for further investigations on those inputs that appear to be a problem for the particular infrastructure enterprise concerned.

An extension of this process is to link remuneration levels or, alternatively, the enterprise's entire funding level to performance. This provides an explicit incentive for both managers and staff to try and improve efficiency levels. However, such schemes have to be carefully designed so that the incentives provided are, in fact, compatible with sustained performance improvement. For instance, simply providing rewards on the basis of absolute improvements in performance would penalise those organisations already operating close to best practice. This is because those organisations starting from a low efficiency level can make easy catch-up gains but those operating close to best practice have already exhausted these easy gains and hence have less scope to improve performance. Linking rewards or funding to how much the proportional gap to best practice is reduced would remove some of this problem but special allowance

would still have to be made for those already at best practice. Failure to do this could lead to 'gaming' by the lead enterprises where they purposefully reduced their productivity one period in order to collect bonuses for improvements in the next period. Bogetoft, 1995, explores the link between incentives and performance measurement in detail.

There is often little competition within particular infrastructure enterprises for the provision of various services. Extending efficiency measurement downwards through the enterprise from major business units right through to a relatively disaggregated level provides a ready means of promoting internal 'yardstick competition'. It provides a focus for workers in those business units to take pride in their achievements, meet specified goals and outperform their counterparts. This can be used by management as a means of improving overall economic efficiency, involving employees to a greater extent and promoting participation of employees in putting forward ideas on how to improve performance.

It should be remembered at all times that the objective of performance measurement is not simply to form a report card and grade people but to help facilitate a program to improve efficiency. The best atmosphere is one where managers of individual enterprises or business units work together, share ideas and mentor each other. Most enterprises have at least some aspect of their operations that others can learn from and sharing information is the best way of transferring best practice.

When used properly, performance measurement and benchmarking encourage enterprises to become 'learning organisations'. This is a term coined by Senge, 1990, to describe organisations which have been successful in getting their employees to think and work as teams, adapt readily to change and take a system-wide view in problem-solving. By recognising that every enterprise can learn something from another's operations and highlighting the need to continually be improving performance to remain in front, benchmarking encourages a culture which emphasises the need to learn and have an open mind. It also promotes decentralisation of decision-making as production team members become involved in modifying processes to improve performance.

The philosophy behind benchmarking is well summarised by the Innovation Network (1993, p. 2) as follows:

> As a change management tool, benchmarking broadens the horizons of your management and employees by helping them realise that "there can be a better way." The successful benchmarker sheds the "not invented here" syndrome by ridding itself of such notions as "We can't learn anything from others," "We're as good as you can get," "There is no one outstanding," or "We're unique and can't be compared to anyone else." Instead, a philosophy emerges characterised by attitudes like "We'll borrow shamelessly," "We can learn something from anyone," and "We don't have all the answers." It forces people to talk openly and honestly with their counterparts.

Demonstrating that better performance is possible and is actually being achieved elsewhere also makes it easier to implement change in an organisation. If management can point to actual examples of superior performance or where different techniques have proved successful, then staff are less likely to dismiss proposed changes as being inappropriate or impractical.

Identifying major gaps in performance levels can also force an organisation to fundamentally rethink how it does things. There has been much focus recently on the 'continuous improvement' approach to managing government organisations but this approach typically limits managers to looking at small changes to what they currently do. If a quantum leap is needed rather than an incremental change then benchmarking can help managers focus on the need for a fundamental rethink of how the organisation operates.

Finally, we need to recognise that no single performance measure or technique can provide the whole answer. All quantitative analysis involves significant assumptions and limitations. Consequently, we should not rely solely on one measure but use the results of several sources, both quantitative and qualitative, to make an overall judgment of how a particular enterprise is performing and what needs to be done. It is important not to become too technique-driven and to always provide a 'sanity check' for empirical results to make sure they make sense.

However, conversely, we should not be afraid of applying and developing new techniques and approaches. In regard to infrastructure services we are often faced with a dearth of key data. This often leads to calls not to apply new approaches as the data is not of sufficient quality. But a useful start can usually be made on performance measurement with data that is currently available — waiting for the perfect data is a recipe for indefinite inaction and the use of available data is itself a catalyst for the development of better quality data. The appropriate management response in this situation is best characterised by the saying 'never underestimate the symbolic importance of actually doing **something**'.

Conclusions

The main conclusions from this paper are;

- infrastructure performance measures are used by a range of stakeholders including enterprise managers, governments, regulators, business users and the community;?
- performance monitoring is considerably more complex for government enterprises than for the private sector due to less clear lines of accountability, thin markets or the complete absence of markets in some instances;
- the most effective information systems will concentrate on a few comprehensive summary measures accompanied by a small range of specific measures which provide supplementary information on key areas of performance either not included in the summary measures (e.g. service quality) or which explain reasons for poor performance;
- the best comprehensive summary measures are the economic indicators of total factor productivity, data envelopment analysis and the economic rate of return;
- the most effective basis for comparison is the use of panel data, including leading overseas enterprises, which permits performance levels as well as rates of change to be compared relative to international best practice;

- the range of supplementary measures included is likely to evolve through time as more experience is gained, better quality data becomes available and perceived problem areas change;
- the specific measures used to supplement the comprehensive summary measures may vary depending on the level of business unit being examined and the purpose of the measurement exercise;
- for infrastructure enterprises it is important to include supplementary measures of service quality and effectiveness; and
- a useful start can be made on performance measurement with data that is currently available — waiting for the perfect data is a recipe for indefinite inaction and the use of available data is itself a catalyst for the development of better quality data.

Bibliography

Bogetoft, P., 1995, 'Incentives and Productivity Measurements,' *International Journal of Production Economics* 39, pp 67–81.

Brealey, R. and S. Myers, 1984, *Principles of Corporate Finance*, McGraw-Hill, New York.

Carter, N., 1991, 'Learning to Measure Performance: The Use of Indicators in Organisations,' *Public Administration* 69, pp 85–101.

Commonwealth Department of the Treasury, 1990, *Financial Monitoring of Government Business Enterprises: An Economic Framework*, Treasury Economic Paper Number 14, AGPS, Canberra.

Klein, R. and N. Carter, 1989, 'Performance Measurement: A Review of Concepts and Issues', in Public Finance Foundation, *Performance Measurement: Getting the Concepts Right*, London.

Lawrence, D., 1992, 'Total Factor Productivity — A Useful Indicator of Overall Economic Performance', in Steering Committee on National Performance Monitoring of Government Trading Enterprises, *Measuring the Total Factor Productivity of Government Trading Enterprises*, Canberra.

Lawrence, D., 1993, 'What Performance Measures Should We Use?', in Swan Consultants (Canberra), *Measuring the Economic Performance of Government Enterprises*, Proceedings of a Conference held at the Sheraton International Airport Hotel, 12 February, 1993.

Lawrence, D., P. Swan and J. Zeitsch, 1991, 'The comparative efficiency of state electricity authorities,' in Johnson M., P. Kriesler and A. Owen (eds), *Contemporary Issues in Australian Economics*, MacMillan.

Senge, P., 1990, The Fifth Discipline: The Art and Practice of the Learning Organization, Currency Doubleday, New York.

Shinnar, R., O. Dressler, C. A. Feng and A. I. Avidan, 1989, 'Estimation of the Economic Rate of Return for Industrial Companies,' *The Journal of Business* 62(3), pp 417–45.

Smith, P., 1990, 'The Use of Performance Indicators in the Public Sector,' *Journal of the Royal Statistical Society* 153(1), pp 53–72.

Swan Consultants (Canberra) Pty Ltd, 1991, *An International Benchmark for Australia's Electricity Supply Industry*, Report prepared for the Business Council of Australia.

Swan Consultants (Canberra) Pty Ltd, 1992a, 'Performance Measures for Australia Post: 1975–76 to 1990–91', in Steering Committee on National Performance Monitoring of Government Trading Enterprises, *Measuring the Total Factor Productivity of Government Trading Enterprises*, Canberra.

Swan Consultants (Canberra) Pty Ltd, 1992b, *International Benchmarks for the Queensland Electricity Supply Industry*, Report prepared for the Queensland Electricity Commission.

The Innovation Network, 1993, 'Applying Benchmarking to Higher Education: Part I Definitions', Internet: http://www.innovnet.com/ben1.htm.

Zeitsch, J. and D. Lawrence, 1996, 'Decomposing Economic Inefficiency in Base-Load Power Plants', *Journal of Productivity Analysis* 7, pp 359–78.

Zeitsch, J., D. Lawrence and J. Salerian, 1994, 'Comparing Like With Like in Productivity Studies — Apples, Oranges and Electricity,' *Economic Record* 70(209), pp 162–70.

Incentive regulation in water — case study

Alan Booker, Office of Water Services, Birmingham, United Kingdom.

Background

Incentive regulation embraces a wider agenda than setting price caps. It should include the collection and publication of comparative performance information, and deciding appropriate penalties for poor performance. This session concentrates entirely on the approach to setting price caps.

The general approach to setting price caps in the UK is discussed. This is followed by tracing the history and development of the approach to setting price caps in the UK. This was initially done by the Government, as part of the process of privatising the former water authorities in 1989, the Periodic Review of price limits by the Director General of Water Services followed in 1994. The preparation for the next Periodic Review of price limits by the Director in 1999 is well under way. The state of the art in the UK is contained in current consultation documents about regulatory methodology and process.

Medium term price caps

Price caps are a benefit sharing mechanism. Periodically the economic regulator sets forward looking limits on prices by reference to a basket of tariffs. Those forward looking price limits will make assumptions about costs and the scope for efficiency improvements. The assumed efficiency improvements are built into price limits, therefore, creating the incentive for companies to improve efficiency even more than the Director's judgements. It gives the customer the benefit of those efficiency savings from year to year. At the end of five years the extra savings, known as out-performance, are returned to customers by way of further price reductions which might be one-off price cuts or returned gradually to customers by price cuts over a number of years. The second method would be known as a glide path.

Out-performance can accrue from lower expenditure than that assumed in operating expenditure, capital expenditure or capital maintenance. Typically, there will be a period of five years between price limit reviews to create an incentive on companies to reduce costs continuously and to provide in the medium term a stable environment. Price limits may be set for a period longer than five years. In addition to affording customers the protection of price limits which will, in a period of stable obligations, reduce charges to customers, it is also important to maintain and protect levels of service so that charges are not being reduced at the expense of reduced standards of service. An important part of an incentive regulatory regime will be to monitor performance of the regulated companies and to penalise companies for poor performance.

One of the important functions of a regulator who operates an incentive regime, is to promote economy and efficiency measured by reference to inputs and outputs. Input costs should be reduced continuously for the same or higher outputs.

In such a regulatory environment it is important for the regulator to define the functions of the regulated companies in terms of their outputs for customers. The regulator should have a duty to ensure that companies are able to finance their functions. This does not mean at the expense of services and it does not mean that inefficient companies should be underwritten by the economic regulator at any price. Far from it. The regulator may have a duty to ensure a reasonable return on a company's capital. But it is the management of the company's responsibility to achieve, at least, the return on capital which the regulator judges to be reasonable. Incentives also need sticks as well as carrots. Sticks would be in the form of penalties which poor performance would attract, whilst carrots would normally be the opportunity to earn a reasonable return or to out-perform the efficiency improvements which the regulator will have built into forward price caps, and earn a higher than reasonable return.

The regulator will need to undertake sophisticated financial modelling in order to set price caps and test out the robustness of a range of options, against a number of financial indicators such as gearing, interest and dividend covers. He will need to do this in advance of coming to sound decisions on price caps, which he will be able to explain publicly and justify when challenged through an appeal mechanism if necessary.

There are other essential features of incentive regulation. The first is the annual monitoring of company performance to ensure that companies maintain or improve levels of service. The information needed for that has to be provided by companies annually and some form of independent reporting is essential to ensure its credibility.

The second related feature is comparative competition. The regulatory incentives provide a surrogate for normal market competition. Comparative competition is used in two ways. The first is through the publication of league tables of performance against key indicators. The second is the use of comparative unit cost information as an input to econometric models to inform the regulators' decisions about the scope for further year on year cost reductions. Econometric modelling is important in helping the regulator to make proper allowance for explanatory factors in his efficiency analysis.

Initial price caps at privatisation — 1990–1995

In the UK regime the price adjustment factor or price cap (K) is the amount by which prices of a basket of tariffs is allowed to increase year on year by reference to a general price index. The retail or consumer price index is probably the least controversial and best understood index available. Price caps are set on the best information about costs which the regulator has. In 1989 the water companies were asked for estimates of their costs. These costs were scrutinised by consulting engineers and then reduced on the basis of the best available evidence about the scope for efficiency savings. The result was a real increase in charges over that period of about 5 per cent per annum or 25 per cent in real terms of the quinquennium.

However this increase has to be seen in the context of the substantial increases required in drinking water quality and environmental quality being demanded as a result of European directives. These improvements required an increase in the capital expenditure of the water industry in the UK from £1 billion a year to £3.5 billion a year.

The first asset management plan or AMP1 was really a bottom up approach to determining functions and costs. In simple terms companies were asked what they needed to do and how much it would cost and price limits were set by Government to allow companies to do what they thought was needed. This process gave results which allowed the water and sewerage companies flotation to take place and investor confidence to be established in an industry which no one in the late 80's seriously thought could be sold off to hard nosed financiers.

There was also a concept developed and embodied in the companies licence of appointment which gave further protection to their income stream. This was known as cost pass through, whereby any unforeseen costs which companies had to incur to finance their obligations, allowed them to come to the regulator with a case for a further increase in their price limits.

Further aspects of the initial price cap were that the financial indicators were extremely loose. This meant that some companies did not need to borrow to finance their capital expenditure programmes which could all be financed from customer charges. There had been no direct customer involvement in setting price limits, and it was only after the event when privatisation had been successfully achieved that customers started to feel the impact. In some cases they experienced very substantial price rises which they laid at the door of privatisation.

In fact the water companies were able to exert undue influence over Government in the setting of initial price limits.

Increasingly, over the period 1990 to 1995 there was customer unrest, high levels of complaints about charges, a great deal of complaining to members of Parliament with the ensuing political fallout, particularly in those areas which were hit worse than average and which also happened unfortunately to be relatively low income areas. By 1995 affordability had become a real issue for a significant number of low income customers, located in particular, but by no means entirely in the South West of the UK. Average incomes are lower than the national average and a major environmental clean up programme was needed to improve bathing water quality.

The regulator had a major problem looming ahead of the first Periodic Review of price limits in 1994 which would put in place price caps for the period 1995–2000.

First Periodic Review — 1995–2000

The methodology and the process adopted at this first Periodic Review of price limits was a considerable development on the initial price caps. The cost of achieving even higher quality standards was identified at an early stage as the main upward pressure on price limits. During the process of consultations which preceded the review it became clear that there was merit in identifying the components of the K factor as -X to represent the scope for efficiency improvements and +Q, the impact of new quality obligations.

Out-performance in terms of the actual return on capital, was returned to customers over the course of the whole ten years for which price limits were set. The glide path was not a straight line. It removes something like 75 per cent of the out-performance in

the second quinquennium. Return on investment had been running at the level of around 13 per cent throughout the first five years against a real weighted average cost of capital of between 5–6 per cent which was determined by the Director following consultation.

The issue of the cost of quality attained the status of a major national debate. In the early part of the first quinquennium around 1991–92, Office of Water Services (Ofwat) decided that it would not be acceptable to allow the price escalator to continue to accelerate to absorb continuous increases in drinking water and environmental quality standards. Following an initial scene-setting publication by Ofwat, a quadripartite arrangement was established under the aegis of the Department of the Environment. The economic regulator, the two quality regulators, the Drinking Water Inspectorate and the National Rivers Authority came together with the water companies to examine the costs of achieving specific standards associated with the European Directive on Urban Waste Water Treatment, and the impact on customer bills.

This quadripartite process resulted in public advice in an open letter from the economic regulator to Government as to the options and likely costs, and their implications for customers' bills of quality improvements being adopted at different rates. The Government responded by indicating that these obligations should not go further nor be achieved more quickly than necessary to just meet EC standards. Guidance was given to the Director General as allowed in the water legislation. This guidance set out the part to be played by not only the economic regulator, but the quality regulators, the companies, and by the Department of the Environment in ensuring that quality standards were increased but at a rate which customers could afford. The need for a political process to balance quality and affordability was, therefore, formally established.

As part of the first Periodic Review the cost and output matrix was defined. Costs were set out in terms of operating expenditure, capital maintenance expenditure, capital expenditure and its associated return on capital. Outputs were defined in terms of base service provision, service enhancements, providing for growth and enhancements to quality. All output elements other than quality was wrapped up in the -X element of the price cap.

The debate between the regulator and the companies about cost of capital was intense and at times heated. However, at the end of that debate the companies accepted that the provision of water and sewerage services is a low risk business attracting relatively low risk capital. The figure of 5–6 per cent real was fixed as the average weighted cost of capital allowing for significant additional borrowing to finance capital expenditure. Headroom of up to 1 per cent above that was built in to reflect the uncertainties both generally and for individual companies.

In order to complete the return on capital calculation asset values were determined. This was done on the basis of flotation values of the companies at privatisation making appropriate allowance for debt or cash, in the companies balance sheets, and for subsequent investment on new assets.

AMP2 was significantly different from AMP1. It was defined as a strategic business plan covering not only underground assets, but also surface assets. It was anticipated

that companies would assemble their AMP2 submissions to the regulator with an executive overview, which he said he would read personally and which would give the Director a strategic overview of the business plans. In the event, and disappointingly, the AMP2 submissions or strategic business plans turned out to be little more than overbids for resources to carry out dubious improvements which Ofwat had to set about cutting down to size. A typical example of attempted gold-plating.

The process associated with the first Periodic Review was a significant development on the process for initial price caps. It involved the collection of a considerable quantity of information on costs over time about the quality obligations. It also involved a number of discrete publications consulting in public on individual parts of the methodology and cumulating in a substantial publication in the autumn of 1993 setting out the approach to be used by Ofwat in setting price caps at the determination in 1994. There was also a substantial effort made to involve customers in the process. Companies were required to undertake market research to find out what their customers were willing to pay for improvements in standards of service. Companies undertook an examination of customer preferences. In some cases this was quite innovative. The customer service committees and in particular the chairmen of those committees were involved in the price setting process. For instance, the chairmen attended the formal due process meeting between the Director and the companies, at which draft determinations were discussed, and at which companies were asked to formally make known to the Director their key concerns arising from draft determinations. Following that due process the chairmen had the opportunity to put customer views about affordability and willingness to pay to the Director.

When the Director published the new price limits at the end of July 1994, together with a detailed explanation of his reasoning behind them, relating to the methodology published a year previously the CSC Chairmen also published a statement to say that they were satisfied that the outcomes were consistent with the companies obligations and the Director's duties.

The publication explaining price limits also broke new ground identifying the X and Q components of K and giving an indicative split of K between water and sewerage.

Following the 1994 determination many commentators acknowledged publicly that the thorough open and transparent process which had been adopted achieved a considerable advance in regulatory process.

The outcomes of the Review were that the escalator had been slowed considerably from +5 per cent per annum to +1 per cent. Quality improvements were still being maintained. A new climate had been established with companies subject to significantly higher efficiency expectations. The incentive ratchet was tightened.

Second Periodic Review — 2000–2005

Incentive regulation was working well but customers were still dissatisfied with the fact that companies were making significant profits and top management was paying itself high levels of remuneration. The period since 1995 has largely been characterised by sniping at the first Periodic Review by customer pressure groups, and by the

Parliamentary opposition, and sniping at the fat cats — the senior directors of water companies and shareholders.

One of the first jobs of the new Government was to implement a special one-off tax on all the utility companies to put right what it regards as the bad deals struck by its predecessor when selling off the utilities. In a sense this move signals a break with the past and a coming to terms with utility privatisation.

Work is already well under way for the second Periodic Review. The timetable has been established and consultation documents have been published covering the approach and methodology and the business planning and consultation process.

Broadly the methodology is captured in RPI - P_0 - X + Q ± V ± S. P_0 is the initial reduction in prices resulting from out-performance during the second quinquennium. X is the forward looking expression of scope for efficiency in the third quinquennium. Q as before reflects the upward pressure of higher environmental and drinking water quality standards. V represents the impact of the supply/demand balance. It can be either positive where new resources are required, or negative in those cases where companies have an abundance of resources and might be expected to export some of their surplus. S represents the service factor which can clearly either result in enhancements or a controlled reduction in standards.

The Director has indicated that he wants to see an initial price cut followed by four years of generally stable prices. This means that any upward pressure on price limits should be contained within the scope for cost reductions from efficiency improvements, i.e. -X should be equal or greater than +Q ± V ± S. The Director has also indicated that in his view price limits should be determined for a period of five years rather than ten years.

Non statutory quality improvements may be an issue for the 1999 Periodic Review. The new Government wants to put the environment at the heart of its policies, and may be more interested in achieving further quality improvements at the expense of stable prices, which the Director has indicated he would like to see. There are also resource issues arising from climate change and the intention of the now Environment Agency (formerly the National Rivers Authority) to review abstraction licences and to reduce abstractions from some environmentally sensitive sources.

Other methodology developments relate to capital maintenance where in 1994 the approach was less well developed than other elements of the matrix. Ofwat has undertaken a considerable amount of work in order to bring the costs of capital maintenance under similar econometric modelling as opex. The principles underpinning capital maintenance are the maintenance of serviceability for customers and the requirement to achieve broad equivalence over time between expenditure and accurate provision for current cost depreciation.

Other approaches to cost estimates are likely to follow similar methods to 1994. Although the process of firming up on costs is likely to be more rigorous in certain respects and better informed.

The final difference in methodology may result from the Director's ability to reprofile capital expenditure and efficiency improvements over the five years of the

determination in order to reflect the learning gained from observing the behaviour of companies over the second quinquennium. The experience is that following agreement of the next set of price limits, companies delay capital expenditure for as long as possible and reduce operating costs as quickly as possible in order to out-perform price limits and retain the fruits of out performance for as long as possible.

One idea proposed in the consultation paper is that companies should have to actually deliver quality improvements before they receive the benefit in price limits.

The business planning and consultation process is more elaborate than in 1994. There is still significant effort required of companies to consult their customers. However, much of the information required for setting price limits is likely to be published as are draft determinations. All this public information will be subject to analysis and challenge. Business planning contains a number of iterations to allow the impact of new obligations both on quality and the supply/demand balance to be costed and the interaction with other outputs to be taken into consideration in an endeavour to achieve stable prices following the P_0 adjustment.

AMP3 will now not be produced until after the determination has been concluded and will reflect the outputs which are being financed. It will become the monitoring document during the period 2000–2005.

Arguments about the commercial confidentiality of information will undoubtedly create a number of issues in the run up to the 1999 Periodic Review. However, the initial expectation is that the burden of proof will reside with the companies to demonstrate to the Director that specific information should not be published on commercial confidentiality grounds. Their arguments for non publication would themselves be published. The process is moving continuously in the direction of more and more openness, more and more transparency and greater opportunity for analysis and challenge by commentators, academics and pressure groups. The great danger is that this process becomes so open and transparent that it will require a legal process to control it.

If incentive regulation is to continue to provide stability with flexibility and incentives to improve performance the UK needs to be wary of how process is developing.

The approach and methodology to determining price caps should not be capable of being developed much further. However, the development of process might be open-ended and could be counter-productive in the overall context of incentive regulation. Time will tell.

Section III. Price structure

Basics of rate design — pricing principles and self-selecting two-part tariffs

Sanford V. Berg, Public Utility Research Center, University of Florida, United States.

Overview

This overview of the basics of rate design in network industries should be read in conjunction with Mark Jamison's, 1997, more comprehensive write-up on rate structure. After describing some of the key concepts, we will examine in some detail the rationale behind rate structures that can reflect incremental costs, while still facilitating the recovery of fixed (often joint or common) costs. In competitive industries, overhead or joint costs cannot be 'allocated', they must be 'recovered where they can' based on a rate-making philosophy that recognises this fact. The second part of this paper describes the basic issues and problems of recovering these costs, with special reference to electric utilities in an increasingly competitive environment.

One approach is to adopt Two-Part Pricing schemes, one part reflecting capacity and distribution system charges and the other based on on-peak and off-peak energy costs. This approach eliminates distortions caused by average cost pricing, enabling customers to face the true costs of additional electricity purchases. Retail utilities have incentives to create contracts with their suppliers that better reflect cost-causation. During the transition; to this new pricing structure, integrated utilities could base the capacity and system charge price on a Historical Customer Baseline (HCB), so that such customers continue to contribute a 'share' of overhead (or non-usage sensitive) costs. On-peak and off-peak prices are then set at incremental cost. Cost recovery is achieved via appropriate calculation of the monthly fee. Of course, if competitors can offer a better 'deal', the HCB would need to be modified to minimise uneconomic bypass of the incumbent. (See Della Valle and Bidwell, 1995).

A second approach focuses on differential demand elasticities. Customers who have demands which are insensitive to price have few alternative sources of supply. Higher prices to these 'inelastic' demanders represent one way to obtain revenues that can be applied to shared resources (overhead costs). While Ramsey pricing (the technical term for this approach) raises political issues, the outcome may be preferable to losing those customers who face competitive alternatives (including co-generation)[1]. But first, we need to review the basics of rate design.

[1] See Stephen J. Brown and David S. Sibley, The Theory of Public Utility Pricing (Cambridge University Press) 1986, for a thorough discussion of Ramsey pricing.

Cost causation and price signals

When considering efficient pricing and investment, if neoclassical economists agree on one point, it is the following: rate design matters[2]. A strong case can be made for regulators to allow some discretion in rate design, since firms have far greater information on cost structures and demand patterns than is available to the regulatory. When cost allocation manuals substitute for estimates of forward-looking incremental costs, inefficiencies arise. In recognition of the important role of price signals, a wide range of rate designs has been analysed. The following represents a partial listing of the topics addressed in the literature on optimum pricing and capacity:

Marginal cost pricing. The allocative efficiency consequences of such pricing are well known. For example, the financial viability of the firm may require a subsidy, or more complicated rate designs (such as multipart pricing). Furthermore, short run and long run marginal costs will differ — so while the former serve as the standard for pricing decisions, the latter are relevant for comparing alternative investment patterns — as when alternative providers have different production technologies.

Cross subsidisation and regulation as taxation. Cross subsidisation can be a deliberate regulatory objective as some customers cover the incremental costs of serving favoured customers. Alternatively, it can stem from inappropriate allocations of fixed or variable costs. A related concern is transfer pricing which increases the reported costs of the regulated firm. This device might be used to shift profits from the regulated firm to an unregulated subsidiary. Such shifting raises a dilemma, because when regulators mandate complete separation among business units in response to this problem, the firms can lose economies of scope. The result is higher costs.

Discriminatory pricing and demand separation. The ability to separate markets and prevent resale is facilitated by customers being hooked up to utility distribution systems. Since consumers with inelastic demanders are often the ones regulators are trying to protect from monopoly power, commissions often will overlay cost allocation regulation upon rate level regulation — preventing 'undue' discrimination.

Ramsey pricing. If the firm can identify different customer groups and charge different prices to the various customer classes, Ramsey pricing can be utilised to minimise economic misallocations. However, such a pricing policy (charging more to those with relatively inelastic demands) might still be viewed as unduly discriminatory, even though the firm does not realise excess profit. Citizens might prefer other price configurations for the multi-product natural monopolist. Furthermore, there is no guarantee that Ramsey prices are sustainable in the long run: some coalition of customers could end up paying more than the stand-alone costs of serving them — leading to self production and the loss of their business. The technical literature on Ramsey pricing is ably summarised in Sheshinski, 1986, and Braeutigam, 1989. Mitchell and Vogelsang, 1991, apply much of this literature to telecommunications pricing.

---

[2] Some of the material in this section originally appeared in Berg and Tschirhart, 1995, pp. 310–330. For an excellent survey of rate design, see Arnold, et.al., 1995.

Predatory pricing and market dominance. If high cost suppliers are driven from the market due to entry by a multiproduct natural monopoly, resource allocation is improved. However, these suppliers may claim predatory pricing if the output is produced by an unregulated subsidiary or by the regulated firm. Producers of substitute products could argue that revenues from the utilitys captive consumer groups (or regulated products) cover costs associated with products subject to competition. However, one ought not accept fully-distributed costs as indicators of subsidisation. Furthermore, when a group of firms is under industry-wide regulation and when price is based on industry average costs, the presence of high cost firms could increase the profits of efficient suppliers (Daughety, 1984). The removal of such cost-based pricing would reduce prices and lead to bankruptcy or consolidation of inefficient firms. Such a development would again bring forth charges of predation, but these charges would be groundless.

Cost allocation regulation. A multiproduct firm is charged with allocating its total costs, including common costs, over its various products in an effort to ensure that revenue from the sale of each product covers its allocated cost (Braeutigam, 1980). This area may be one of the more under-analysed in the field of regulatory economics. Regulators have often viewed fully-distributed cost allocations as techniques for ensuring that customer groups are not unfairly burdened with shared costs. Besides leading to potentially undesirable prices and cross subsidies, there is the danger that separations procedures and cost allocation manuals may foster an unwarranted feeling of accomplishment among regulators. Sweeney, 1982, finds that output-based allocation schemes can yield perverse results: we end up with prices such that one or more of them can be lowered — improving welfare without decreasing the monopolist's profit. In addition, we can have relatively high prices in unregulated markets. Finally, Cabe, 1988, illustrates that any output vector can be achieved by some fully-distributed cost method.

Peak load pricing and intertemporal patterns of demand. This literature has a rich history. The early contributions by Boiteux, 1960 and Steiner, 1957, stimulated analyses of intertemporal demand patterns. The production technology (involving fixed or variable coefficients and with and without scale economies) drew upon actual engineering studies of cost structures. Diverse technologies, interdependent demands, selection of rating periods and other issues were addressed, as economists began to characterise realistic demand and cost conditions. Rate design in such situations must take these factors into account. Sweeney explains these results by noting that because regulated products are permitted to obtain a 'fair return' on shared input, output reductions in unregulated markets allow more of the common cost to be shifted to regulated markets. As a result, greater profits are earned in these regulated markets. In one sense, these results may appeal to regulators. The prices are high in the unregulated markets, thereby quelling fears of cross subsidies from the regulated markets. Also, competitors in the unregulated markets would be pleased, since the monopolist is apparently not relying on profits from the regulated markets to predatory price in their markets. In the long run, the monopolist's prices could be undercut in at least some of the unregulated markets, perhaps even driving the monopolist out entirely. The advantages of natural monopoly production for multiple markets are then lost, and the regulator's optimism regarding FDC procedures proves short-lived.

Pricing with random demand and supply. The intertemporal issues noted above have also been addressed in the context of uncertainty. Consumers value reliability of service, which will be affected by the interaction of price (announced in advance) and uncertain demands (driven by weather, seasonal conditions, and hourly factors) and production capabilities (related to unplanned outages).

Nonlinear pricing and interpersonal patterns of demand. Whether one is considering pricing entry and rides in an amusement park or access to and usage of a telephone system, multipart pricing offers a viable option for enhancing revenues. Much of the literature on multipart pricing and nonlinear outlay schedules is surveyed by Brown and Sibley, 1986. The pattern of individual demands proves to be important for the development of first and second-best rate designs involving fees and usage charges.

Sustainable pricing. Faulhaber, 1975, showed that a natural monopolist was not necessarily immune to entry under certain cost structures. This insight raises a dilemma for regulators and implies significant information requirements for optimal pricing decisions. A related issue is the role of the incumbent firm as the supplier of last resort (Weisman, 1988).

Quality of service. Service quality also raises a number of important issues for analysts. The quality level provided under competition, monopoly, or regulation has received substantial attention. In practice, regulators tend to utilise pass/fail standards. While such standards are clear and precise, problems arise in using them to monitor and reward quality. For example, by evaluating performance relative to a pass/fail cut-off, distinctions among various levels of sub-standard and super-standard performance are ignored: utilities have little incentive to exceed targets. In addition, the targets themselves are often somewhat arbitrary, having arisen from a chaotic process reflecting historical engineering capabilities, political pressures, and administrative happenstance. Consumer valuations of different quality dimensions and firm knowledge of emerging technological opportunities are not likely to be reflected in current pass-fail standards.

Recovery of joint costs in a competitive environment

The brief overview of rate design concepts provides a foundation for addressing the recovery of joint costs in a competitive environment. Electric utility managers understand that the industry has rapidly moved from local monopolies to one that is customer-driven. Co-generation and competition via open transmission policies have disrupted traditional pricing arrangements. Cost allocation manuals are becoming increasingly irrelevant as the electricity industry becomes more competitive. This section outlines two approaches that enable the recovery of joint costs in a competitive environment. One approach involves utilising Two-Part Pricing schemes, so the customers' share of joint costs is less dependent on total consumption. Such rate designs better reflect cost causation on the margin, while permitting recovery of some fixed costs. Another approach involves lower prices to those demanders with alternative sources of supply. Although higher prices to inelastic demanders — those without options — raises some tough political issues, those customers would be even worse off if business and other large customers abandoned their traditional suppliers. Thus, Ramsey pricing involves price discrimination (or price differentiation, if the former term seems too value-loaded).

Electric utilities are going to have to generate value for customers by devising new rate designs which create win-win opportunities. Both Two-part pricing (using Historical Customer Baselines) and Ramsey pricing represent innovative ways to recover joint costs. First, some background material needs to be reviewed.

Background — recent trends

In the U.S., the electric energy industry lags behind telecommunications in terms of competitive pressures, but regulatory roadblocks to competition at various stages of production have been demolished in a number of countries. In the U.S., since most regulatory authority is vested in the states, the system is conducive to regulatory innovation. The process of evolution within the American regulatory environment is driven by wider adoption of approaches that have been successfully implemented in a few states. This heterogeneity also results in confusing and sometimes contradictory state regulatory regimes. For investor-owned utilities, Rate Of Return (ROR) on rate base regulation has characterised the industry, with customer-class cost allocation rules, fuel adjustment clauses, and management audits further constraining prices and revenue requirements. For municipals and other publicly-owned utilities, cost allocations have tended to establish revenue targets across customer classes, with prices reflecting some type of fully distributed cost.

The evolution of US regulatory policy illustrates changing attitudes towards the efficacy of competition in promoting efficiency. At the same time, concern over environmental impacts has placed new objectives onto the regulatory agenda; the new instruments for achieving new objectives raise complex issues. For example, state-mandated conservation programs will come under pressure, especially if retail wheeling is widely adopted. The costs of DSM programs cannot be spread across a set of captive customers: larger customers, especially, will face choices they are currently denied. Most industry observers expect vertical disintegration and partial deregulation to continue. The implications for incumbents are mixed. Non-generating distribution systems are in a position to 'shop around'. Integrated suppliers are likely to face revenue erosion as competition becomes more widespread.

The interests of various constituencies are tough to reconcile. The National Association of Regulatory Utility Commissioners (NARUC) wants to preserve the flexibility of states so that state PSCs can craft policies which fit their unique circumstances. Groups benefiting from current state regulations want to retain 'local' control — preserving their relative benefits. The American Public Power Association (APPA) supports the agenda of municipally-owned utilities. The National Rural Electric Cooperative Association (NRECA) seeks retention of rules that assist rural electric utilities. At the national level, FERC oversees wholesale and transmission issues, while state PSC regulate facility additions and retail rates. The conflicting pressures and overlapping jurisdictions make coherent policy development very difficult.

One conclusion is clear, price signals are being given greater prominence, although policy makers (both regulators and municipal authorities) tend to avoid dramatic changes in rate design for fear of political repercussions. Historically, prices for different customer groups were set using cost allocation procedures. Revenue 'requirements' were determined from top down — with minimal attention to

incremental cost causation. Today, prices and incumbent investments in generating capacity are constrained by competitive alternatives — induced by regulatory promotion of co-generation and independent power producers (IPPs). Thus, in non-core (industrial) markets, customers have alternatives in the form of self generation or geographic re-location. When revenues from some customer groups fall short of 'allocated' costs, utilities experience financial pressures. Core (residential) customers can flex their political muscle to avoid rate increases, resulting in realised returns becoming a residual. For IOUs, rates of return were never 'guaranteed'; rather, they were 'allowed'. However, returns have become more problematic in a world where traditional entry restrictions are being set aside. These developments will constrain municipal utilities and REAs as well.

Deregulation and emerging competition will tend to promote least cost supply. Some vertically-integrated suppliers have low incremental costs relative to their potential competitors. If they have high average (embedded) costs, they will come under financial pressure. IOUs will find cut dividends and change their financial structures (towards more equity). For municipal utilities, it may be difficult to meet interest coverage or continue to transfer traditional amounts to city coffers. However, low incremental cost suppliers are still going to be able to compete for business. On the other hand, high incremental cost suppliers will be in trouble in a competitive environment. 'Stranded investment' is just another word for generating capacity that cannot yield cash flows for covering fixed costs when electricity markets are opened up. Book value exceeds market value of capacity.

National regulatory policy has leaned in the direction of pro-competitive market structures at the generation level. Since PURPA's promotion of co-generation via qualifying facilities (QFs) and of IPPs, national policy has continued to view wholesale competition as stimulating real savings for final demanders. Non-utilities supply almost ten per cent of all electric power in the U.S., and between 1991 and 1994, they built over half of all new capacity. *The Energy Act 1992* created Exempt Wholesale Generators (EWGs) as another vehicle for introducing new players into the game. Since access to transmission can be mandated by FERC, terms and conditions of transmission access has become a significant regulatory issue. Ultimately, large buyers may gain access to alternative suppliers via the transmission network: retail markets will change dramatically. While the problems for network coordination, construction, and reliability are substantial, the trend seems irreversible.

In 1994, both California and Michigan established programs designed to promote more competition. Larger customers who have the ability to shop will tend to pay market-based (incremental cost) prices, leaving core (residential) customers at risk for covering the costs associated with higher cost capacity. The fear of so-called 'stranded investment' blunts efforts to open up local markets. 'Securitisation' represents one mechanism for permitting some cost-recovery by high-cost firms. The short run impacts of competition differ from the long run impacts. In the short run, the efficiency gains may not be substantial, given the demand elasticities — though the monetary transfers could be significant. Over the long run, the movement away from cost-based regulation for IOUs is likely to further stimulate cost-containment and improved price signals. No utilities will be insulated from these pressures.

PURPA-induced competition in the wholesale market for electricity has increased the importance of transmission access as utilities try to find the lowest cost suppliers whose generating facilities may be located far from the utilities' retail markets. The provision of *the Environment Protection Act 1992* requiring utilities to offer wheeling to third parties for a fee is possibly the biggest change in the industry in more than fifty years.

Reduced demand growth, nuclear plant cost overruns, environmental costs, and continued low natural gas prices have led to excess and high cost capacity whose economic value is lower than book value. The resolution of the stranded investment problem has been linked by some to the terms and conditions of transmission access. There had been a 'regulatory compact' under which capacity was built and changing the rules of the game is perceived as unfair. As Costello, Burns, and Hegazy, 1994 note, vertically-integrated utilities, conservationists, and environmentalists tend to oppose retail wheeling. For the former, monopoly franchises are lost as competitors threaten to take away customers. The two latter groups fear reductions in (or elimination of) utility-funded Demand-Side Management (DSM) programs. Also, the forms of Integrated Resource Planning (IRP) which emphasise environmental costs above and beyond those addressed in national laws are threatened. Those supporting immediate retail wheeling argue that sunk costs ought to be ignored for policy purposes — leaving investors holding the bag. Large industrial and commercial customers do not want to bear transition costs.

Continued regulatory and legislative debate can be expected on transmission access and pricing, bidding procedures, setting new price regulations, and devising alternative regulatory constraints. We can already see the outlines of changes that are altering the regulatory landscape. Some believe that competition has become an objective — rather than a mechanism for achieving economic objectives. Certainly, national legislation and FERC have promoted entry into generation markets as a way to keep energy costs down. With this thrust has come pressure for transmission access at a fair price.

Multipart pricing and the promotion of efficiency

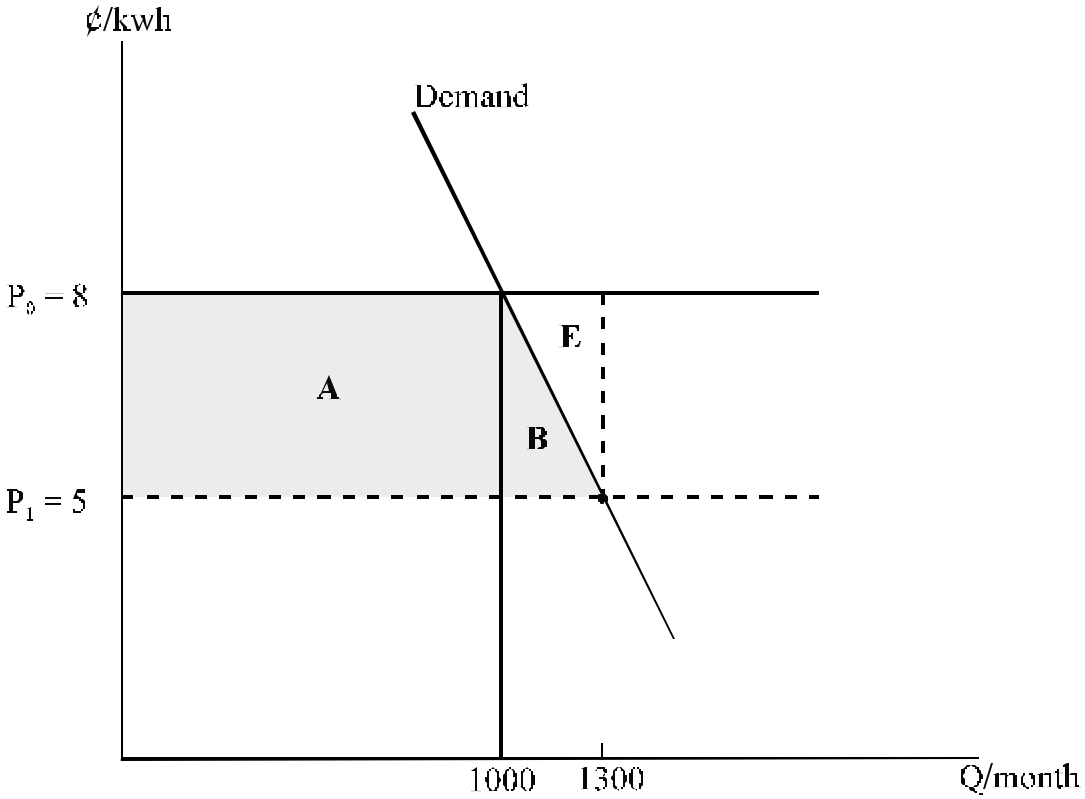
In the short run, with capacity costs fixed, changes in the wholesale pricing structure can involve particular customers or customer classes benefiting at the expense of others. Whether the process is a zero-sum game depends on the nature of rate re-structuring. If the savings obtained by winners is roughly equal to the additional outlays required of losers, then the objective of net revenue neutrality sows the seeds of conflict. For example, lowering the price to one group and raising it for another can have this characteristic. However, multipart pricing enables the supplier to create win-win options — bringing the marginal price down to incremental cost, while recovering current capacity costs via fixed monthly fees.

It should be noted that cost allocations which are currently used may seem reasonable and consistent with industry practice. Nevertheless, these allocations often are quite arbitrary — reflecting some view of fairness rather than cost causation. Evidence from other industries suggests that competition will force marginal price towards incremental cost.

A diagrammatic representation may help explain the win-win aspects of multipart pricing. Figure 1 depicts a demand curve. At lower prices, the customer is willing to

buy more electricity. At very high prices, customers will only apply electricity to very high valued uses. If price is quite low, then thermostats may be adjusted to give greater comfort, more electricity-intensive machinery might be utilised, and energy-conservation activity is less cost-effective from the standpoint of the buyer. In the short run, customers are not likely to be able to make substantial behavioural or operating adjustments, but the change in consumption will be greater as customers have more time to adapt to a permanent price change. Greater long-run responsiveness means that the efficiency gains from improved price signals are greater when consumers have time to make adjustments.

Figure 1: Benefits from a price change



Utilities are used to thinking in terms of a customers load shape and how this influences the system load. However, the load shape is a function of the price structure. Time-of-use pricing will alter the hourly pattern of electricity consumption — with that pattern changing more dramatically as customers have more time to adjust to the new price structure. Responsiveness of customers is characterised by economists in terms of demand curves. Greater long run responsiveness means that the consumers have time to make adjustments.

The Law of a Downward Sloping Demand has theoretical and empirical support. Utilities recognise that price influences consumption in the way described above. The Law's Corollary of Greater Responsiveness with Longer Adjustment Time has also been verified. The position of the demand curve is affected by other factors outside the

utilitys control. If the price of substitutes decreases, demand for electricity shifts in. If the prices of appliances that use electricity (complements) fall, then the demand for electricity shifts out. Weather conditions also affect the hourly load and monthly consumption. In the Figure, if price is $.08/kwh, then the customer depicted here consumes 1000 kWh. This could be broken down to hourly consumption, but this simple example illustrates the impact of a price reduction. If price falls to $.05/kwh, more than 1000 kWh would be demanded with a lower price. Note that if incomes rise, or average family size increases, or square feet per house increases, or temperatures are less moderate, the demand schedule will shift out. The hypothetical demand curve depicted in Figure 1 holds all these other factors constant, so that monthly consumption depends on price.

In this example, if price per kWh is $.08, then 1000 kWh are purchased, for a total consumer outlay of $80. If price were $.05, then 1300 kWh would be purchased, for a total outlay of $65. If demand had been more responsive to the price reduction (so that consumption rose to, say 2000 kWh, then total expenditures by this customer would have risen to $100. Thus, an increase in outlays does not necessarily imply a reduction in customer satisfaction. In this case, the price reduction induced additional consumption, and kWh were applied to valued uses by the customer!

In the case of the demand curve depicted in Figure 1, the price reduction from $.08 to $.05 yielded an improvement for the consumer. Analytically, this gain could be broken into two parts. The first part reflects the $.03 is saved on each of the 1000 units that used to be purchased at the higher price (area A = $30). Furthermore, 300 additional units are purchased when the price is only $.05. Economists identify area B as reflecting the benefits (above the outlays) associated with this additional consumption. Area B is $4.50 (the area of this triangle is half the base times the height).

Thus, the price reduction benefits the customer by $34.50. The $15 reduction in outlays (from $80 to $65) is not a good indicator of the consumer benefits associated with the price reduction. This point is very important, because rate design that focuses on outlays rather than customer satisfaction is likely to miss some win-win opportunities. In a competitive environment, suppliers cannot afford to ignore opportunities.

So far, we have not considered the firm. If its incremental costs were $.08, then a price reduction to $.05 is a losing proposition. The cost of serving the customer is $104 ($.08/kwh times 1300 kWh), but the revenue from the customer is only $65 ($.05/kwh times 1300 kWh). Underpricing electricity relative to its cost hurts the supplier more than it benefits the customer! From the diagram, the customer gains A+B, while the firm loses A+B+E. If, during peak periods, the price is below incremental costs, the utility ought to revise its prices (if the metering costs are not small relative to the savings).

In the case of the price reduction, what if the incremental cost were $.05? The $.08 price was high relative to the cost of additional kWh. Now the customer gains (A+B), which is more than the supplier loses (-B) from the price reduction! This observation suggests that a win-win option is possible. The utility could offer the customer a multipart price instead of the uniform price of $.08/kwh. The rate structure could be a $30 monthly fee (regardless of units consumed) and a per unit price of $.05. Since area

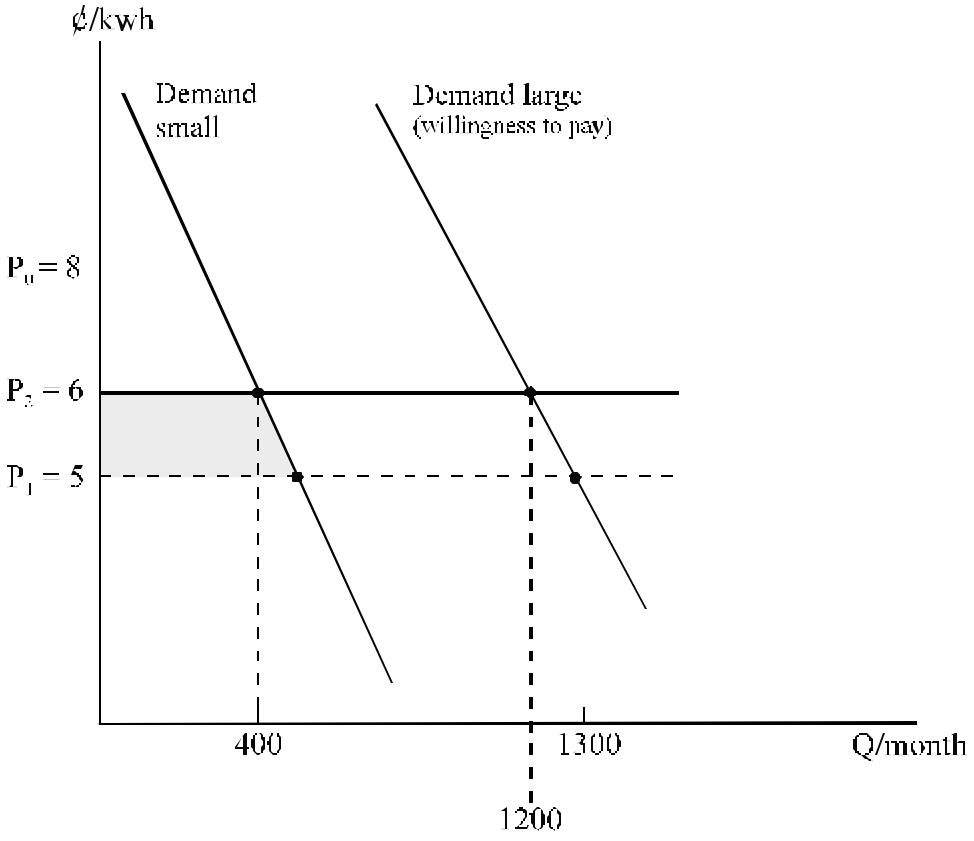
A is $30, it is clear that the customer is better off by area B ($4.50) under this alternative rate design. And the firm is no worse off. So long as the monthly fee is less than $34.50 (and per unit price is $.05), the customer is better off under the multipart scheme than paying a uniform per unit price of $.08.

Return to the $30 fee case, where total customer outlays now equal $95 and incremental cost is $.05. If the total bill is divided by the 1300 kWh, the average price is about $.073. Why not just set a price of $.073 and avoid the slightly more complicated pricing scheme? After all, customers look at their total bills. The response to this question is that the combined gains to the customer and the supplier would be less if price were only lowered from $.08 to $.073 than if the $30.00 monthly fee were imposed in conjunction with $.05/kwh. By himself, the customer is better off by more than $7.00 with the $.007/kwh price reduction. That per unit saving times 1000 kWh happens to be greater than area B. But all of that gain is essentially balanced by a net revenue loss experienced by the supplier! That price reduction is not a win-win outcome. The multipart scheme keeps the supplier whole, while making the customer $4.50 better off than before. Furthermore, the price of $.073/kwh is inefficient. It discourages consumption that is worth more than the resources that would have gone into the production of additional kWh (i.e. the price of $.073 is greater than incremental cost, $.05).

We saw that setting the marginal price equal to incremental cost increased consumption to 1300 kWh. The customer valued that additional consumption more than society valued the resources that went into creating the additional kWh. Thus, incremental cost pricing promotes the efficient use of society's resources. If price is above incremental cost (as is the case with much off-peak consumption), we are under-consuming electricity. If price is below incremental cost (as can be the case with on-peak consumption), we over-consume electricity. Multi-part pricing combined with peak load pricing can make both the firm and the customer better off. Peak load pricing by itself may benefit customers and/or the supplier.

The firm lacks information on the full nature of customer willingness to pay. Billing records can provide clues regarding potential consumption patterns, but optimal rate designs can facilitate both cost-recovery and efficiency. Figure 2 illustrates the benefits from utilising self-selecting two part tariffs. Two demands are depicted in the Figure; the supplier does not know which customer has which demand. As shown in the example, Option 1 has a $10 monthly fee, but a marginal price of $.05 per unit. The larger demander will select this option, since Option 2 (no fee, but a per unit price of $.06) yield less consumer surplus. The smaller demander selects the second option. Such price options enable the supplier to extract more consumer surplus than under uniform pricing — which enhances the financial viability of a firm under competitive pressure. Such rate designs are especially important if there are substantial fixed costs.

Figure 2: Self-selecting two part tariffs



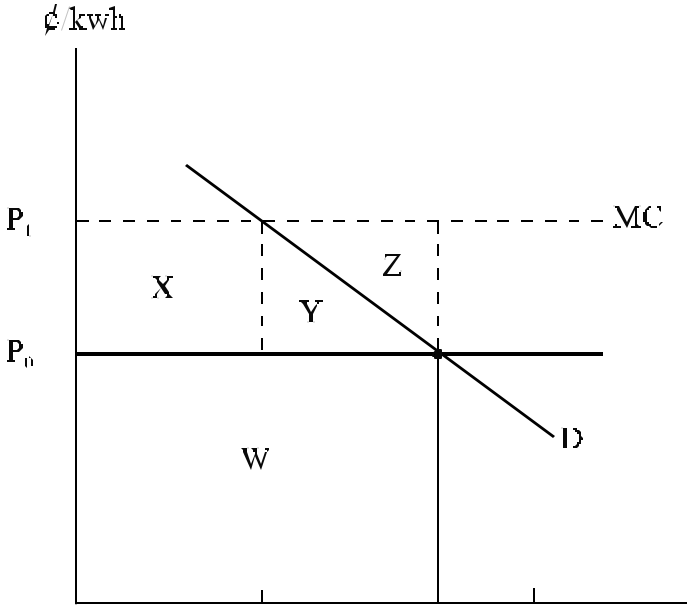
Option 1	**Option 2**
$F_1 = \$10$	$F_2 = 0$
$P_1 = 5¢$	$P_2 = 6¢$

D_{large}: if select option 2, lose more than \$12 of surplus (vs F_1 is only \$10)

D_{small}: if select option 1, gain less than \$10 of consumer surplus (vs F_2 = \$10)

Figure 3 illustrates the problem of pricing below incremental cost. Each unit sold is priced at less than the cost of production. The supplier would be better off making a deal in which he gave the demander a lump sum credit to his bill of X+Y, while raising the price to P1. The customer is no worse off than initially (less electricity is consumed, but the customer is indifferent to P0 for each kWh versus a rebate of X+Y per month and a higher price of P1).

Figure 3: Pricing below incremental cost



Note, these observations regarding multipart pricing are strengthened when longer run adjustments are taken into account. Demand is more elastic (or responsive to price changes) when customers have more time to adjust their energy-using equipment. If price increases, the firm may have few alternatives in the short run. But soon, energy-conserving investments can be implemented, and consumption drops more dramatically than in the initial months. The time for adjustment depends on the nature of the industrial, residential, or commercial demand. Utility managers who understand the role of price signals can promote the efficient conservation of energy.

The precise amount to be included in the fixed fee is not a simple calculation. Were recent years 'normal'? What if the level of this historical contribution is no longer sustainable in a competitive market? The answers to these questions require substantial analysis. Suffice it to note that such monthly fees can be calculated that would leave the firm better off than before. Bringing the incremental price in line with incremental cost is a potential win-win move. Rate designers ought to consider this addition to their price portfolios in a competitive era.

Ramsey pricing

As noted earlier, Ramsey pricing corresponds to price discrimination such that total revenues equal total costs. The ability to separate markets and prevent resale is facilitated by customers being hooked up to utility distribution systems. Faulhaber and Baumol (1988, p. 594) cite Ramsey pricing as '... a clear example of a principle that derives from the [economic] literature and has (recently) achieved a good deal of attention among government agencies.' They note that it has been discussed in many courts, state commissions, the Federal Communications Commission (FCC), the Federal Energy Regulatory Commission (FERC) and the Interstate Commerce Commission (ICC), as well as in other countries. Faulhaber and Baumol also

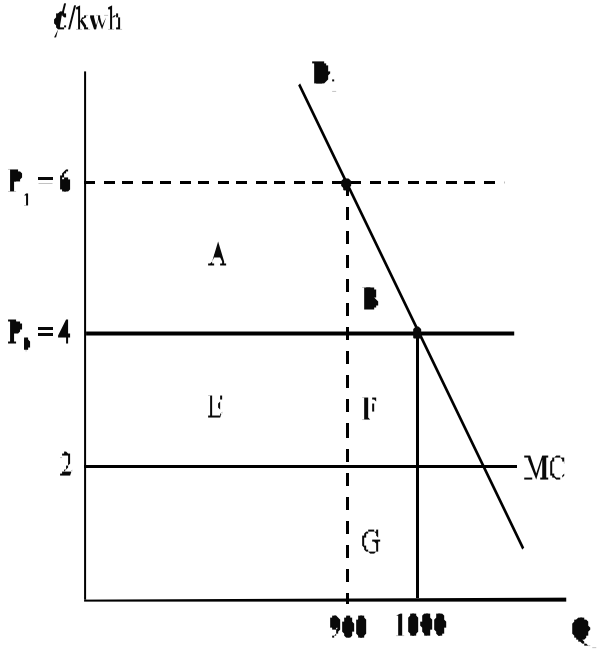
highlighted the stand-alone cost test as an example of a contribution of economic theory to regulatory pricing practices. The test places a ceiling on rates. No consumer or group of consumers ought pay more for service than the cost of serving them apart from all other consumers. Although this approach can be traced to pricing practices established for the Tennessee Valley Authority under the name 'separable costs and remaining benefits' (EPA, 1975, Straffin and Heaney, 1981), only in the last two decades has it become prominent in the literature where it has been related to the core concept in game theory (Faulhaber, 1975; Sorenson, et al., 1976, Sharkey, 1982). The theory also provides regulators with a rigorous definition of cross subsidisation that eschews the arbitrary definitions associated with fully distributed cost pricing practices. Faulhaber and Baumol indicate that both the FCC and ICC have considered using stand alone cost tests in rate making. Since consumers with inelastic demands are often those regulators or municipally-owned utilities are trying to protect, firms often overlay cost allocation procedures onto price level decisions — limiting 'undue discrimination'.

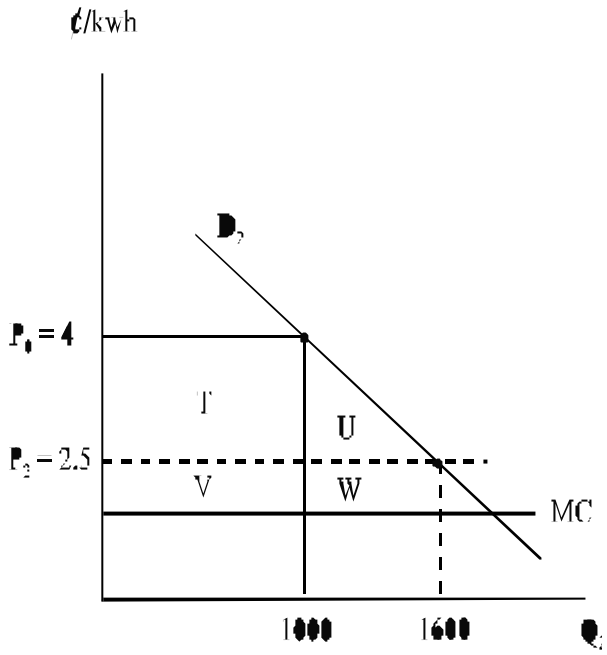
Ramsey Pricing is related to marginal-cost pricing in that prices are a percentage deviation from marginal costs, where the percentage is inversely proportional to demand elasticities: more elastic (responsive) demands have lower price-cost margins. In Figure 4, two demands are illustrated. If price is $.04 in each market, then the total contribution towards fixed costs is E+F ($20 in market 1) and T+V ($20 in market 2). So the base case involves 1000 kWh monthly consumption in each market, with $40 in revenue above incremental cost going towards fixed costs. If price is increased to $.06 in the relatively inelastic market 1, the firm gains A ($18) and loses F ($2), for a supplier gain of $16 in this market. Although the firm no longer receives revenues of F + G ($4) due to the 100 kWh reduction in quantity consumed, the supplier also avoided the cost (G) of producing that output. Thus, the loss of F is recorded. The customer is worse off: given by -(A+B), or $19. The numbers indicate that societal welfare has declined by $3, as price has been increased far above incremental cost ($.02). However, if the price is reduced to $.025 in market 2, then this customer gains more than the utility loses, and the potential for an overall welfare improvement exists.

On net, the reconfiguration of prices can improve societal welfare so long as gains (and losses) experienced by either customer are valued equally. To see this, consider Market 2 in greater detail, where price is reduced from $.04 to $.025, and consumption increases from 1000 kWh to 1600 kWh. The supplier loses T ($15) but gains W ($3) when it reduces price to $.025. There is a supplier loss of $12 in market 2. However, customer 2 gains T and U, where U is $4.50. Clearly, the customer gains (T+U equals $19.50) are more than the supplier's losses. The common sense explanation is that bringing price closer to incremental cost increases economic efficiency in this market (here, by $7.50). When the two markets are taken into account together, the supplier is actually better off (+ $18 – $12) while customer welfare (weighted equally for winners and losers) has also increased (down $19 in market 1 and up $19.50 in market 2).

This example is meant to illustrate the benefits of pricing closer to incremental cost in elastic markets — balancing off the inefficiencies of raising price further above incremental cost in another market. The formula for calculating optimal markups in the two markets involves markups that are inversely proportional to the respective demand elasticities.

.





Price increase
Supplier: + A - F
+ $18 - $2

Customer: - A - B
- $18 - $1

Price decrease
Supplier: -T + W
-$15 + $3

Customer: +T +U
+$15 + $4.50

Electric utilities have charged different prices to different customers for decades. Ramsey Pricing involves charging more to those with relatively inelastic demands. Commercial customers often do not have co-generation opportunities and are often the ones hit with the highest prices relative to incremental cost of service. Residential customers have political clout, so that despite relatively inelastic demands, their price-cost margins are often smaller than for other customer groups. Industrial customers, on the other hand, may be footloose in the long run: firms can move production capacity to other locations. Alternatively, industrial customers may have self-production as an option. In either case, these customers have relatively more elastic demands. Electric utilities have responded to such situations by offering lower prices (via 'cooked' cost allocations) or interruptible rates at discounts that might be greater than the savings warranted by the extent of actual interruptions.

Competition is likely to attack those customer classes with greatest price-cost margins. Thus, commercial accounts would appear to be vulnerable to entrants who have access to transmission and distribution facilities. All utilities will have to respond to such threats by re-structuring their rate designs. The presence of alternatives makes customer demands more elastic. So utilities will act in such a way as to reduce the prices quoted to such customers. From the standpoint of social efficiency, this restructuring is appropriate if the market demand of such customers is relatively elastic.

If the customers market demand for electricity is actually relatively inelastic — but it becomes elastic with the availability of competitive options, then the reduction of such prices may not promote social efficiency. However, from the standpoint of public policy, this possibility is probably swamped by the view that competitive pressures will be more effective in promoting cost containment than regulation. In terms of economic theory, the social efficiency losses (and gains) associated with welfare triangles are dominated by the large rectangles reflecting cost savings associated with improved incentives for cost containment. Public policy has supported increased competition in generation not just because of the associated rate restructuring (such that prices track costs), but also because costs are likely to be lower with competitive pressures.

Concluding observations

What are the implications of competitive trends for utility pricing?

Utilities that understand their cost structure and are successful at cost containment will be in a better position to develop prices that enable them to survive in a competitive world.

Utilities that understand their customers' actual (and potential) consumption patterns (and valuations) have an advantage over potential rivals.

Market intelligence will become a major factor in decision-making. Major customers will be lost and gained on the basis of the types of contracts that are developed.

New skills will be required of utilities. During the transition to competition, utilities must restructure themselves to provide the information and internal incentives required to compete effectively with their rivals.

These points need to be underscored. If the supplier does not know its own incremental cost, it cannot be sure whether additional sales are financially desirable. One implication of this point is that an obligation to serve high cost customers with low prices will need to be replaced by other funding mechanisms, if 'universal service' is to remain a public policy objective. Similarly, if the supplier does not know how the customer is likely to respond to new prices, financial planning and capacity decisions become problematic. Potential load shapes become relevant for decision-makers, since new rate designs will induce changes in consumption patterns. If simplicity is one casualty of competitive pressures, utility account managers will have to explain the benefits of more complicated rate structures to their customers. Competition will make life harder for infrastructure executives. The possibility of competition also complicates rulings on the price and quality of interconnection or an entrant's access to information possessed by an incumbent.

Monopoly suppliers are able to dictate price. An unregulated monopolist has an incentive to know its cost structure and demand patterns. However, a regulated monopolist (or a government enterprise) whose prices are partly the result of political compromises, has weakened incentives for cost containment and an inadequate understanding of cost-drivers. So long as total revenues cover total costs, there may not be pressure for a non-profit oriented firm to identify incremental costs. Some customers may be paying far more than cost of service while others could be paying less that incremental cost. However, the prices, based on (politically acceptable) cost allocations are not necessarily sustainable under competition — where price tends to track the incremental cost of electricity. Cost allocations which were unrelated to cost causation were possible in the absence of customer choice. But the world of electric utility and telecommunications monopolies is rapidly disappearing.

New price structures can offer win-win opportunities, such that both the supplier and customers can be better off than before. Of course, if competition drives the average price too low relative to the utility's average cost, most of the benefits from rate re-design will be captured by customers — if the utility tries to successfully retain its customers. The principles of rate design identified here are crucial to the recovery of joint costs in a competitive environment. Both managers and regulators need to understand the efficiency implications of alternative rate structures.

Bibliography

Arnold, Ron et.al, *Issues in Infrastructure Pricing, Bureau of Industry Economics*, report 69, Canberra, August 1995.

Berg, Sanford V. and J. Tschirhart, 1995, 'Contributions of Neoclassical Economics to Public Utility Analysis,' Land Economics, August, 71:3.

Berg, S. V. and J. Tschirhart, 1988, *Natural Monopoly Regulation: Principles and Practice*, New York, Cambridge University Press.

Braeutigam, R. E., 1989, 'Optimal Policies for Natural Monopolies,' Handbook of Industrial Organisation, vol. II, eds. R. Schmalensee, and R. D. Willig, North Holland, Elsevier Publishers B.V.

Brown, S. J. and D. S. Sibley, 1986, *The Theory of Public Utility Pricing*, New York, Cambridge University Press.

Burness, S. and R. H. Patrick, 1991, 'Peak-Load Pricing with Continuous and Interdependent Demand,' Journal of Regulatory Economics 3 (March), 69–88.

Crew, M. and P. K., 1986, *The Economics of Public Utility Regulation*, Cambridge, Mass., MIT Press.

Della Valle, A. P. and M. Bidwell, 'Restructuring Rates Creates Value and Reduces Stranded Costs,' The Electricity Journal, December 1995, 19–25.

Environmental Protection Agency, Water Planning Division, 1975, *Comparative Analysis of Cost Allocation Mechanisms*, Washington D.C.

Faulhaber, G. R., 1975, 'Cross-Subsidisation: Pricing in Public Enterprises,' American Economic Review 65, pp. 966–977.

Faulhaber, G. R. and W. J. Baumol, 1988, 'Economists as Innovators: Practical Products of Theoretical Research,' Journal of Economic Literature, 26 (no. 2, June), pp. 577–600.

LaMorte, Michele, 1983, 'Overview of Reports from the Electricity Utility Rate Design Study,' in Innovative Electric Rates by Sanford Berg, Lexington Books, 1983.

Malko, J. R. and P. R. Swensen, 1989, 'Pricing and the Electric Utility Industry,' in Public Utility Regulation, eds. K. Nowotny, D. B. Smith, and H. M. Trebing, Boston: Kluwer Academic Publishers.

Panzar, J. C. 1976, 'A Neoclassical Approach to Peak Land Pricing,' Bell Journal of Economics 7, pp. 521–30.

Sharkey, W. W., 1982, *The Theory of Natural Monopoly*, New York, Cambridge University Press.

Sherman, R., 1989, *The Regulation of Monopoly*, New York, Cambridge University Press.

Sheshinski, E., 1986, 'Positive Second-Best Theory: Brief Survey of the Theory of Ramsey Pricing,' In Handbook of Mathematical Economics, vol. 111, ed. K. J. Harrow, and M. D. Intriligator, North Holland, Elsevier Publishers B.V.

Sorenson, J., J. Tschirhart, and A. Whinston, 1976, 'A Game Theoretic Approach to Peak-Load Pricing,' Bell Journal of Economics 7, pp. 497–520.

Straffin, P. D., and J. P. Heaney, 1981, 'Game Theory and the Tennessee Valley Authority,' International Journal of Game Theory 10 (no. 1), pp. 35–43.

Valuation and costing issues in access pricing with specific applications to telecommunications.

Henry Ergas, University of Auckland, New Zealand.

Normatively, regulation can be viewed as an implicit or explicit contract between the regulatory authority, consumers and the regulated supplier(s)[1]. The essence of this contract is that the authority, acting as a consumer agent, commits to setting prices on a basis which recoups the long-run costs of efficient supply, where efficient supply is defined with reference to that which would prevail were the market at issue contestable. This, in turn, implies that the supplier derives a stream of residual income dependent on the costs it incurs relative to the prices which would be established under contestability.

In practice, the regulator cannot know the prices and costs associated with a purely hypothetical contestable market. In these circumstances, a pure cost-plus version of the regulatory contract, such as that involved in rate of return regulation without regulatory lag, would provide little incentive for cost-minimisation by the regulated supplier. Hence, if such incentives are to be provided, the expected residual income to the regulated supplier (and the expected path of prices to consumers) must — if opportunities to increase productivity are taken up — exceed that which would accrue to the supplier in the perfectly contestable counterfactual. The regulatory contract will therefore centre on an agreement between the regulator and the supplier about a residual income formula which balances the allocative efficiency costs of supra-normal margins with the gains in technical and dynamic efficiency which the prospect of supra-normal earnings can bring.

The fact that the regulated output is an input into a more-or-less competitive downstream market does not materially alter this conclusion, although it does make it more difficult to discern the ultimate welfare consequences of particular paths of prices[2]. There is, in particular, no case for manipulating the upstream price so as to artificially encourage downstream entry. It is true that the regulator, were it an omniscient social planner armed with perfect information, might balance incentive effects in the upstream market with competitive effects in the market downstream, and carry out the type of second-best analysis conventionally associated with infant industry protection. However, it is readily shown that the knowledge required to carry out this kind of analysis is very great[3]; and since regulators do not have anywhere near

---

1 See Goldberg, 1976, Williamson, 1976, and Schmalensee, 1979.

2 Economic analysis shows that welfare losses arise when taxes are imposed in upstream rather than downstream markets — that is, when intermediate inputs, rather than final outputs, are taxed. Even putting aside any consideration of competition effects, this complicates the setting of regulated prices for a supplier of intermediate inputs. (This is taken up again in the discussion below). The point being made in the text, however, is a different one and relates to whether the upstream price should be manipulated in such a way as to artificially encourage downstream entry.

3 See, for example, Damania, 1996.

perfect information, and nor do their proximate and ultimate supervisors, this kind of social engineering, like infant industry protection generally, is a policy that 'invites ultimate disappointment and (what might be worse than the mortification of non-arrival) misery on the way'[4].

Putting the temptations of social engineering aside, three issues need to be addressed in defining the expectation of residual income to the regulated supplier. These are: the definition and measurement of the residual income stream; the allocation issues associated with joint and common costs; and the basis for evaluating the resulting rate of return in terms of market benchmarks for the cost of capital.

Subsequent sections of this paper deal with each of these. A final section summarises and concludes.

Asset valuation and the definition and measurement of the residual income stream

I take it as given that the relevant income concept is that of Hicksian, standard stream income — that is, the maximum amount which could be taken out of the regulated enterprise by the owners in a given period without impairing their ability to take the same amount in all future periods of equal length[5]. Residual income, is in other words, defined in terms of the level of distribution to owners consistent with capital maintenance. This, essentially economic, approach to income determination is consistent with the fundamental requirement of company law (Corporations Law, s. 201(1)) that 'No dividend shall be payable to the shareholder of any company except out of profits'.

Operationalising the Hicksian approach requires a definition of capital maintenance. Attempts to do so in terms of the maintenance of physical operating capability suffer from serious, in my view fatal, conceptual weaknesses: they are, in particular, of no use when the asset base of the enterprise whose income is being measured is changing[6]. Rather, as Hicks himself emphasises, an economic concept of income can only be given substance in terms of maintaining the value of owners' equity[7]. Thus defined, the

---

4 Oakeshott, 1996, at p. 31.

5 Scott, 1976, at page 2. The generally accepted ex-post counterpart to Hicks' definition is 'the amount the corporation can distribute to the owners of equity in the corporation and be as well-off at the end of the year as at the beginning'; Alexander, 1950, at p. 15.

6 As Lemke notes: 'When the firm is in the process of changing the nature or composition of its physical capital (operating capability), it is nonsensical to talk about maintaining it', Lemke, 1982, at p. 320, emphasis in original.

7 Hicks contrasted his own views, which he considered consistently 'fundist' (i.e. concerned with maintaining 'the fund' advanced by the purchasers of residual income rights), with those of Pigou, who most clearly expressed what Hicks termed the 'materialist' approach to capital maintenance. See Hicks, 1974. The essence of the materialist view is that changes in asset prices, to the extent to which they do not force changes in physical operating capability or alter the physical service potential of assets, are not treated as forming part of current income but rather flow directly through to an adjustment reserve in the equity section of the balance sheet. Defences of the physical maintenance concept can be found in Revsine, 1973, and — in a much modified and qualified form — in Scott, 1989. Strong, and in my view convincing, criticisms are Sterling, 1982, at pp 3–58 and Carsberg, 1982, at pp 58–74.

capital maintenance standard determines both the measurement of the income flow and the resulting changes in the balance sheet.

In particular, since the decision to put an asset into operation must take account of anticipated changes in the asset's value, these changes are properly charged as depreciation against income from the asset in the period in which they are expected to occur[8]. Economic income, as Hicks put it, is then obtained by defining 'depreciation of the original stock of capital as the difference between the total value of the goods comprising that original stock as it is at the end of the year (C_1) and the value (C_0) **which would have been put upon the initial stock at the beginning of the year if the events of the year had been correctly foreseen, including among those events the capital value C_1 at the end of the year'**[9]. The Profit and Loss statement should, in other words, record a charge against operating income for correctly anticipated changes in asset values, leaving a Balance Sheet value of assets which reflects their replacement cost.

Basing the charge against operating income on the period-on-period change in replacement cost raises the issue of how 'replacement cost' is to be assessed when the asset which should be put in place at some future date differs from that currently used. In particular, it is sometimes argued that the replacement value should reflect the level of outlays required by the set of techniques yielding a service flow equivalent to those of in-use assets at the lowest monetary charge. This corresponds to valuation on a Modern Engineering Equivalent Replacement Asset (MEERA) basis.

Where in-use assets are still being traded — for example, because a programme of continued investment in those assets is in train — the prices at which those assets are obtained will provide the best reflection of their value to the entity. In particular, those prices will embody both the efficiency gains obtainable from the next generation assets and the adjustment costs the entity would incur in shifting from this generation to the next generation[10]. In contrast, optimised or MEERA prices may overstate the value of

---

8 In terms of Alchian's approach to the definition of costs, the firm incurs an exposure to the change in asset prices as a result of beginning the production programme at that time rather than postponing it to a future date (when the changes in asset prices will have occurred). The consequences of that change are therefore a charge correctly attributable to consumption in the period leading up to the change in prices. See Alchian, 1977, at pp 273–300.

9 Hicks, 1942, at p. 177, emphasis in original. Hicks repeats this definition in Hicks, 1973, at pages 164–6 and emphasises that it 'is rigorously forward-looking'. Indeed, he doubts whether any other rigorous definition of depreciation can be given consistent with his definition of income. A similar definition, cast in terms of social income, can be found, along with a useful contrast of capital maintenance concepts, in Triplett, 1996, at pp 93–115.

10 Were prices higher than this amount, the enterprise would simply shift to the next generation. In practice, adjustment costs, and hence the optimal lag in changing from current to next generation technology, are likely to be significant, as is apparent from the extent of technical inefficiency (in the sense of gaps between the production possibility frontier and actual factor usage) in even very competitive markets subject to rapid technological change; see for example, Caves, 1992.

next generation assets by excluding the capital costs involved in shifting from one vintage to the next[11].

Moreover, from a practical point of view, when optimised or MEERA values are used to value in-use assets, consistency between the Profit and Loss statement and the Balance Sheet can only be maintained by imputing notional adjustments to operating outlays. In particular, any savings in operating costs that would be achieved using the next generation assets need to be added back to actual (in-use technology) cash operating inflows[12], as is obvious when the next generation involves a lower (higher) initial cost but a higher (lower) per-unit cost[13]. It is, to my mind, questionable whether this can be done to an acceptable level of accuracy in complex systems with rapidly changing technologies[14].

In short, (1) assets should be valued at replacement cost, using, whenever possible, entry prices for in-use assets[15]; these will incorporate the effect of gains in efficiency associated with the arrival of new vintages; and (2) the period charge for depreciation should reflect the anticipated period-on-period change in replacement cost, adjusted for decay and deterioration[16,17].

---

11 See Revsine, 1979, at pp 306–322; Brooks, 1993, at pp 246–257; and Hulten and Wykoff, 1996, at pp 10–23. MEERA or optimised prices will overstate economic value unless the supplier of the next generation assets bears the adjustment costs involved in shifting users of current generation assets to the next generation assets. In contrast, the efficiency gains, netted against the transition costs, set a ceiling on the price of current generation assets. Current prices for the current generation assets will therefore provide a conservative basis for their valuation.

12 This is because the use of 'optimised' values involves deducting depreciation on the basis of the next generation asset. A consistent income statement can then only be obtained if current outlays are adjusted to reflect the level which would prevail were that asset in place. Revsine, 1979, at pp 306–322 provides worked examples.

13 In accounting terms, the initial cost is a non-current cost while the per-unit cost is a current cost. Adjusting one without adjusting the other would make little sense.

14 One can but sympathise with the view of the Bell System technical staff who wrote, of optimised asset valuation, that '.. for a large utility like a telephone company, even the visualisation of a complete property built of the most modern facilities would be a staggering job.' American Telephone and Telegraph Company, 1957, at p 121.

15 By convention, 'entry' prices are the prices at which goods can be purchased, while 'exit' prices are the prices at which they can be sold. Consistently with common sense, increments in asset prices should be valued using exit prices (the highest price the firm could obtain for its now more valuable assets) while decrements should be valued at entry prices (the lowest price at which the firm could obtain the now less valuable assets).

16 Decay is the decline in expected useful life; deterioration is the decline in the value of the services achieved at each point in remaining life. (The classic 'one-hoss-shay' depreciation of simple growth models arises when assets decay but do not deteriorate before exhaustion (retirement); with a zero discount rate, this gives rise to straight-line economic depreciation.)

17 This is subject to the following conditions: (1) No asset should be valued at more than the net realisable value associated with its flow of services. This condition, although not of practical significance in a growing network, means that in a declining network assets which will ultimately not be replaced should be valued at the higher of economic value and scrap value. (2) For all assets, differences between anticipated and actual changes in replacement values are taken directly to shareholders' equity.

This approach is consistent with observed valuation behaviour in capital markets. Although studies find mixed results with respect to the market response to current cost accounting information, there is growing evidence of the predictive ability of the 'clean surplus' equation, which defines corporate income in terms of the sum of this-period earnings and changes in asset values[18]. Investors, in other words, appear to assess returns on the basis of Hicksian income as defined above, and could be expected to cast a forward-looking contract, be it with a firm's managers or with its regulators, on that basis[19].

It remains to contrast the valuation approach set out above with three others: that based on historical costs; Optimised Deprival Value; and valuation based on exit costs.

Historical costs, as George Stigler put it, 'have powerful sway over untutored minds'[20]. Despite this, income flows derived from historical cost accounting will coincide with economic income only under extremely restrictive circumstances. It is possible to manipulate historical cost data so as to make the accounting rate of return on a project equal to its economic rate of return; however, this can only be done by taking account of changes in asset prices over the life-time of the project (and hence no longer corresponds to historical cost accounting as conventionally defined)[21]. Prices based on historical asset costs, and which reflect historical cost depreciation[22], will, as a general matter, bear no relation to those which would prevail in an unregulated market, since decisions in such a market will be made looking to the future, rather than to the past pattern of outlays[23].

---

18 On clean surplus, see Ohlson, 1995, Kwang, Munro and Peasnell, 1995, and O'Hanlan and Peasnell, 1996. Australian evidence on market response to changes in asset prices is discussed in Whittred, Zimmer and Taylor, 1996, at pp 114–117 and at p. 259.

19 The theoretical case for the role of changes in asset prices and of depreciation in valuation is set out in Feltham and Ohlson, 1996.

20 Stigler, 1987, at p. 111.

21 The basic result on equivalence of accounting and economic rates of return is set out in Edwards, Kay and Mayer, 1987, at pp 12–31. The dependence of this result on the treatment of changes in asset values is demonstrated in Peasnell, 1982, and Brief and Larson, 1992.

22 Under historical cost accounting, depreciation is not a valuation process but rather one of allocation of costs previously incurred on the basis of some concept of 'matching' the flow of services from those costs (which are then said to 'expire') with that of current revenues. This is inevitable as 'in an historical-cost based accounting system, a valuation approach to depreciation would be inconsistent with the measurement of non-current assets. Adoption of an allocation approach to depreciation is unavoidable given the decision to measure non-current assets at historical cost and to maintain intact money capital.' Peirson and Ramsey, 1994, at p. 16.

23 This point was apparent to the classical economists. Thus, Wicksteed noted that cost of production 'in the sense of the historical and irrevocable fact that resources have been directed to this or that special purpose, has no influence on the value of the thing produced. (What affects supply is anticipated cost) in the sense of the alternatives still open which now be relinquished in order to produce this specific article' since it is this, and only this, which 'influences the craftsman in determining whether he shall produce it or not'; Wicksteed, 1910, at p. 380. The best elaboration of this point remains Buchanan, 1969, which also contains a detailed discussion of the development of economists' view of costs.

Two claims are nonetheless made on behalf of historical cost accounting.

The first is that only the use of a historical cost standard can guarantee capital maintenance — i.e. can ensure that the value of investors' initial outlays is preserved[24]. In principle, regulators, given a free hand, could set prices so as to secure virtually any concept of capital maintenance. From an economic point of view, however, the question is whether the capital maintenance goal selected has any normative justification. The replacement cost approach set out above values assets at current entry prices — and hence reflects the asset valuations consistent with the output prices which would prevail in a contestable market — while achieving financial capital maintenance through the depreciation charge to period income. In contrast, maintenance of historical capital, achieved through a backward-looking depreciation charge, will generate a balance sheet which serves merely as a 'repository of unamortised costs'[25], and hence reflects no more than 'a mélange of variously dated amounts, of qualitatively disparate significance'[26]. It is difficult to find an efficiency justification for setting prices on this basis.

The second, somewhat stronger claim, points to practicality. Historical costs, it is said, are what firms do; they are 'objective' and easy to ascertain. In contrast, replacement valuation relies on judgements which are 'subjective', outside of firms' normal experience, and inevitably contentious[27]. Two points need to be made in this respect.

To begin with, even in principle, historical cost accounting involves a broad range of inherently arbitrary judgements and allocations. For example, there is no clear basis in historical cost accounting for determining the choice of the depreciation schedule; rather, this is done with reference to a notion of 'matching' (or 'cost attachment') which is broadly incapable of falsification[28]. There is, in other words, no analytical basis in historical cost accounting for the capital consumption charge imposed against revenue in any particular period, making the apparent 'objectivity' of this charge entirely illusory.

---

24 Professor Hausman, has argued, for example, that: 'Even if actual historical network investment decisions were always completely efficient at the time they were made, improvements in technology will **always** guarantee that a totally new, hypothetical network will have a theoretical lower cost than the actual network in place (or otherwise the older technology could be used in the hypothetical network). Thus, basing cost on the current most efficient technology will impart a downward bias on estimates of actual network costs, causing an economic loss to the LECs which made the historical investment' Hausman (Affidavit), at note 4 on p. 7.

25 Coase, 1973, at p. 120.

26 Chambers and Wolnizer, 1990, at p. 360.

27 See, for a recent expression of these views, King, 1996.

28 The classic demonstration of the arbitrary nature of these allocations is in Thomas, 1969.

Second, precisely because historical cost accounting is so arbitrary, it is not in fact what firms do or are required to do. Thus, entities subject to the Corporations Law are required to account in conformity with Australian accounting standards. AASB 1010 requires that non-current assets not be shown at above recoverable amount. In conjunction with s. 294(4) of the Corporations Law, this provision requires Directors to review the replacement value of non-current assets relative to their earnings ability, and report accordingly[29]. Consistent with the comprehensive income definition embodied in AASB 1018, AASB 1010 also requires that downward revaluations flow directly to the Income Statement, where they are treated as an expense.[30] With the number of Australian companies revaluing their assets increasing substantially in recent years, these provisions have been of growing significance[31]. The requirement to maintain accounts on a basis consistent with current asset valuations is even clearer in respect of government business entities and forms a central part of National Performance Monitoring[32]. Viewed together with mark-to-market valuation for current assets, and for particular classes of reporting entities (notably for insurance funds, including those involved in superannuation[33]), it seems reasonable to conclude that in Australian accounting 'too many instances of the use of market prices as the basis of valuation currently exist for anyone to say that market prices do not constitute sufficient objective evidence'[34].

Optimal Deprival Value (ODV) valuation, also known as valuation on the basis of 'value to the owner', differs from the replacement cost approach set out above in two respects.

To begin with, ODV, at least in its implementations to date, rests on a concept of physical rather than financial capital maintenance, and hence corresponds to the entity, as against proprietorship, approach to financial accounting[35]. For the reasons indicated

---

29 The interaction of AASB 1010 with s. 294(4) is set out in ASC *Practice Note 21: Value of Non-Current Assets*. See also Perkins, 1996, at 74–75. Note also that AASB 1021 requires annual reviews of likely useful life, and that the effects of resulting changes must be reported in the Profit and Loss statement in accordance with AASB 1018.

30 Persumably on the basis of conservativism, under Australian accounting standards, asset revaluations do not flow to the P&L but rather must go directly to an asset revaluation reserve, which is available for distribution on realisation.

31 Whittred, Zimmer and Taylor, 1996, at pp 252z–3.

32 See Steering Committee on National Performance Monitoring of Government Trading Enterprises, 1994.

33 Under AASB 1023, these are now required to report all assets at market values on balance sheet date.

34 Godfrey, Hodgson, Holmes and Kam, 1994, at p. 476. The authors go on to note that 'it may be true that in some cases they are not reliable, but we cannot demand that in all cases a firm must be a direct participant in the transaction before revenue or gain can be recognised'.

35 It is therefore curious that the Public Sector Accounting Centre of Excellence of the Australian Society of CPA's, in its useful study of asset valuation approaches for GTE's (*Asset Valuation by Government Trading Enterprises*, March 1996), argues that ODV accounts should record holding gains on debt. Even if there were such gains (which assumes that creditors do not fully price protect), they would be transfers among stakeholders in the entity, and hence would not be viewed as income in an entity (as against proprietorship) accounting approach.

above, this approach breaks down when the composition of the assets held and services provided by the firm changes substantially over time.

The second difference lies in the role and nature of optimisation. In theory, ODV bases valuation not solely on MEERA but also on re-engineering the system being valued to eliminate redundant assets. It is a matter of some debate as to how far this is actually taken in ODV implementation[36]. Nonetheless, two points need to be made.

First, the areas where ODV has been applied — primarily high-voltage electricity transmission — have highly characterised technologies, in the technical sense of having a clearly defined best-practice framed by tight physical design constraints. Technological change tends to be relatively slow and incremental, rather than radical, in nature, and relative input prices and implied service prices are broadly predictable. As a result, there is a high level of agreement as to the 'right approach' to be adopted to service provisioning in particular instances, facilitating the kind of optimisation envisaged in ODV[37].

Second, even within these systems, optimisation has very largely concentrated on stranded assets — that is, on assets which clearly will not be replaced, and where the optimisation involves bringing forward in time the recognition of the asset's writing-off. This is a particularly simple version of system redesign, and one that raises issues which are primarily commercial rather than technical.

In contrast, in telecommunications, where technological change is faster and more radical, there is far greater uncertainty as to the network designs which will ultimately prove successful, increasing the error associated with hypothetical efforts at optimisation. Moreover, when there is uncertainty as to the 'best bet', yet investment decisions are irreversible, the appropriate metric for determining the efficient path of investment is not the Net Present Value of expected cash flows (or even less of future gross costs) but rather the option value of each of the alternative paths[38]. However, even the most sophisticated of the current network optimisation models are far from

---

36 It has, for example, been suggested that the stream of service costs associated with the asset values generated by the ODV valuation of Transpower in New Zealand are too high to be consistent with effective optimisation; see Cox, Spiller and Teece, 1994.

37 Equally, the fact that these systems involve many assets which are extremely long-lived, and for which reproduction costs may not be available or meaningful, makes it more sensible to use MEERA for class-of-asset valuation than it would otherwise be.

38 Kester shows that project selection on the basis of the NPV metric may yield extremely misleading results under these circumstances: see the examples given in Kester, 1993, pp 187–207; see also Trigeorgis, 1988.

being capable of carrying out this kind of analysis[39], and hence cannot provide a reliable indication for asset valuation purposes of the gap between the 'network as it is' and the 'network as it ought to be'. Lastly, the fact that telecommunications networks generally require two-way, any-to-any, compatibility means that the adjustment costs involved in shifting vintages can be considerable. Efficiency, even when narrowly defined as cost-effectiveness, may therefore require a network design which seems quite at odds with that of a 'green-field' alternative; but this too is poorly captured by drawing-board contrasts between actual and (hypothetical) best-practice.

As a result, ODV-style optimisation cannot readily be implemented in the circumstances of the telecommunications industry. Rather, as suggested above, current price information should be used to value assets onto a replacement basis, with anticipated reductions in asset values being treated as part of period costs.

A final approach is that of **exit price valuation**, which is often referred to as Continuously Contemporary Accounting (CoCoA)[40]. The criticisms of CoCoA are well-known: not the least of these is the point that if the economic value of firms simply equalled the sum of the current resale prices of the individual assets under their control, firms as such would cease to exist[41]. Further, CoCoA seems to be based on a confusion as to the nature of opportunity costs: there is no sense in which the opportunity cost of using an asset for any one purpose can generally be taken to be equal to the salvage value of the asset; rather, it is equal to the value of the asset in the next most-highly valued use, which may well be its deployment for some other purpose within the firm. Moreover, even if the disposal stand-point is adopted, exit prices are in practice highly sensitive to the level of aggregation (the 'bundles' of assets being priced), to the timing of sale, and to the information asymmetries familiar from the 'lemons' problem in the market for used cars; as a result, exit prices are difficult to define and calculate[42]. Because of these issues and others, a positive profit under CoCoA valuation merely indicates that the firm should retain its current assets, but says nothing as to whether it should replace these assets over the longer term.

---

39 Some of the analytical requirements for even rather simple models of network evolution under uncertainty are set out in Adjali, Fernandez-Villacanas and Gell, 1994, making the gap between these requirements and current implementations all the starker.

40 Exit price accounting is primarily associated with Chambers, 1966, and Sterling, 1970.

41 After all, the resale (exit) prices are presumed to be certain, while any income to the residual claimant is not. As a result, all assets would be sold.

42 King, 1996, suggests the use of exit prices (salvage values) as an alternative to ODV (which he considers impractical), but does not consider the standard issues raised in this respect in accounting theory. See, for example, Godfrey, Hodgson, Holmes and Kam, 1994, at pp 181–202; Whittred, Zimmer and Taylor, 1996, at pp 577–580, who conclude that while exit price accounting has some attractions, it is impractical and hence not used in contracts; and Thomas, 1969, at p. 67 and following (although it is fair to note that Thomas, despite his earlier strong criticisms of exit prices, later gave highly qualified support to the use of CoCoA as being less harmful than conventional historical cost accounting — see American Accounting Association, Florida, 1974).

Despite these weaknesses, the main feature of CoCoA of relevance here is its treatment of highly specialised, 'nonvendible' assets (such as ducts). In CoCoA, these are subject to immediate expensing, on the grounds that the firm, in acquiring these assets, forecloses the opportunity of adapting to changing circumstances through either the postponement of decision or (subsequent to purchase) the resale of the assets in question. In this, CoCoA parallels the practice common for many years in railroad valuation, and which has re-emerged in recent years as 'infrastructure accounting'[43]. Two points need to be made in this respect.

First, infrastructure accounting offers only a very partial valuation framework, because it deals mainly with the nonvendible component of the firm's asset base. The valuation method to be used in respect of the remaining assets is not determined within the infrastructure accounting approach.

Second and most important, the immediate expensing of nonvendible asset outlays, if reflected in prices, entails a form of short-run marginal cost pricing, in which prices rise steeply as capacity constraints are approached[44]. Simulation studies suggest that such pricing can provide for welfare gains when compared to smoothed, or Boiteaux, prices based on approximations to long-run incremental costs[45]. However, maintaining these welfare gains when the access service is required to break-even will generally entail a lump sum charge which will be the higher, the greater the lumpiness of capacity, the difference between capacity goods in the size of efficiently-scaled lumps, and the extent of economies of scale in capacity addition[46]. The effect of these lump-sum charges on the cost structure of entrants, combined with the inherent variability of congestion-related charges, may make them unacceptable. Moreover, the calculation of short-run marginal costs in a network environment will generally require imputing shadow capacity prices at each node in the network — a task of very great complexity

---

43 See, for example, Office of Water Services, 1994.

44 See especially Ng, 1987. Somewhat unusually, Hausman says that 'short-run marginal costs do not account for the cost of capital at all' (Professor Jerry A Hausman *In the Matter of Implementation of the Local Competition Provisions in the Telecommunications Act of 1996*, Federal Communications Commission, CC Docket No. 96–98, at p. 8). This is not true with goods subject to congestion, in which short-run cost rises to the lesser of the value of lost demand, or the full cost of the next lump of capacity, as supply constraints are approached. As Hotelling put it 'when a train is completely filled, and has all the cars it can haul, the marginal cost of carrying an extra passenger is the cost of running another train.' (Hotelling, 1938, at p. 264). In switched telecommunications networks, where the blocking probability rises rapidly as the ratio of the offered load to capacity rises, the congestion cost element in short-run marginal cost can be triggered even at what seem to be low levels of capacity utilisation.

45 Boiteaux prices involve smoothing the path of costs over time; they are obtained by pricing as if (lumpy) capacity were perfectly divisible — that is, pricing at the average incremental cost of capacity plus variable costs. Since Boiteaux prices are defined as the minimum of the cost function, Boiteaux pricing requires that capacity be fully utilised, which in turn requires that unit charges are calculated at a level of output defined in terms of integer multiples of the asset with the largest minimum efficient lump size. As a result, at output levels other than that corresponding to full utilisation, Boiteaux prices will fall short of recouping costs.

46 See especially Park, 1989; and Park, 1994.

in fully-interconnected star or mesh network structures. As a result, estimates of these costs can be subject to considerable error.

Overall, three conclusions flow from the analysis set out above.

First, asset values should be set on the basis of replacement costs, subject to the constraint that no asset be valued in excess of its recoverable amount. So as to maintain consistency between the balance sheet and the income statement, replacement costs should be calculated by reference to the current purchase prices of the best in-use assets. Since these prices will reflect expectations of efficiency gains available from next-period technology, the use of these prices provides the most reliable means of effecting optimisation.

Second, anticipated changes in replacement cost should be included in the period depreciation charge, and hence netted from income in the calculation of economic profit.

Third, assets with indefinite lives may be treated separately and expensed in their period of acquisition. However, this entails acceptance of charges which may be highly variable over time, and may differ substantially from node to node in the network. Moreover, as congestion is reached, this approach may give rise to contentious, and fundamentally irresolvable issues with respect to cost responsibility, as traffic-generating units seek to be treated as infra-marginal so as to escape the charge bearing on marginal traffic.

The treatment of joint and common costs

However complex it may be, asset valuation is obviously only a part of the process of defining service costs. In practice, asset services flow to a range of products, both at any one point in time and across points in time, and cost definition requires some means for identifying these flows. These issues, classically dealt with in the setting of cost recognition rules, will not be canvassed here; rather, I will take as given 'best practice' analysis of cost causation, and focus on the problems associated with the recovery of those costs which, despite such analysis, cannot be unambiguously allocated to particular service flows.

The problems posed by these costs, which can loosely be referred to as common costs, are clearly of considerable practical significance. Thus, Professor Hausman states that 'fixed and common costs are typically estimated at about 50 per cent or more of total Local Exchange Carrier (LEC) costs.'[47]. Moreover, it seems likely that the share of

---

[47] Professor Jerry A. Hausman *In the Matter of Implementation of the Local Competition Provisions in the Telecommunications Act of 1996,* Federal Communications Commission, CC Docket No. 96–98, at note 1 on p. 4. Professor Hausman's approach is unusual in grouping 'fixed' costs (which may well be product-specific) with 'common' costs (which by definition are not). Moreover, in the long run, all the costs are variable, but not all are product-specific. Professor Hausman then states, at paragraph 10, that TSLRIC will not recover 'fixed and common costs which arise from network economies of scale and scope'. Either this statement is wrong, or Professor Hausman is actually referring to joint and common costs. Fixed costs, and those associated with economies of scale, are included in TSLRIC.

costs which are truly common is rising over time, as the transition to multi-service, integrated networks increases the economies of scope attainable in telecommunications networks[48].

Further issues of common cost recovery arise from the costs of non-commercial service obligations. These costs are not caused solely by the requirement to provide designated services throughout Australia at geographically averaged prices; they also arise from the constraints which the current price caps impose on price rebalancing as between the provision of access on the one hand and of usage on the other. There seems little doubt that the latter costs are very much greater than the former[49]; yet it is only the former that are captured in the explicit funding mechanisms de signed to recoup the costs which non-commercial obligations impose. The unfunded liability this creates can be viewed as a cost which is common to, and must be recouped across, the range of services provided.

In short, common costs are likely to account for a substantial part of the total resources deployed in the telecommunications network. Barring direct subsidies (which are clearly not a feasible political option at this stage), these costs must be recovered if the access provider is to break-even.

Efficient recovery of these costs will require a mark-up over the attributable long-run costs of each service, including access. This is for two reasons.

First, recovery of the entirety of these costs at the final service (retail) layer would only be efficient if the access service provider could directly or indirectly tax competitors' revenues at that layer by an amount just sufficient to recoup the cost shortfall. A tax only applied to the access service provider's final output would distort choice toward competitors' offerings, and could result in substantial inefficiency. Since such a competitively neutral tax on final output is not feasible under the regulatory arrangements, some part of the relevant contribution must be obtained through the price of access.

Second, if access is priced only at Total Service Long Run Incremental Costs (TSLRIC), access providers will, under plausible assumptions, lack incentives to fully

---

48 The increased ability to exploit scope economies is emphasised (although without being referred to as such) in, for example, the analysis of ATM networks in Linberg, 1995; of optical transport in Gerstel, 1996; and of intelligent networks in Magendanz and Popescu-Zeletin, 1997.

49 See, for example, the estimates of inherited price distortions presented in Industry Commission, 1997, at p. 99 and at pp 149–156. These confirm the earlier results presented in Ergas, Ralph and Sivakumar, 1990, derived from reworking the data set used in the landmark study by the Bureau of Transport and Communications Economics, 1989. The Ergas-Ralph-Sivakumar estimates, when combined with those from BTCE, imply that transfers between services are nearly ten times greater than those between geographical areas.

achieve economies of scope in the provision of access services[50]. Rather, they may opt for technologies which increase total costs, so long as a larger share of these costs is directly attributable to the access service[51]. Given that the regulatory arrangements may create other incentives which run in the same direction[52], it would seem undesirable to compound the efficiency-reducing effects through the system of access price regulation.

Given that a mark-up over long-run attributable costs is required, the allowable nature of this mark-up needs to be determined.

In principle, the mark-up should be sensitive to demand conditions, with the wedge between prices and attributable costs being set so as to minimise the mark-up's effect in suppressing demand[53]. The practical implementation of this principle needs to be modified to take account of the consequences of the constraints on service price rebalancing, which mean that loss of contribution in the (relatively elastic) trunk services cannot be recouped through raising charges for rentals and local calls[54]. As a result, although the unconstrained second-best mark-up would be greatest on inputs to the local access service, and lowest on those to the trunk service, the opposite pattern may need to prevail.

Within each service, the mark-up across customers should also be such as to minimise the efficiency costs of the price distortion. This, in turn, suggests that the contribution

---

50 Hausman put this proposition as follows: 'If the LECs do not recover all of their joint and common costs, they have an incentive to use technology with reduced economies of scale and scope but higher per-unit (in case of LRIC) or per-service (in case of TSLRIC) costs, because the latter costs will be more fully captured by prices set at incremental cost. This action may be rational for the firm but it raises social costs and deprives society of productive efficiencies.' Professor Jerry A Hausman In the Matter of Implementation of the Local Competition Provisions in the Telecommunications Act of 1996, Federal Communications Commission, CC Docket No. 96–98, at p. 5.

51 For example, assume that there are two services, 'access' and 'usage'; that all costs are fixed; that there is only one provider of access but that usage is contestable; and that there are two alternative technologies, one which allows substantial economies of scope (and lower total costs (TC)), and one which does not. Let TC1 and TC2 refer to total costs for the first and second technology respectively, and similarly for LRIC(access). If access prices are set at LRIC, and TC1 - LRIC1 (access) > LRIC2 (usage) then, noting that competitors downstream cannot be assumed to be operating at more than LRIC2(usage), the access provider will choose technology two (since its 'costs' in the downstream market otherwise exceed those of competitors).

52 Thus, incentives to increase the attributable costs of access may arise if access prices are effectively subject to some form of lagged rate of return regulation, and access demand has a low firm- and product-elasticity of demand, while retail prices are subject to incentive regulation and are more elastic.

53 As Baumol and Sidak succinctly put it 'where 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 prices'; Baumol and Sidak, 1994, at p. 50.

54 A useful survey of research on demand elasticities for telecommunications services is provided in Industry Commission, 1997, at pp 31–42 and 133–148.

to common costs should be recovered to at least some extent through a higher charge on infra-marginal units. For example, a non-linear pricing scheme, in which fixed fees played a part, could well provide a less distorting means of pricing the access service than would a uniform price. The effects of such schemes are admittedly complex when the number of consumers is endogenous, and the complexities are all the greater when the consumers of the service are competitors in a downstream market. However, assuming that the downstream service is at least reasonably potentially competitive, a multi-part tariff which is available to all purchasers of the input is unlikely to distort competitive conditions, so long as the purchase price at the top end of the discount scale does not fall below attributable cost[55].

Experience in competitive markets is relevant here. In effect, studies find (1) that firms, in setting internal transfer prices, take account of the need to recover common costs, and hence mark-up over attributable costs; (2) that complex, non-linear charging arrangements are used to minimise the efficiency costs of the mark-ups; and (3) that these patterns of charging are consistent with the need to align the incentives of managers at decentralised levels within the firm with overall value maximisation, notably in the taking of investment decisions[56]. At the same time, similarly complex, discriminatory pricing arrangements are widely used in respect of third parties, including in markets for intermediate inputs where the supply of these inputs is characterised by economies of scale and scope[57]. All of this merely reinforces the conclusion that efficient pricing of access services will require prices which depart significantly from uniformity.

Determination of the cost of capital

The last issue to be considered here is the determination of the cost of capital. In a relatively capital-intensive activity such as telecommunications, the approach taken to determining the cost of capital can have a major impact on estimated service costs. This makes it all the more important that the approach adopted to be firmly based.

The Capital Asset Pricing Model (CAPM) provides the primary point of reference in this respect. The CAPM has figured in AUSTEL's regulatory price determinations; and it is put forward for special consideration in the Australian Competition and Consumer Commission's guidelines for Part IIIA undertakings, as well as for access pricing in telecommunications.

The prominence of the CAPM sits uneasily with its performance. It has been known for some years that the CAPM does not perform well in explaining behaviour in capital markets, although it is admittedly difficult to frame models which discriminate between testing for the CAPM and testing for capital market efficiency. Further, it is also

---

55 The relevant tests are set out in Baumol and Sidak, 1994, at p. 73; and in Baumol, 1996.

56 For empirical studies see, for example, Fremgen and Liao, 1981; Joye and Blayne, 1990; and Shih, 1996, 178–195. The relevant theory is surveyed in Hemmer, 1996.

57 The results set out in Dolan and Simon, 1996, provide a wealth of examples in this respect.

known that the CAPM has difficulty in explaining relatively widespread patterns of corporate behaviour, for example the taking of insurance against what appear to be diversifiable risks[58]. Last but not least, the analytical bases of the CAPM have come under sustained criticism in recent years, and the results of that criticism suggest that the CAPM may be especially inappropriate in telecommunications.

In essence, in the CAPM, uncertainty only affects investment through the expected covariance of a firm's earnings with those of the market (which is defined as the collection of assets which would figure in a fully diversified portfolio). An increase in the covariance increases the riskiness of investment, increasing the required rate of return. Conversely, a change which leaves this covariance unchanged will not alter the firm's incentive to invest or its desired level of the capital stock even if it increases the expected variance of the firm's earnings.

However, this view of the relation between the expected pattern of future earnings and the required rate of return does not hold when investment is irreversible. At its simplest, this is because irreversibilities make returns to investment asymmetric[59]. If outcomes fall short of expectations (say because competitors' entry costs fall more rapidly than anticipated), investors are stuck with the low returns[60]. Conversely, if outcomes exceed expectations, the incentive is to invest more, thereby limiting the rise in the marginal product of capital[61]. As in the CAPM, an increase in uncertainty will, under these circumstances, lead to a rise in the required rate of return and a reduction in the desired capital stock. However, in contrast to the CAPM, so long as the change increases the expected variance of the firm's earnings, the discouraging effect on investment will occur independently of any change in the correlation between the uncertainty bearing on the firm and that bearing on the market as a whole.

Empirical tests confirm this result, showing that it is the variance of earnings, rather than the covariance with market returns, which determines required rates of return on investment[62]. The results of these tests suggest that CAPM-based estimates of capital costs will be inaccurate, and may be seriously so in an industry as capital-intensive as telecommunications.

Two further points are important in this regard.

---

58 See, for example, Stedman and Kim, 1993, at pp 120–139.

59 See generally Dixit, 1989; and Pindyck, 1988.

60 Even if a firm can resell its capital stock, the same market conditions reducing the attractiveness of those assets to the firm may well reduce their attractiveness to potential buyers. As a result, even industry-specific, but not firm-specific, investment may be irreversible in the face of industry-specific shocks. See Abel et al, 1996.

61 This simple model is set out in Leahy and Whited, 1996.

62 See for example Howe and Vogt, 1996; Hurn and Wright, 1994; Leahy and Whited, 1996.

First, greater competition, to the extent to which it increases the variance of expected earnings, will increase the required rate of return on investment[63].

Second, the extent of these effects is likely to rise as advances in telecommunications technology increase the minimum efficient scale of capacity increments. While new technologies do offer some scope for modularity, the trend, at least in the fixed network, is to provision capacity in ever larger lumps[64].

Given these considerations, the determination of capital costs will need to draw on an option value framework, rather than on the CAPM. While analytically challenging, considerable experience has by now accumulated in this type of analysis, making it ripe for use in the costing of access services.

Conclusions

This paper has touched on only a few of the many issues which arise in the costing of access services. The main points made can be summarised as follows.

First, asset valuation should be based on replacement costs, with anticipated reductions in asset values being charged against period revenues. Rather than system optimisation of the kind envisaged in ODV, the effects of improvements in technology should, where reasonably possible, be addressed by pricing assets at entry prices for the best in-use vintage. The major constraint here lies in the possible error involved in estimating the relevant prices. Unfortunately, the higher the rate of technological change — and hence the greater the gap between historical and current costs — the greater may also be the difficulty and error estimating replacement costs entails. This can severely limit the additional informational the analyst — be that analyst a regulator or investor — can derive from replacement cost valuations.

Second, access prices will need to contribute to the recovery of joint and common costs. The contribution sought for these costs from access services will need to reflect regulatory price distortions, notably so as to avoid inefficient entry. To minimise the resulting economic costs, access prices should be structured in such a way as to secure the greatest contribution from infra-marginal traffic. This will entail a move away from uniform prices.

Third, the cost of capital benchmarks used to evaluate the reasonableness of access charges need to reflect the effect of irreversibilities in investment. Cost of capital

---

63 Thus, Caballero and Pindyck, 1996, find, using U.S. data, that a doubling of industry-wide uncertainty raises the required rate of return on new capital by about 20 per cent.

64 Fibre optic transmission, for example, is being provisioned in cables with up to 288 fibres, with single-fibre-pair WDM multiplexers available capable of providing 40 Gb/s over 600 kilometres. Assuming that each fibre supports 20 wavelengths, and each wavelength is modulated at 2.48 Gb/s, a node joining the cable to others would handle over 6000 wavelengths and have an offered load of some 15 Tb/s. Optical servers capable of handling this throughput would dwarf even the largest current exchanges.

estimates derived from the CAPM are likely to understate these effects, and hence to underestimate required rates of return.

Ultimately, the regulatory contract is by its nature incomplete[65]. Preserving the ability to respond to change requires that the parties retain a degree of flexibility, rather than seeking to anticipate, at the outset, all possible contingencies. In turn, the discretion which contractual incompleteness vests in the regulator, and the resulting scope for the regulator to act opportunistically[66], inevitably creates an added element of uncertainty. Reducing uncertainty is not a sensible goal in and of itself[67]; but nor is any purpose served by introducing more uncertainty than is needed[68]. A clear commitment to addressing the issues set out above in a manner consistent with economic efficiency could only help to achieve a desirable balance in this respect.

Bibliography

Abel, A. et al, 1996, 'Options, the Value of Capital and Investment', *The Quarterly Journal of Economics*, vol. 106, pp 754–777.

Adjali, I., J. L. Fernandez-Villacanas, and M. A. Gell, 1994, 'Markets Inside Telecommunications', *BT Technology Journal*, vol. 12, pp 30–40.

Alchian, A., 1977, 'Costs and Outputs', in *Economic Forces at Work*, Liberty Press, Indianapolis, pp 273–300.

Alexander, S. S., 1950, 'Income Measurement in a Dynamic Economy', *Five Monographs on Business Income*, American Institute of Certified Practising Accountants: Study Group on Business Income, New York.

American Accounting Association, 1974, *The Allocation Problem: Part II*, Florida.

American Telephone and Telegraph Company, 1957, *Depreciation: History and Concepts in the Bell System*, New Jersey.

---

65 See Crocker and Masten, 1996.

66 A common form of opportunistic conduct by regulators is the attempted expropriation of sunk assets, which (by reducing the return on these assets below that corresponding to long-term replacement valuation) takes from the supplier the quasi-rents it would otherwise secure. Extreme forms of this have at times prevailed, notably but not solely in developing countries, often under the pressure of political populism.

67 See for example Raz, 1977, stressing the trade-offs between certainty, both as a goal in itself and as a means to other goals, and other objectives of public policy.

68 Thus, the evidence suggests that unstable and opportunistic behaviour by U.S. regulators, notably in disallowing the recovery of investments, has been associated with increases in the cost of capital to regulated utilities and reductions in their investments and service quality. See, for example, MacAvoy, 1992, at pp 53–57 (on the experience of the electric utilities); and Loudder, Khurana and Boatsman, 1996.

Baumol, W., 1996, 'Predation and the Logic of the Average Variable Cost Test', *Journal of Law and Economics*, vol. 39, pp 49–72.

Baumol, W. and G. Sidak, 1994, *Towards Competition in Local Telephony*, The MIT Press for the American Enterprise Institute, Cambridge.

Brief, R. P. and R. A. Larson, 1992, 'The Role of the Accounting Rate of Return in Financial Statement Analysis', *The Accounting Review*, vol. 67, pp 411–426.

Brooks, L. D., 1993, 'Capital Budgeting for Replacements With Unequal Remaining Life', in R. Aggarwal (ed.), *Capital Budgeting Under Uncertainty*, Prentice-Hall, New Jersey, pp 246–257.

Buchanan, J. M., 1969, *Cost and Choice: An Inquiry in Economic Theory*, Markham, Chicago.

Bureau of Transport and Communications Economics, 1989, *The Cost of Telecom's Community Services Obligations*, AGPS, Canberra.

Caballero, R. and R. Pindyck, 1996, 'Uncertainty, Investment and Industry Evolution', *International Economic Review*, vol. 37, pp 641–660.

Carsberg, B., 1982, 'The Case for Financial Capital Maintenance' in Sterling, R.R. and K. W. Lemke (ed.s), *Maintenance of Capital: Financial versus Physical*, Scholars' Book Co, Houston, pp 58–74.

Caves, R. E., 1992, *Industrial Efficiency in Six Nations*, MIT Press, Cambridge.

Chambers, R., 1966, *Accounting, Evaluation and Economic Behaviour*, Prentice-Hall, New Jersey.

Chambers, R. J. and P. W. Wolnizer, 1990, 'A True and Fair View of Financial Position', *Company and Securities Law Journal*, pp 353–368.

Coase, R. H., 1973, 'Business Organisation and the Accountant', in Buchanan, J. M. and G. F. Thirlby, *L.S.E. Essays on Cost*, Wiedenfeld and Nicolson, London, pp 97–132.

Cox, A., P. T. Spiller, and D. Teece, 1994, *Transmission Pricing Issues in New Zealand*, LECG, Berkeley.

Crocker, K. and S. Masten, 1996, 'Regulation and Administered Contracts Revisited', *Journal of Regulatory Economics*, vol. 9, pp 5–39.

Damania, D., 1996) 'The scope for collusion in a vertically integrated telecommunications industry', Information Economics and Policy, vol. 8, pp 141.

Dixit, A., 1989, 'Entry and Exit Decisions Under Uncertainty', *Journal of Political Economy*, vol. 97, pp 620–638.

Dolan, R.J. and H. Simon, 1996, *Pricing*, The Free Press, New York.

Edwards, J., J. Kay, and C. Mayer, 1987, *The Economic Analysis of Accounting Profitability*, Clarendon, Oxford.

Ergas, H., E. Ralph, and S. Sivakumar, 1990, *Reforming Australian Telecommunications*, MONICT, Melbourne.

Feltham, G. and J. A. Ohlson, 1996, 'Uncertainty Resolution and the Theory of Depreciation Measurement', *Journal of Accounting Research*, vol. 34, pp 209–233.

Fremgen, J. M. and S. S. Liao, 1981, *The Allocation of Corporate Indirect Costs*, National Association of Accountants, New York.

Gerstel, O., 1996, 'On the Future of Wavelength Routing Networks', IEEE Networks, vol. 10, pp 14–20.

Godfrey, J., A. Hodgson, S. Holmes, and V. Kam, 1994, *Accounting Theory*, Jacaranda Wiley, Milton.

Goldberg, V., 1976, 'Regulation and Administered Contracts', *Bell Journal of Economics*, vol. 7, pp 426–448.

Hausman, J. A., *In the Matter of Implementation of the Local Competition Provisions in the Telecommunications Act of 1996*, Federal Communications Commission, CC Docket No. 96–98.

Hemmer, T., 1996, 'Allocations of Sunk Capacity Costs and Joint Costs', *The Accounting Review*, vol. 71, pp 419–432.

Hicks, J. R., 1942, 'Maintaining Capital Intact', Economica, vol. 12, pp 177.

Hicks, J. R., 1973, *Capital and Time — a Neo-Austrian Theory*, Clarendon, Oxford.

Hicks, J. R., 1974, 'Capital Controversies, Ancient and Modern', *American Economic Review*.

Hotelling, H., 1938, 'The General Welfare in Relation to Problems of Railway and Utility Rates', Econometrica, vol. 6, pp 242–269.

Howe, K. and Vogt, S., 1996, 'On 'q'', The Financial Review, vol. 31, pp 265–286.

Hulten, C. and F. C. Wykoff, 1996, 'Issues in the Measurement of Economic Depreciation', *Economic Inquiry*, vol. 34 pp 10–23.

Hurn, A. and R. Wright, 1994, 'Geology or Economics?', *Economic Journal*, vol. 104, pp 363–371.

Industry Commission, 1997, *Telecommunications Economics and Policy Issues: Staff Information Paper*, AGPS, Canberra.

Joye, M. P. and P. L. Blayne, 1990, *Cost and Management Accounting Practices in Australian Manufacturing Companies*, Monograph no. 7, Accounting and Finance Foundation, Sydney University.

Kester, W. C., 1993, 'Turning Growth Options Into Real Assets' in R. Aggarwal (ed.) *Capital Budgeting Under Uncertainty*, Prentice-Hall, New Jersey, pp 187–207.

King, S. P., 3rd Quarter 1996, 'Efficiency and Access: Analysing the Draft Access Code for Australian Electricity Transmission', *The Australian Economic Review*, pp 292–298.

Kwang, M. F. C., J. W. Munro, and K. V. Peasnell, 1995, 'Commonalities Between Added Value Ratios and Traditional Return on Capital Employed', *Accounting and Business Research*, vol. 26, pp 51–67.

Leahy, J. V. and Whited, T. M., 1996, 'The Effect of Uncertainty on Investment', *Journal of Money, Credit and Banking*, vol. 28, pp 64–83.

Lemke, K. W., 1982, 'Financial versus Physical Capital Maintenance: A Review of the Arguments' in R. R. Sterling, and K. W. Lemke (ed.s), *Maintenance of Capital: Financial versus Physical*, Scholars' Book Co., Houston.

Linberg, B. C., 1995, *Digital Broadband Networks and Services*, Mc-Graw Hill, New York.

Loudder, M., I. Khurana, and J. Boatsman, 1996, 'Market Valuation of Regulatory Assets in Public Utility Firms', *The Accounting Review*, vol. 71, pp 357–373.

MacAvoy, P., 1992, *Industry Regulation and the Performance of the American Economy* Norton, New York.

Magendanz, T. and R. Popescu-Zeletin, 1997, *Intelligent Networks: Basic Technology, Standards and Evolution*, Thomson, London.

Ng, Y. K., 3rd Quarter 1987, 'Equity, Efficiency and Financial Viability: Public-Utility Pricing With Special Reference to Water Supply', *The Australian Economic Review*, pp 21–35.

Oakeshott, M., 1951–52; 1996 edition, *The Politics of Faith and the Politics of Scepticism*, University Press, Yale.

Office of Water Services, 1994, *Setting Price Limits for Water and Sewerage Services*, HMSO, London.

O'Hanlan, J. and K. V. Peasnell, February 1996, *'Measure for Measure' Accountancy — International Edition*, pp 44–46.

Ohlson, J. A., 1995, 'Earnings, Book Values and Dividends in Security Valuation', *Contemporary Accounting Research*, vol. 11, pp 661–687.

Park, R. E., 1989, *Incremental Costs and Efficient Prices With Lumpy Capacity: The Single Product Case*, RAND, Santa Monica.

Park, R. E., 1994, *Incremental Costs and Efficient Prices With Lumpy Capacity: The Two Product Case*, RAND, Santa Monica.

Peasnell, K.V., 1982, 'Some Formal Connections Between Economic Values and Yields and Accounting Numbers', *Journal of Business Finance and Accounting*, vol. 9, pp 361–381.

Peirson, G., and A. Ramsey, 1994, *Depreciation of Non-Current Assets*, Australian Accounting Research Foundation, Melbourne.

Perkins, S. J., 1996, *AASB Accounting Standards Handbook*, CCH, North Ryde.

Pindyck, R., 1988, 'Irreversible Investment, Capacity Choice and the Value of the Firm', *American Economic Review*, vol. 78, pp 969–985.

Raz, J., 1977, 'The Rule of Law and Its Virtue', *The Law Quarterly Review*, vol. 93, pp 195–211.

Revsine, L., 1973, *Replacement Cost Accounting*, Prentice-Hall, New Jersey.

Revsine, L., 1979, 'Technological Change and Replacement Costs', *The Accounting Review*, vol. 54, pp 306–322.

Schmalensee, R., 1979, *The Control of Natural Monopolies*, Lexington Books, Lexington.

Scott, M. Fg., 1976, *Some Economic Principles of Accounting*, Institute for Fiscal Studies, London.

Scott, M. Fg., 1989, *A New View of Economic Growth*, Clarendon Press, Oxford.

Shih, M., 1996, 'Optimal Transfer Pricing Method and Fixed Cost Allocation', *Abacus: A Journal of Accounting*, vol. 32, pp 178–195.

Stedman, B. and Y. Kim, 1993, 'Managerial Risk Preferences and Investment Decisions Under Uncertainty' in R. Aggarwal (ed.), *Capital Budgeting Under Uncertainty*, Prentice-Hall, New Jersey, pp 120–139.

Steering Committee on National Performance Monitoring of Government Trading Enterprises, 1994, *Asset Valuation by Government Trading Enterprises*.

Sterling, R., 1970, *Theory of the Measurement of Enterprise Income*, University of Kansas Press, Kansas.

Sterling, R. R., 1982, 'The Concept of Physical Capital Maintenance' in R. R. Sterling, and K. W. Lemke (ed.s), *Maintenance of Capital: Financial versus Physical*, Scholars' Book Co., Houston, pp 3–58.

Stigler, G., 1987, *The Theory of Price*, 4th edition, Macmillan, New York.

Thomas, A. L., 1969, *The Allocation Problem in Financial Accounting Theory*, American Accounting Association, Evanston.

Trigeorgis, L., 1988, 'A Conceptual Framework for Capital Budgeting' *Advances in Futures and Options Research*, vol. 3, pp 145–167.

Triplett, J. E., 1996, 'Depreciation in Production and in Income and Wealth Accounts: Resolution of an Old Debate', *Economic Inquiry*, vol. 34, pp 93–115.

Whittred, G., I. Zimmer, and S. Taylor, 1996, *Financial Reporting: Incentive Effects and Economic Consequences*, Harcourt-Brace, Marrickville.

Wicksteed, P. H., 1910, *The Common Sense of Political Economy*, Macmillan, London.

Williamson, O., 1976, 'Franchise Bidding for Natural Monopolies — in General and With Respect to CATV', *Bell Journal of Economics*, vol. 7, pp 73–10.

Regulatory techniques for addressing interconnection, access, and cross-subsidy in telecommunications

Mark A. Jamison, Public Utility Research Centre, University of Florida, United States.

Section I — introduction

Liberalisation of telecommunications markets worldwide has created the need to interconnect rival networks, prompted regulatory requirements that incumbent service providers provide network access to new entrants,[1] and increased regulatory interest in issues of cross-subsidy. Network interconnection and its pricing are important because of their strong link to the development of competition. Network access allows entrants to expand service more quickly than if they had to duplicate network components that are either costly or effectively impossible to duplicate. Ensuring that cross-subsidy does not occur is important to facilitate efficient competition and protect customers in residual monopoly markets from financing an incumbent's competitive efforts.

Prices and terms[2] for interconnection and access have been the most controversial interconnection issues. Prices are controversial because they affect profitability and the development of competition. In general, higher interconnection and access prices are thought to favour incumbents because higher prices preserve or enhance incumbents' revenue streams and raise entrants' costs. Conversely, lower prices are thought to favour entrants because lower prices mean lower entry costs, allow entrants to use incumbents' networks and pay less than what it would cost the entrants to build their own networks, and decrease incumbents' revenues. Terms for interconnection and access are controversial because they define what it is that service providers receive when they obtain interconnection with another provider's network. For example, terms that require an entrant to obtain interconnection at a high level in the network hierarchy force an entrant with an extensive network to purchase more transport and switching than is necessary.[3]

---

1 This paper uses the term interconnection to include the physical interconnecting of networks and the exchange of traffic. This paper uses the term access to mean the sale of essential inputs. In many jurisdictions, this would be called the provision of unbundled network elements. For purposes of this paper, the company selling the inputs may or may not be vertically integrated into the downstream market.

2 'Terms' refers to all non-price aspects of interconnection, such as allowed Points Of Interconnection (POIs) and the settlements process. 'Settlements process' refers to the payment system by which money from retail customers is distributed among service providers and by which service providers compensate each other for interconnection services.

3 'Network hierarchy' refers to the levels of switching in a network. In most telecommunications networks, a tandem switch is the highest level of switching and connects lower level switches called central offices. Central offices are the switches that connect directly to customers. Calls between local calling areas generally route through a tandem. If a service provider interconnects at a rival's tandem, then the service provider generally purchases switching at the tandem and the central offices, and purchases transport between the tandem and the central offices. If this service provider were instead to interconnect at a central office, then the service

Cross-subsidy enters the debate in two ways. First, there is the issue of how to fund Universal Service Obligations (USOs). Traditionally, incumbents have funded these obligations by charging high prices to some customers (or for some services) and using the profits to fund below-cost prices required by the USO. The second cross-subsidy issue is the amount and type of price flexibility the incumbent should be permitted in markets with actual or emerging competition.

This paper describes tools that regulators have used to address these issues, and the strengths and weaknesses of these tools. In addition to this Introduction this paper has three sections. The next section describes tools for regulating interconnection and access prices. Section three discusses cross-subsidy issues. The last is the conclusion.

Section II — interconnection and access

This section describes tools for regulating prices for interconnection and access. A recent international survey of interconnection policies found that all of the countries surveyed expect service providers to negotiate interconnection and access agreements.[4] The scope of the negotiations varies across countries. In Canada, for example, the Canadian Radio-television and Telecommunications Commission treats interconnection and access as normal tariffed services, and requires negotiations for only a few technical details. At the other extreme, New Zealand and Sweden give regulators almost no authority over interconnection and access arrangements. New Zealand has no industry regulator, relying instead upon competition law and the threat of creating an industry regulator to police interconnection. Sweden has a regulator — the Ministry of Transport and Communications — but gives the regulator no authority over interconnection. Sweden limits the regulator's role to expressing opinions on fairness of proposals if negotiations fail.

It remains to be seen whether negotiations have a long-term role in interconnection and access. Some economists have concluded that, once competitive service providers become established, interconnection negotiations will become a vehicle for collusion.[5] This indicates a long-term role for regulation to limit collusion by regulating interconnection prices.

When regulators become involved in interconnection pricing, whether through settling negotiation disputes or through normal tariffing procedures, they generally consider three basic approaches to price setting: (1) the Efficient Component Pricing Rule (ECPR); (2) cost-based pricing; and (3) demand-based pricing or Global Price Caps (GPCs). This section explains these three approaches. Each approach has its own subsection.

---

provider would only purchase switching at the central office and would purchase no transport unless necessary to get from the service provider's network to the central office.

4 Jamison, Mark A., 'International Survey of Interconnection Policies', 31 March, 1998, (unpublished).

5 See, for example, Laffont, J.J., and Tirole, J., 'Creating Competition Through Interconnection: Theory and Practice,' *Journal of Regulatory Economics*, No. 10, 1996, pp 227–256; and Buehler, Stefan, 'A First Look at Interconnection Regulations in Switzerland,' 1998, (unpublished).

Efficient component pricing rule

The ECPR, which is also called the Baumol-Willig rule, recommends that entrants pay incumbents their opportunity costs. In other words, the prices the incumbent would charge to competitors would ensure that the incumbent would make the same amount of profit regardless of whether it succeed in the competitive portion of the market.

The ECPR formula for setting interconnection and access prices (called wholesale prices in the formula) is:[6]

Wholesale price = Retail price - [Retail TSLRIC – Wholesale TSLRIC]

Or alternatively,

Wholesale price = Retail markup (above Retail TSLRIC) + Wholesale TSLRIC

For example, assume an incumbent would receive $200 000 serving a group of customers and incur costs of $130 000 to do so. Further assume that providing interconnection and access to a competitor that would serve these customers would cost $110 000. The set of ECPR-based interconnection prices would be a set of prices that would generate $180 000 in revenues.

The basic theory behind the ECPR is that, if the incumbent receives the same profits from interconnection and access as it does from sales of the retail product, then competitors can enter the market only if they are more efficient in providing retail functions than is the incumbent. The following example illustrates this idea.

Example 1. There are three products — Lines (L), Retail Calls (RC), and Cable Television (CT). Lines are a necessary input RC. So a firm that sells RC must either build L or purchase L. There are two firms, INCTEL and NEWTEL. INCTEL is the incumbent and owns lines.

The costs of producing these three products are as follows: *Cost*(L) = $100, *Cost*(RC) = $200 (which includes purchasing L for $100), *Cost*(CT) = $100, *Cost*(L + RC) = $180, *Cost*(RC + CT) = $275 (which includes purchasing L for $100), *Cost*(L + CT) = $200, and *Cost*(L + RC + CT) = $300. Based on these results, the most efficient market structure is for one firm to produce L and another to produce RC + CT. This market structure costs only $275. This $275 results from adding the costs of L and RC + CT, and netting the $100 payment for L.

In this example, if INCTEL produced both L and RC, the ECPR would set the final product price at $180 and line price at $100.[7] The $100 represents the difference

---

[6] Explained in more detail below, Total Service Long Run Incremental Costs (TSLRIC) is an incremental cost concept. It represents the additional cost of providing the current or anticipated volume of a service versus not providing the service in the long run. In other words, TSLRIC includes all of the usage costs and fixed costs that are involved in providing the service and that would not be incurred if the service provider did not provide the service. TSLRIC does not consider forgone profits as a cost.

[7] Using the first ECPR formula shown on the previous page, the line price would be found as follows: $180 - ($180 - $100) = $100.

between the total cost of $180 and the $80 incremental cost of RC. Under the restrictive assumptions of the example, this sends correct price signals to the market because NEWTEL has lower incremental costs for producing RC and is able to reflect those lower costs in its price for RC. NEWTEL's incremental cost of producing RC is $75, net of the $100 payment for L. Because NEWTEL can charge a price for CT that is equal to the $100 stand-alone cost of producing CT, NEWTEL is able to charge $175 for the final product RC. This price is lower than would be INCTEL's price, so the ECPR results in an efficient market structure.

The ECPR's efficiency claim is based on an outdated assumption about telecommunications markets; namely, that new competitors are fringe competitors that can offer only some subset of what the incumbent produces.[8] The ECPR is inefficient if either the incumbent or its competitors cannot charge stand-alone costs for their other products. Modifications to Example 1 illustrate why the ECPR is inefficient in either of the circumstances just described.

Example 2. The first modification to Example 1 is to constrain the maximum price for L. Assume that rivalry creates a $90 maximum price for L.[9] The ECPR would have INCTEL price its products in one of two ways: (1) RC's price would be the $90 maximum price for L plus the $80 incremental cost of adding RC, or $170;[10] or (2) L's price would be RC's $180 price less the $80 incremental cost of adding RC, or $100.[11] Either price structure puts INCTEL out of business regardless of whether it is efficient (even though the firm is inefficient in the example). So the ECPR results in an inefficient market outcome.

Example 3. The other modification to Example 1 is to assume that rivalry creates a $90 maximum price for CT. The ECPR would have INCTEL price RC at $180 and L at $100. NEWTEL must be able to price RC in a way that covers the $100 that it must pay for L, and the remainder of the firm's costs that are not covered by the $90

---

8 Willig, Robert, 'The Theory of Network Access Pricing,' in *Issues in Public Utility Regulation*, 109–52, (H. Trebing ed. 1979) at 139.

9 Jamison, Mark A., 'General Conditions for Subsidy-Free Prices,' *Journal of Economics and Business*, 48:371–85, 1996, explains how multilateral rivalry creates this situation.

10 Manipulating the ECPR formula gives this result. From the formula:

Wholesale price = Retail price - [Retail TSLRIC – Wholesale TSLRIC]

We can solve for "Retail price" to obtain:

Retail price = Wholesale price + [Retail TSLRIC – Wholesale T SLRIC]

Then applying the numbers from the example, we obtain:

$170 = $90 + [$180 - $100].

11 Applying the ECPR formula:

Wholesale price = Retail price - [Retail TSLRIC – Wholesale TSLRIC]

gives this result as follows:

$100 = $180 - [$180 - $100].

maximum price for CT. This means that NEWTEL's price for RC is $185. This $185 price is higher than the price the ECPR would have INCTEL charge for RC even though NEWTEL is more efficient. This shows that the ECPR would allow INCTEL to retain the market for RC even thought NEWTEL is more efficient. The efficiency loss is at least the $5 by which INTEL's incremental costs of producing RC exceed NEWTEL's incremental costs of producing RC.[12]

The appropriateness of the ECPR is also based on some other assumptions that do not fit today's markets:

- there are no sunk costs and no monopoly profits;[13]
- there is no discrimination against the entrant in price or quality of interconnection;
- the margin between the incumbent's input price and retail price reflects the incumbent's economic costs of producing the retail product;[14]
- the retail market is homogeneous (i.e. identical products);[15]
- entrants are price takers (i.e. they have no market power);
- regulators are able to perfectly regulate the incumbent.[16]

One advantage of the ECPR is that, because competition would have no impact on the incumbent's profits, the incumbent would be less likely to try to protect markets from competition, except the market for interconnection and access. The interconnection and access markets would become the source of the incumbent's profits. However, with the possible exception of the U.S. policy on setting wholesale prices for resale, no regulator appears to have adopted the ECPR for interconnection or access.[17] And in the U.S., competitors have complained that incumbents are protecting markets, even those where the competition is only from resellers.

---

12 INCTEL's incremental cost of producing RC is calculated as:

$$Cost(L + RC) - Cost(L) = \$180 - \$100 = \$80.$$

NEWTEL's incremental cost of producing RC is calculated as:

$$Cost(CT + RC) - Cost(CT) - Payment(L) = \$275 - \$100 - \$100 = \$75.$$

13 Tye, William, 'The Pricing of Inputs Sold to Competitors: A Response,' *Yale Journal on Regulation*, 11 (1994):203.

14 Kahn, Alfred, and Taylor, William, 'The Pricing of Inputs Sold to Competitors: A Comment,' *Yale Journal on Regulation*, 11, 1994: 225.

15 Willig, Robert, *supra* at 138; and Armstrong, M., and Doyle, C., *Access Pricing, Entry and the Baumol-Willig Rule*, Discussion Paper No. 9422, University of Southampton (no date).

16 Mitchell, Mitchell; Neu, Werner, et. al., *The Regulation of Pricing of Interconnection Services*, 1995 (unpublished).

17 Jamison, 1998.

Cost-based prices

Regulators' options for cost-based prices for interconnection and access are similar to those for other products — regulators must choose between accounting approaches and economic approaches, and between having some contribution to shared costs and no contribution to shared costs. The difference is whether to include a subsidy amount to cover USO costs.[18]

Accounting approaches include Fully Distributed Cost (FDC) and embedded direct analysis (EDA). FDC allocates and assigns costs by account to service categories. Cost assignments are generally restricted to direct costs. Allocation factors that are generally believed to be related to cost causation and reasonable, form the bases for the cost allocations. Generally the factors are usage measures (volumes of demand). EDA is just like FDC, but without the allocation of corporate overheads.

It is generally believed that FDC simply distributes common costs[19] and that services continue to cover their incremental costs. This is not correct. In reality, FDC distributes all accounting costs, including costs that are incremental to only a single service. This misunderstanding results from accountants and non-accountants using the same words to mean different things. In general usage, direct cost means all of the costs caused by the service (or services) in question and not caused by any others. However, in the context of FDC, direct cost refers to the cost of inputs that are only needed to provide a specific service or set of services, and that have their own identity for accounting purposes; i.e. their own account or sub-account. So, for example, if a company were required to install ISDN lines to satisfy an USO requirement, the lines themselves would not be considered a direct cost of ISDN. Instead, they would be considered either a shared cost or a directly attributable cost. If line costs are considered as directly attributable costs, spreading the company's total line costs across all lines, and then multiplying the result by the number of ISDN lines would estimate ISDN line costs. So if the ISDN lines cost more or less than the company's average

---

[18] Extensive discussion of these approaches can be found in Mitchell, Neu, et. al., 1995; Arnbak, J. et. al., 'Network Interconnection in the Domain of ONP: Study for the DG XIII of the European Commission,' 1994 (unpublished); and Jamison, Mark, 'A Competitive Framework for Pricing Interconnection in Global Telecommunications Markets,' Denver Journal of International Law and Policy 23(3):513–33, 1995.

[19] In accounting, common costs are the costs of inputs that are shared by more than one output; e.g. a telecommunications central office switch. In general usage (and in economics), common costs are costs that are not changed if the service or services in question change, including going to zero production. Joint costs are the costs of inputs that, once placed into production, necessarily produce more than one product in fixed proportions. There are very few joint costs in telecommunications. Shared costs is a general term for common and joint costs. There are two types. Shared incremental costs are shared costs that are specific to only some services. For example, some consumer services may have shared costs in consumer billing, but these costs are not shared with business services. Overhead shared costs are costs shared by all services. These are costs that do not change or go away unless the company goes out of business. The classic example is the president's desk, but it's not a perfect example because the desk's cost tends to grow with the company.

line, the costs allocated to ISDN might be less or more than what the company actually spent.

Accounting approaches have all of the benefits and suffer from all of the deficiencies of accounting approaches for rate design in general. The general benefits of FDC are that: (1) FDC-based prices add up to the total revenue requirement under rate of return regulation; (2) FDC can be simple to implement, appear fair, and be easy to understand, although the minutia can create bureaucratic inertia; and (3) if costs can be traced, FDC may encourage companies to be responsible for service-specific investments. The disadvantages of FDC are that: (1) it may be unfair because volumes drive cost allocations;[20] (2) there is a lot of discretion, so widely varying results can be justified; (3) costs are historical rather than forward looking; (4) FDC may assign overhead costs to new services that have not yet established a market; and (5) FDC can result in a cross-subsidy.[21]

In the case of interconnection, FDC has an additional deficiency. If accounting costs are much greater than economic costs, competitors end up providing to the incumbent positive cash flows that the incumbent can then use to finance competitive pricing responses. On the other hand, if accounting costs are much lower than economic costs, the interconnection price makes the competitor's services appear to be much more efficient than they really are.

Economic-cost approaches to pricing interconnection and access use either TSLRIC (or TELRIC) or TSLRIC + contribution (TSLRIC+C). To estimate these, analysts use engineering process models to model the way telecommunications firms incur costs. The models isolate service costs by examining how the network changes when services change.

The economic-cost approach, and specifically TSLRIC+C,[22] is the most popular approach for pricing interconnection and access.[23] There may be several reasons for this, but the most prevalent appears to be that this approach promises to prevent discrimination and cross-subsidies. The conventional wisdom is that prices that are cost-based are non-discriminatory and are subsidy free.

---

[20] Having volumes drive cost allocations creates problems because: (1) customers in non-competitive markets have to carry the full cost of the company if it has problems in competitive markets; and (2) the company's competitive operations have to bear increasing loads of cost if the company is successful.

[21] When two or more services are responsible for the costs in an account, the account is allocated among them based on relative use. The actual costs they cause may be greater than their relative use of the account. For example, if capacity drives costs for central office switching, and a large business customer has primarily peak demand (20% of the total) and very little off-peak demand (only 1% of the total), FDC could allocate only 1% of the central office costs to this customer even though this customer caused 20% of costs.

[22] Prices equal to TSLRIC are unlikely to be sustainable because there is little assurance that they will be subsidy-free. See Jamison, 1996.

[23] Jamison, 1998.

However, the real effects of TSLRIC+C are sensitive to the methods used to determine contribution and to estimate the TSLRIC. If the contribution is comparable to what the incumbent can expect from other products on average, then the contribution should be both sustainable and consistent with competitive market outcomes. Basing the contribution on Ramsey-pricing principles (a demand-based approach discussed below) may promote allocative economic efficiency if competition does not affect demand elasticities, but may not be sustainable or free of cross-subsidies.

Getting the TSLRIC estimate correct has proven to be very difficult. Typically, the engineering process models estimate the cost of a newly constructed firm established to serve current demand with growth factors to estimate spare capacity for future demand. Implicit in these models is the assumption that the plant constructed is either used for its entire depreciated life, or is part of the growth-based spare capacity for some portion of its depreciated life. These assumptions ignore a common business event — demand either evaporating or diminishing after the plant is placed to serve the demand. Unless this plant is fungible in that it can be immediately be either used to serve someone else or become part of growth-based spare capacity, then the engineering process models underestimate the company's actual economic costs. This underestimate makes shareholders bear all of the risk of projects that do not fit the engineering process model's assumptions. This is a greater risk than occurs under rate of return regulation and generally under price cap regulation.

Demand-based prices

The demand-based approach to interconnection and access pricing uses Ramsey-Boiteux pricing principles to promote consumer and producer welfare. This is also called the **Optimal Access Pricing Rule** or **GPCs**.[24]

With Ramsey-Boiteux pricing, customers are charged different prices based on their responsiveness to price changes. 'Responsiveness' is measured in terms of how much customers change the amount they purchase. Customers who do not respond very much to price changes are said to have inelastic demand. Customers who respond a lot are said to have elastic demand. 'Break even' means that the company's revenues equals its economic costs. This is also called the inverse elasticity rule because prices are increased in inverse proportion to the customer's elasticity of demand. The objective of Ramsey-Boiteux pricing is to deviate as little as possible from the consumption mix that would occur if prices were equal to marginal cost.

There are two demand-based approaches: (1) regulator-set prices (mentioned in the context of cost-based pricing); and (2) GPCs. The regulator-set prices requires knowledge of service provider costs and of demand elasticities for the service provider's and the competitor's markets. GPCs treats interconnection as a product and places it in a global price cap basket with exogenous weights. If exogenous weights cannot be determined, then forecasted demand and estimates of market share may be substituted.

[24] See Laffont and Tirole, 1996.

The benefits of demand-based approaches are:

- They promote allocative efficiency if competition does not affect demand elasticities and if the incumbent does not engage in strategic pricing.
- GPCs eliminate or substantially reduce incentives for exclusion and cross-subsidies.
- With GPCs, increasing the weights can effect lower interconnection prices.
- Prices can reflect (but not equal) marginal costs while also allowing the incumbent to cover its total cost.

The problems with using these approaches are:

- GPCs require price cap regulation of all prices, including prices in competitive markets.
- Entrants must behave competitively for GPCs to work. [25]
- It is unclear how dynamic weight updating should be done with GPCs.
- Customers and politicians often oppose these approaches because they give the highest mark-ups over marginal cost to the customers who have the least ability to protect themselves.
- GPCs allow incumbents to benefit from predatory pricing and cross-subsidy if entrants can be kept from markets by short-lived price reductions in competitive markets.

Section III — cross-subsidy issues

This section describes tools regulators use for preventing subsidies to competitive markets and for funding USOs. These tools involve applications of the ECPR, cost-based pricing, and price caps.

Defining cross subsidy

The first issue to confront on cross subsidy issues is to decide what is meant by cross subsidy. There is often general agreement that cross subsidies are problematic, but there is generally wide disagreement on what constitutes a cross subsidy. There are four basic views, although there are many flavours of each.

- **The public policy view**. From a public policy perspective, cross-subsidisation occurs in a regulated industry when the regulated firm uses revenues from one market to keep operations in another market financially viable. The cross subsidy is considered anti-competitive if the cash flows from non-competitive to

[25] Laffont and Tirole, 1996, point out that entrant market power in final product markets distorts optimal pricing. However, this problem may not be unique to global price caps.

competitive markets. The cross subsidy is considered an USO if the cash flow: (1) goes the other way; (2) occurs only because regulatory rules create it; and (3) would not occur absent the government policy and/or if the funding markets were competitive. In a nutshell, the public policy view is that the cash flow is a cross subsidy if fully competitive markets would not allow it.

- **The cost allocation view**. In more general usage, if a service's prices do not make a reasonable contribution to overhead costs, it could be argued that the service is not carrying a fair share of the overheads and is therefore being subsidised.

- **The Baumol-Faulhaber view**. Baumol and Faulhaber[26] have taken the view that cross-subsidisation occurs when prices for a service do not cover the service's incremental cost and the company still earns a normal profit (i.e. zero economic profit) overall. This implies a maximum price of stand-alone cost. This is a popular view among economists.

- **A more comprehensive economic view**. More recent economic studies have shown that cross-subsidisation occurs when prices for a service are higher than would be charged by the next most efficient competitor and the company still earns a normal profit.[27] A variation on Example 1 illustrates this view.

Example 4. There are three products — lines (L), switching (S), and cable television (CT). There are also two firms, INCTEL and NEWTEL. INCTEL provides L and S, while NEWTEL provides CT. Assume that either can adopt any technology, so that neither has an inherent cost advantage. The costs of producing these three products are as follows: *Cost*(L) = \$100, *Cost*(S) = \$110, *Cost*(CT) = \$150, *Cost*(L + S) = \$180, *Cost*(S + CT) = \$250, *Cost*(L + CT) = \$235, and *Cost*(L + S + CT) = \$340. Based on these results, the most efficient market structure is the current market structure — one firm producing L + S and another producing CT. This market structure costs only \$330.

Table 1 illustrates the subsidy-free prices for Example 4. The first row shows minimum and maximum subsidy-free prices under the Baumol-Faulhaber view.[28] These match the incremental costs of L and S (\$70 and \$80 respectively) and the stand-alone costs of L and S (\$100 and \$110 respectively). This view ignores the possibility of NEWTEL offering either of the services. The second row shows minimum and maximum subsidy-free prices under the more comprehensive view.[29] This framework

---

[26] Baumol, William J., 'Minimum and Maximum Pricing Principles for Residual Regulation', *Eastern Economic Journal*, 5 (1–2), pp 235–248, January 1979; and Faulhaber, Gerald R., 'Cross-subsidisation in Public Enterprises', *American Economic Review*, 65(5), pp 966–977, 1979.

[27] Jamison, 1996.

[28] The minimum prices are calculated as follows. For L, the floor is *Cost*(L + S) - *Cost*(S) = \$180 - \$110 = \$70. For S, the floor is *Cost*(L + S) - *Cost*(L) = \$180 - \$100 = \$80.

[29] These price ranges are calculated as follows. For L, the floor is *Cost*(L + S) - *Cost*(S + CT) - *Cost*(CT) = \$180 - \$250 - \$150 = \$80, and the ceiling is *Cost*(L + CT) - *Cost*(CT) = \$235 - \$150 = \$85. For S, the floor is *Cost*(L + S) - *Cost*(L + CT) - *Cost*(CT) = \$180 - \$235 - \$150 = \$95, and the ceiling is *Cost*(S + CT) - *Cost*(CT) = \$250 - \$150 = \$100. For a comprehensive explanation of these calculations, see Jamison, 1996.

incorporates NEWTEL's potential production. This forces INCTEL's minimum prices to be greater than incremental cost and INCTEL's maximum prices to be below stand-alone cost.

Table 1. Subsidy-free prices under the two economic views

	Subsidy-free prices	
	Minimum	**Maximum**
Lines		
Baumol-Faulhaber view	$70	$100
Comprehensive view	$80	$85
Switching		
Baumol-Faulhaber view	$80	$110
Comprehensive view	$95	$100

Detecting and remedying cross subsidy

Regulators have used various devices for detecting and dealing with cross-subsidy issues. Generally, detecting a cross-subsidy is little more than applying the definition chosen. However, it is not always simple. For example, if the regulator chooses the cost allocation view, then the regulator must decide which cost allocation is the appropriate standard. Generally this involves decisions on mechanics of FDC. As was explained above, there are large numbers of reasonable options, so the detection can become quite involved. Also, the more comprehensive economic view requires large amounts of information for detecting cross subsidy. Fortunately, there are some fairly simple tools, which will be discussed later, which take care of this view's cross subsidies without getting into detection.

Techniques for remedying cross-subsidy concerns vary depending on whether the concern is which anti-competitive cross subsidies or USOs, but fall into the general groups of FDC, incremental cost, and imputation (which is actually the ECPR in reverse).

The fully distributed cost technique

In instances where regulators use earnings to regulate prices, such as in the case of rate of return regulation or assessing earnings during price reviews (as is done in the UK), there is really little way to avoid using FDC. The basic technique is to either base regulated prices on their FDCs or to require competitive prices to cover their FDCs. FDC has some clear drawbacks:

- FDC's underlying theory is that regulators can use accounting records to determine the costs caused by particular services. Unfortunately, FDC allocates costs by account. For the reasons explained above, the costs allocated to a service may be less than, or even a lot more than, the costs the service actually caused.

- FDC formulas shift costs to non-competitive markets. This happens for two reasons. First, the accounting records on which FDC is based do not show why costs were incurred. So it is at best difficult for regulators to prevent companies from acting on the incentive to shift costs incurred for competitive services into prices for non-competitive services. Second, usage-based allocators shift costs to non-competitive markets when companies lose market share in competitive markets. This shifts the risk of cost recovery from shareholders to captive customers.

- FDC restricts regulated companies' abilities to innovate and respond to competition in four ways.

    a. First, regulatory processes to approve investments and new services cause delays. In the US, local exchange carrier (LEC) video dialtone services are a recent example. Prior to passage of the *Telecommunications Act of 1996*, LECs had to get approval of the FCC before constructing facilities for video dialtone. This prior approval was required to prevent cost shifting through the accounting process.

    b. The second reason FDC limits innovation is that it creates rigid structures and procedures. For example, US Federal Communications Commission (FCC) rules contain artificial distinctions between switched and non-switched services. The FCC also requires uniformity across LECs in how they provide and measure costs for some non-regulated services.

    c. The third way that FDC limits innovation is that it constrains management thinking about services and markets. Service development, introduction, and marketing follow the accounting framework because companies must conform their businesses to the regulatory structure.

    d. FDC limits how regulated companies respond to competition by sending false cost signals to management. When a regulated company gains or loses customers in a competitive market, changes in FDCs affect this company's bottom line, not the costs caused by the gain or loss of customers. This sends false signals to management because changes in FDC may be greater or less than the costs actually caused by the change. If the FDC change is too large, management will be discouraged from pursuing customers. If the FDC change is too small, management will be encouraged to over invest in the market. Both actions cause a loss of economic efficiency and could harm the long-term financial interests of the company.

Despite these problems, FDC does have its benefits. Section II listed these benefits, so they are not repeated here.

Regulators could also apply FDC to measure costs for USOs. With this approach, the difference between the price the service provider is allowed to charge in a market and the FDCs of the market is treated as the cost of the USO. This, in effect, is setting prices (actually price plus subsidy) on FDC. As a result, this method suffers from all of

the problems and enjoys all of the benefits of using FDC for interconnection and for protecting against anti-competitive cross subsidies.[30]

The stand-alone cost technique

Several regulatory techniques fall under the rubric of incremental cost. These include setting maximum prices at stand-alone cost and setting minimum prices at incremental cost, generally TSLRIC. The Baumol-Faulhaber view of cross subsidy forms the basis for these techniques. So to the extent that this view is out of date, these tools are also out of date. However, because they are still in use, they are discussed below.

The stand-alone cost approach allows service providers to increase prices in non-competitive markets up to stand-alone cost — the total cost of a specialised company producing only the service or services in question. For example, the stand-alone cost of providing water to residential customers would be the total cost of a company that provided only sufficient pumping, processing, distribution, etc. to serve residential customers and produced nothing else.

This approach has been used at least twice. The US Interstate Commerce Commission used stand-alone cost tests to determine if captive rail shippers were paying too high prices. More recently, the Independent Pricing and Regulatory Tribunal of New South Wales used stand-alone costs as the standard for maximum prices for contract services in gas. The application distributes overhead costs among customers, so the effective price ceiling is below stand-alone cost.

Aside from the problems with the underlying theory, stand-alone cost suffers from a practical problem. It effectively assigns all common costs to residual customers, the very customers regulation generally intends to protect. So applying this tool would appear to conflict with basic regulatory mandates.

TSLRIC price floors

Regulators frequently TSLRIC-based minimum prices in competitive markets. There are two benefits from using this approach. First, the practice is well established, so there are many examples to follow. Second, if properly implemented, it ensures that captive customers do not cover costs incurred only to produce the competitive services.

Unfortunately, there are several problems. The first is that TSLRIC fails to protect against cross subsidy. This happens for two reasons: (1) TSLRIC ignores strategic pricing. Through strategic pricing, a dominant firm can drive more efficient, non-dominant competitors from markets without lowering prices below incremental cost in competitive markets. (2) TSLRIC is based on the Baumol-Faulhaber definition of cross subsidy.

[30] A more complete discussion of measuring costs for USOs can be found in Jamison, Mark A., 'Estimating Costs for Universal Service Obligations,' *Telecommunications Journal of Australia* 47(1), pp 51–58, 1997; and Jamison, Mark A., 'Principles for Estimating Universal Service Costs,' Paper Presented at the Telecommunications Policy Research Conference, September 1997 which can be found at http://www.cba.ufl.edu/eco/purc/primary/mjamison.htm.

Another problem with TSLRIC price floors is that they fail to protect captive customers because they shift costs to non-competitive markets. This happens in two ways. First, as with stand-alone cost, TSLRIC price floors place all of the responsibility of covering common costs on captive customers. In effect, customers in non-competitive markets ensure the company is financially viable because the only costs that are put at risk of non-recovery are those that can be avoided by exiting the competitive market. Second, the incremental cost studies that form the bases for the price floors sometimes omit costs caused by a competitive service. For example, studies done after a service is developed would generally omit development costs. These development costs remain in the regulated company's overall cost and are potentially covered by other services.

TSLRIC floors also restrict regulated companies' abilities to innovate and respond to competition. This happens because the price floors can remove companies' abilities to price below incremental costs. This is overkill because there are legitimate reasons, as well as anti-competitive reasons, to price below incremental costs. Examples of legitimate reasons include filling product and market gaps, maintaining stakes in strategic markets, expanding markets, and engaging in price wars.

The last problem with TSLRIC price floors (or any price floors) is that they increase the regulator's role in competitive markets. Intuitively, we would expect that regulation should decrease as competition increases. However, because these price floors apply to competitive markets and the number of stakeholders is higher in competitive markets than in non-competitive markets,[31] the amount of regulatory oversight actually increases as competition increases.

Regulators frequently use incremental cost techniques for dealing with USOs. Examples include the UK and Australia. With this approach, the difference between the price the service provider is allowed to charge in a market and the TSLRICs of the market is treated as the cost of the USO. This, in effect, is setting prices (actually price plus subsidy) equal to TSLRIC. This has many of the problems and benefits of using TSLRIC for interconnection and protecting against anti-competitive cross subsidies, so these are not repeated here. There is one exception. The Baumol-Faulhaber view may actually be appropriate in this context. As long as providers are not competing for the USO, and as long as the USO does not hinder USO providers in their competing in other markets, then the Baumol-Faulhaber assumptions would appear to hold. This appears to be the case in the UK and in Australia.[32]

Imputation techniques

The last approach for remedying cross-subsidy is imputation. Generally, imputation is a method of setting price floors for a company's competitive services when that company also provides essential, non-competitive inputs for those competitive services. The objective is to prevent a price squeeze by forcing a service provider to charge itself the same price for non-competitive, essential inputs that it charges its competitors.

---

31 Competitive markets have all the stakeholders of non-competitive markets, plus the new competitors.

32 Jamison, 'Estimating Costs', 1997, and Jamison, 'Principles', 1997, contain more complete discussions of this point.

Imputation is basically a reverse application of the ECPR. Advocates generally suggest one of two basic methods. One method includes two elements: (1) incumbent's own input prices; and (2) the incumbent's TSLRIC of being a competitive service provider in addition to an input provider. This system can be illustrated with the following formula:

$$\text{I-Floor} = \text{Price}_{\text{input}} + \text{TSLRIC}_{\text{competitive}}$$

Where:

I-Floor = the price floor for the incumbent's competitive service;

$\text{Price}_{\text{input}}$ = the incumbent's regulated prices for essential, non-competitive inputs; and

$\text{TSLRIC}_{\text{competitive}}$ = the incumbent's cost of providing the competitive services, over and above the cost of providing the essential, non-competitive inputs.

The other method differs from the first method in that this second method adds an adjustment for economies of sequence[33] that the incumbent might have from being both an input provider and a provider of the downstream product. Two formulas are generally proposed for this method. One formula uses an implicit adjustment where the incumbent would include in the imputation study the difference between the TSLRIC of input services and the TSLRIC of the incumbent's own competitive services. This formula is:

$$\text{I-Floor} = \text{Price}_{\text{input}} - \text{TSLRIC}_{\text{input}} + \text{TSLRIC}_{\text{competitive}}$$

where:

$\text{TSLRIC}_{\text{input}}$ = the incumbent's TSLRIC for the essential, non-competitive inputs; and

$\text{TSLRIC}_{\text{competitive}}$ = the incumbent's TSLRIC for downstream services.

Incumbents generally prefer this method, which is called the lost contribution method. The other formula includes an explicit adjustment for economies of sequence. This formula is:

$$\text{I-Floor} = \text{Price}_{\text{input}} + \text{TSLRIC}_{\text{competitive}} - \text{TSLRIC}_{\text{vertical}}$$

where:

$\text{TSLRIC}_{\text{vertical}}$ = any efficiency gains the incumbent receives from being both an input provider and a provider of the final product.

---

[33] Economies of sequence are where it is cheaper for a company to both produce an input and use it to produce the final product than to produce only the input, sell it to someone else, and have the input buyer produce the final product.

This formula is mathematically the same as the previous formula. Competitors prefer it because it is easier to review assumptions about vertical integration.

Price cap techniques

An indirect method of remedying cross-subsidy concerns is through price caps for non-competitive services. Sometimes called cost-based price caps, this method applies the comprehensive view of cross-subsidy by ensuring that some, but not all, common costs are covered by prices in non-competitive markets. As described by Trebing[34] and Jamison,[35] this system would: (1) deregulate all prices except prices for basic utility service, interconnection, and access; and (2) use cost-based price caps as the regulatory limit on these prices. No price floors or cost allocations would be used for any services. Establishing cost-based price caps involves two steps: (1) choosing a target level for the prices; and (2) estimating this target price level. The target price level should cover the services' TSLRIC and provide a limited (generally average) contribution to shared costs. Proxy costs should be used to estimate the price level. This method, in effect, kills two birds with one stone. First, it protects customers in non-competitive markets by ensuring that they cover no more common costs on average than the competitive services. Second, it should help prevent cross-subsidisation by limiting non-competitive profits that the incumbent could use to finance competitive operations.

Conclusion

This paper describes techniques regulators use to address interconnection, access, and cross-subsidisation, and discusses their advantages and disadvantages. None of the tools discussed has a clear advantage over all others in all situations. As a result, regulators will generally find it necessary to assess the applicability of these tools in the context of regulatory objectives, institutional abilities, and markets. However, that some techniques have things in common with other techniques, and that some techniques conflict with others, helps narrow the choices.

For example, and as has already been discussed, the use of earnings in price reviews and in rate of return regulation implies that some form of FDC is already in use for cross-subsidy issues. Substituting TSLRIC or imputation-based price floors for the FDC cross-subsidy may bring about the worst of both worlds by increasing regulators' oversight of competitive markets and giving incumbents the opportunity to use the earnings monitoring to recover competitive costs that are not captured in the TSLRIC estimates.

---

34 Trebing, Harry, 'Public Control of Enterprise: Neoclassical Assault and Neoinstitutional Reform', *Journal of Economic Issues*, 18, 1984: 353.

35 Jamison, Mark, Pricing and Deregulation during Telecommunications' Transition to Competition, Presented at the Consortium for Research on Telecommunications Policy conference, May 1996; at the 1996 Annual Meeting of the American Law and Economics Association, May 1996; and at the 1997 Annual Meeting of the American Economic Association, January 1997.

Likewise, adopting ECPR for interconnection implies adoption of imputation for price floors, and vice versa. Mixing with other mechanisms would generally be ineffective. For example, basing interconnection and access on TSLRIC makes imputation unnecessary because the company cannot lower retail prices even near the imputation level without earning losses because common costs would be unrecovered.

In countries where competition is expected to be strong in at least some markets, some form of price caps would likely be preferred, especially something along the lines of the cost-based price caps. An earnings monitoring (which is implicit in FDC) provides incentives and opportunities that directly conflict with what regulators expect from competitive markets. Incumbents have the incentive and opportunity to shift costs from competitive to non-competitive markets. They also have little incentive to become more efficient in the competitive markets because high profits are taken away and low profits are compensated. In contrast, the price caps look just like competitive market pressure to the incumbent. The incumbent cannot increase the price caps to protect overall profits while competing aggressively in competitive markets, cannot use the non-competitive markets to cover common costs that competitors must cover in their competitive operations, and can keep every reward from competitive success that any other firm can keep.

Regulating access to railway infrastructure

Stewart Joy, Stewart Joy Associates Pty Ltd., Melbourne, Australia.

Introduction

The ideal railway infrastructure ownership model

Infrastructure[1] is just one input to the railway process, but its high threshold cost and lumpy increments mean that with public railways optimal allocation of capacity among users is a complex process. In addition, given the history of past generous state support there is often a need to match infrastructure capacity with total demands, and to work toward efficient costs[2], and this confuses pricing.

Where optimal decision making about capacity is the primary goal in managing railway capacity, there is an **a priori** preference for vertical integration. Where the main user also manages the infrastructure, it is best placed to resolve trade-offs between infrastructure inefficiencies and marginal train operating costs. But recent experience with government-owned railways has been that the predominant user's commercial interests have often been allowed to interfere with efficient allocation of the infrastructure resources, and infrastructure ownership has been used to preserve monopoly interests in train operations[3].

A further regulatory complication arises in Australia because, although railway *infrastructure* is a clear concept, some rail users are attempting to use the Trade Practices Act and the National Competition Council's (NCC) processes in attempts to broaden this to include readily reproducible aspects of above-rail operations. This paper distinguishes access to infrastructure from the renting of other railway assets, such as rolling stock, which is readily reproducible and transferable, or terminals, the commercial provision of which is limited only be land values, which are in turn able to be traded off against access costs. The NCC, in its recent decision in the Carpentaria Transport case, made a correct decision that there can be no natural monopoly in these assets. An attempt by a potential user to avoid paying **current** costs, i.e. replacement costs, does not raise an issue of national importance which would merit recourse to the Australian Competition and Consumer Commission (ACCC). (NCC and ACCC interest would be aroused, however, if a railway owning such assets gave privileged access to them to one of a group of competitors).

---

1 Includes all below rail assets, and the operational control of their use.

2 In Australia, every mainland state public railway **network** has excess capacity. Only the intercapital routes operate at near to their effective capacities, because they are required to be equipped to handle peak within-the-day demands. On most rural lines which do not carry mineral exports one train per day, or less, is the normal output.

3 This issue should not be confused with a track authority charging monopoly prices to cover inefficiencies or stranded costs. See the section titled **stand alone cost**.

Rail infrastructure owners in Australia are finding themselves under three main pressures, detailed below.

Access user expectations

Actual and potential rail users have an over-ambitious view of the smallness of the prices at which they should be able to rent route capacity for their trains. But the new rules provide for fair prices for access, and these rules do not automatically confer rights to transfer profits (or contributions over marginal costs) from infrastructure owner to user by forcing owners to share their economies of scope with users.

Isolation of redundant costs incurred by traditional railway organisations

Specialist users, without the burdens of past public ownership and related costs, have highlighted railway landlords' 'other' costs[4], and the desire of these railways to recover some of those costs from infrastructure users, even though their access needs may not cause the costs. Highlighting inefficiencies, broadly defined, in the landlord railway was the reason for the recent development of track authorities in Britain[5], Sweden[6], and Germany[7] and will shortly be the cause of an upheaval in France. A lack of confidence in conventional railway managements' readiness to deal fairly with competing railway operators, has meant that most governments which have adopted an infrastructure access regime have chosen to use a separate infrastructure owner.

Whether placing infrastructure in the hands of third parties, public or private-owned, will result in more efficient rail operations is not yet clear. The trade-off between infrastructure economies from vertical integration and allocational economies from independent control of access has not yet been measured with any precision, but governments (and the European Community administration) have generally opted for the independent structure, avoiding the task of detailed surveillance of the railway operational organisations.

---

4 Note the readiness of private railway operators to pay very large premia for the right to own the former loss-making Tasmanian and South Australian operations of Australian National. This represents the cost savings the new owners expect from higher labour productivity; there is not significant new traffic potential on the purchased routes.

5 For an account of the development of rail infrastructure access in Great Britain, see Stewart Joy, 'Public and Private Railways': *Journal of Transport Economics and Policy*, vol. 32, Part 1.

6 The situation in the first track authority, in Sweden, often quoted in support of separation, was more complex. The 'Banverket' was formed primarily to allow local passenger railway companies to operate lines which SJ, the government railway, could not operate economically, and to treat rail infrastructure proposals similarly to those for roads. The Swedish Government perceived a problem of SJ resistance to allowing other, lower-cost, operators to try out on routes where SJ had failed to cover costs or attract subsidies from local bodies.

7 Germany has created DB Netz AG as a subsidiary of the railway holding company to own and operate infrastructure, selling access to other DB subsidiaries and outside users. Whilst the Federal Transport Department wants DB Netz to be taken out of the railway holding company, the DB management has so far convinced the Minister for Transport that non-discriminatory access already exists.

Exploiting NCC procedures to force sale of reproduceable railway capacity

Specialist users are attempting, via applications for NCC declarations, to force railways to sell readily reproducible services, with no natural monopoly characteristics, at **less than stand alone cost**[8]. This is an attempt to use the legislative provisions relating to below-rail costs, on which there may be natural monopoly, to force below-cost sale of rail services when there is nothing preventing the applicant from supplying its own. The fact that a larger railway, whose scope of operation might result in lower costs than a marginal user can achieve, is not a ground for forcing sale of above rail service at the provider's **costs**, rather than the stand alone cost which a new user should incur. Such a remedy is not available to participants in other industries.

Goals of infrastructure pricing

These are:

- to recover the **efficient** cost of its provision;
- in a manner which interprets to users the **marginal costs** of their usage, and thus leads to optimal capacity changes by the owner;
- to leave the cost of excess capacity with the agency **responsible for its retention or eradication;** and to
- maximise the **cost-covering** use of the infrastructure[9].

It is complicated in practice by:

- high claimed threshold costs of retaining infrastructure ready for use, which do not match notions of 'efficient' unit costs on higher volume routes;[10]
- infrastructure owners' attempts to exercise monopoly powers over users, which risk the businesses of both, because of one or more of:
    a. unsustainable revenue goals unrelated to efficient costs — attempts to recover the costs of excess capacity or inefficient maintenance methods;[11]

---

8 See, for example, the NCC's Carpentaria Transport and SCT/Westrail cases, in which the NCC rightly rejected the claim that the applicants should be able to use s. 44f of the Trade Practices Act to force a railway to sell above rail services at less than stand alone cost, i.e. transferring profit from landlord to tenant.

9 This may be a short-run or a long-run concept, depending upon the level of cost which the surplus over other costs can cover. See Stewart Joy: 'Pricing and Investment in Railway Freight Services', *Journal of Transport Economics and Policy*, vol. V, no. 3, September 1971.

10 These fixed costs for routes with very low usage are presently determined under conditions of government ownership, which may not be the most efficient form of management. In NSW, these are claimed to be more than four times as high as are experienced by US short line operators, who operate similar rolling stock over similar terrain.

11 This appears to have been the case with SCT's early attempts to operate in NSW. See 'Private operators fight for slots on systems', (Australian Financial Review, 7th October 1997), where

b. attempts to protect owners' vertically integrated businesses;[12]

c. flawed efforts to extract all of the margin between users' revenues and their other costs, in order to recover revenues in excess of long-run marginal costs; [13]

d. unwarranted price discrimination through resale prohibition on low-price users[14] (as in 'C' above) without regard to cost differences.

- attempts to force each user to accept all risk of volume variations by other users [15], so that the infrastructure owner has no risk at all, and may recover more than one risk premium in respect of the same capacity;
- failure to price excess capacity optimally in the short run (i.e. at the maximum price which gains the traffic and is greater than Short Run Marginal Cost (SRMC) and less than Long Run Marginal Cost (LRMC), and to eliminate the capacity in the long run. [16]

This last issue is the crux of railway infrastructure economics. Where excess capacity exists, it is best to use it until the underlying assets expire, provided that all other costs are recovered.

Basic cost characteristics for rail infrastructure pricing

Railway infrastructure is characterised by:

- natural monopoly — less cost to use an existing route more intensively (or expand its capacity) than to construct a parallel route;
- long lived assets (rail on Melbourne — Albury standard gauge, one of the heaviest trafficked interstate routes, has lasted over 30 years);

---

the company claims that RAC originally demanded an access fee between four and five times what it was paying on other systems.

12 In both Queensland and Western Australia and with the private railways in Tasmania and South Australia, there may be a need for regulators to investigate the railways' total costs in order to decide the fairness of any proposed access fees. It would violate the intent of the Competition Principles Agreement if a landlord railway was to charge a user more for access than it charged (or recovered) internally for the same traffic, and internal recoveries can only be determined by reference to the other costs of the transit.

13 This may have happened in NSW. See earlier note.

14 This is ultimately self-defeating, because the low-price user can sell an integrated service direct to the client of the landlord being discriminated against. This was a nub of RAC/SCT disputation (late 1997), because on routes in NSW it appears that NRC has a much more favourable access price from RAC. This problem will be eliminated with the creation of Track Australia on interstate routes, with its open pricing.

15 This requires an **ex post** pricing system, where user A's charge is varied retrospectively if other users' revenues do not fulfil the owner's expectations. NSW Freight Corp and State Rail Authority claim that this was originally applied to them by the NSW RAC.

16 The normal test for excess capacity is whether SRMC < LRMC. (See George H Borts, Increasing Returns in the Railroad Industry, *Journal of Political Economy*, August 1954).

- negligible marginal costs of use, relative to annual costs of route, for a small number of trains operating in otherwise unused paths;

- (in many cases) sunk costs for subgrade[17] written off a hundred years ago, but (where margins over all other costs are sufficient) the sub-grade, non-replaceable component of stand alone cost (see the section titled **marginal cost has time and output dimensions**) may be relevant for major traffics[18].

We now consider particular types of cost.

Efficient cost

Most Australian public railway owners concede that their infrastructure costs do not yet match even best Australian practice; this is confirmed by their actions in contracting out track maintenance. Users may therefore rightly demand to be sold access on the basis of **efficient** costs, not those pertaining under old methods, because unless sale of capacity is forced on this basis, there is no pressure on owners to improve their operations, and owners are able to use their monopoly to force the support of inefficient practices.

Marginal cost has time and output dimensions

- **time**, for the cost of —

  a. keeping the line open (mainly fencing and inspection);

  b. rectifying natural damage (flooding and similar);

  c. (over 20 years or more) timber sleeper replacement;

  d. by the shift — train control;[19]

- **output** — rail wear, rectifying fastenings (tightening and replacing) — primarily a function of gross tonne kilometres output, maximum axle loads and wagon suspensions[20], (but in localities with large temperature variations, this may also be partially a time cost, regardless of traffic variations).

---

17 Subgrade is the infrastructure below the ballast — the 'earthworks', including tunnels and bridges.

18 For example, the stand alone costs of many of the export coal flows should include construction costs, including return on investment. But note that this may imply preferential treatment for some of the lower-volume (or more distant) mines whose cash flows, net of all other costs, could not recover their share of stand alone cost for the rail route.

19 But this cost has an annual increment — you cannot just buy one shift of train control. Usually the incremental change possibility for train control is limited to decisions whether to operate on weekends or not.

20 Conventional three-piece bogies under freight vehicles cause considerably more damage to track than more sophisticated designs.

Stand alone cost

The upper bound of costs for infrastructure pricing is **stand alone cost**. This is an artificial construct which can be applied in two ways:

- **ex ante** stand alone cost — (very rare) hypothesises the construction of a new railway to find whether this is less costly to a major[21] shipper than paying a rental to use an existing line. Very few present-day railway owners would try to set an asking price this high (although many did just this in fuelling the drive to construct parallel routes in Nineteenth Century USA). Note that we are comparing **cost** of a new route with **price** to use an existing route.

- **ex post** stand alone cost — (general use) hypothesises the adaptation of an existing route to carry just the subject traffic, and assesses the costs of that.

Practical applications of stand alone cost are:

- **Ex parte 347** in the United States — used to assess the infrastructure component of a rate (haulage component is usually easily assessed)[22]. But note the complication of apparent rail inefficiencies where a haulage program is used to produce a required mix of coal for export, as in NSW Hunter Valley, where optimal train (and route) efficiency is in conflict with port terminal mixing requirements[23].

- An unambiguous Australian definition proposed on behalf of one infrastructure supplier is — 'the hypothetical cost which would arise if an individual operator were the only user of a particular part of infrastructure and the existing infrastructure configuration was **rationalised** to satisfy only that individual operator'.

This restrictive definition is useful, and relies upon the user's needs rather than the scale of provision by the landlord. It rejects any payment for unused capacity – either actual capacity on a route, or in administrative support which would not be required to handle the user's traffic if it managed the route.

Note the distinction which must be made between users of routes whose traffics would not bear full replacement cost and those traffics which, in the absence of existing routes, would justify new construction. In Australia, the former group might cover East Coast inter-capital routes. For the other group, in the Hunter Valley if there was no railway between some of the major mines and the port terminals, it would pay those major coal shippers to build one in order to export their production.

---

21 That is, a shipper whose traffic could justify a separate route.

22 This is the short title of a US Interstate Commerce Commission Decision. In some cases, the application of **ex parte 347** has been adjusted according to the Baumol-Willig rule, by adding to the access fee the profit that the landlord would have made if it were handling the traffic. This profit may include the economies of scope available to the landlord's larger operation.

23 In the Hunter Valley, rail operational efficiency is traded off against port terminal capacity and sorting cost, by having the operational pattern of the trains substitute for physical sorting and mixing of coal within the port terminal.

Accounting cost

The only accounting concepts useful in infrastructure pricing are:

- accrual basis for expected renewals;
- to separate recoverable costs from the inefficient component of costs in accounts, to establish responsibilities and avoid taxing users for stranded costs.

Regulators should reject the notion of 'recovery of accumulated losses'[24] in favour of **future stand alone cost** (assessed on the basis of the cost of modern construction). The value of a railway route can only be its **present worth**, not past expenditure. There is no need to reward a railway for having had infrastructure available; the need is only to pay the cost of providing it **as needed**. This may leave stranded some 'accounting' costs, but these are not relevant to future pricing, and are the responsibility of whoever caused them, not future users.

Regulators need to distinguish stranded costs from sunk costs[25]. Optimal resource utilisation should not be inhibited by attempts to recover mistaken investments of the past; only future resource usage is relevant. Where the volume of traffic indicates that a new route might be appropriate, then the modern-day equivalent of the sunk cost might be relevant to charging for access. But stranded costs are the responsibility of whoever allowed them to be created, and should not be allowed to influence future use of the infrastructure.

Basic price structures

Basic pricing needs are:

- **predictability** for users, with the owner taking the risks of under-utilisation and elimination of unused capacity and cost (because only the owner can eliminate those redundant costs);
- **transparency**, sufficient to give each user certainty that a competitor (i.e. with similar traffics) does not have access to a different charge not related to cost differences (see below).

Regulators should reject any pricing basis which attempts to load back to users the costs of unneeded capacity unrecovered from other users — these represent unused capacity which should be eliminated.

---

[24] This was the consultants' proposal put to NSW coal companies for NSW SRA coal routes in the late 1980s. This may have been instigated by the railway because it gave a very high value relative to then current revenues. Modern railway construction costs are much lower in real terms than those of even fifty years ago. Of course, land costs would have been less (in real terms) for a government railway than landowners would claim from a modern, private, constructor, who may not have powers of compulsory acquisition.

[25] An Australian example of sunk rail infrastructure cost is the Webb Dock railway in Melbourne, built in the early 1980s, very rarely used, and now obsoleted by the Victorian Government's preparedness to approve the operation of super trucks on strictly defined routes which can carry multiple containers between ship and rail terminals.

Transparency, particularly in specifying the components of the costs being represented by a proposed charge, is vital in supporting the option of challenging the charge. Single-offer, non-negotiable charges are **prima facie** evidence of an attempt to exploit a monopoly situation.

Confidentiality and resale-ban clauses imposed on a buyer are a negation of transparency, as are contractual penalties for breaching confidentiality clauses.

Linear tariffs

Linear tariffs, based on train-km. or some similar measure (or a combination) appear to be simple, but provide no rationing capability where competing demands exist. An example likely to lead to distortions is the present UK Railtrack tariff for passenger train operating companies, which is effectively linear due to being based on rigid train formations[26], with charges based on train miles (about 90+ per cent of cost) and gross tonne miles of the standard formations (remainder) without regard to peak requirements, quality differences or deferability of demands. This British access fee, although a practical solution to an urgent need for a readily understandable tariff, is an extreme example of poor linear tariffs, due to the wide range of differences in cost causation between trains. It will be a source of considerable difficulty in future.[27]

Track Australia, now Australian Rail Track Corporation (ARTC) has a tariff basis broadly similar in principle, but in a coarse way its division into four quality classes probably brings prices nearer to costs, reflecting differing demands[28], and allows users to choose an appropriate price/quality combination. Note that this unit tariff does not resolve the issue of bidding for peak times. By the nature of much of the traffic on its routes, with many users desiring the same departure and arrival times for time-critical freight, ARTC will eventually need a bidding process or a volume-based proxy. At present, existing large users appear to have grandfather rights to prime times. These may be justified in terms of their respective volumes and the implied contract lengths.

Multipart tariffs

These may be based on combinations of time of day, required train speed and priority, maximum axle loads and other factors, have not been effective in present railway access practice due to some owners' very ambitious views of the size of their fixed costs, which generally assume little variability with output at the route level. This is because many owners seem unable to envisage long-run changes in which assets and their costs vary, whereas users will claim a much shorter time horizon and seek to pay based on a very short run marginal cost. Of course, users' preference for short

---

26 This method is acceptable at present only because, for passenger trains, the Railtrack charge is passed through to government in the agreed subsidy to the train operating company.

27 See Stewart Joy: 'Public and Private Railways', *op cit.*, sections 8–10.

28 This paper uses 'Track Australia' to refer to the Australian Rail Track Corporation, the inheritor of the practices of the Australian National Railway track access unit. Present tariffs of Track Australia, appear to be based on long run efficient costs; i.e. the Commonwealth is accepting the cost of any short run inefficiencies. Some users question whether TA tariffs adequately provide for rerailing costs, and this will be determined when other maintenance economies are achieved under contracting.

contractual commitments militates against the use of multi-part tariffs which envisage long-run changes.

This issue of agreeing appropriate time horizons will recur as tariffs are refined. Some users have claimed, on the basis of present ARTC charges, that the appropriate time horizon on other systems should not include rail renewal. Where that claim is made, those users should be invited to stand aside to see if competitors are prepared to pay a higher charge for the capacity they use and the priorities they have. If the market cannot support rail renewal costs, retention of routes is an issue for governments and direct subsidies may be needed.

Note that the costs or savings from capacity or other output quality changes are not symmetrical. There are small cost savings from reduction in capacity (closure of loops), compared with relative high costs of expansion of capacity (a number of new complimentary loops, or lengthening of loops, over critical sections of route). Multi-part tariffs may come into wider use if users are given the opportunity to propose pricing structures which would enable them to optimise their combinations of operations and infrastructure. The ideal is to offer each user an option to propose a multi-part tariff which allows it to minimise the owner's costs, by covering strictly defined fixed costs (which cover peak requirements), and with the incremental components covering wear and tear. But this requires a flexible approach on behalf of the owner.

One alternative is to give a **predominant** user freedom to choose from all paths to set basic capacity and to pay for it and usage cost, and for the owner to then sell marginal paths **within that capacity** at some price[29] between marginal cost and stand alone cost, on behalf of the predominant user[30]. There is room for a wide range of product quality in terms of time of day, or acceptable delays, and appropriate prices. The predominant user would then control and pay the cost of any capacity spare after all demands are met. This approach is sometimes resisted by owners, who wish to:

- limit predominant users' influence over capacity decisions;
- retain control of total capacity in hopes that it can be sold, (where potential revenues may exceed cost of retention in the interim, particularly if some of that cost can be extracted from other users); and
- deal in capacity with marginal users at prices greater than MC and on a basis where they keep the revenue, rather than rebate it against the predominant user's charges (which may have been based on full cost cover)[31].

---

29 If the 'predominant user' is in competition with other users, e.g. carrying similar freight (or passenger flows), the only commercially fair pricing basis for the other user is a price near to the **average** access cost incurred by the predominant user.

30 Difficult commercial relations between National Rail Corporation and competitors suggest that NRC would not be an acceptable 'predominant user' on interstate routes, even where it generates most of the demand.

31 This applies with Railtrack in Britain, where the revenue from additional trains is not reimbursed to users which may already be (together) paying full cost.

A practical way — so far unused — to implement this alternative would be for the predominant user to retain ownership and responsibility for all costs, and for the track authority to act as its sales agent in selling unused capacity, passing through the revenue earned (less selling costs). This would have been a preferable basis, for example, for CityRail track in NSW, where the needs of CityRail's suburban and interurban trains determine the capacity[32]. But the track authority must have power to enforce the access deals it makes, i.e. to prevent the owner from providing less advantageous access than contracted. This approach would prevent a vertically integrated railway from discriminating against its competitors, but leave it to make optimal decisions on capacity and quality.

Ramsay pricing

Ramsay pricing implies the recovery of common costs according to the inverse of buyers' respective demand elasticities. But these are not known with precision, or at all. In regulatory affairs, the term is usually a euphemism for what a seller or buyer thinks the market will bear, and has little specific relationship to relative demand elasticities. There is nothing wrong with bargainers trying to extract each other's surpluses from a negotiation, but it should not be dignified by a name which suggests some intellectual purity to the approach. The asymmetry of risk exposures between a user, most of whose costs are future cash costs in respect of varying demands, and an owner, most of whose costs are sunk, suggests that unless most of a user's surplus is retained, the traffics may not move and both parties will be worse off.

It is common in rail access disputes for both sides to claim that 'Ramsay pricing' should be used, and for each to swear as valid a theoretical justification for charging efficient SRMC, and less than efficient Fully Distributed Costs, respectively. In these cases, one side offers what it feels like paying, and the other demands what it feels like charging, with no supportable cost basis offered in either case.

In a perfect world, where opportunistic claims are absent, this issue relates primarily to recovery of common costs, and if these are not paid by one or more users, or subsidised by government on the basis of external benefits, the facility may not be available. In practice, costs in excess of efficient costs are sometimes attempted to be recovered under the guise of Ramsay prices[33], which is no more than an attempt to tax a user to pay for the owner's inefficiencies.

Peak load pricing

---

[32] The present arrangement in NSW, where RAC owns the infrastructure, but pays CityRail to control its use (signalling operation and train control), then sells back this service to CityRail, with each adding a profit margin, and then with CityRail billing Treasury for its below rail costs, is an unnecessary administrative jumble.

[33] The argument generally offered is: 'If you want to use the railway, you must take it as it comes, and pay to cover excessive costs while we improve it!' But if these costs can be recovered by a monopolist public-sector owner, there is little incentive for it ever to make the improvements toward efficient costs.

This is a necessary investment indicator where more than one user demands to use peak paths, which in turn determines route capacity. It will eventually be required on inter-capital routes in Australia .

In the UK, grandfather rights have been given to original franchisees on the basis of the timetable at privatisation. There is already some evidence, where potential demands indicate marginal revenues exceed the marginal costs of train operation, of franchisees seeking (and being prepared to pay for) additional paths to block out competitors with service, but see the section titled **important role of the right to interconnect**.

Interconnection

Examples and options

Provided that common safety standards are observed, the only interconnection issue in railway operations is in gaining the right to run a train over a route owned by another party, at a specified time[34]. This right has been exercised on railways for nearly 150 years, under different names, with or without payment:

- ‘running powers’, of which examples in Australia are the rights of early Hunter Valley and Illawarra coal railways to run on government routes to port.

- ‘trackage rights’ in the US, with or without an arbitrary payment[35], and sometimes with the owner allowing the user to repair a route to protect the user’s operations[36].

There is always a possibility of a vertically integrated railway using, either deliberately or through bureaucratic inertia, spurious mechanical or operating standards to prevent access. The Australian proposals that safety certification in one State should apply in all States[37] should control this possible barrier to entry, but practice is so far falling short of the legislative and contracted intent. Different approaches to deciding who must be certified are causing unnecessary problems, and in some places appear to be used for purposes of commercial constraint. Some of these differences may be just bureaucratic confusion, such as the Victorian track authority’s requirement that the charterer of the train be safety-certified, in addition to the operator and the providers of

---

34 Note that in the short run a railway interconnect demand is deferable — leave the train off route until path available — but this is practicable only if the deferment does not itself use up route capacity. In the long run, users will demand certainty of access to minimise their own costs of crew and trains, and to be able to offer an acceptable quality of service.

35 In the US, haziness on costs sometimes leads to a relaxed view of the required price — a commuter railway in North East US ‘cannot remember’ the basis of charging Amtrak for access — ‘probably car-miles, inflated by some series’.

36 Inter-railroad relations are not always as sweet as this. In Britain, the rule of priority was generally that the first train which offered (at or after the scheduled time) at a junction received first access, but in the US a tower operator might give absolute priority to his own railroad’s trains. In the US, haziness on costs sometimes leads to a relaxed view of the required price — a commuter railway in North East US ‘cannot remember’ the basis of charging Amtrak for access — ‘probably car-miles, inflated by some series’.

37 See, *National Guidelines if Rail Safety Accreditation Applications,* Rail Safety Intergovernmental Agreement Working Group, May 1997.

locomotives and rolling stock. More serious is the refusal of the NSW Rail Access Corporation to deal with an operator, already safety-certified by three regimes, on the ground that it is not certified in NSW[38].

It is ironic that the Commonwealth Department of Transport and Communications, which readily accepts the certification of about one hundred countries in respect of airline operators wishing to fly into Australia, cannot achieve similar cooperation between the five railway-owning states in Australia.

Important role of the right to interconnect

In vertically integrated railways, the right of an outsider to have access, either imposed on terms by a regulator or arbitrator, or under an access undertaking, puts a limit on the profits which can be extracted by the landlord. Examples are Queensland, Western Australia, and the railways in Tasmania and South Australia recently sold by the Commonwealth, where the access cost (however determined), plus haulage cost, will set the maximum a landlord railway can charge for its own services, because the landlord will be able to charge no more for above-the-rail services than the costs (including normal profit) of a potential user.

This right, and its effect on landlords' total prices, should be enhanced in Australia by the way in which rail safety accreditation is intended to be administered, where an accreditation in one state is proposed to apply elsewhere, so that a competent operator should not be refused access on spurious safety grounds[39]. This creates a market in which competent operators will intermediate between major shippers and railway landlords. In the end, the threat of using the best operator's costs will determine rates.

However, a further regulatory issue arises with a vertically integrated railway owner, in that for traffics for which users require access prices, regulators will need to have surveillance over **all other costs and prices** of the owner for those traffics. This is in order to prevent the owner from setting an unrealistically high price for access in order to protect its terminal and haulage operations by making a potential user's access plus terminal/haulage costs uncompetitive. Such activity may be used to protect a highly profitable (i.e. overpriced) business of the owner.

Incentive vectors

These are arrangements which encourage users to frame their demands in ways which minimise the net costs of both owner and user. For the owner:

- Use of **efficient** stand alone costs prevents recovery of extraneous or inefficient costs, and focuses attention on their elimination;

---

38 See, National Competition Commission: Declaration of the Hunter Railway Line Application by the NSW Minerals Council Ltd.

39 In the UK, safety accreditation is based on both rolling stock and management systems, and an operator on one route does not have an automatic right to run elsewhere. In addition, new powered rolling stock requires route-specific safety approval.

- Requirement that contracted payments must cover capacity expansion protects against wasteful investments — i.e. risk for this to be taken by the potential user(s).

For the user:

- Pricing based on a two part tariff — train path plus rail wear — gives incentive to maximise train length (subject to capacity of loops).

- 'Train path' cost should represent a rational division of stand alone cost for the total traffic, including capacity enhancement if for a peak path[40].

- Share stand alone cost on rational basis if all users equal in demands (or differences unmeasurable)[41].

- If users' demands differ, offer priorities of access according to preparedness to pay, with compensation in the form of lower prices for those users prepared to avoid peak times and possible reduction in total cost from capacity reduction as users rationalise their demands.

- Long term (multi-year) commitments, with right of resale by user, give to the **owner** security in making renewal investments and in dealing with other users, and to the **user**, the ability to recover some of the cost to it of any false projections of demand. It is also reasonable to determine peak priorities in terms of length of contract offered by users[42].

- Right of resale in the hands of a large user which commits to large volume at lower rate gives protection to casual users who can buy resold capacity. Such a secondary market provides a satisfactory control over quality and prices offered by infrastructure owner to small users. The large volume users effectively has a take-or-pay contract, and is forced to seek sub-contract users[43].

Some potential problems requiring regulatory oversight —'overhead costs', resale prohibitions and denial of transparency

'Overhead costs'

These will be a continuing problem in NSW, where the Railway Access Corporation has tried to recover costs from users which have no relationship to the actual cost of the access required, but simply reflect the costs of failure to run down past organisational

---

40 The user of the peak path, to the exclusion of another peak bidder, should be prepared to pay the fee, plus at least the cost to the second bidder of any additional costs it incurs.

41 See the section titled **problems with multi-users**, below.

42 Note that this raises issues of enforcing contract length undertakings. In the absence of underwritten guarantees, it is fair for an access seller to give preference to a user of greater financial strength.

43 Note the argument elsewhere in this paper, that if a large (and therefore low marginal price) user is prevented from on-selling access, it will seek to use its capacity to compete directly with the smaller potential users.

structures or of new structures created[44]. This is a problem of the way in which RAC was set up. The stand alone cost test, which should be the basis of any access charging, precludes any of these extraneous costs, as they add no value to the access required but are more in the nature of a tax levied to cover the cost of self-inflicted difficulties. Provided that it keeps its own administrative costs at present levels, the creation of ARTC to run the interstate routes protects access charges for those traffics from arbitrary allocations of 'overheads'.

The issue may also arise in Queensland and Western Australia, where it is possible that the landlord railway might argue that specific users should pay their 'share' of general railway overheads, even if these are largely incurred on other routes and traffics.

Resale restrictions and denial of transparency

Resale prohibitions are a defence of price discrimination. Because secret prices will generally be seen as an attempt to charge discriminatory prices unrelated to cost variations, it is difficult to justify this practice by a publicly-owned authority. With secret prices, there is no guarantee that similar prices are charged to similar users. Secret prices, and any restrictions on resale, are basically a protection for unused capacity, inefficient operation, or the capricious exploitation of a weak bargainer.

Restrictions on resale are sometimes attempted to be justified on grounds that price discrimination is required to enable the owner to extract total revenues sufficient to cover total costs from users with varying elasticities of demand. But there is no automatic indicator which assures any of the users that total revenue does not exceed total **efficient** cost.

There is a particular problem where an excessive proportion of the owner's total cost is at the overhead level, and cannot be attributed to any particular route. The owner then feels justified in recovering these 'overheads' as and where it can; some may in fact be phantom costs, i.e. a tax on particular rail users.

Attempts to charge prices unrelated to marginal costs risk misallocation of resources by either causing loss of traffic, or diverting users' profits to support internal inefficiencies. The NSW RAC's opaque prices may be rejected by the NCC in considering its application for approval of its pricing regime. In any case, comparison with the TA tariff may force reasonable prices on the RAC.

Note that secret prices eventually result in a shift of profits from the infrastructure owner to the user favoured with a lower access price, because the favoured user is able to approach another user's clients with a lower total price. This may be the case in NSW, where NRC apparently has a favourable price basis and is able to compete for the traffic which SCT and TNT hope to carry.

Opaque prices, which are different from tariffs elsewhere, will always be challenged. Resale prohibitions are a defence of price discrimination, and will similarly be challenged.

---

[44] See note xxxx, re. the SCT claims.

The primary tests on confidentiality clauses for a regulator are whether —

- for the **landlord**, they have any purpose other than to deny users evidence for appeal to regulators; and
- for the **user**, whether there is any benefit of secrecy other than an advantage over other users.

In practice . . .

Government influences

Some Governments may want railway network owners to recover their stranded costs from users. If this approach continues in NSW it will lead to a sub-optimal usage of the extensive network. State government-owned railway networks have excess capacity on most routes, and may incur inefficient costs in maintaining that capacity. So the stranded costs comprise:

- excess physical capacity (which may not offer significant savings from immediate rectification);
- excess human resources capacity and inefficient technologies.

Contract users will decline to pay for either. If the owning Government is unprepared to act, it is preferable in these cases that it identifies the costs of both excesses and pays that directly to the infrastructure owner. That way, the government gains knowledge of the efficiency of its authority. Note that abstraction of traffic to more efficient operators will create new forms of stranded cost if government railways do not reduce all other costs in step with any traffic losses.

It is only possible legally or commercially for a regulator to force contract users to pay for the excesses from within the user's profit margin. If a regulator attempts to impose this on a user, abstracting its profit in favour of the landlord, the asymmetry in risk acceptance will eventually cause the user to withdraw and the landlord to lose the total revenue. (The traffic may transfer to a user with a more favourable access arrangement).

The worst solution is to force residual government-owned railway operations to pay for the excess, and to subsidise them accordingly. That approach, which was tried initially in NSW, meant there was no control over the infrastructure owner, and the subsidy to the government operator was not controllable on any objective basis[45]. Later, RAC arranged to receive subsidy of a 'basic fixed cost' of $177 m. per annum, which is now paid direct from government, instead of through FreightCorp and SRA[46]. The main

[45] It also led to confusion in that FreightCorp, having paid for all the infrastructure cost, was entitled to resist attempts by RAC to sell capacity on the line concerned to FreightCorp's competitors. RAC was in the position of being able to charge twice for the same capacity.

[46] A realistic estimate of the true basic fixed cost of RAC (at government railway efficiencies) is about $7000/route km on country lines, or about $35 m. per annum. RAC's 'fixed' cost obviously includes a great deal of controllable, variable cost. In fact, the subsidy is a very large

virtue of this approach is that presumably FreightCorp and its competitors will pay the same usage charge. But this is difficult to verify with RAC's opaque pricing.

In Queensland, the Government intends paying the 'basic infrastructure cost' on lines other than heavy traffic mineral lines, with Queensland Railways rebating the infrastructure component[47] of any access revenues. For purposes of transparent accounting, and to prevent QR making a profit at the expense of its subsidising government, this will require access charges in Queensland to be itemised as to their cost components. The 'basic infrastructure cost' is yet to be determined in detail.

Problems with vertically integrated railways

We have already noted that vertical integration is potentially the least-**cost** solution in terms of railway resources expended, because it avoids the duplication of management and other transaction costs which a track authority entails[48,49]. Provided that realistic access prices and conditions are offered to contract users, vertical integration leaves all of the issues of capacity determination and efficiency with the railway operator which benefits from them. ('Realistic prices' means those which a regulator would agree, i.e. within the stand alone/marginal cost envelope, based on the expected life of the traffic and using efficient costs).

But unless the access pricing system is open to evaluation, the preference to retain a vertically integrated railway may be based more on a desire to protect the vertically integrated railway's profits. There is always a temptation that preference for vertical integration may be based on a commercial decision to attempt to provide access to others at a high price, to prevent competition with the services of the landlord railway. The critical issue for a regulator is how much of the economies of vertical integration are passed through to end-users[50]. The purpose of an access regime is to ensure that

proportion of total non-metropolitan costs, items, and is thus largely a contribution to cover variable costs (which should be billed to traffics) inefficiencies.

47 That is, covering any costs already paid for by the Government.

48 Overseas experience with track authorities is discouraging so far on this point. The Swedish Banverket has offered very little gain beyond facilitating access for minor users. Railtrack and the train operating companies in Britain must use a cumbersome system of penalties for delays which is often self-defeating, but in this case, to meet other Government objectives, separation was unavoidable to meet other Government objectives. In Britain, whilst economies may be obtained, they will not necessarily be passed on in full. It might be noted that (March 1998) the share market expects Railtrack to earn super profits from the efficiencies it gathers but does not pass on to users (and the subsidising government). The share market values Railtrack shares at over £10, compared with a flotation price of less than £2.

49 In some cases, claims that the transaction costs of managing access separately can exceed the gains to users may be valid. NSW Rail Access Corporation incurs very high administrative costs, and its practice of buying-in and reselling services from other NSW rail instrumentalities, each of which adds a profit margin, probably creates phantom costs. In contrast, the published price list of AN's Track Access unit does not appear to support heavy overheads, profits, or (possibly) even the long run replacement cost of the infrastructure.

50 'Separation reduces the net worth of a railway', i.e. it reduces cost recovery, which can be either monopoly profits or cover for inefficient costs elsewhere. Quotation is from paper by the Executive Manager, Strategic Issues, Queensland Railways, to the Australian Railway Association Track Access Seminar, 4 June 1997. But while there is no good case for enforcing access on a basis which simply transfers monopoly profit from owner to access-user, there is

competition is possible, to maximise that pass-through, and that requires the regulator to consider the end-market for the service. Where the owning, vertically integrated, railway is faced with a demand for access, evaluating that request also means evaluating the owner's pricing of the traffic concerned, to ensure that its price recovers both its full above rail costs and infrastructure costs on the same basis as it seeks to charge to other users.

The regulatory issue is in determining whether an access request on a vertically integrated railway **will benefit end-users**, or simply represents a conversion from owner's margin into user's profit. This is the argument used by Queensland Railways in the infrastructure component of the Carpentaria case, and by most US railroads. Note that just because an owner is using differential prices for different traffics in order to recover the total cost of providing a route, there is no valid argument for a user to be able to avoid such discrimination, because if it succeeded, the owner would be failing to recover an appropriate level of costs.

equally no case for enabling a landlord to mulct end-users by protecting it from competition form (possibly) more efficient access-users.

Finding stand alone costs

These are easy to determine for the track modulus and train control, but a little more difficult for the subgrade. Stand alone costs for the subgrade are relevant only where, **in the absence of the present route, the shipper would have built its own railway.**

An area of inevitable argument in determining stand alone cost is engineering and administrative overheads. Railway systems and track authorities are accustomed to simply pro-rating these on some simple basis, like tonne-kilometres, train km or track km, but stand alone cost implies assessing **the specific cost to manage the route concerned**, and not just sharing historically developed costs. Items like 'unallocated (to routes) overheads' are inadmissible; if they cannot be specifically attributed to a particular route, they do not represent any value added in respect of that route and are therefore not a component of stand alone cost.

Note that stand alone costs should, **where the market will bear it**, include a return on capital for the full construction cost of the route required by the user.

Problems with multi-users

These arise only where the landlord attempts to discriminate, using separately bargained and 'secret' prices. There are two dimensions —

- different prices for the same services
- different train priorities.

But see the note on secrecy in the section titled **resale restrictions and denial of transparency** above. The only fair and workable multi-user arrangement is to have open prices, paid for predetermined access for specific times and priorities. Multi-user pricing will be a problem in NSW, where existing levels of costs are sought to be recovered through price discrimination, but not with the inter-capital network which will come under the control of ARTC. ARTC's problem may come on the cost side, where its viability depends on reducing costs to the levels represented by its published charges. This may require transfer of the interstate routes in NSW to TA, in order to give it control over maintenance efficiency. Alternatively, TA may elect to pay for interstate route capacity in NSW on the basis of the costs it would incur if it were the owner, leaving inefficiencies with RAC.

A multi-user issue which arises in Australia relates to imbalance between respective users, where one potential user may demand a lower access price (based on short run costs) in competition with a larger user such as NRC or FreightCorp. The principle which should be observed here is that the potential operator should not have access to a lower price than existing operators are paying.

The workable solution where multi-users are interdependent is to calculate stand alone costs for their joint demands, and then to divide them among those users in a manner which gives no user a competitive advantage over any other. This appears to be the ARTC solution, with its published tariffs, which will provide a preferable solution in NSW for interstate operators.

The new private railways

There are now privately-owned railways in Tasmania and South Australia. These have purchased from the Commonwealth the streams of profits they believe they can earn from the infrastructure and rolling stock. This brings two problems in dealing with the major shippers:

- The shippers have recently been arguing with the former owners of the routes that they should pay only the 'cost' of moving their traffics, and now the Commonwealth has recognised and sold the profits concerned, **which exceed that cost**.

- In most cases the monopoly the Commonwealth has sold is focussed on ownership and remaining life of specialised wagons, because locomotives are readily transferable and another accredited operator can gain access at stand alone costs.

- It will be interesting to see how regulators treat the return on investment required by the new owners, in respect of their purchase payments to the Commonwealth. South Australia has avoided this problem in respect of the most important private railway route, between Leigh Creek and Port Augusta, by taking ownership of the route rather than having the bidder buy it from the Commonwealth.

The main advantage available to a landlord railway is its ability to share fixed costs between its own users, and benefiting from sales of spare capacity (at stand alone cost) to competitors.

Conclusions

Some of the critical points we have argued here are that —

- Rail access prices should be set between short run marginal cost and stand alone costs.

- Short run marginal cost lower than long run marginal cost implies excess capacity for which users will resist paying.

- If users can only pay less than long run marginal cost, either they must cease operations when the unpaid-for assets fall due for replacement, or subsidies will be required.

- The Australian access regime cannot be used to **force** large railway operators and infrastructure owners to sell capacity and other services at less than stand alone cost, and that the economies of scope remain with those whose other businesses generate them.

- Infrastructure owners may choose to sell capacity at less than the buyer's stand alone cost, where the sale provides net revenue to them, but they cannot be forced to distribute their economies of scope to competitors.

- Secret prices imply price discrimination based on non-cost criteria.

- Lack of transparency implies a refusal to justify levels of cost.
- Resale restrictions for access imply price discrimination based on non-cost criteria and are inherently unstable due to low-fee users being able to divert traffic from high fee users.
- 'Overheads' which would not be saved in the absence of a particular usage are not a justifiable element in stand alone costs because they add no value to the usage.

A primer on access regulation and investment[1]

Joshua S. Gans and Philip L. Williams, Melbourne Business School, University of Melbourne, Australia.

This paper reviews recent advances in regulatory theory concerning the effect of access pricing regulation on incentives to invest in infrastructure. We demonstrate that regulation has a dual role of ensuring that investment costs are themselves shared by multiple users of a facility and that potential providers compete to invest in a timely manner. Our paper, therefore, provides a rationale for using fixed access charges to allocate investment costs so as to ensure timely investment and competition.

This paper provides a discussion of the relationship between access regulation and investment in essential facilities. Regulation of access terms and prices affects the return a facility provider can expect to receive on its investment. Hence, expectations of the nature of regulation affect investment incentives. A regulator, therefore, has the power to influence investment indirectly by pre-committing to an access pricing regime. However, uncertainty and inappropriate signals can potentially have an adverse influence on investment. So the regulator must take care when stating regulatory policy and applying regulatory instruments.

In Australia, Part IIIA of the *Trade Practices Act (1974)*, gives the Australian Competition and Consumer Commission (ACCC) the power to determine prices of access to essential facilities. These powers are constrained somewhat, however. For example:

> s44W(1): The Commission must not make a determination that would have any of the following effects ...
>
> (d) resulting in the third party becoming the owner (or one of the owners) of any part of the facility, or of extensions of the facility, without the consent of the provider,
>
> (e) requiring the provider to bear some or all of the costs of extending the facility (or maintaining extensions of the facility).

This indicates that the ACCC does not have absolute power to impose any form of pricing.

In this paper, we wish to review how regulation affects investment in infrastructure that involves sizable sunk costs. In particular, we suppose that the infrastructure produces its services by way of a natural monopoly technology. In addition to the sunk costs of its provision, there are constant marginal costs involved in its use. Moreover, the infrastructure effectively has unlimited capacity. It is not certain that the ACCC has the

---

[1] We thank David Briggs for helpful comments. Responsibility for all views expressed in this paper lies with us. All correspondence to: Joshua Gans, Melbourne Business School, 200 Leicester Street, Carlton Victoria 3053; E-mail: J.Gans@mbs.unimelb.edu.au; Fax: (03) 9349 8133.

power to regulate access to such infrastructure. The Trade Practices Act requires that in order for a service to be declared it must be 'uneconomic for anyone to develop another facility to provide the service.' (s. 44G(b)). For a natural monopoly technology, a single provider is economically efficient. However, it may be commercially viable to duplicate the facility allowing an entrant to bypass an incumbent. In this paper, we will cast a broad net concerning the powers of the ACCC in this regard and assume that it can regulate the pricing of access to all facilities with a natural monopoly technology. As we demonstrate below, regulation can improve social outcomes in this instance.

Regulation serves to affect investment indirectly. By committing to an access pricing structure that it will implement whenever access is sought, the ACCC influences expectations. Here we focus purely on the dynamic aspects of regulation and hence, assume that the regulator has complete information so that the optimal form of pricing takes to the form of a two part tariff (King and Maddock, 1996, Chapter 5). The first part of the tariff sets the usage charge equal to short-run marginal cost. Note that under incomplete information this is not necessarily the preferred solution (Laffont and Tirole, 1993; Armstrong, Cowan and Vickers, 1994; Vickers, 1995; and Armstrong, Doyle and Vickers, 1996). Here, however, we abstract from such concerns to focus on dynamic issues. The second part of the two part tariff sets a fixed charge for access. Under complete information, the key regulatory choice that influences investment is, therefore, the choice of the fixed charge.

The determination of fixed charges has always been a contentious issue in regulation. In the past, its choice has been seen as arbitrary — essentially, redistributing income from access seeker to provider — without any real efficiency consequences. However, from the point of view of market participants, the level of fixed charges is a contentious issue. This is because providers realise that it affects the overall return on their investments and access seekers realise it influences their incentives to enter markets and compete with incumbents.

We contend that the use of fixed charges can have a key role in determining investment incentives. To this end, whenever we refer to access charges below it will concern only fixed charges. To have an effect on investment, regulation must modify incentives. A higher fixed access charge raises the incentives of firms to invest. However, it cannot be too high or seekers might have an incentive to duplicate the facility. For providers and seekers, therefore, a formula that determines their expected access charge will form a critical part of their perceived returns from their actions.

This paper will demonstrate that appropriate regulatory policies can have a beneficial effect on investment incentives. We demonstrate this by comparing regulation to what occurs without regulation. In the absence of regulation, access providers are torn between their incentive to maximise the use of the infrastructure and a desire to limit competition downstream. After all, intensive use of a natural technology facility reduces long-run average costs for **all** users. However, allowing your downstream competitors to have access to the facility reduces your ability to earn monopoly rents downstream (Rey and Tirole, 1996). In general, in an unregulated environment, providers will limit optimal use of the facility so as to limit profit-reducing competition downstream.

Below we distinguish between two important cases. The first concerns a situation in which use of the investment is **non-rival**. In this case, alternative users of the infrastructure do not compete directly downstream. This might include access to facilities such as towns, ports and rail lines for mines that are world price-takers or situations in which there is high product differentiation downstream. In this situation, providers do not fear the rent dissipation caused by downstream competition and hence, have an incentive to optimise the use of the facility. Even here regulation can have the beneficial role of allowing sunk investment costs to form part of access pricing — something that would not occur in an unregulated environment. Hence, access pricing can beneficially accelerate investment decisions to the benefit of all.

In contrast when the use of the investment is **rival**, the regulatory issues are more complex. When there is some competition downstream, a vertically integrated provider may restrict access so as to maintain monopoly rents. Here the regulator has a dilemma. By forcing the provider to grant access, it reduces the returns a provider can earn on investment. So if the expected access charge is too low, investment might be deterred. However, it is too high, alternative providers might race to investment. In this case, investment might take place with an inefficiently inferior technology.

The non-rival case

There are extensive deposits of iron ore in Western Australia's Pilbara region. Most of these were discovered in the early 1960s. These deposits vary in quality and in location. There were no existing facilities to exploit these resources. That is, there were no towns, ports, railways and the like. They were developed by pioneer investors under considerable uncertainty regarding their actual future use.

These facilities involve natural monopoly technologies. Investment in them continues. Users of the facilities are unlikely to compete directly with one another. If a potential user of one of these facilities came to the NCC seeking access or the ACCC for access terms, what should these regulators do? How should it determine any fixed access charges? Can it have any favourable role here? The answer is yes and this can be demonstrated by the following simple example.

Investment costs

Suppose that today (at date 0), a potential provider is considering making an investment in, say, a rail line. The current cost of that investment would be $100. However, if it waited five years, those costs will fall to $77.88 and if it waited ten years, they would fall to $60.65. This is because of technological progress that results in a continuously falling cost of production of rail lines and engines. In present value terms, therefore, the rate of cost reduction is even greater. These costs are illustrated in Table 1 assuming an annual interest rate of 5 per cent.

Table 1: Illustrative infrastructure costs

Time of construction	0	5	10
Present value at time 0	$100	$60.65	$36.79
Current cost	$100	$77.88	$60.65

Stand-alone choices

The firm can earn total profits (over time) of $100 from the use of the infrastructure, regardless of when it invests. With this, what are its incentives to invest in the infrastructure on its own? That is, what happens if it does not believe there will be any other users of the facility. Its **stand-alone** returns are listed in Table 2. Those figures tell us that the firm will earn the greatest return (in present value terms) by waiting ten years to invest. When it is the only user of the facility, it pays the firm to wait until the costs of investment are less prohibitive.

Table 2: Stand-alone incentives

Time of construction	0	5	10
Present value of profits	$100	$77.88	$60.65
Net present value of investment	$0	$17.23	$23.86

Social optimum in the presence of multiple users

Suppose now that another user of the rail exists. While the provider knows about the potential existence of such a user, they cannot get together prior to the investment to form a joint venture. Instead, that user will seek access at some time after the rail line is constructed. Suppose that this user can earn total profits from the use of the rail of $50. Recall that in the non-rival case, allowing the user to have access does not diminish the profit opportunities of the provider. From a social point of view, it is now optimal for the rail line to be built in five years. This can be seen from the calculations in Table 3.

Table 3: Social incntives

Time of construction	0	5	10
Present value of profits	$150	$116.82	$90.97
Net present value of investment	$50	$56.17	$54.18

No regulation case

However, it is unlikely this outcome will be achieved when there is no regulation. Consider what happens in this case. First, the provider chooses when to invest. Then the seeker chooses when to seek access. Finally, the seeker and provider negotiate over the access charge. The provider's choice in the first stage is contingent on its expectations of when the seeker will seek access and the outcome of resulting negotiations.

In economics, negotiations can only be over variables that can be altered at the time of negotiation. As the investment has already taken place at this time, the infrastructure costs themselves cannot play a role in negotiations. They are sunk. This weakens the provider's position in negotiations, lowering the access price it receives. If the provider

wishes to terminate access negotiations, it can only lose. By denying the seeker the use of the infrastructure it can gain nothing and lose whatever access charge it might receive. For the seeker, if it terminates access negotiations, it might be able to duplicate the facility and still earn some profits. Table 4 depicts the seeker's bypass incentives. Notice that the seeker never earns a positive profit by bypassing the provider's investment. In general, this will not be the case, so that option will provide the seeker some bargaining power.

Table 4: By-pass incentives

Time of construction	0	5	10
Present value of profits	\$50	\$38.94	\$30.32
Net present value of investment	-\$50	-\$21.71	-\$6.47

When the by-pass option is not attractive to the seeker, both parties gain nothing by terminating negotiations. However, if an agreement is reached for a charge of p, the provider receives p above the amount it would otherwise earn, while the seeker receives \$50 - p. In total they jointly create \$50 in additional profits by allowing the seeker to use the facility. Thus, under many bargaining games (e.g. Nash, Shapley or Rubinstein — see Gans and Williams, 1997), the resulting agreement would have p = \$25.

Note that with this expected price, it is always in the interests of the seeker to seek access as soon as possible, i.e. as soon as the investment takes place, as it earns \$25 sooner rather than later. The provider builds this and the expected price into its timing calculations. Its new anticipated returns from investment are depicted in Table 5. In this example, while the provider earns more from investing at each date (as compared with the stand-alone case), it still chooses to provide the infrastructure in ten years. So, in the absence of regulation, investment is delayed relative to the social optimum (Gans and Williams, 1997).

Table 5: Incentives under no regulation

Time of construction	0	5	10
Present value of profits	\$125	\$97.35	\$75.81
Net present value of investment	\$25	\$36.70	\$39.02

The potential inefficiency arising in the no regulation environment comes from the fact that investment costs are sunk. Therefore, they do not enter into any negotiations over access. Hence, as the provider only claims a proportion of seeker value in those negotiations, it does not internalise the full social benefits of the infrastructure when making its timing choice. Indeed, its private incentive is less than the social incentive, leading to delayed investment.

The potential for 'racing'

The above analysis makes a critical assumption: that the seeker cannot, for some reason, not actually become the provider itself and sell access to the other firm. This might be because the seeker does not have good access to capital. It is worth considering what occurs if this assumption is relaxed. In this situation, the 'small' firm can invest first and pre-empt the 'large' firm. In this case, the tables on negotiations are turned. However, if the 'larger' firm is the access seeker rather than the provider, it does have a credible by-pass option. To see this, suppose that the infrastructure is built in ten years. The large firm could ensure itself a profit in cash terms at that time of $100 minus $60.65 (the current cost of investment in ten years). Unless that firm receives at least $39.35 after paying for access it will find it profitable to by-pass the 'small' provider.

Gans and Williams, 1997, demonstrate that it is reasonable to expect the access charge in such situations to be the minimum of half the seeker's profits and the investment costs incurred in by-passing the provider. That is, p = min[$50, $60.65] which, in this case, will equal $50. Alternative specifications such as pure Nash bargaining yield slightly different results but with the same qualitative implications. As it turns out, in this example, despite the by-pass option being positive for the seeker, it never binds so p = $50 regardless of the time chosen by the 'small' firm. Table 6 depicts the 'small' firm's incentives to invest given this expected access charge.

Table 6: 'Small' firm incentives under no regulation

Time of construction	0	5	10
Present value of profits	$100	$77.88	$60.65
Net present value of investment	$0	$17.23	$23.86

The key point here is that there is the potential for both firms to race in order to be the first to provide the infrastructure. By doing so, that firm avoids having to pay the other for access and, moreover, receives an access payment. In the unregulated case, each firm, regardless of their size, can expect to receive an additional $75 (= $50 + $25) by becoming the provider rather than the seeker. This prize raises their investment incentives by giving them a reason to pre-empt the other. However, this prize must be traded off against the actual investment costs. Providers have to pay these while seekers do not.

To see this, consider the 'large' firm's decision. It knows that if it does not invest in five years, then the 'small' firm will surely invest in ten years time. In present value terms, the difference is $6.38. Hence, the 'large' firm is potentially better off by pre-empting the other firm. Similarly, the 'small' firm using the same reasoning will gain $2.07 by investing early. Both would be individually better off if they could invest later, but the 'race' between them speeds up their choices. However, this advantage only goes so far. Pre-empting the other firm at time turns out to be too costly for both firms.

What will happen in equilibrium? Gans and Williams, 1997, demonstrate that the equilibrium under racing speeds up investment relative to the situation in which only one firm can possibly be the provider. While this might align the timing choice closer to the social optimum it could actually do too much. A firm might invest with an inefficiently costly technology in order to become the provider. Hence, investment might take place too soon. So while, in the absence of regulation, racing might occasionally help, it could also lead to over investment.

Optimal use of regulation

In Australia, access regulation is not necessarily mandatory. Instead, Part IIIA of the Trade Practices Act sets up a system of 'regulation by negotiation.' That is, when infrastructure is provided, a seeker first approaches the provider to privately negotiate access terms. If they fail to reach an agreement, the seeker can either by-pass the facility or approach the NCC seeking a declaration. If a declaration is granted, the parties may negotiate a price or the parties might approach the ACCC to impose a regulated access regime. The point is that regulation and expected regulatory prices play a key role in the outcomes negotiated by the provider and seeker. The seeker can either by-pass the facility or seek a regulated solution. The provider can also go to regulation by refusing the grant access. It is more difficult for a provider to prevent by-pass.

For the purposes of this paper, we shall simplify the regulatory process by assuming that when regulation is a possibility, by whatever means, the stated regulatory price is actually imposed. For regulation to be most effective, therefore, pricing policy must be stated prior to access being sought and indeed, prior to investment being made. That is, it is important for the ACCC to establish pricing guidelines that are more precise than those it has currently adopted.

While any regulatory price should have a usage charge set equal to short-run marginal cost, what should the fixed access charge be? One possibility is for the regulator to use the access charge to allocate the investment costs between the provider and the seeker. That is, if these costs are C, then $p = \gamma C$ where $\gamma < 1$. Notice that this brings those costs into the price — a situation that did not arise in the unregulated environment.

When considering investment cost allocation, the first question to ask is: what basis should be used to value those costs? For instance, if infrastructure was built in five years for a cost of $77.88 and access is sought at that time, it is clear what the value of the infrastructure is. However, if access is only sought in ten years, the issue is more contentious. One could use a historical cost methodology which sets the asset valuation at the actual costs incurred less depreciation. As there is no depreciation in our example, this would be $77.88. Alternatively, one could use a replacement cost methodology that values the investment at the amount it would cost to replace the investment at the latest technology. In this case, replacement cost would be $60.65.[2] Gans and Williams, 1997, demonstrate that any optimal regulatory policy should provide incentives for an access seeker to do so as soon as possible. If it is paying a share of historical costs, a seeker will choose to seek access at the earliest possible date,

[2] See King, 1996, for an excellent discussion of issues in asset valuation.

as delay does not alter the amount it pays. On the other hand, under replacement cost, a seeker might have an incentive to delay so as to reduce the charge it has to pay. Nonetheless, it will be demonstrated below that it is possible to find a cost sharing rule that overcomes this difficulty.

Gans and Williams, 1997, demonstrate that by setting ? equal to the relative share of the seeker in total profits then, regardless of the method of asset valuation, investment takes place at the socially optimal time and access is sought immediately. Recall that, in our example, we want investment to take place and access to be sought in five years time. If ? = 1/3 = ($50 / ($50 + $100)) when the seeker is 'small' and 2/3 when it is 'large', then this occurs. Note that if investment takes place in five years, under historical cost, the seeker always seeks immediately. Under replacement cost, by seeking access in five years rather than ten years a 'small' seeker saves $0.64 in present value terms while a 'large' seeker saves $1.31. So access is sought immediately regardless of which firm is the provider.

Given this, a provider now expects to receive an access payment immediately. In effect, a 'large' provider expects to incur 2/3 of the investment costs whenever it invests, while a 'small' provider expects to incur 1/3 of those costs. Indeed, each firm is now indifferent between being a provider and being a seeker. So there is no incentive to pre-empt the other. Regulation removes any adverse incentives associated with this.

What happens to actual timing under these conditions. Table 7 depicts the relevant payoff calculations for the 'large' and 'small' providers. Notice that both find it optimal to choose to invest in five rather than ten years. In equilibrium, one of them will do so. Notice also that each firm earns more in present value terms than it would have under the no regulation case. This is because investment is taking place at a time that maximises their joint returns.

Table 7: Incentives under regulation

Present value of profits	0	5	10
Large firm	$33	$37.44	$36.13
Small firm	$17	$18.72	$18.06

Using a cost sharing rule based on relative profits, the regulator can ensure that infrastructure is provided and used at the socially optimal time. It does this by removing pre-emption incentives and aligning individual firm timing choices with the social optimum. Regulation allows the actual investment costs to be shared — something that cannot occur without regulation unless firms enter into a joint agreement prior to investment taking place. However, if this could actually occur then there would be no access issue to be concerned about.

This point is worth emphasising. The optimal regulatory price is the same price that the firms would agree upon if they were to enter into a joint venture. In that venture the public good aspects of the infrastructure are internalised. So the venture would choose to invest at the socially optimal date as we have assumed there are no other

beneficiaries. Gans and Williams, 1997, demonstrate that this is exactly the same as a Lindahl equilibrium in public goods. That is, if we asked — as did Lindahl, 1919, — if the firms were to nominate a time to invest, under what cost allocation rule, would they agree to the same date? It turns out that the cost allocation rule based on relative profitability would achieve this outcome. Therefore, the role of regulation in a non-rival environment is to achieve an outcome that mimics what would occur if the relevant parties had been able to form a joint venture agreement.

The rival case

The access issues that are perhaps more common are where alternative users of a facility compete in some downstream market. In some cases, this competition might be weak. For instance, a seeker might produce a product differentiated from the provider's. Hence, the profits of the provider may only be slightly affected. Here the provider might still grant access in the absence of regulation because industry profits are enhanced by so doing. In many cases, however, the competitive effects are more extreme. The provider might be relinquishing a monopoly position by granting access. As monopoly profits generally exceed total industry profits under more competitive conditions, the provider will not grant access unless required to.

If industry profits were the only concern, then it would be best to allow a monopoly if that is profitable and to use the optimal pricing discussed in the previous section when multiple firms maximise industry profits. However, in general, we are concerned about competition because allowing for it reduces deadweight losses. Hence, consumer surplus under competition is greater and, moreover, the increment to consumer surplus more than offsets any losses in industry profits. So, in general, access regulation is designed to allow for competition where it would not arise in an unregulated environment.

As identified earlier, however, if industry participants are to provide the infrastructure they will only take into account their individual profits when determining when to invest. But timely competition will, therefore, reduce those incentives. To see this suppose that firm *A* receives a profit of $100 if it competes with another firm *B* and $200 if it is a monopolist. Suppose that firm *B* earns $50 under competition and $150 if it is able to act as a monopolist. Note that even if the provider were to receive a payment equal to the profits of its rival under competition the most they could earn would be $150. However, a monopolist could earn more than this. If it is firm *A*, it could earn $200. Hence, if it was guaranteed a monopoly firm *A* would invest earlier than if it expected competition.

This issue arises in many contexts. Take, for instance, the issue of Local Number Portability (LNP) in telecommunications. To provide LNP one must invest in digital switching technology and other hardware. However, LNP also potentially has a competitive impact by allowing consumers to easily switch between telecommunications firms. This increases the intensity of competition those firms face and hence, reduces industry profits. An entrant is more likely to desire LNP than an incumbent but often it falls on the incumbent to make the investment. Without compensation, however, the incumbent will wish to delay the introduction of LNP. This can be solved by requiring the entrant to bear costs greater than would be freely

negotiated; but for the regulator such burdens might reduce the possibility of entry itself.

How should the regulator manage the tension between providing investment incentives and allowing for timely competition? In Gans, 1998, it is demonstrated that appropriate access pricing can be used to create competition for the provision of infrastructure. Where in Section I, regulation removed pre-emption incentives, here regulation can manage them. In particular, an appropriate access pricing formula can ensure that infrastructure is provided at the socially optimal date (including consumer and producer surplus) despite competition being expected immediately. And all this without the use of government subsidies.

It is beyond the scope of this survey to describe exactly how this optimal 'race' is achieved. But basically, it requires that the cost of investment at the socially optimal date is less than total industry profits under competition at that date. Given this feasibility condition, the access charge is a function of the (replacement or historical) cost of the investment with factors adjusting for the rate of technological progress and the flow of social surplus. Effectively, the share of costs paid by the provider falls relative to the seeker over time. If this rate is aligned properly, it can be such that the equilibrium in the 'racing' game between market rivals coincides precisely with the socially optimal date. No participant can gain anything from pre-empting their rivals at this point. And to delay beyond this point is to invite pre-emption and a higher cost share.

Practically this suggests that in order to encourage optimal competition for provision, access seekers should pay a charge reflecting a greater share of investment costs if they do not seek access immediately or are late into the industry. This gives providers a sufficient bonus to ensure they invest in a timely manner. Thus, for issues such as LNP, this suggests that entrants should contribute relatively more (taking into account their size and profitability) than the incumbent provider. To do otherwise would send a poor signal to incumbent providers: diminishing their incentives to invest in new infrastructure (Gans and King, 1998).

Conclusion

Determining the level of fixed charges in any optimal pricing regime has always been a difficult issue. Economic theory has had little to say regarding their level. However, recent research has demonstrated that by appropriately designing access pricing formulae, regulators can improve on unregulated outcomes. In the non-rival case, regulated prices can include sunk infrastructure costs and hence, ensure that private investment incentives are aligned with their social levels. In the rival case, regulated prices can change the nature of the 'race' between market participants to provide infrastructure. Such prices can effectively allow regulators to **both** ensure timely investment and competition.

In order to implement such policies, regulatory clarity is essential. In order to influence investment incentives, firms must form expectations as to the regulated price that will be applied. Thus, it is critical that the ACCC use its guidelines and its early access pricing decisions to send clear signals to market participants.

Bibliography

Armstrong, M., S. Cowan and J. Vickers, 1994, *Regulatory Reform: Economic Analysis and British Experience*, MIT Press: Cambridge (MA).

Armstrong, M., C. Doyle and J. Vickers, 1996, 'The Access Pricing Problem: A Synthesis,' *Journal of Industrial Economics*, 44 (2), pp 131–150.

Gans, J. S., 1998, 'Regulating Private Infrastructure Investment: Optimal Pricing for Access to Essential Facilities,' *mimeo*., Melbourne.

Gans, J. S. and S. P. King, 1998, 'When Being First Doesn't Pay,' *Australian Financial Review*, Friday 30th January, p. 32.

Gans, J. S. and P. L. Williams, 1997, 'Access Regulation and the Timing of Infrastructure Investment,' *Working Paper*, no. 13, Melbourne Business School.

Katz, M. L. and C. Shapiro, 1987, 'R & D Rivalry with Licensing or Imitation,' *American Economic Review*, 77 (3), pp 402–420.

King, S. P., 1996, 'Asset Valuation and Access to Essential Facilities Under Part IIIA of the *Trade Practices Act 1974* (Cth), in M. Richardson (ed.),' *Deregulation of Public Utilities: Current Issues and Perspectives*, Centre for Corporate Law and Secutirites Regulation, Melbourne, pp 94–116.

King, S. P. and R. Maddock, 1996, *Unlocking the Infrastructure*, Allen & Unwin: Sydney.

King, S. P. and R. Maddock, 1997, 'Regulation by Negotiation: A Strategic Analysis of Australia's Essential Facility Access Regulation,' *mimeo.*, ANU.

Laffont, J. J. and J. Tirole, 1993, *A Theory of Incentives in Procurement and Regulation*, MIT Press, Cambridge (MA).

Lindahl, E., 1919, *Die Gerechtigkeit der Besteuering*, Gleerup: Lund, Part I, Chapter 4, 'Positive Losung,' trans. by E. Henderson and reprinted as 'Just Taxation — A Positive Solution,' in *Classics in the Theory of Public Finance*, R. A. Musgrave and A. T. Peacock (eds.), Macmillan: London, 1958.

Rey, P. and J. Tirole, 1996, 'A Primer on Foreclosure,' *Handbook of Industrial Organization*, vol. 3, Elsevier, Amsterdam, (forthcoming).

Vickers, J., 1995, 'Competition and Regulation in Vertically Related Markets,' *Review of Eocnomic Studies*, 62 (1), pp 1–18.

Section IV. Financial techniques

The cost of capital and access arrangements

Kevin Davis and John C. Handley, Department of Accounting and Finance, The University of Melbourne, Australia.

Introduction

The recently introduced Part IIIA of the *Trade Practices Act 1974* establishes a legal regime under which third parties can in certain circumstances obtain a right of access to services provided by certain essential facilities on fair and reasonable grounds. The policy underlying the regime is set out in the report by the Independent Committee of Inquiry on National Competition Policy (the 'Hilmer Report') which sees access arrangements as being critical to the promotion of competition in some markets. Essential facilities exhibit natural monopoly characteristics, in the sense that they cannot be duplicated economically and may occupy strategic positions in an industry such that access by the third party to the facility is required if it is to be able to compete effectively in a particular upstream or downstream market[1]. Examples of essential facilities include infrastructure assets such as road and railway lines, electricity transmission grids and communication services. Access may be obtained by either ministerial declaration or by the owner or operator of the facility giving a written undertaking to the Australian Competition and Consumer Commission (ACCC) to provide access to a third party. The ACCC may be required to determine access prices in relation to the arbitration of an access dispute or to assess the terms and conditions of an access undertaking from an access provider. To do so, it is necessary to determine the rate of return necessary to compensate supplies of capital (debt, equity or hybrids) to an access provider for the provision of those funds and reflecting the risk involved. The paper thus analyses certain issues associated with cost of capital methodologies in the context of these access arrangements, including:

- possible alternatives to the Capital Asset Pricing Model (CAPM) approach in determining the cost of equity and their feasibility;
- factors affecting the derivation of beta risk and the feasibility of international benchmarking of beta risk;
- determination of beta risk for vertically integrated entities operating in different markets with different contestability/competitive conditions;
- the impact on beta risk of a firm subject to price regulation and the impact of subsequent increases in competitive conditions;

[1] The Hilmer Report, 1993, p. 240.

- the impact on the cost of equity under the CAPM framework in relation to levered and unlevered betas, tax/dividend imputation, and imputation and levered betas;
- treatment of depreciation in the cost of capital;
- relationship between beta risk, pricing principles and asset valuation methodologies and the use of nominal and real values of the Weighted Average Cost of Capital (WACC) in relation to the asset valuation approach; and
- treatment of stranded assets and the determination of an appropriate rate of return.

Alternatives to the CAPM

The cost of equity capital for a company is the rate of return required by investors to provide equity capital to the company[2]. It is a market determined rate reflecting the return required on an investment of equivalent risk. In this context the cost of equity capital is an ex-ante or forward looking concept which measures the return expected by investors on their investment[3]. Accordingly, the terms 'cost of equity capital' and 'required rate of return on equity' may be used interchangeably. There are six principal alternatives to the capital asset pricing model that might be used to estimate a company's cost of equity capital:

- comparable earnings;
- discounted cashflow;
- price earnings ratios;
- risk premium;
- arbitrage pricing theory; and
- the Fama-French model.

Comparable earnings

The comparable earnings method provides for a company's cost of equity capital to be based on the return on equity for a sample of 'comparable companies' where for each company in the sample, the return on equity is calculated as the accounting return on the company's book value of equity. Although this method has been used in relation to public utilities in the USA[4], it suffers from two deficiencies: difficulties in identifying

---

2 In this paper, the term 'company' is used generically to refer to access providers which operate in either the public or private sectors and includes non corporate entities.

3 In an ex-post context the cost of capital measures the realised rate of return to equity investors over some prior period. In many cases, expectations of future returns are based on historical returns.

4 Kolbe, Read and Hall, 1986.

comparable companies, and the use of book values of income and equity when the cost of capital is a market related concept[5].

Discounted cashflow

According to Myers and Borucki, 1994, discounted cashflow is the most widely used method to estimate the cost of equity capital for regulated firms in the USA. The discounted cashflow approach defines the cost of equity capital as the discount rate which equates the present value of expected future dividends with the current share price. Algebraically,

$$P_0 = \frac{D_1}{(1+k_e)} + \frac{D_2}{(1+k_e)^2} + \frac{D_3}{(1+k_e)^3} + \ldots \quad (1)$$

where

P_0 is the current share price;

D_t is the expected dividend at time t; and

k_e is the cost of equity capital.

If dividends are assumed to grow at constant rate of g in perpetuity then equation (1) simplifies to:

$$P_0 = \frac{D_1}{k_e - g} \quad (2)$$

which can then be rearranged to give:

$$k_e = \frac{D_1}{P_0} + g \quad (3)$$

i.e. the cost of equity capital is equal to the sum of the current dividend yield and the expected long term growth rate in future dividends. Subsequently, in order to use this method in practice we require estimates of the company's expected future dividend stream and its current share price. Major limitations with this method are the requirement to forecast the company's expected future dividend stream and the need for the company to be listed.

Empirical research in the USA by Hickman and Petry (1990) provides evidence that the dividend growth model performs poorly. While that study used estimated discount rates (from the CAPM etc.) and growth rates to calculate value, the accuracy of the reverse approach of inferring the cost of equity capital from current value and growth assumptions is also called into question. Further problems with this approach lie in the fact that the assumed growth rate, g, is not independent of the assumed value creation ability of the company. It is possible that a growth rate assumed could imply an ability

[5] For further discussion, see Myers, 1972.

of the company to generate returns on investment in excess of the required returns, thereby biasing the estimate of the required rate of return.

Price earnings ratios

This method provides for a company's cost of equity capital to be derived from the Price Earnings (PE) ratios of a sample of comparable companies. By definition, a company's PE ratio is equal to the ratio of its current share price to its earnings per share and therefore is a derived figure. In the simple no growth case where the comparable company is expected to produce a constant stream of earnings and dividends, then the company's PE ratio is equal to the reciprocal of its cost of equity capital:

$$\frac{P_0}{E} = \frac{1}{k_e} \qquad (4)$$

However a complication arises in using PE ratios of (apparently) comparable companies where those companies differ in terms of growth opportunities. It is, for example, possible to demonstrate that the PE ratio can be written as:

$$\frac{P_0}{E} = \frac{1}{k_e} + \frac{PVGO}{E} \qquad (5)$$

where PVGO (the present value of growth opportunities) represents the extent to which the company is expected to be able to generate earnings in excess of the required rate of return on its current asset base and through future investment opportunities. Using PE ratios from comparable companies to derive a cost of equity capital may therefore incorporate the possibility of assuming that returns in excess of or below the required rate of return can be earned and thus potentially bias the result.

Risk premium

This method involves deriving an estimate of the extra return required on equity investments over and above the return required on debt investments. In estimating the risk premium, reference is usually made to some study of long term historical rates of returns such as Officer, 1989, and Ball and Bowers, 1986, in the case of Australia and Ibbotson Associates, 1991, for the USA[6]. The return required on debt investments is usually estimated by the yield to maturity on some traded long term government bond. The risk premium plus the current yield to maturity on the long term bond is then used as an estimate of the average cost of equity for all companies. A subjective adjustment may then be made to reflect any difference in the risk between the average company

[6] It should be noted that the Officer, 1989, and Ball and Bowers, 1986, studies relate to returns prior to the introduction of the dividend imputation system. Further, both Ball and Bowers, 1986, and Officer, 1989, provide evidence that the market risk premium is non stationary over time being sensitive to the period over which the premium is estimated. For example according to Officer, 1989, the historic market risk premium over the period 1882 to 1987 averaged 7.9% per annum and varies from 0.4% per annum for the ten years from 1968 to 1977 to 11.87% for the ten years from 1978 to 1987. It is possible that higher or lower figures could result if different ten year periods were selected.

and that of the particular company in question. Although this approach recognises that the cost of equity capital for a company should be related to a benchmark return in the capital markets it provides no guidance as to the size of the risk adjustment for a particular company.

Arbitrage pricing theory

The Arbitrage Pricing Theory (APT) developed by Ross, 1976, is based on the idea that in competitive financial markets arbitrage will ensure equilibrium pricing according to risk and return. Similar to the CAPM, investors are assumed to hold well diversified portfolios and accordingly the only risk that matters is systematic risk i.e. the risk that cannot be diversified away. However, whereas the CAPM assumes the only source of systematic risk which effects expected returns is market risk, the APT recognises that there may be more than one source of systematic risk (referred to as 'risk factors') which influences expected returns on assets. Assuming there are say n risk factors then the cost of equity capital for a company is given by:

$$k_e = r_f + \beta_1\lambda_1 + \beta_2\lambda_2 + \ldots + \beta_n\lambda_n \qquad (6)$$

where for i = 1,2, ... ,n:

r_f is the risk free rate of interest;

λ_i is the expected risk premium associated with factor i; and

β_i is the sensitivity or beta to factor i.

The concepts of risk premium and beta in equation (6) correspond to similar notions used in the CAPM and therefore the APT may be thought of as being equivalent to a multi-beta CAPM[7]. The main limitation with the APT is not so much in having to estimate betas, the risk free rate and the risk premia as in actually identifying the relevant risk factors. Unfortunately the APT provides no guidance as to the identity of the relevant factors nor as to how many factors there are[8].

[7] It should be stressed that this interpretation is heuristic only as the CAPM and the APT are distinct and separate theories based on different underlying assumptions.

[8] A number of studies have sought to address this deficiency on both theoretical and empirical grounds including Chen, Roll and Ross (1986) and Eltin, Gruber and Mei (1994).

Fama-French model

The most recent development in research relevant to estimating the cost of equity capital (and one which is currently subject to a large amount of contention) stems from the recent work of Fama and French, 1993. Fama and French, 1993, have developed a three factor model of security prices which they claim is superior to the CAPM. They suggest that the cost of equity capital for a company is related not only to market risk (as predicted by CAPM) but also to company size (measured by the market value of the company's equity) and to the ratio of its book value of equity to its market value of equity. Algebraically,

$$k_e ? r_f ? ?_1(r_m ? r_f) ? ?_2 SMB ? ?_3 HML \qquad (7)$$

where

r_f is the risk free rate of interest;

$r_m ? r_f$ is the expected risk premium on the market;

SMB is the expected risk premium associated with the size factor (equal to the expected return on a portfolio of small stocks less the expected return on a portfolio of large stocks);

HML is the expected risk premium associated with the book-to-market factor (equal to the expected return on a portfolio of high book-to-market stocks less the expected return on a portfolio of low book-to-market stocks); and

$?_i$ is the sensitivity or beta to factor i.

Thus the model suggests that the cost of equity capital is higher for smaller companies and for those companies with higher book-to-market ratios. To a certain extent this model may be regarded as an extended-CAPM or as a three factor APT model. There are currently two main limitations with the Fama-French model. Firstly, the model is in its infancy and is not universally accepted with its validity subject to challenge[9]. Secondly, Fama and French, 1992, actually derived the model empirically, rather than starting from a theoretical base, relying on previous research which documented empirical contradictions of the CAPM to identify the two new risk factors. It is therefore likely that the size and book-to-market factors are simply proxies for some other as yet unidentified underlying risk factors which may effect expected returns[10].

---

[9] See for example, the discussion in Fama and French, 1996.

[10] Fama and French, 1996, suggest that HML may proxy for relative distress as weak (strong) firms tend to have high (low) book-to-market ratios and therefore are penalised (rewarded) by the market with higher (lower) costs of capital.

Factors affecting the derivation of beta risk and the feasibility of international benchmarking of beta risk

The total risk of a company's equity returns can be partitioned into two components; systematic risk and unsystematic risk. Systematic risk is due to risk factors which affect **all** companies and therefore cannot be diversified away. The CAPM assumes there is only one source of systematic risk, referred to as 'market risk', which collectively accounts for all those risk factors which affect the overall market as a whole such as changes in the general level of economic activity and political developments. Within the CAPM framework, market risk is also referred to as 'beta risk'. Unsystematic risk is due to risk factors which are unique to a particular company (or industry) such as labour disruptions and technological breakthroughs. An investor can eliminate unsystematic risk by diversification. The concept of diversification is based on the common sense notion of spreading risk across a number of assets or investments rather than investing in a single asset and thereby 'putting your eggs all in one basket'. Diversification involves combining assets in such a way as to reduce the unsystematic risk faced by the investor. Ultimately the only risk that remains in a well diversified portfolio is systematic risk. Since the CAPM assumes investors hold well diversified portfolios (and therefore have eliminated any unsystematic risk) then the only risk that is relevant in determining a company's cost of capital is the only risk that remains being its systematic risk. In this case investors who choose not to diversify will not be compensated (in the form of a higher required rate of return) for bearing unsystematic risk which is more risk than they otherwise need to bear. Specifically, according to the CAPM, the cost of equity capital is given by:

$$k_e = r_f + \beta(r_m - r_f) \quad (8)$$

where

r_f is the risk free rate of interest;

$r_m - r_f$ is the expected risk premium on the market; and

β is the company's beta.

A company's beta is a **relative** risk measure. It reflects the sensitivity of the company's systematic risk to the systematic risk of the market as a whole. By definition, the beta of the market is one[11]. In general, companies which are more risky than the market will have betas greater than one while companies which are less risky than the market will have betas less than one. Therefore, the broad principle to bear in mind in estimating a company's beta is that beta reflects the **systematic** risk of the company relative to the market. **It is stressed that the beta of a company reflects only its systematic risk and not its total risk.**

As mentioned above market risk collectively accounts for all those risk factors which affect the overall market. Although it is impossible to identify all the possible

[11] The beta of a portfolio is equal to a market value weighted average of the betas of the individual companies in the portfolio.

economic determinants of a company's beta, three factors appear to play a significant role: financial leverage, operating leverage and the sensitivity of a company's cashflows to the general level of economic activity. The impact of each of these determinants on a company's beta is quite intuitive. Financial leverage measures the extent to which a company has funded itself with debt. Since debtholders have a fixed contractual claim on the cashflows of the company, then the higher the company's financial leverage the greater the risk to its equity investors and the higher its beta. The impact of financial leverage is considered further in section titled **Leverage and Betas**. Operating leverage measures the extent to which a company's cost structure is fixed. Since by definition fixed costs do not change with changes in production levels, then the higher the company's operating leverage the higher the risk to its equity investors and the higher its beta. The sensitivity of a company's cashflows to the general level of economic activity measures the extent to which the fortunes of the company are dependent on the state of the business cycle. The more cyclical the company the higher the risk to its equity investors and the higher its beta[12].

Betas may be estimated by three different methods. Firstly a fundamental approach looks at factors such as industry, company size and growth prospects to subjectively derive an estimate of the company's relative risk and hence its beta. Secondly, regression techniques may be used to estimate beta on the basis of past share prices. Thirdly, reference may be made to a published beta risk service (which in turn generally derive their estimates using regression techniques based on past share prices). It should be noted that it is well documented in the finance literature that beta estimates are sensitive to a number of factors including the estimation period, the frequency of data used and the particular regression technique adopted (for example whether the standard method of Ordinary Least Squares is used or whether a technique which seeks to adjust estimates for thin trading in securities markets such as the Scholes-Williams procedures is used).

In addition, a number of specific factors need to be taken into account in estimating a company's beta including:

- where the company operates in more than one line of business i.e. a multi-business company;
- where the company is subject to price regulation; and
- where the company is unlisted.

Issues associated with multi-business companies are discussed in the section titled **Determination Of Beta Risk For Vertically Integrated Entities Operating In Different Markets With Different Contestability/Competitive Conditions** and the impact of price regulation is discussed in the section titled **Price Regulation, Competition and Beta**. The issue of unlisted companies is discussed below.

When a company is unlisted there is the obvious problem in that there is no share price history on which to base the beta estimate. In this case, guidance may be provided by examining the betas of other comparable listed companies (i.e. companies with the

[12] See for example Brealey and Myers, 1996, for further discussion.

same risk profile) and then making subjective adjustments to take into account any remaining differences between the comparable companies and the company in question[13].[14] In this regard it may be appropriate to focus on the average beta for the sample of comparable companies rather than on the beta for any one particular company in the sample since this reduces the likelihood of estimation error. There may of course be difficulties in identifying comparable companies in Australia and in some cases they may not even exist. This leads to the issue of the suitability of international benchmarking. Beta estimates of similar companies in different countries may be sought particularly from the USA and UK with their large capital markets and rich data sources. However caution needs to be applied in the employment of such international comparisons. By definition, beta is a relative concept and therefore is market specific. The beta of an Australian company reflects its systematic risk relative to the Australian market whereas the beta of a USA company reflects its systematic risk relative to the USA market. Complications may arise in international beta comparisons due to different structural and institutional differences between the markets particularly in respect of alternative capital structures and tax systems. However casual empiricism suggests there are broad similarities once the effects of leverage are removed, for example, utilities tend to have lower than average betas in most countries while airlines tend to have above average betas in most countries. Care also needs to be exercised in using international comparisons in industries which are subject to regulation as any differences in the regulatory environments need to be taken into account. Finally, it is considered worthwhile to briefly refer to the issue of how precise beta estimates should be. From a theoretical perspective, beta is the key determinant of a company's cost of capital but from a practical perspective, this importance is somewhat reduced when one takes into account the relative amount of subjectivity that exists in the other input variables particularly the market risk premium. For example, as mentioned in the section titled **Risk Premium**, the expected risk premium is usually based on some empirical study of long term historical rates of return with typical values presented being in the order of 6 to 8 per cent per annum. Given that there is 2 per cent variation in generally accepted values for the market risk premium, it is suggested that estimating beta to beyond two significant figures is attempting to incorporate a degree of precision which just isn't there.

Determination of beta risk for vertically integrated entities operating in different markets with different contestability/competitive conditions

In the section titled **Factors Affecting The Derivation Of Beta Risk And The Feasibility Of International Benchmarking Of Beta Risk** we stated that the beta of a company is a measure of its systematic risk relative to the systematic risk of the market as a whole. A number of complications may arise in relation to companies which operate in more than one line of business. One issue that needs to be addressed is whether the beta for such a multi-business company can be applied to an individual business unit of the company. The answer lay in comparing the systematic risk of the

---

[13] Usually, the most critical of these is a difference in leverage.

[14] In a valuation context, Lonergan, 1994, p. 27 notes that the use of comparable listed companies for the valuation of unquoted shares has been endorsed by the courts.

individual business unit to the systematic risk of the company as a whole. In general if the business unit is just as risky as the company then the beta of the business unit will be the same as the beta of the company. However if the business unit is more or less risky then it would be inappropriate to use the company's beta as an estimate of the beta of the business unit.

A second albeit related issue concerns determining the beta for a multi-business company. If we regard a multi-business company as being equivalent to a portfolio of its underlying business units then the systematic risk of the company reflects the average systematic risk of all its business units. Since the beta of a portfolio is equal to a weighted average of the betas of the individual companies in the portfolio, then similarly the beta of a multi-business company is equivalent to the weighted average of the betas of its underlying business units. Algebraically, for a company consisting of k business units:

$$\beta_e = \omega_1\beta_1 + \omega_2\beta_2 + \ldots + \omega_k\beta_k \quad (9)$$

where for i = 1,2, ... ,k:

β_e is the beta of the company;

β_i is the beta of the ith business unit; and

ω_i is the weight applied to the ith business unit.

In theory, ω_i is a market value weight representing the equity value of the ith business unit relative to the equity value of the company. In practice, however, there may be difficulties in not only determining the equity values of the individual business units but also in actually identifying the actual business units as distinct and separate operations[15]. This leads to a further issue which specifically relates to the subject of this paper. If the access provider is a single purpose company which does not compete in any other market, then the position is quite clear: the relevant beta to be estimated is the company beta. However if the access provider is a vertically integrated entity operating in different markets with different contestability/competitive conditions, an issue arises as to whether it is the beta of the entire entity or just the beta of the business unit which operates the essential facility that is required[16]. In either case, the appropriate principles outlined in this section may be used. It is considered worthwhile stressing that beta relates only to the company's systematic risk and not its total risk.

---

[15] Empirical support comes from Fuller and Kerr, 1981, who have shown that the beta for a multi-business company is equal to a weighted average of the betas of its business units. In this case however the weights were based on business unit sales rather than business unit equity values.

[16] The Trade Practices Act states that the matters to be taken into account by the ACCC in making a determination include the legitimate business interests of the provider and the provider's investment in the facility and the direct costs of providing access to the service (s. 44X).

Price regulation, competition and beta

If a company is subject to price regulation then its systematic risk may be different to the systematic risk of an otherwise equivalent unregulated company. The specific impact of price regulation on a company's beta is dependent on the nature of the regulation. In general if the price regulation serves to reduce the impact of market risk on the company's equity returns (such as from transferring risk from the company's shareholders to its customers) then the company's beta should be lower than it otherwise would in the absence of the regulation. Similarly if the price regulation serves to increase the impact of market risk on the company's equity returns then this should result in a higher beta compared to an otherwise equivalent unregulated firm. As any subsequent increase in competitive conditions is likely to expose the company's equity investors to greater market risk then the company's beta should be closer to that which would prevail if the company was completely unregulated. Again it is stressed that in considering the impact of regulation on the company's risk, the relevant measure to focus on is the company's systematic risk and not its total risk.

The CAPM, dividend imputation, leverage and betas

Vigorous debate has occurred in Australia over the appropriate form of the CAPM following the introduction of the dividend imputation tax system. Part of the debate has reflected differences of opinion (and confusion) about the best way to measure rates of return, whereas part has reflected the ultimately empirical (but difficult) issue of the value investors place on imputation tax credits. The problems are best seen by initially examining the standard CAPM equation for a classical tax system, given by equation (8) and repeated here:

$$k_e = r_f + \beta(r_m - r_f) \qquad (8)$$

Recall that the required return on equity of a company exceeds the risk free interest rate by a risk factor equal to the market risk premium multiplied by the company's beta. Note that under a classical tax system the CAPM equation given above is for rates of return calculated on an after company tax but before personal (investor) tax basis. In principle asset pricing is based on an equilibrium relationship after personal taxes. However, if all rates of return are subject to the same personal tax rate t_p then the after personal tax CAPM relationship derived from equation (8) is:

$$k_e(1 - t_p) = r_f(1 - t_p) + \beta(r_m(1 - t_p) - r_f(1 - t_p)) \qquad (10)$$

which on cancelling out the common term $(1 - t_p)$ leaves the original equation (8). In practice, most classical tax systems have involved somewhat different personal tax treatment of equity income to interest income because of preferential treatment of capital gains. While the CAPM can be amended to allow for that difference, in practice this complication has been largely ignored.

Following the introduction in Australia of the imputation tax system in 1987 (and inclusion of capital gains in the income tax base in 1985) it is not possible to ignore such complications. Equity income is now formally subject to different personal tax treatment to interest income. Most importantly, equity income taking the form of franked dividends is 'grossed up' for the calculation of assessable income for tax purposes and a tax credit of equivalent dollar value given to the investor. This has had a number of consequences relevant for using the CAPM including:

- there are different ways in which equity returns can be calculated, depending upon at which point in the taxation process they are considered;
- the definition of cash flows used in any valuation process must be consistent with the way in which equity returns have been calculated;
- it is conceptually possible to define the company tax rate in (at least) two ways. One, used here, is to define it by reference to tax actually paid by companies. The other, advocated by Officer, 1994, is to define it using that part of tax paid by companies which is not offset by reduced personal tax payments due to imputation tax credits received;
- the value of imputation tax credits will differ between investors. Most Australian taxpayers will benefit from an equivalent reduction in tax payable and value $1 of imputation credits as being worth $1. Foreign taxpayers, in the absence of a market for the sale of imputation credits, may value them at zero[17]. This raises the question of what value should be placed on imputation credits in the derivation of the CAPM;
- equity income can take three forms (franked dividends, unfranked dividends, and capital gains) each subject to different tax treatment, introducing the complication of allowing for these differences; and
- while the CAPM takes the market risk premium, defined above as $r_m - r_f$, as an externally given parameter, this may have changed as a result of the introduction of imputation.

Terminology and measurement

Debate on the CAPM under imputation has intermingled two separable issues. The first concerns that of whether equity returns (and the cost of equity capital) should be measured in a way which adds in the value of imputation credits or not. The second concerns the value which investors place on imputation credits.

To examine these issues, consider for example a company with an initial and end of year share price of $10.00 which pays a franked dividend of $0.64 (when the company

[17] The value of imputation credits to a foreign shareholder is dependent upon the tax laws in the shareholder's home country.

tax is 36 per cent) to a shareholder on a personal tax rate of 40 per cent. It is possible to define the shareholder return in, at least, four different ways:

Definition	Calculation	Rate%	Comment
Fully grossed up	(dividend+credit)/price = (0.64 +0.36)/10.00	10.00	Relates to cash flows prior to company tax payments
Partially grossed up	(dividend +value of credit)/price = (0.64 + 0.5x0.36)/10.00	8.20	(Value of credit assumed to be 0.5) Relates to cash flows grossed up by some proportion (here 0.5) of company tax payments
Ungrossed	(dividend)/price = 0.64/10.00	6.40	Relates to cash flows after company tax payments
After investor tax	(dividend+credit)$(1-t_p)$/price = (0.64+0.36)(0.60)/10.00 =	6.00	Relates to cash flows after both company and personal tax payments

Conceptually, all of these definitions are useable, but each has different implications for the methods to be used in valuation and in formulating the CAPM. First, each relates to a different measure of cash flows. Second, the current relevance of historical values for the market risk premium (derived under a classical tax system) for each definition needs to be carefully considered.

If returns are measured on an 'ungrossed' basis, it is to be expected that the market risk premium will have declined following the introduction of imputation. This is most easily seen by examining the risk premium after personal tax, under the simplifying assumption that returns on the market take the form of fully franked dividends, post imputation. Assume that the risk premium was 8 per cent per annum prior to imputation, the risk free interest rate is unchanged at 6 per cent per annum, the company tax rate is unchanged at 36 per cent, and that the marginal investor's personal tax rate is 40 per cent. The after tax investor risk premium prior to imputation was then:

$$r_m(1-t_p)-r_f(1-t_p)=(r_m-r_f)(1-t_p)=0.08\times 0.6=4.8\% \quad (11)$$

Assume that the value of the after investor tax risk premium has remained unchanged since the introduction of imputation (and there are no strong arguments which have been advanced to suggest otherwise). Under imputation, the after tax risk premium is calculated as:

$$\frac{r_m(1-t_p)}{1-t_c}-r_f(1-t_p) \quad (12)$$

where the $1/(1-t_c)$ term reflects the grossing up of taxable income associated with the receipt of franking credits. Equating this expression to the historical after tax risk premium of 4.8 per cent we are able to calculate an implied value for r_m and thus for $r_m - r_f$ as follows:

$$\frac{r_m(1 - t_p)}{1 - t_c} - r_f(1 - t_p) = \frac{r_m(0.6)}{0.64} - 0.06(0.6) = 0.048$$

$$r_m = 0.0896 = 8.96\% \qquad (13)$$

i.e. the market risk premium (measured by $r_m - r_f$) has fallen to 2.96 per cent (although note that the 'grossed up' return on the market has remained at 8.96/0.64 = 14 per cent, equivalent to the pre imputation value, and the market risk premium using that definition of returns has remained at 8 per cent).

The import of this example is clear. If returns are measured without grossing up for franking credits, the market risk premium utilised in the CAPM should be lower. If returns are grossed up, the historical value for the market risk premium may still be appropriate for use in the CAPM where returns are measured on a 'grossed up' basis.

In practice, matters are complicated relative to this example by two issues. First, market returns and returns for individual stocks can comprise a mix of franked and unfranked dividends and capital gains. Second, franking credits may not be fully valued by some (notably foreign) investors. It is however possible to derive an estimate of the size of the market risk premium following the introduction of imputation by examining the composition of returns on the market portfolio[18].

The significance of franking credit valuation

Officer, 1994, and McKinsey & Company, 1994, have argued that franking credits may have a value to investors of less than their dollar amount. Underpinning this view is the argument that foreign investors may not be able to utilise franking credits to reduce tax in their home jurisdiction. Hence, they appear to suggest that returns should be 'partially' grossed up by the **value** of the franking credits to the investor. If a $1 franking credit has a 'value' of $0.40, ($\gamma$ = 0.4), a franked dividend of $64 with a franking credit of $36 would be regarded as having a value of $(64 + 36 x 0.4) = $ 78.40. Note that this approach requires that valuation of any cash flow series requires those cash flows to be similarly amended. For example, if a cash flow of $c is generated on which company tax is paid leaving $c(1-t) available for distribution as a franked dividend, the cash flow needs to be augmented by the value of the franking credit which is γtc. Hence the cash flow to be evaluated is $c(1-t(1-$\gamma$)).

The benefit of this approach is that **if the true value of γ can be established**, the historical value of the market risk premium can be used in the CAPM, and the standard CAPM used (with returns defined in this partially grossed up fashion). However, there are a number of problems with this approach. First, γ is not readily determined. Second, it is not clear whether γ is an economy wide measure or a measure applicable to individual companies. Third, in practice γ generally takes on values of

[18] See for example by Van Horne et. al., 1995.

approximately 0 or 1 for investors who are non-taxpayers or Australian taxpayers respectively. Some average value of ? can be calculated but has no theoretical justification for being included in an asset pricing equation such as the CAPM.

If the cost of equity capital is to be measured on an 'ungrossed' basis (i.e. using the cash value of dividends plus share price capital gains) as is common in international practice, it is necessary to recognise that the return required by shareholders will reflect the type of dividends paid. A return of 10 per cent per annum consisting entirely of a franked dividend, is more valuable to an Australian taxpayer than a 10 per cent per annum return consisting of an unfranked dividend. If it is assumed that franking credits are fully valued by investors, it is possible to derive expressions for the cost of equity capital for companies expected to provide returns taking different forms. Van Horne et. al., 1995, demonstrate that the CAPM takes the following forms:

- for returns in the form of franked dividends:

$$k_e^f = r_f(1 - t_c) + \beta(r_m - r_f(1 - t_c)) \qquad (14)$$

- for returns in the form of unfranked dividends:

$$k_e^u = r_f + \beta\left(\frac{r_m}{1 - t_c} - r_f\right) \qquad (15)$$

- for returns in the form of unfranked dividends and capital gains:

$$k_e^{u\&g} = \frac{r_f(1 - t_c)}{z} + \beta\left(\frac{r_m(1 - t_p)}{z(1 - t_c)} - \frac{r_f(1 - t_p)}{z}\right) \qquad (16)$$

where $z = (1 - D)t_p + Dt_g$, D is the proportion of returns in the form of unfranked dividends and t_g is the effective tax rate on capital gains. For companies of equivalent risk (i.e. equal ? 's),

$$k_e^f = k_e^{u\&g} = k_e^u \qquad (17)$$

i.e. investors will require the lowest rate of return on companies expected to provide returns in the form of franked dividends, and require lower returns from companies which provide some return as capital gains rather than as unfranked dividends.

A potential complication arises in the case where the owners of an Australian tax paying company are foreigners who are unable to utilise the franking credits generated. It might be argued that the required return for equity of such owners is higher than that prevailing in the Australian market, because of the lack of value of franking credits to foreigners. Consequently, prices should be set to allow the achievement of the required rate of return of the foreign owners. To the extent that there is no substantive difference between operational efficiency under domestic or foreign ownership, this

argument is not supported. In that case, the activity can be undertaken by domestic owners for whom the cost of equity is lower. There is no rationale for compensating foreign owners for their (tax induced) disadvantage by allowing prices to be set higher than would be the case for an Australian owner with a lower cost of equity. The tax disadvantage need not be borne for the activity to be undertaken efficiently, and there is thus no need to compensate owners who elect to take on that disadvantage.

In the case where the foreign owners have some particular skills or expertise which enable them to operate the activity with greater operational efficiency than domestic owners, the issue is more complex. If the skills or expertise are of a form such that they can be subcontracted from foreign suppliers without the necessity of foreigners providing the equity finance, there is again no apparent argument to justify pricing based on the higher cost of equity capital. However, if the skills and expertise are of a form where ownership is a necessary condition of supply (for example where it is necessary to protect commercial information), the situation is different. Even allowing for a higher cost of equity, the required price may be lower than if the activity were undertaken by the less efficient domestic owner. It would appear necessary to examine these matters on a case by case basis.

The cost of equity and the WACC

If the cost of equity capital is measured on an 'ungrossed' basis, then the conventional approach to measuring the WACC can be followed. The WACC is given by:

$$WACC = k_e \frac{E}{D + E} + k_d (1 - t_c) \frac{D}{D + E} \qquad (18)$$

where $k_d(1 - t_c)$ is the company's after tax cost of debt and D/E is the company's debt to equity ratio measured in market value terms. This discount rate is used to evaluate cash flows after company tax **where the cash flows and tax are estimated <u>as if</u> the company were unlevered**. This measure of cash flows is used because:

- the focus is upon the returns to all providers of capital (debt and equity) and thus needs to consider all cash flows available to them; and
- the company tax benefits gained by the company from the tax deductibility of interest is taken into account in the use of an after tax cost of debt and thus should not be considered in measuring cash flows.

Leverage and betas

A simple approach to the impact of leverage on the systematic risk of a company (as measured by its beta (β)) is to note that the required return can conceptually be broken up into three components:

Required return = risk free rate + business risk premium + financial risk premium

In this formulation, the risk free rate is that required by investors in risk free assets and can be thought of as an indication of pure time preference. For assets involving risky

returns, investors will require a premium over the risk free rate to compensate. The business risk premium measures this and can be thought of as that associated with an unlevered company with assets of a particular level of systematic risk. Since the company has no debt, the systematic risk of the assets is equal to the systematic risk of the company's equity. The business risk premium depends upon two factors, those being the market risk premium and the ? of the assets. Each must be estimated, and although most attention is commonly given to the accuracy of the estimate of ?, both are equally important. In fact, there are probably greater errors associated with mis-estimation of the market risk premium than with ?.

The financial risk premium reflects the impact of leverage on the systematic risk of the company. For a more highly leveraged firm, the volatility of asset returns (and systematic risk) becomes magnified when equity returns associated with those leveraged assets are considered. The precise nature of the relationship depends upon the tax system in place. In a world without taxes, the required return relationship is:

$$k_e ? r_f ? ?_u(r_m ? r_f) ? ?_u(r_m ? r_f)\,D/E \qquad (19)$$

where

$?_u(r_m ? r_f)$ is the business risk premium;

$?_u(r_m ? r_f)\,D/E$ is the financial risk premium; and

D/E is the company's debt to equity ratio measured in market value terms.

From equation (19) the beta of the leveraged company β_L is:

$$\beta_L = \beta_u (1 + D/E) \qquad (20)$$

In a classical tax world, the required return relationship is:

$$k_e = r_f + \beta_u (r_m - r_f) + \beta_u (r_m - r_f)(1 - t_c) D/E \qquad (21)$$

From equation (21) the beta of the leveraged company β_L is:

$$\beta_L = \beta_u \left(1 + \frac{D}{E}(1 - t_c)\right) \qquad (22)$$

Under the imputation tax system, the required return relationship if imputation credits are fully valued is:

$$k_e = r_f (1 - t_c) + \beta_u (r_m - r_f (1 - t_c)) + \beta_u (r_m - r_f (1 - t_c))(1 - t_c) D/E \qquad (23)$$

From equation (23) the beta of the leveraged company β_L is:

$$\beta_L = \beta_u (1 + D/E) \qquad (24)$$

In practice, the appropriate adjustment for leverage under the Australian tax system is likely to lie somewhere between the classical and pure imputation adjustments described above. That is:

$$\beta_L = \beta_u \left(1 + \frac{D}{E}(1 - \theta t_c)\right) \qquad (25)$$

where $0 < \theta < 1$. Unfortunately there is little empirical evidence on whether these adjustments apply in practice and thus there is little guidance on what value θ might take. However, not too much should be made of this, since the impact on required rates of return is relatively minor compared to other possible estimation errors in calculating required rates of return. For example, if $\beta_u = 1$, $D/E = 1$ and $t_c = 30\%$ then the effect on β_L of choosing different θ values is as given below:

θ	0	0.5	1
β_L	2	1.85	1.7

Since this is being applied to a market risk premium in the order of 5 per cent per annum, the difference between the case where θ=2 and θ=1.7 is equivalent to a rate of return difference of 10 per cent versus 8.5 per cent. While this is not insignificant,

estimates of the market risk premium (pre imputation) are typically presented as being of the order of between 6–8 per cent, i.e. having a greater margin for error. Clearly, on a theoretical basis a leverage adjustment should be undertaken, but concern over the precise nature of the tax adjustment probably involves concern of an unnecessary degree of precision.

Pricing principles and asset valuation methodologies

Beta, pricing principles and asset valuation

Once an acceptable cost of equity capital has been derived it may be appropriate to compare it to the company's economic rate of return defined as:

$$ERR = \frac{d + \Delta E}{E} \qquad (26)$$

where E is the market value of the company's equity and d is the value of dividends paid over the period. This in turn can be linked back to the company's cash flow, cf, by noting that:

$$E = MVA + G \qquad (27)$$

and

$$d + \Delta MVA = cf + ED \qquad (28)$$

where

MVA is the market value of assets in place, G is the value of growth opportunities and ED is economic depreciation (the decline in the market value of assets in place), so that:

$$ERR = \frac{cf + ED + \Delta G}{E} \qquad (29)$$

In practice, accounting data provides us with a measure of the accounting rate of return defined by:

$$ARR = \frac{d + RE}{BVA} = \frac{cf + AD}{BVA} \qquad (30)$$

where AD is accounting depreciation, BVA is book value of assets in place, and minor differences between accounting accruals and cash flows are ignored. Since accounting relationships will provide information on the links between product price charged and ARR, it is possible to impute an acceptable price once an acceptable ARR is determined. The problem in practice is that ARR and ERR can differ significantly, and the required ERR cannot be simply translated into a required ARR. The differences arise from several factors.

Growth opportunities

Stock market value and return on equity reflect the value (and change in value) of growth opportunities available to the company. These are the opportunities available to the company to invest in (and/or earn on current assets) a rate of return in excess of that required by investors. Where such opportunities exist, the stock market value can exceed the market value of assets in place. Consequently, the base value upon which a return is required differs between accounting and economic rate of return measures, even where assets in place are marked to market value. However, if the company in question can earn only the required rate of return, the two valuations should coincide — as the value of growth opportunities will be zero. Thus if the assets in place are marked to market value, there will be no difference between the bases of the rate of return calculations.

Non verifiable asset value

The problem referred to here is that where such assets do not have alternative uses from which a competitive market valuation can be obtained, their mark to market value cannot be determined independently of the prices which the company can charge. Thus, for example, if the pricing regime allows a stream of profits in perpetuity of $10 and the required rate of return is 10 per cent, the asset value is $100, whereas a profit stream of $20 would cause the asset value to be $200.

Depreciation

Divergence between accounting and economic depreciation creates two potential problems. First, because of the cumulation of differences, the book value of assets may differ from the mark to market value and thus affect the base on which returns are to be calculated. Second, in any period accounting income and economic income may differ.

The implication of the above discussion is as follows. Given the determination of an acceptable return on equity capital, the regulatory setting of prices is hampered by differences between accounting income and economic income, and by the problems of obtaining independent market values for assets in place. If asset values can be determined, then price determination can proceed by using economic depreciation rather than accounting depreciation in the determination of income[19].

Value added approaches

A fashionable approach to measuring corporate performance is to estimate value added by the company, based on an estimate of the return achieved over and above the cost of capital used by the company. Thus, for example, temporarily ignoring complications associated with accounting information:

$$\text{Value added} = [\text{return on investment} - \text{WACC}] \times \text{capital invested}.$$

[19] If prices are set with the objective of ensuring a return equal to the required return, there will be no discrepancy between stock market value and mark to market value of assets.

Such measures, it is argued, provide a single performance measure incorporating the effects of operating, investment, and financing decisions. In the current context, the relevance of such an approach is that a competitive industry would achieve a value added measure of zero since it would just achieve a return sufficient to compensate suppliers of capital for the risk borne.

In practice, there are various approaches to the measurement of value added, recognising the problems of converting accounting information into a form compatible with economic concepts. Stern-Stewart Inc. advocate a concept known as Economic Value Added (EVA) which involves adjustments to accounting data specific to the user in order to measure return on investment, and which uses a book value of invested capital. The Boston Consulting Group utilises a method of 'cash on cash', preferring to derive explicit estimates of cash flow generated and the initial cash flow commitment of investors. A.T. Kearney (Australian Financial Review, 26th August 1996, p. 14) use a similar approach, calculating measures of shareholder value creation and economic profitability.

In the case of access arrangements, where the facilities are most likely to involve significant depreciable assets which may have limited alternative uses, the adjustments required to accounting data are likely to be quite specific to the case involved, and potentially complex.

Nominal and real values of WACC

As a general rule, the valuation of an asset may be undertaken using either nominal or real data provided both the cashflows and discount rate are defined on a consistent basis. Specifically this means applying a real discount rate to cashflows expressed in real dollar terms or applying a nominal discount rate to cashflows expressed in nominal dollar terms. Because of difficulties associated with forecasting future inflation levels it is often easier to forecast cashflows in real terms. However there are two complications which should be kept in mind if real values are to be used. Firstly, because tax depreciation is based on nominal historical costs, distortions may occur in the calculation of depreciation tax shields. Secondly, similar distortions in tax liabilities may arise from the fact that the valuation of inventories is also based on nominal and not real costs.

Stranded assets and an appropriate rate of return

Where investments are undertaken and prove to be a failure, economic income can be determined by recognising the loss in market value when it occurs. This means that the market value of remaining assets will then exclude that asset, and reflect the market value of capital to shareholders. The negative return associated with the asset will be recorded in the period in which it occurs, depressing the return to shareholders below that expected.

The possibility of such negative returns occurring raises the question of whether 'fair pricing' needs to allow for such losses. For example, including the original (depreciated at accounting standards) value of stranded assets in the asset base on which the required return is applied would allow shareholders to recoup the loss over

time (by passing it on to customers). There is no general case for such an approach, since such losses can be thought of as reflecting unsystematic risk and thus not relevant for determining required rates of return. Over time, greater than anticipated returns on some assets can be expected to offset such losses. While the owners of the business may not be sufficiently diversified to make the loss associated with this risk irrelevant, from the perspective of the capital market as a whole such risks are diversifiable and thus not priced. If the owner chooses to be less than fully diversified, they are entitled to bear the upside and downside of such non-systematic risks, but cannot be expected to warrant a higher rate of return in a competitive market place. Only if the regulatory process distorts the non-systematic risk facing an undiversified owner, for example by inhibiting positive unsystematic returns while permitting negative outcomes might a case be made.

Changes in the WACC and value of assets in place

Regulatory determination of an appropriate price is further complicated by the fact that the WACC can change over time. Consequently, the WACC prevailing today may not have the same value as the WACC prevailing when an operational asset was first purchased. This raises the questions of whether the current or original WACC value should be used in regulatory determination of prices and how the asset value should be calculated as the base for application of the WACC.

A useful way to address this question is to ask whether owners of the assets should bear the risk of changes in the WACC. The answer to this is yes. Investment decisions involve undertaking activities which rearrange the timing of cash flows and one risk involved is that of market changes in the value of time (interest rates). An increase (decrease) in market interest rates leads to a decrease (increase) in the market value of an asset which has a set of future operating cash flows expected (and unaffected by the change in interest rates).

This does not mean that the future cash flows associated with the asset should change (or be allowed to change) when interest rates change. The risk borne by the owners of the asset is that the present (market) value of the predetermined (risky) future cash flows will change. Consequently, regulatory price determination should not be conducted in such a way as to lead to changes in allowed prices on production arising out of past investments when interest rates change.

Implementing an approach which achieves this outcome is difficult. If market (reproduction) values of assets were available the current WACC value could be applied to the asset base calculated in that way to determine allowable prices. Implemented correctly, this would lead to no changes in allowable prices from output arising from past investment decisions arising from a change in market interest rates and the WACC. Owners of the organisation would bear the risk of changes in discount rates via changes in the market value of assets in place and realised returns (incorporating asset value changes) would move inversely with unanticipated changes in discount rates. Using market (reproduction) values would also impose technological risk upon owners of the assets, which would seem appropriate. Consider, for example, a situation where a new discovery meant that existing assets were no longer the most efficient available. Their lower market value combined with current WACC would lead to a reduction in allowable prices reflecting the benefits of the new discovery.

Unfortunately, utilising market (reproduction) values involves considerable complexities when there is no market based estimates available and where those values depend upon discretionary determinations of future cash flows by the regulatory authorities. Nevertheless, the conceptual appeal of the approach suggests that it warrants continued attention as a method to be aimed for.

An alternative approach might be to utilise accounting values of asset value and to apply a required rate of return based on the WACC prevailing at the time the assets were purchased. This is conceptually simple (although it may require more from accounting systems than they can deliver) and has one advantage and one drawback. The advantage is that changes in the WACC will not lead to changes in allowed prices. Owners will still bear the risk arising from changes in discount rate, although these will not be recognised in the accounting rates of return used. The disadvantage is that this approach will shield owners from technological risk, unless asset valuations take account of such effects. Consider a situation where a new innovation significantly reduces the cost of production relative to using existing assets. Unless the asset valuation approach uses a (now lower) replacement cost value reflecting the new technology, the allowable price will not be changed and the benefits of the new technology not passed on to consumers. Since the objective of the regulatory pricing policy is to mimic to some degree a competitive outcome, this appears inappropriate and shield owners from risks which they can be reasonably expected to bear.

Bibliography

Ball, R. and J. Bowers, 1986, 'Shares, Bonds, Treasury Notes, Property Trusts and Inflation: Historical Returns and Risks', *Australian Journal of Management*, 11, December, 117–137.

Brealey, R. A. and S. C. Myers, *Principles of Corporate Finance*, 5th Edition, McGraw-Hill, 1996.

CCH Australia Limited, *Australian Trade Practices Reporter*.

Chen N., R. Roll and S. A. Ross, 1986, 'Economic Forces and the Stockmarket', *Journal of Business*, 59, 383–403.

Eltin, E. J., M. J. Gruber and J. Mei, 1994, 'Cost of Capital Using Arbitrage Pricing Theory: A Case Study of Nine New York Utilities', *Financial Markets, Institutions and Instruments*, 3, 3, 46–73.

Fama, E. F. and K. R. French, 1993, 'Common Risk Factors in the Returns on Stocks and Bonds', *Journal of Financial Economics*, 33, 3–56.

Fama, E. F. and K. R. French, 1996, 'Multifactor Explanations of Asset Pricing Anomalies', *Journal of Finance*, 51, 55–84.

Fuller, R. J. and H. S. Kerr, 1981, 'Estimating the Divisional Cost of Capital: An Analysis of the Pure-Play Technique', *Journal of Finance*, 36, 997–1009.

Hickman, K. and G. H. Petry, 1990, 'A Comparison of Stock price Predictions Using Court Accepted Formulas, Dividend Discount and PE Models', *Financial Management*, Summer, 76–87.

Ibbotson Associates, *Stocks, Bonds, Bills and Inflation*, 1991 Yearbook, Chicago, 1991.

Independent Committee of Inquiry, *National Competition Policy*, August 1993.

Kolbe, A. L., J. A. Read Jr. and G.R. Hall, 1986, *The Cost of Capital: Estimating the Rate of Return for Public Utilities*, MIT Press.

Lonergan, W., 1994, *The Valuation of Businesses, Shares and Other Equity*, 2nd Edition, Longman Australia.

McKinsey & Company, 1994, *Capturing Value from Dividend Imputation*, McKinsey & Company Inc.

Myers, S. C., 1972, 'The Application of Finance Theory to Public Utility Rate Cases', *Bell Journal of Economics and Management Science*, 3, 58–97.

Myers, S. C. and L. S. Borucki, 1994, 'Discounted Cashflow Estimates of the Cost of Equity Capital — A Case Study', *Financial Markets, Institutions and Instruments*, 3, 3, 9–45.

Officer, R. R., 1989, 'Rates of Return to Shares, Bond Yields and Inflation Rates: An Historical Perspective', in *Share Markets and Portfolio Theory*, R. Ball, P. Brown, F. J. Finn and R. R. Officer (eds), 2nd edition, University of Queensland Press.

Ross, S. A., 1976, 'The Arbitrage Theory of Capital Asset Pricing', *Journal of Economic Theory*, 13, 341–360.

Van Horne, J., J. Wachowicz Jr., K. Davis and M. Lawriwsky, 1995, *Financial Management and Policy in Australia*, 4th Edition, Prentice Hall Australia.

Section V. Non-price issues

Environmental externalities, congestion and quality under regulation

Peter Forsyth, Monash University, Melbourne, Australia.

Introduction

In is becoming increasingly recognised that problems of quality degradation, congestion and environmental externalities can be present, and often exacerbated, when industries are subjected to price regulation. A price regulated firm has an incentive to downgrade quality of its output, since it can save on costs, but need not lose much revenue, if it lowers quality For example, service quality has become an issue with privatised, regulated, electricity distributors. Congestion is an aspect of quality, and regulated firms often operate congestible facilities. This is an issue for busy airports which are subject to price regulation. Congestion also involves an element of externality, since one user imposes costs on others through lowering the quality of service they receive. As with other industries, regulated industries can generate negative environmental externalities. This could pose difficulties for regulating electricity industries if greenhouse gas emission limits are imposed.

It is convenient to begin with the problem of quality reduction under regulation; how this comes about is discussed, and options for addressing it are examined. Congestion is best treated as an aspect of quality. Finally, the control of externalities in the context of the regulated firm are discussed.

Quality under regulation

Imposing price controls, of the price-cap form, on a firm will give it an incentive to undersupply quality (Rovizzi and Thompson, 1992, Armstrong, Cowan and Vickers, 1994, pp 173–74). The mechanism is quite straightforward. The firm can gain additional revenue by adding to quality;it will also face additional costs. The profit maximising firm will equate the additional revenue from additional quality to the additional cost. However, the additional social benefits from additional quality will include the addition to consumer's surplus, and will exceed the addition to revenue. Hence the firm subject to a price control will undersupply quality.

An unconstrained monopoly would not face the same incentives. If it increased quality, it would face an upward shift in its demand curve, and it would be able to increase its price, with a consequent impact on profits. It depends on the circumstances whether a monopoly will choose the optimal level of quality (Spence, 1975); it could choose a higher or lower level, though its choice will be around the optimal level. Price control limits the ability of the firm to gain full advantage from its quality choices by stopping it from charging more when it delivers higher quality.

This situation can be compared to that of a rate of return regulated firm. This firm can pass on the costs of higher quality, and thus does not have an incentive to skimp on quality. It too does not have an incentive to choose the level of quality that consumers are willing to pay for, and depending on the nature of the regulation, it may have an incentive to oversupply quality, especially if the allowable rate of return exceeds the firm's cost of capital. Higher quality may necessitate a larger capital base, and thus higher absolute level of profit. Rate of return regulated firms are often accused of 'gold plating', and supplying a higher level of quality than the consumers require. The publicly owned firm has ambiguous incentives towards quality. It may wish to maximise size of operations, and this might be achieved by supplying too high a level of quality. It might be under pressure from its owners to minimise (politically inconvenient) risks, and may play safe by providing too high a quality. On the other hand, public firms may not be responsive to consumers, and may take an easy way out by supplying too low a level of quality. While public firms face some pressure to take quality and its costs into account, and many strive to provide the quality which consumers want, there is no reliable mechanism to ensure that owners and managers seek the ideal quality/cost solution.

Congestion

Congestion can be regarded as an aspect of quality which depends on the relationship of demand to capacity. As demand grows relative to capacity, the quality of the service received by the users declines; they may face delays in obtaining service, the service may become less predictable, or facilities may be overcrowded. The costs of congestion are borne by the users, and the users are in the hands of the facility provider to supply sufficient capacity to enable an efficient level of quality. When congestion mounts, it can be lessened by pricing to reduce usage, by non-price rationing (e.g. slot controls at airports), or investment to increase the capacity of the facility. In a situation of growth of demand, there will come a point where the last option, which is the most expensive, will be warranted.

A price regulated utility cannot use price to lessen congestion. It could implement non-price methods of rationing demand, but it does not have an incentive to do this; it will gain additional revenue and profit if it increases throughput, and it will have an incentive to allow extra demand even though users face a high cost in terms of congestion. It will not have an incentive to invest to lessen congestion, since it will have to pay for the investment, but its customers will enjoy the benefits of reduced congestion, and it will be unable to recover the costs of investment via higher prices. Thus a price regulated utility will allow congestion to be too high, and will under-invest in capacity (for an application to airports, see Forsyth, 1997). Unconstrained profit maximising firms, by contrast, will take congestion into account because they will be able to charge users more if they supply a less congested service.

Access Pricing

The possibility of quality degradation when access prices are regulated is issue which has not been much attention, though its consequences could be considerable. In Australia there has been a shift towards breaking up hitherto monolithic utilities, and making it possible for competitors to be able to use essential facilities owned by the incumbent monopolist by setting up an access regulatory regime (King and Maddock,

1996). A utility which is required to provide access to its competitors will face the same sorts of incentives, if access prices are regulated, as would a utility which is selling in the final goods and services market. It will be able to lower costs if it reduces quality, and it will be unable to profit if it increases quality since it will not be permitted to charge more. Thus it will have an incentive to supply an inefficiently low level of quality to its competitors.

However, there is a further effect which could prove even more serious in the access situation. A facility provider will not only save costs by supplying low quality services to its competitors, but it will also be able to lessen their effectiveness as competitors. At the extreme, it could supply such low quality services that its competitors were unable to compete, and its monopoly in the final goods and services market would be restored, even though it is required to provide access. By supplying low quality services, the utility either increases the costs of its competitors, or lowers the quality of the services they are unable to sell, or both; whatever happens, the utility will be able to sell its output at a higher price than if it were forced to supply its competitors with the same quality of service that it has been supplying to itself. The utility can use quality degradation as a device to partly or fully get around the access price regulation.

If the utility is not price regulated, and is allowed to negotiate with access seekers on access price, it will not have this incentive to lower quality. It will only sell to its competitors at prices which are profitable to itself. It will either not lose if competitors are selling comparable quality services, and it can gain from selling to them if they are more efficient than itself in producing the downstream service. The utility will not have an incentive to lower quality if an access price set at the Efficient Components Pricing Rule (ECPR) level, (see Baumol and Sidak, 1994, for a discussion of ECPR) since it will not lose if a competitor replaces it to some extent in the final goods and services market.

It is difficult to obtain empirical evidence on how serious a problem this is likely to be. Access seekers often complain that the standard of service they are being supplied with is lower than that supplied by the facility owner to its own divisions, and that their ability to compete is thus restricted. It has also been maintained that there are fewer disputes over quality of service in the New Zealand telecommunications industry, where access prices are set by negotiation (with some government pressure), than telecommunications industries where access prices are regulated.

Valuing quality

If the quality problems identified in the preceding sub-sections are to be addressed, the regulator needs to know the value the users put on quality. When there is no regulation, firms solve the quality choice themselves; perhaps by a method of trial and error, they find out what the market is prepared to pay for. This method is no longer available, since the nexus between prices, revenues and quality is now broken. The regulator must use other information to try to determine what quality level to impose.

Typically, there will be an attempt to identify what physical aspects of quality are important; depending on the nature of the utility, there will be a range of quality indicators which are relevant. Some of these may have been collected for some time, perhaps well before corporatisation or privatisation of the utility. However, because

quality issues become more critical under new arrangements, it is often the case that a whole new range of indicators will need to be developed. These will cover the nature of the product, such as impurities in water, delays in obtaining service, as with airport delays, delays in getting connections re-established, as in telecommunications, service reliability, as in electricity, and safety aspects. It is usually fairly easy to determine which are indicators are most appropriate for a particular utility; the main problem is that the list is likely to be a long one.

The next stage is to determine what trade-offs users are prepared to put on different aspects of quality and price. In some cases, users themselves may be poorly informed about an aspect of quality, and may not be able to choose optimally. For example, users are typically not informed about the disease risks in the water they drink; they just presume that it is safe. Where externalities may be linked to quality, for example, there may be public health aspects to water quality, users preferences will need to be supplemented with information about the costs of the externalities. For the most part, the regulator's task is to determine what trade-offs users would be prepared to make.

One option is to assume that existing levels of quality are appropriate, and try to ensure that these are maintained. This may be an inappropriate assumption where there the utility had no strong incentive to offer the optimal level of quality previously; as noted before, public or regulated firms can have incentives to supply excessive or deficient quality. Another option is to use yardsticks; to find out the quality levels comparable utilities are offering, and to use these as the standards. Regulators which are doubtful about whether existing quality levels are appropriate can undertake surveys of users, to try to determine their willingness to pay for different quality levels. Finally, there may be some synthetic means of estimating what users are prepared to pay for some aspects of quality. For example, it is possible to make estimates of the cost of delays to aircraft caused by congestion at airports using measures of the value of time of passengers and aircraft operating costs. None of these methods is perfect, but it should be borne in mind that unconstrained firms also face imperfect information on what users are prepared to pay for quality, and thus their decisions can be subject to a degree of error as well.

Regulating quality

Once the regulator has decided what values to put on different aspects of quality, there are several ways in which it might impose its preferences. A simple means is to have quality standards. Various indicators are developed, and particular standards are established. Data on performance are collected regularly, and these are compared to the standards. Simple monitoring may sometimes be sufficient to induce the utility to maintain adequate standards. Often this will be insufficient, and standards will have to be enforced. Fines might be imposed for failure to achieve standards. Conformity with standards may be made a condition for renewal of the licence to operate. An approach which is gaining in popularity is the contract with the customer, whereby the utility contracts to supply the customer with a stated standard of service, and is required to pay the customer specified amounts should it fail to meet the standards. These methods have the advantage of practicality, though they do involve the regulator in becoming much more directly involved in the operations of the utility; this is not 'light handed' regulation.

Another approach is to try to give the utility an incentive to choose the optimal level of quality by including quality directly in the price-cap. This is done by taking some indicators of quality, and by varying the allowable price according to performance in the quality indicators; a utility which increases quality will be allowed a higher price. This approach is not often followed in its pure form, though approximations to it exist. In the UK water utilities were allowed a higher price-cap in recognition of investments they would be making to improve water quality. Also in the UK, airports have been allowed higher price-caps on condition that they made investments which lessened congestion for aircraft and passengers (Forsyth, 1997).

It may well be that some forms of quality are so important that they warrant specific regulation by a body separate from the regulator. For example, airport congestion can be alleviated by restriction use of the airport at busy times, and by allocation slots for use in some manner; these slots could be given to existing users according to some rule, or they could be auctioned. Whichever approach is adopted, there will need to be some body charged with the slot management of the airport. Such a body will require skills quite separate form those of price regulating, though it will need to operate in close cooperation with the price regulator.

Externalities and regulation

The nature of the problem posed by externalities is well known. When externalities are associated with a production process, they involve either a cost or a benefit to the economy at large, but not to the producer. The producer does not take them into account when determining how much to produce, and thus either produces too much, when the externality is a cost, or too little, when the externality is a benefit. To correct the externality, it is necessary to induce the producer to alter its production, perhaps by quantitative controls or by imposing a tax or subsidy related to the externality. When the producer is subject to regulation, such as price-cap regulation, there are a number of issues which arise concerning how externalities can be controlled within this environment. These are considered in this section.

Another well known problem which is common to all externality problems concerns valuation of the externalities, or establishing what trade offs with other variables are appropriate. To reach an efficient solution, it is necessary to put a value on the externality. Since, by its very nature, the externality is not something traded in a market, there will be no established market valuation for it. A range of techniques for estimating the value of externalities have been developed; these include estimating the implicit valuation through hedonic means (e.g. measuring the cost of aircraft noise from house prices), contingent valuation methods (such as surveys of values people put on environmental features), or assessments of the cost of eliminating the externality. These will not be explored here; it suffices to note that all of the problems of establishing a value for externalities are normally present when regulated firms are the generators of externalities.

Techniques of control

There are several ways in which externalities are controlled. Three general approaches, which encompass a wide range of controls, are arbitrary standards, quantity constraints and taxes.

Arbitrary standards and prohibitions are simple, and often effective means of controlling externalities, though they are blunt in their effects, and do not usually result in an optimal level of output being achieved. Some examples of these include airport night curfews, requirements to put cables underground and bans on effluent discharges. Several of these are all-or-nothing controls, and there need be no attempt to make it possible for the firm to create the externality if it is prepared to pay for doing so. This type of control can be enforced by fines for not conforming, or by refusing licences to produce if conditions are not met.

A second type of control is the quantitative limit. A firm or industry is permitted to create externalities up to a defined limit; beyond this it is not allowed to create any more externalities. This limit may be enforced through heavy fines or licence conditions. Examples might include limits on emissions of air impurities, limits on emissions of greenhouse gases and limits on wastewater put into a river. In principle, the level of the limits would be set optimally, such that the gains to the producer from an additional unit of externality created are equal to the additional environmental costs imposed. In some cases, the limits may be determined by a jurisdiction higher than that of the regulator; for example, allowable greenhouse gas emissions may be set by international negotiation. Quantitative limits may be set directly for a specific firm, or they may be set for all producers of the external effect which is to be controlled.

Even if the level of the externality has been set optimally, there need be no certainty that the rights to create the externality will have been allocated optimally between the producers. For example, if all producers are required to cut back the externality by equal amounts or in equal proportions, this may impose significantly higher costs on some producers than others. A means to minimise the costs of meeting the constraint is to have tradeable permits to create the externality (Wills, 1997, Chapter 15). Permits to create the externality which add up in total to the allowable level of the externality are allocated, in some way, to the producers; once allocated, these can be traded amongst the producers such that those which suffer the highest costs in reducing the externality can purchase more permits, thus minimising the cost of achieving the overall target. If there is no means for minimising the costs of eternality reduction such as by tradeable permits, the imposition of quantity constraints can result in two sources of inefficiency; through the overall quantitative limit being set too high or low, and through the costs of achieving the constraint being higher than necessary.

A third method of control is to impose a tax on the production of the externality; ideally this tax would equal the marginal cost of the externality. This is the Pigovian approach, and it is often taken by economists as the preferred option. It has the advantage that it will induce the producer to choose the most efficient level of production, taking into account private and external costs. An example of this type of tax would include a noise charge for aircraft; if charges depend on the noisiness of the aircraft, the airline would take the noise generated into account as a cost of operation, and it would have an incentive to substitute quiet for noisy aircraft.

Prices versus quantities

The normal preference of economists is for price methods of intervening in markets to correct for market failures; in this case, the price method would involve setting a tax equal to the marginal cost of the externality. When certainty is present, such a tax would result in an efficient level of the externality, and minimisation of the costs of not producing it. However it need not be the case that price methods are best when uncertainty is present; even a tax set at the optimal level might perform worse than a quantitative limit (Weitzman, 1974). This presumes that there is an efficient allocation of the quantitative limits amongst producers; this may not always be the case.

The preference for price or quantitative methods of control depends on how uncertainty impacts on demand and supply, and on the shape of the relevant curves. Looked at another way, it depends on whether the marginal cost of the externality or the optimal level of the externality are better known.

A price method of control would be superior if the marginal cost of the externality is constant and known fairly accurately. In this case, uncertainty would mainly impact on demand; which could be high or low. In either case, it does not matter how high demand is, since the marginal cost of the externality depends little on the amount of the externality created. A tax set equal to the marginal cost of the externality will induce firms to make efficient production decisions, by effectively internalising the externality.

By contrast, suppose that demand is uncertain, and the marginal cost of the externality is also quite uncertain. This could be because it depends on how much of the externality is created; suppose that the marginal cost is low up to a point, but if more of the externality is created, the total and marginal cost rises very rapidly. In this situation, there is an effective natural limit to the externality that can be created. The objective of the control is to induce the producers to choose to create the efficient level of the externality. If a quantity limit is set at or about the critical level of the externality, beyond which it becomes very costly, the producers will choose to create the efficient level of the externality.

If a tax were imposed, it could well get matters badly wrong. Suppose a low tax were levied; if demand were high, there would be far too much externality created Alternatively, a high tax would excessively discourage production when demand turns out to be low. Even a tax which is optimally designed to take account of uncertainty and which maximises expected net benefits will be too high sometimes and dangerously low at other times. A quantitative limit will ensure that production is as close as practical to the optimum.

The problem with the quantitative limit is that it may not be allocated efficiently amongst producers. A limit might be of the form of an allowable total amount of pollution to be allocated amongst a number of firms. These rights to pollute will be valuable. If they are allocated on an arbitrary basis, such as pollution emissions by firms in the past, the cost of conforming to the limit will probably not be minimised. If the rights are tradeable, and an active market develops, they will be allocated so that the costs of conforming to the overall limit is minimised. For efficiency purposes, it is important that rights be traded and firms adjust to the costs of the externality; the rights

can be allocated initially using several mechanisms, with different implications for distribution. Giving the rights to producers will be in their interests, but auctions of the rights would result in the revenues going to the government.

There are likely to be real situations where the quantitative approach is preferable to the price approach. One of these would be where there is a requirement to reduce greenhouse gas emissions. Suppose that Australia commits itself to specific greenhouse gas emission targets. The electricity industry is one of the main generators of these emissions, but other sectors, such as transport contribute. Different types of electricity generators contribute to emissions differently; brown coal contributes the most per unit of electricity produced, black coal the next, gas rather less, and hydro not at all.

The response of the electricity and transport industries to a carbon tax is not known; thus it is very difficult to set a carbon tax at the level which will meet the target. If an overall limit is established and enforced, and tradeable rights to create emissions are established, the target can be met exactly. A market for the rights will develop, and the price of the rights will be established. Electricity producers, along with other creators of gas emissions, will adjust to the cost of creating emissions, and this will lead to shifts within the industry. For example, there will be a greater reliance on gas and black coal at the expense of brown coal. There will also be trading in rights between electricity and other industries. How individual firms are affected will depend on how the rights are allocated; brown coal producers may gain, even though they reduce their own output, if rights are given to existing producers rather than auctioned.

As long as an efficient market in rights comes about, a system of quantitative limits and tradeable permits to emit are likely to be the most efficient practical means of achieving gas reduction targets. Price solutions, such as carbon taxes, will not be as effective achieving the targets. Whichever approach is used, there will be issues posed for regulators.

Price regulation with externalities

The issue for the regulator is how to allow for the cost of externality controls in the price regulation framework. Externality costs will raise costs to the firm; it will have to pay externality taxes, it will have to pay for tradeable permits purchased, and it may adjust its production process to lessen externalities created, though this will be at some cost to itself (for example, a brown coal electricity generator may switch some of its production to gas, but at a cost). There is also the possibility that the regulated firm may gain from the externality controls, if it is allocated more tradeable permits than it needs for itself, and it sells them.

If the system of externality controls is in place before regulation commences, it may be straightforward to allow for. If price-caps are being used, the initial price-caps can be set at a level which allows the firm to earn a profit under current externality controls, the costs of which are known, and to which the firm has adapted. Problems emerge when externality controls are altered, or introduced for the first time. This need not pose a problem if there is rate of return regulation; the costs of the externality controls will be an allowable cost for the firm which will be permitted to pass them on to users, whatever their level is. However, price-caps are now a preferred method of regulation,

and there are difficulties in handling externality charges when these are in place. The price-cap will have to be altered; this may happen when the 'X' factor is subject to its periodic review, but if the externality charges are large and suddenly imposed, it may be necessary to have a special review.

One option is to allow full pass through of externality charges. This, however, would create very poor incentives; the firm would have no incentive to reduce its creation of externalities, since it can pass on all of the charges associated with them to users. An efficient response to the imposition of externality would normally involve the firm reducing the externality, though at the expense of some increases in cost of production, such that the total of the two is minimised.

Another option is for the regulator to make a forecast of the likely cost increase imposed on the firm created by the externality control, and to adjust the price-cap so as to enable the firm to remain profitable. This would involve the regulator in estimating the minimum cost response by the firm to the control; the maximum cost to the firm would be that if it continues the previous level of externality creation, and does not adjust. In many cases, the firm will be able to adjust, especially over the long run, and reduce the cost of meeting the externality controls. It is likely, in many cases, the regulator will not be able to make an accurate estimate of the cost increase; it will not know how the firm will adjust, and it may not know the unit cost of the externality (for example, if there are to be tradeable permits, it will not know what these will sell at). If it chooses this approach, there is the possibility that the regulator will impose a price-cap which enables the firm to make large profits or forces it into losses. There will be some opportunity for the regulator to revise its price-cap in the light of additional information; such as when trading in externality permits becomes established and prices are known, or when substitution possibilities open to firms become clearer. However, price-caps can impose considerable risks on the firm.

An alternative approach, which may impose less risk on the regulated firm, would be to allow partial pass through of the externality cost. This gives the firm some incentive to reduce costs through reducing the externalities it creates, but protects it somewhat against the risks if externality costs turn out to be very high. This would be an example of moving the regulatory system towards one of mixed or sliding scale regulation, under which allowable prices are set partially with reference to the firm's actual costs, and away from pure price-caps (see Mayer and Vickers, 1996 and Crew and Kleindorfer, 1996). This lessens the risks to the firm but also lessens its incentives to minimise costs.

In some cases, the regulator may be able to resolve the problem by shifting the externality charge on to the users, rather than the regulated firm. This would be appropriate if it is the users who are the primary creators of the externality, and not the regulated firm; if so, the users would be able to alter their behaviour to lessen the overall cost. Thus airports are associated with aircraft noise; however, the airport itself has little control over the amounts of the externality created. Noise charges can be levied directly on airlines, and if properly structured (with lower charges for quiet than noisy aircraft), they will have an incentive to minimise the overall costs by substituting quiet for noisy aircraft. This is happening at Sydney airport, where differential charges levied on airlines are giving them an incentive to lessen aircraft noise (Forsyth, 1998).

Except in cases where it is feasible to handle the problem in this way, there is no simple solution to the problem of incorporating externality controls into price-cap regulation. This is so regardless of the form of the control; whether tax, quantity limit or arbitrary restriction. The problem is not great if there is little uncertainty about the cost of the externality control and how the firm can adjust to it. The problem is also not great if the externality costs are large as a proportion of the firm's total costs. However, there can be cases where the externality controls involve large and uncertain costs; this would be so if greenhouse gas emission controls are imposed on electricity generators. It is difficult to allow for large uncertain costs within a regulatory framework which avoids imposing too much risk on the firms but also preserves their incentives to minimise costs.

Separate regulators

It is quite likely that there will be separate economic (or price) and environmental regulators. The price regulator is not likely to have expertise in environmental matters, and the environmental regulator is not likely to have expertise in economic aspects of the industry's operation. Problems can develop as a result of this separation.

One obvious source of possible difficulty lies in the two regulators having different objectives. This will translate as different weights being put on to aspects of the firm's operations, and different trade offs for objectives. At the extreme, this could lead on regulator to be encouraging what the other is attempting to discourage. Inconsistencies can develop, and these can make it difficult for the firm to meet all of its requirements simultaneously.

At a more practical level, problems can develop as a result of having separated regulation, and what is essentially a two stage regulatory process. Ideally, the regulated firm should be induced to produce and price at a level which maximises overall net benefits, balancing environmental as against other costs. However, a two stage process involves one regulator setting its conditions, and then another. The environmental regulator may not have a good idea of the costs of meeting its requirements when it is setting them, and the economic regulator may not have a good idea of the environmental consequences of its rules. Thus, even if the two regulators are agreed about objectives and trade offs, their separate actions may make it difficult to achieve an overall efficient solution; there is no body charged with overall optimisation.

As an example, consider airport noise. The airport may be subject to a price regulator and to an environmental regulator concerned about noise. The price regulator may impose rules which encourage use of the airport in the off peak; which may be a time when noise costs are high. The environmental regulator may change flight paths to lessen noise; this may reduce the capacity of the airport, and result in delays and necessitate investment, which will add to the cost base of the airport. The price regulator will need to revise its controls in the light of this. Neither regulator is attempting to determine the best solution overall; both are solving part of the problem. Good communications between the regulators will help, though neither one is able to take all aspects into account, and to implement a solution which is the best feasible overall.

Conclusions

Problems of quality choice, congestion and eternalities exist with the utility and transport industries, but these become acute when these industries are subject to regulation, especially of the price-cap form.

In the case of quality, price-caps create an incentive for the firm to supply a less than optimal level of quality. This is especially a problem when access prices are being regulated. The firm being required to provide access at a regulated price can save on costs by lowering quality, but, in addition, it can make it more difficult for its rivals to compete, by supplying a lower level of quality than it does to itself. Congestion is essentially another aspect of quality; one which depends on the relationship of demand to capacity. The firm cannot convert reductions in congestion into higher revenue because its price is capped. Thus the firm has an incentive to provide too little capacity and allow congestion to be inefficiently high.

Arguably the most difficult problem posed for regulators is that of evaluating quality, to determine the appropriate level for the firm to provide: this is difficult because regulation breaks the nexus between price and quality. Once this is done, quality can be monitored, standards enforced, and incentives for providing appropriate quality levels can be built into the price-cap. This is sometimes done in an ad hoc way, with firms being offered easier price-caps if they invest to improve quality or reduce congestion.

Externalities are present with regulated utility and transport industries, and as always, they can be difficult to evaluate. Of the available methods of control, economists normally prefer price methods, such as Pigovian externality taxes, though under uncertainty, quantity limits can be preferable. Problems are created for regulators when firms are subjected to externality charges or must purchase permits. If full pass through is permitted of the costs to the firm of the charges or permits, it will face no incentive to minimise overall costs, including externality costs. If no pass through is allowed, and there is considerable uncertainty about the costs of the externality, very large risks would be imposed on the firm. Mixed systems of regulation, which permit partial pass through of the externality charges, may strike the best balance. A further problem for regulation is the likely existence of separate price and environmental regulators. Even if these have similar objectives and trade offs, there will be a problem of how their separate decisions can be coordinated to secure the best overall result.

Bibliography

Armstrong, M., S. Cowan and J. Vickers, 1994, *Regulatory Reform: Economic Analysis and British Experience*, Cambridge Mass., MIT Press.

Baumol, W. and G. Sidak, 1994, *Toward Competition in Local Telephony*, Cambridge, Mass., MIT Press.

Crew, M. and P. Kleindorfer, 1996, 'Incentive Regulation in the United Kingdom and the United States: Some Lessons', Journal of Regulatory Economics, May, pp 211–225.

Forsyth, P., 1997, 'Price Regulation of Airports: Principles with Australian Applications', Transportation Research-E, 33:4, pp 297–309.

King, S. and R. Maddock, 1996, *Unlocking the Infrastructure*, Sydney, Allen and Unwin.

Mayer, C. and J. Vickers, 1996, 'Profit Sharing Regulation: An Economic Appraisal', Fiscal Studies, February, pp 1–18.

Rovizzi, L. and D. Thompson, 1992, 'The Regulation of Product Quality in the Public Utilities and the Citizen's Charter', Fiscal Studies, 13(3):74–95.

Spence, M., 1975, 'Monopoly, Quality and Regulation', Bell Journal of Economics, 6:417–29.

Weitzman, 1975, 'Prices vs. Quantities', Review of Economic Studies, 41:477–91.

Wills, I., 1997, *Economics and the Environment : A Signalling and Incentives Approach*, St Leonards, Allen and Unwin.

Information disclosure in the Australian regulatory environment: legal and good practice aspects

Luke Woodward, General Counsel, Australian Competition and Consumer Commission.[1]

Introduction

It is obvious that decision makers need evidence upon which to base their decisions. Further, if you asked a decision maker whether, in making a decision, they wanted good evidence[2] or would accept inadequate evidence, the answer would be obvious. And you would expect a similar response if you asked the decision maker if they felt that they should make their decision for good reasons.[3]

But what then is good information?[4] How do you get it? How do you use it? What are good reasons? And what is the link between evidence and reasons?

These five questions are the focus of this session. There is no easy answer to these questions this paper's objectives are to outline the key issues, and to provide some tips.

It is probably useful to outline what this session does not cover. It doesn't cover specific types of information that may be required for specific regulatory functions such as the types of cost and revenue data to be collected under a record keeping rule.

In addition, it does not cover the information aspects behind the design of a regulatory regime. For example, this paper does not address the information based arguments for ring fencing, structural separation or accounting separation nor does it address the information arguments that would favour price caps over rate of return regulation.[5]

This paper assumes that the regulatory framework is given. However, this is not to ignore the critical importance of information considerations in regulatory design. Before working through the questions outlined above, it is worthwhile to recognise that the principles which govern efficient regulatory design will also be critical in a

---

1 This paper has been prepared with the assistance of Harriet Glen, Australian Competition and Consumer Commission (ACCC), Legal Unit.

2 'Good evidence' refers to evidence which is adequate and appropriate for the decision being made.

3 Similarly, 'good reasons' refers to reasons which are adequate and appropriate for the decision being made.

4 'Evidence' is a sub-set of 'information' and is defined as information which in a information which in a particular context tends to establish a relevant fact: See Byrne & Heydon, *Cross On Evidence Australian Edition* (Butterworths, NSW, 1991) para 1001. However, for the purposes of this paper, 'information' and 'evidence', are used interchangeably except where the distinction is significant.

5 See King & Maddock, *Unlocking the Infrastructure — The Reform of Public Utilities in Australia*, (Allen & Unwin, NSW, 1996), Chapter 4; Viscusi, Vernon & Harrington, *Economics of Regulation and Antitrust* (MIT Press, Cambridge, 2nd ed 1995) Chapters 11 & 12.

balanced and efficient implementation of a given regulatory scheme. These principles include:

a focus on the perceived problems and the related purpose of the proposed regulation;

consideration of alternatives for dealing with the problems; and

assessment of costs and benefits of those alternatives.[6]

There will always be trade offs between the cost of obtaining information (including the cost to the provider and the regulator) and the utility and value of that information. It is easy to err in getting too little at small cost or of getting more than is useful at a high cost. This should be at the forefront of the regulator's considerations when deciding whether and how to use information powers.

I. What is good information?

Information needs differ for different regulatory functions

An obvious but important starting point is that different regulatory functions have different information demands.

It is useful to characterise different types of regulatory functions and decisions. While there are a number of ways this can be done, the following characterisations are designed to highlight functions which are likely to raise discrete information issues.

Primary functions

Statutory exemptions and inclusions.

> This refers to threshold determinations such as ACCC authorisation decisions which operate to exempt particular conduct from a legislative prohibition; National Competition Council (NCC) access declarations; or ACCC acceptance of access undertakings which operate to oust the declaration/arbitration regime.[7]

Arbitration.

> Most of Australia's modern infrastructure regulators have a power to arbitrate over access disputes in relation to declared access services. For example, the ACCC has arbitration powers under Part IIIA (Access to Services) and Part XIC (Telecommunications Access Regime). The NSW Independant Pricing and Regulatory Tribunal (IPART) has (or will have) power to arbitrate access

---

6 See for example Australian Industry Commission, Regulation and its Review 1995–1996, (1996); Victorian, Law Reform Committee; Regulatory Efficiency Legislation (Discussion Paper, May 1997); Aust, Council of Australian Governments; Principles and Guidelines for National Standard Setting and Regulatory Action by Ministerial Councils and Standard Setting Bodies (1995?).

7 Trade Practices Act 1974 (Cth) ('TPA') Part IIIA.

disputes in relation to certain services in the electricity and gas industries.[8] The Queensland Competition Authority (QCA Act) has power to arbitrate access disputes in relation to services provided by (as yet to be specified) government monopoly and near monopoly business activities.[9] Similar arbitration functions are proposed for the ACT Independent Pricing and Regulatory Commission.

Enforcement.

The ACCC has primary enforcement powers for a breach of Part IV of the TPA (Restrictive Trade Practices) and Part XIB (Telecommunications Competition Rule). Most state regulators do not have primary enforcement functions although some have ancillary enforcement powers (such as non-compliance with a regulatory decision). For example, section 35 of the *Office of the Regulator-General Act 1994* ('ORG Act') gives the ORG a quasi-enforcement function as it has the power to issue orders in relation to non-compliance with a determination or licence conditions. A breach of these orders carries a substantial penalty.

Setting rates, terms and conditions and their review and monitoring.

This is the primary function of most state regulators (see for example Part 3 of the ORG Act). The ACCC has only limited powers in respect of surveillance and monitoring of prices under the *Prices Surveillance Act 1983* (Cth) and the *Airports Act 1996* (Cth). The ACCC does have some interim pricing powers in the telecommunications industry (*Telecommunications Act 1997* (Cth) s. 41) and a potential role in Ministerial determinations prescribing certain retail telephony prices.

Ancillary powers

Information gathering, record keeping rules and reporting requirements. There is a wide range of such powers across regulatory agencies and across functions within agencies. Many agencies have powers to compel the production of information relevant to a function or power (e.g. TPA s. 155; QCA Act s.185; and s. 40 of the proposed IPARC Act (ACT)). Express powers for record keeping and accounting separation requirements can be found, for example, in Part XIC of the TPA and Part 5 Division 9 of the QCA Act. Often these requirements may be imposed in licence conditions. The ACCC has also indicated that it would require some form of accounting separation to be included in arrangements proposed under a Part IIIA access undertaking.[10]

---

8 Pursuant to Pt 4A of the Independent Pricing and Regulatory Tribunal Act 1992 ('IPART Act').

9 Pursuant to Pt 5 Div 5 of the Queensland Competition Authority Act 1997 ('QCA Act').

10 Aust, ACCC, Access Regime, (November, 1995); Aust, ACCC, Draft Guide to Access Undertakings under Part IIIA, (1996).

Registration of agreements.

> Registration of agreements can perform dual policy goals, as a means of public policy oversight of significant private transactions, and as a means of capturing information relevant to other regulatory functions or powers or for public disclosure.

Information repository (such as public registers and tariff filings).

Each of these functions or powers raise different issues.

For example, one can contrast a statutory exemption/inclusion decision with an arbitration. In an arbitration, the role of the arbitrator is to resolve a dispute between parties. Hence information disclosure will be determined largely by that role. The decision maker may take a somewhat passive role as to the presentation of evidence as there will be two opposed parties seeking to put forward their best case. The role of the arbitrator is similar to a court which tries to ensure fair disclosure of relevant information between the parties so as to facilitate the proper presentation of each side's case. The arbitrator may be able to rely to a large degree on specific requests from the parties for disclosure. In the regulatory context, a key role of the arbitrator may be to balance information asymmetries.[11] This is the adversarial model.

An exemption/inclusion decision is not a simple adversarial process. In some cases there may be limited opposition to a particular application, but that is not determinative. This means that the decision maker is far more responsible for independently and actively considering what its (as opposed to the parties') information requirements are. This is the inquisitorial model.

Some regulatory arbitrations do have a special place for the decision maker in independently assessing public policy elements.[12] So in these cases, the decision maker approaches the arbitration in a more inquisitorial manner.

Another useful way to distinguish regulatory function is between functions which are adjudicative (e.g. deciding issues between parties such as an arbitration or a contested application for a declaration of access services) and those which are more administrative (e.g. a decision to issue a record keeping rule).

Decisions which have as their primary object, the conferral or denial of substantive rights are likely to be treated as adjudicative. Decisions which are merely ancillary to that, or have as their primary object, the assistance of the particular administrative functions of a regulator are more likely to be treated as administrative.

This distinction will be important in determining the extent to which, for example:

---

[11] For example, the ACCC has the power under TPA s. 152CT(2)(a), to give to a person who it considers has not been negotiating in good faith, a direction to give relevant information to another party: discussed in Asher, *Information Gathering Strategies* (Paper for Presentation at the Public Regulators Forum, 3rd July 1997).

[12] For example, TPA s. 44X in relation to Part IIIA arbitrations; TPA s. 152CR in relation to Pt. XIC (telecommunication) arbitrations; QCA Act s. 120; IPART Act s. 24B(3).

there will be a need to put adverse material to a party;

the decision maker can rely on information obtained outside the process or must contain itself to information and submissions arising in the course of the matter; and

the decision maker must, in general, adopt standards of procedural fairness.[13]

These requirements are less likely to be imposed for administrative functions than for adjudicative functions.

Another example is the difference between the arbitration and exemption/inclusion functions on the one hand and enforcement decisions on the other. As I will discuss later, a decision maker exercising a statutory administrative decision function will, within parameters, be free to determine the nature, degree and type of evidence it will rely on. In contrast, in exercising an enforcement function, where a decision is made to 'prosecute' a suspected contravention, information requirements will necessarily be set by legal rules as to evidence and standards of proof.

General observations

Whatever the function, good information is determined by what is necessary in order to be able to produce cogent and credible reasons for what is being done. Put another way, in the language of administrative law, the information should be sufficient to ensure that the decision is reasonable.

This in turn means that good information is information that is reliable, and, particularly where it is controversial, has been properly tested.

It should go without saying, good information must also be relevant. This is easily said, however in practice, determining what is relevant and what is not, and what is important and what is not, can be quite difficult and a matter for judgement.

Accordingly, in considering what is good information, it is important to begin with the end in mind. That is at the outset consider what information you need to gather to be able to produce good reasons. It will not be helpful, to say the least, to get to the end of an inquiry, begin preparing reasons and to find that a critical part of the puzzle is missing. However, ultimately there is no formula for determining what is good as opposed to inadequate information.

---

[13] The implications of these characterisations are both important and uncertain: see discussion in Aronson & Dyer, *Judicial Review of Administrative Actions* (Law Book Co, Sydney, 1996) pp 517–19, 532–33, 543–44 & 550–51.

II. How do you get information?

Powers to obtain information

Formal powers

The major categories of formal regulatory power to obtain information used in Australia to date are:

Disclosure in arbitrations.

There are two elements to information disclosure in arbitrations.

Firstly, the production of information to the arbitrator and its use and control by the arbitrator. Some regulatory regimes include quite detailed procedures for the production, use and disclosure of such information. For example, the QCA Act and to a lesser degree, the proposed IPARC Act. The IPART Act adopts the procedures under the NSW *Commercial Arbitration Act 1984*.

A related but important issue is the question of disclosure between the parties. Equal and open disclosure between parties to an arbitration is a key principle. It is also important in a regulatory context, to address information asymmetry problems. Likewise, information disclosure can facilitate negotiations. However, in practice, questions of confidentiality make this a difficult area.

Request for specified information in pursuit of another statutory function such as a pricing inquiry or investigation, enforcement investigation[14] or an exemption/ inclusion determination.

Tariff filings and record keeping rules.

Some regulators have the power to require the collection and maintenance of financial and accounting records in a manner designed to facilitate the agency's performance of its functions. For example, Part XIB Division 6 of the TPA provides for the ACCC to make a record keeping rule.

Some regulatory regimes also require that certain tariff information, including proposed changes to a tariff, be filed and that the information be made public. For example, see Part XIB Divisions 4 and 5 of the TPA.

Ring fencing and accounting separation.

Ring fencing is a generic term for a set of arrangements designed to stop or make transparent the flow of information and financial transactions between related entities. It is typically used where one related firm is operating in a non-contestable market, and it provides services to other firms which compete with its related entity in a downstream market. It can also be relevant for multi-product

[14] Eg. TPA s. 155.

firms where one product is non-contestable. Arrangements of this type can be seen in the *Moomba Sydney Pipeline System Sale Act1994* (Cth).

Ring fencing is typically seen as a method of ensuring that a vertically integrated or multiproduct firm does not leverage market power from non-contestable markets. However, it has a number of other important information dimensions.[15] For example, ring fencing can be used to:

create information structures which will produce information in a way that is useable by the regulator;

improve the transparency of internal information such as costs, revenues and business strategies. Financial separation may make it more difficult to conceal anti-competitive access pricing; and

impose limits on the flow of information within firms which would otherwise give rise to competition problems (e.g. where a firm provides access services to another part of the firm and to competitors, it may be necessary to restrict the flow of information within the firm).

Accounting separation is a less intrusive mechanism designed to achieve similar objectives. It merely requires various parts of a firm to account for their operations on a stand alone (or fully distributed cost) basis. An example is the QCA Act which makes provision for accounting separation of the declared service and empowers the QCA to devise a cost allocation manual (see Part 5 Division 9 of the QCA Act).

The legal principles which govern the use of such powers may vary depending on the specific legislation. However, the key principles which govern the ACCC's information gathering powers and which are likely to apply to most, if not all, statutory powers to obtain information are:[16]

the information should relate to a particular matter or a particular function for which there is power to seek such information;

the information sought should be properly identified;

the request should be reasonable. Typically this requires that the decision maker considered the balance between the burden of producing the information and the value of the information to the exercise of the regulatory function. It does not

---

15 This issue is canvassed in a number of ACCC discussion papers: Aust, ACCC, Development of Record Keeping Rules for the Telecommunications Industry,(Discussion Paper 26, August 1997); *Ring Fencing Standards For Access Undertakings*,(Draft Discussion Paper, September 1996); *Accounting Separation in Telecommunications* (Draft Discussion Paper, undated).

16 There are numerous cases on this issue e.g. *Johns v ASC* (1993) 116 ALR 567; Riley McKay P/L v Banneman; *Riley McKay P/L v Banneman* (1977) 31 FLR 129; *Kotan Holdings P/L v TPC* (1991) ATPR 41–134. *Panelboard Pty Ltd v TPC* (1982) ATPR 40-272; *Bond Corp Holdings Ltd v Sulan* (1990) 2 ACSR 97; *ASC v Graco* (1991) 5 ACSR1; See *Spargos Mining NL v Standard Chartered Limited* (No. 2) (1989) 1 CSR 314; *Melbourne Home of Ford Pty Ltd v TPC* (1979) ATPR 40-131, (1980) ATPR 40-174.

mean that a decision maker can not make a request which is burdensome or even oppressive, provided that it has considered that burden and acted reasonably. Of course an unduly burdensome notice may suggest the decision maker has not acted reasonably; and

the information should be sought in good faith and for the purpose of the relevant statutory function or a relevant matter, and not some other function or matter.

Informal powers

Regulatory functions may give the regulator a defacto power to seek information. In particular, where a regulator has the power to make a decision which requires them to be 'satisfied' on some matter, then they can largely determine the information that the applicant will be required to produce. In effect, an applicant can either provide the information or otherwise run the risk of not meeting the necessary standard of proof. Further, the decision maker may be entitled to draw adverse inferences from the failure to provide such information. [17]

This approach may also be possible when dealing with parties who oppose such an application. This is further discussed below in the context of information testing and standards of proof.

The power to obtain information by raising the standard of proof is subject to legal limitations. If an applicant refuses to provide certain information and the regulator consequently decides that the standard of proof has not been met, the decision could be subject to judicial review if 'it is so unreasonable that no reasonable person could have come to it'. [18] For example, if a statute merely requires the regulator to have a 'reason to suspect' in order to make the decision, it would probably be 'unreasonable' to require information which establishes the criteria beyond reasonable doubt.

In addition to the legal restriction of 'unreasonability', practical considerations such as the cost to the applicant or an adverse political environment, may constrain a regulator's exercise of this power.

The apparent authority and role of the regulator may also promote the voluntary disclosure of information. In this situation, and more generally, it is important to have developed standard operating procedures in relation to the requests for information and any restrictions that may be imposed. These procedures and a good track record will help establish the confidence of firms as to the safety of confidential information.

This defacto power to obtain information gives rise to a threshold question of whether to use formal information powers. Why not just ask for the information first? In many cases this should be sufficient. There may be good reasons to avoid using formal

---

17 This is akin to the evidentiary rule in *Jones v Dunkel* which may be summarised as: The unexplained failure by a party to give evidence may, not must, in appropriate circumstances lead to an inference that the uncalled evidence would not have assisted that party's case. However, the rule cannot be employed to fill gaps in the evidence: see Byrne & Heydon, *Cross On Evidence*, Australian Edition, (Butterworths, NSW, 1991), para. 1215.

18 *Associated Picture Houses Ltd v Wednesbury Corporation* [1948] 1 KB 223; *Administrative Decisions (Judicial Review) Act* 1977 (Cth) ss. 5(1)(e) & 5(2)(g).

information gathering powers when the same result can be achieved without them. For example, a cooperative approach may be more flexible and efficient in assisting the regulator to identify precisely what they want.

The key question is whether you can achieve the same result without recourse to formal powers. There are a variety of reasons why an informal arrangement may not be as useful, for example:

voluntary disclosure means that the terms of disclosure and the use to which the information may be put are determined by negotiation or the circumstances of disclosure (e.g. a voluntary disclosure may impose a duty of confidentiality where none would otherwise apply, or the party providing the information may seek to limit the use of the information and its return (this is discussed later));

the information may not be provided on oath or otherwise verified;

if there is non-compliance (which may not be apparent initially)[19] there is unlikely to be any sanction; and

notice of the request may lead to the destruction of the information. Destruction following a statutory request is likely to incur some form of sanction.

Restrictions on the power to obtain information: fishing, catch-alls and regulatory reticence

Fishing

A fishing expedition is almost always not a proper use of statutory evidence gathering powers. A fishing expedition is where the person requests information not in pursuit of a particular kind of fish which it has some relevant basis for believing exists, but where they want 'to drag the pool in order to find out whether there were any such fish there or not'.[20]

I say almost always because, at least in my view, the 'fishing rule' does not appear to be relevant to some specific information powers such as record keeping rules. It seems to me to be entirely appropriate for a regulator to use a record keeping rule to obtain information which will:

disclose as yet non-existent contraventions and assist in enforcement actions;

assist in as yet non-existent arbitration disputes; and

assist the regulator to monitor the activities of the firm from time to time.

Catch-alls

---

19 Where it is apparent initially, then formal powers may be used to obtain information not provided voluntarily.

20 *R v Salem* (1989) 16 NSWLR 14 at 17 per Hunt J cited in Carter, *Subpoena Law and Practice in Australia* (Blackstone Press, NSW, 1996).

'Catch-all' requests are to be distinguished from fishing expeditions in that they are focussed on a particular matter. Their vice is that they are designed to catch all information relevant to that matter. Such requests may be invalid due to their oppressive or burdensome nature.[21] But even if they are not invalid, they are undesirable for a number of reasons, including:

obtaining more information than required or than can be analysed; and

the problems arising out of a failure to conduct an upfront assessment of the key issues and information requirements can be exacerbated by having to wade through a deluge of material, some, if not most, of which is only marginally relevant and of no practical use.

Reticence

At times there may be a self imposed restriction on the use of a power to obtain information, that is, a reticence or reluctance to seek out information. Often this is because the use of compulsory powers is seen to be an aggressive step. This should not be the case. Such powers have been given to be used — appropriately. Courts on the whole are not antagonistic to their use and tend to be pragmatic in considering challenges to their use. The ACCC's experience has been that the considered use of information gathering powers is likely to:

enable you to critically test submissions made;

produce relevant information of which the regulator may not have been aware;

materially enhance understanding of the industry and firm specific factors such as costings and business strategies; and

be expected of regulators by industry and other regulatory observers. The appropriate use of these powers will be a positive signal of the serious approach the regulator takes to its functions.

Where to look and what to look for

These are obvious issues and perhaps general observations are not much assistance in particular cases. However, experience suggests that these questions can be vital, yet are often not given sufficient attention. Some points to consider are:

that the only source of information is in the hands of the party to whom it relates. For example, banks and financial advisers may have information which is useful in testing profitability claims;

sources of information which are not created for the purpose of the particular decision may be more credible, or presentable to the regulator; and

---

[21] See cases referred to at fn. 15.

there will be sources of information which, through organisational necessity, are likely to be full and honest.

Asking for what you want

This is another obvious point, but can also be easily overlooked. It is always useful to ensure that there is a clear case of the information required and that any requests (including a record keeping rule) has actually sought that information.

Making sure you get information in a way you can use

Making sure you get the information in a way you can use requires case by case thought. It is important to think about ways that the information can be delivered which will enhance the capacity to assess it. In response to a poorly framed or focussed request, respondents may produce information that is opaque and difficult, if not impossible, to use.

III. How do you use information?

Testing information

Having determined the particular information needs and obtained the information, the next step is to test the information and assess its reliability and the conclusions that can be drawn from it.

This gives rise to two questions: How information is tested; and to what extent should the regulator test the information? These questions are discussed below.

How can you test the information?

In house testing

It is often possible to test information and submissions quite effectively with a particular party through:

reading material provided thoroughly and cross referencing it against other material provided;

seeking primary source data;

benchmarking against other known indicators; and

interviewing representatives of the party for the purposes of putting questions and seeking clarification.

Experts and auditing

If a regulator does not have the capacity to understand or assess what has been given, then it can seek the help of someone who can. It is important not to ignore or minimise the relevance of information simply because it is not easy to understand.

Market place assessment and testing

In most cases there will be express statutory requirements for consultation. While these should obviously be complied with, the absence of such requirements does not mean that there is no legal obligation to test information. Under general administrative law principles, a decision maker may be required to put facts and contentions to a party that are adverse to that party's interests.

In any event, the practical advantages of such testing are very strong, and warrant its use even in the absence of an express requirement. The ACCC's experience has been that such market testing is vital. It is not safe to assume an understanding of what is being submitted or all the implications and counter arguments. Competitors, suppliers and customers can be the best source of information. When conducting market place inquiries, it is important not to merely seek submissions, but to also seek to have key facts and contentions tested.

To what extent should the regulator test the information?

The extent to which the regulator accept the validity of information without testing it, or may be required to exhaustively test evidence will depend upon the standard of proof provided for in the legislation.

Generally, statutes which confer a power to make a decision also outline:

the criteria for making that decision; and

the standard of proof which the decision maker must apply to the criteria.

There is no legal requirement that when drafting legislation, a particular standard of proof must be applied to a particular type of decision. In theory, parliaments can enact legislation which allows a court to convict someone when it has 'reason to suspect' (as opposed to 'beyond reasonable doubt') that the statutory criteria have been met.[22]

However, in practice, the standard of proof imposed by legislation tends to reflect the type of decision making power conferred upon the decision maker. For example:

Statutes creating criminal offences usually require the decision maker (i.e. the court) to be convinced that the statutory criteria have been proved 'beyond reasonable doubt';

Statutes creating a right to seek a civil remedy usually require courts or tribunals to be satisfied 'on the balance of probabilities' that the criteria have been met;

---

[22] Although there are strong policy and other legal reasons why this is unlikely.

Legislation conferring powers upon regulators to make decisions such as issuing licences and authorising conduct often require the regulator to be 'satisfied' that the criteria have been fulfilled;

'Reason to believe' is generally used when the legislature recognises that a precautionary approach should be taken (e.g. a licence should not be granted if the regulator has reason to believe that the applicant has engaged in undesirable conduct) or where there is a need to seek further information (e.g. power to inspect or to require a party to provide further information);

'Reason to suspect' arises when there is an immediate risk to public safety (e.g. carriage of dangerous goods or road traffic defect notices) or when the legislation confers a preliminary power such as the right to investigate a situation; and

Further, some decisions, typically involving complex policy considerations, do not impose any standard of proof but merely outline the criteria which the decision maker must 'take into consideration' when making a decision.

The standard of proof will obviously influence a regulator's need for information. For example, a regulator who must be 'satisfied' that certain criteria have been met, will require more information than a regulator who merely has to have a 'reason to suspect'.

If the statute merely requires the regulator to have a 'reason to suspect', then it is unlikely that the regulator will be required to stringently test the information in order to perform his or her statutory duty.

However, even though legislation may stipulate the standard of proof, the standard does not apply to every item of information which the decision maker relies upon. In criminal proceedings, for example, in order for a person to be convicted, the court must be satisfied that the prosecution case has been proved beyond reasonable doubt. However, each individual item of evidence does not have to meet this standard.

Consequently, a decision maker has considerable discretion to determine the extent to which individual items of evidence are tested, provided that when the information is combined, the final decision cannot be overturned on the grounds that:

there was no evidence or other material to justify the making of the decision;

an irrelevant consideration was taken into account or a relevant consideration was not taken into account; or

the decision was so unreasonable that no reasonable person could have come to it.

However, it should be noted that even if the decision cannot be overturned on judicial review, it may still have to be justified in the 'court' of public opinion which in turn may require greater testing of the information.

Confidentiality

In my experience, confidentiality issues pose some of the greatest problems for statutory decision makers. Some things to bear in mind are:

Is the information confidential and what is the scope of the duty?

> Whether a duty of confidence exists will depend upon the context in which the information is given and the context in which it is sought.[23] A party giving information voluntarily to a decision maker can impose a duty of confidence on the decision maker in respect of that information. Alternatively, where statutory powers are used, it may be implicit in the operation of the statute that the information is to be kept confidential.
>
> The duty of confidentiality may have two dimensions. Firstly, not to **disclose** the information except as permitted; and secondly, not to **use** the information except as permitted.
>
> It is possible that a statutory decision maker having obtained confidential material for a particular purpose may not be permitted to use that information for another statutory purpose even in circumstances where it is directly relevant or even vital to the exercise of another statutory function. The cases in this area demonstrate the real difficulty that a 'limited use' confidentiality restriction may pose for statutory decision makers. In one case, for example, this limitation was held to extend so as to restrain the use of information in consideration of a possible prosecution under Part V of the TPA.[24]
>
> More recent Court decisions indicate a greater reluctance to impose a duty of confidentiality in a way that precludes the use of information in the course of statutory functions clearly in the public interest.[25]
>
> However, as a practical matter, it is important to be aware of this issue and to resist voluntary disclosure of information with unduly restrictive use provisions, particularly where there may be other means of obtaining the information.

Accurately identifying what is and what is not confidential.

> It is important to ensure that parties identify the specific sections of the information which are claimed to be confidential. Global claims for confidentiality are unhelpful, can hinder your inquiries and are difficult to manage. In addition, for each item of information over which confidentiality is claimed, you should ask the party to state why it is confidential.[26]

---

23 For a full discussion on the law of confidentiality see Meagher, Gummow and Lehane, *Equity Doctrines and Remedies* (Butterworths, Sydney, 3rd ed. 1992) Ch 41.

24 *Castrol v Emtech* (1980) ATPR 40–183 (summarised in attachments).

25 *TPC v Allied Mills* (1981) ATPR 40–204 per Gummow J; *Smith, Kline & French v Dept Health* (1991) 99 ALR 679 (summarised in attachments).

26 General categories may be sufficient.

Testing confidential information.

Testing confidential information is a major legal and management issue. The imposition of a duty of confidentiality inherently limits the capacity of the regulator to test the information. However, the mere existence of such a duty does not necessarily preclude all means of publicly testing the information. For example, it may be possible to disclose the information to certain people or to indicate and seek submissions in relation to the general nature of the material. Even if the duty of confidentiality prevents any form of public disclosure, it should still be possible to test the information internally.

Generally, a decision maker is not prevented from relying on information which is confidential and has not been publicly tested when reaching their decision. However, there can be important exceptions to this, including:[27]

where the information is adverse to a party's interests and has not been put to that party; and

where the veracity of information is uncertain due to the inability to test the information. In this situation, it may be necessary to give the information less weight or disregard it altogether.

Writing up reasons based on confidential information.

Where the reasons are to be published, the decision maker has to decide how to record reasons that are based on confidential information. The failure to adequately explain reasons can lead to a distorted evidentiary picture or, at worst, a decision which appears to be contrary to some of the evidence.

There are some practical ways of dealing with this problem:

One approach is to incorporate the confidential information in the text of the reasons (or preferably, in an attachment). This material can then be deleted from the public version, and its omission can be noted in the public version.

Another approach is to summarise the effect of the confidential information where this will not be inconsistent with the confidentiality constraint. This is the best option if it is possible.

---

[27] See Aronson & Dyer, *Judicial Review of Administrative Actions*, (Law Book Co., Sydney, 1996) pp. 531–550.

Other restrictions and limitations on the use of information

An issue closely related to confidentiality is the use to which information obtained for a particular statutory purpose can be put.

Clearly, information must be sought for a valid purpose (i.e. the particular purpose for which the power is conferred). Difficulties can be encountered where information sought for one purpose for which there is a power to obtain that information, is relevant to another purpose for which there is no such power. In this case, the appearance of impropriety may arise.[28]

But what about information that is sought for a particular and proper purpose, but subsequently turns out to be relevant to another purpose?

There are conflicting public policies:

On the one hand, 'compulsory powers are not to be regarded as encroaching more upon the rights of individuals then is fairly and reasonably necessary to achieve the purpose for which the powers were created'.[29]

On the other, once a person with a statutory function is in possession of information relevant to that function, they should not have to ignore it because it was obtained for another purpose. Otherwise, agencies would be prevented from using the 'fruit' of any poisoned information i.e. agencies could not seek information which they know exists through their other functions. This would lead to bizarre and unsatisfactory results. For example, an enforcement/regulatory agency which has obtained information in relation to one particular function (e.g. an access dispute) which clearly demonstrates a breach of the law (e.g. predatory pricing) would be prevented from not only using that information, but also from obtaining it in some other manner.

The current law in Australia on this issue is not particularly satisfactory. The High Court held in *Johns v ASC*[30] (see attachment) that:

> A statute which confers a power to obtain information for a purpose defines, expressly or impliedly, the purpose for which the information when obtained can be used or disclosed. The person obtaining information in the exercise of such a statutory power must therefore treat the information obtained as confidential whether or not the information is otherwise of a confidential nature.
>
> Such a duty is a duty imposed by statute, not by equity, but the equitable remedy of injunction is available in appropriate cases to enforce such a duty against a public authority.[31]

---

28 *TNT v F*els (1992) ATPR 41 –190 provides a useful illustration of this problem.

29 Maris v Director of the Serious Fraud Office (1993) 3 WLR 1 at 7 cited by Brennan J in Johns v ASC (1993) 116 ALR 567 at 600.

30 At 574–575.

31 At 574–575.

Johns v ASC dealt with the ASC Act which contained a specific legislative framework for the use and disclosure of information. The Court did not appear to fully consider the implications of their broad observations for other statutory powers to obtain information. It is not clear why public disclosure should necessarily be inconsistent with the purpose for which the information is sought, indeed it could be an important and appropriate part of that.

In *TPC v Ampol*, Davies and Burchett JJ made observations which appear to temper the breadth of the High Court's observations in *Johns v ASC*.[32]

The approach in *Johns v ASC* can also be contrasted with the approach in *Smith Kline & French*[33] (see attachment) and the approach taken by courts when dealing with evidence obtained either illegally or through the use of a warrant for a different matter.[34]

---

[32] (1994) 127 ALR 533. Burchett J said: 'But the power conferred by S. 155 of the *Trade Practices Act 1974* is a statutory power conferred in aid of the performance of the tasks of the Commission. As Mason J said in *Pioneer Concrete (Vic) Pty Ltd v Trade Practices Commission* (1982) 152 CLR 460 at 472; 43 ALR 449 at 457: 'Its purpose . . . is to aid the Commission in the discharge of its functions under the Act'. It would be totally inconsistent with a valid exercise of such a power for the Commission to blindfold itself so as to blot out all further examination of the evidence to be obtained, evidence which might demand the exercise of some other power of the Commission in the performance of its duties and to do so in favour of simply referring that evidence to a solicitor for advice about its effect in a predetermined legal proceeding. In my opinion, when the Commission takes up this very great power made available to it by the legislature, it also accepts a responsibility to use the information obtained thereby in the performance of its statutory functions, and it cannot limit the exercise of those functions in advance so as to deny itself the right and duty to make full use of the information, whether that information leads to the pursuit of a previous decision or action, or to its variation, or to its reversal' at 542. As to whether the statutory request gave rise to a duty of confidentiality (which arose in the context of the TPC's claim for confidentiality), Davies J said: 'In my opinion, a s. 155 examination is outside the ambit of communications which are encompassed within the concept of legal professional privilege. The TPC does not conduct such an examination solely to obtain legal advice but in pursuance of its statutory purposes. It does so as a statutory authority established under the Trade Practices Act to undertake the functions which the Act reposes in it. One of those functions is to gather information and evidence with respect to contraventions or possible contraventions of the Act. As the examination is conducted in the exercise of a statutory power and as the examination is coercive, the element of a confidential communication made for the purposes of the giving or receiving of legal advice or the futherance of anticipated existing or contemplated judicial proceedings is lacking. The person who is examined is not in the position of a witness or a potential witness who, by voluntarily making a statement for use in legal proceedings, may, perhaps if the circumstances are appropriate, impliedly accept a duty of confidentiality with respect to the statement given' at 538–539. Of course, ultimately, High Court decisions take precedence.

[33] *Smith Kline & French* (1991) 99 ALR 679.

[34] See Williams, Investigations by Administrative Agencies (Law Book Company, Sydney, 1987) pp 115–131.

Specific issues arising in relation to arbitrations

Most regulators represented here have specific statutory arbitration functions. The rules of procedure governing arbitration are strict and can be complicated in practice. One feature of the arbitration rule, relevant to information control, is that the arbitrator can only act on evidence led at the arbitration. The arbitrator can use their general knowledge as an expert body but should approach the arbitration with no knowledge of the facts of the actual dispute.[35]

IV. Good reasons and the link between evidence and reasons

Why prepare reasons?

In many cases, the statute conferring the power to make the decision will specifically require preparation of written reasons.[36] Furthermore, at the federal level and in some states, there are administrative law requirements for the preparation of written reasons.[37] At common law, the general principle has been that administrative decision makers can not be compelled to publish reasons. However, this is changing[38] and even where the Court may not require preparation of reasons, the absence of any may weigh heavily against the decision maker.[39]

But ultimately, while these legal principles are sufficient, the best incentives to prepare reasons are not legal. They include:

the analytical discipline imposed by preparing cogent reasons ('If you can't explain it, you don't understand it')[40];

the desirability of explaining to the parties the reason for the decision;

the wider credibility of the decision;

the precedent value of the decision and consistency across decisions; and

to demonstrate, on judicial review, compliance with administrative law requirements.

---

35 These comments are based on legal advice given by Mr. Ray Finklestein QC to the ACCC (now Justice Finklestein of the Federal Court). Where there is some knowledge of the facts of the dispute, this should be disclosed at the commencement of the arbitration.

36 E.g. ACCC authorisation decisions or decisions to accept access undertakings.

37 See s. 13 of the Administrative Decisions (Judicial Review) Act 1997 (Cth).

38 See Aronson & Dyer, Judicial Review of Administrative Actions (Law Book Co, Sydney, 1996) pp 581–585.

39 As above, fn. 38.

40 See Robertson, 'Tribunals Conference — Writing Reasons for a Decision' (1996) 47 Admin Review 53.

The key ingredients of good reasons [41]

Some guidance on what good reasons are can be gained from the Commonwealth *Administrative Decision (Judicial Review) Act 1977* ('AD(JR) Act'). Section 13(1) provides that a person entitled to seek review of a decision is entitled, upon request, to a statement in writing:

setting out the findings on material questions of fact;

referring to the evidence or other material on which those findings were based; and

giving reasons for the decision.

Section 5 contains the grounds for review of a decision. A decision should, at least, not fall foul of this section. [42] In short, the decision should not be an improper exercise of power, should be based on relevant considerations and not irrelevant ones, and should be supported by the evidence. Furthermore, the decision should consider to what extent

41 For a fuller discussion of this topic see: Aronson & Dyer, Judicial Review of Administrative Action, (Law Book Co, Sydney, 1996) pp 578 ff; Halsburys Laws of Australia (Butterworths, Sydney, 1991); Allars, Introduction to Au*stralian Administrative Law* (Butterworths, Sydney, 1990) para 10–150 ff; Robertson, 'Tribunals Conference — Writing Reasons for a Decision; (1996) 3 (4) Australian Journal of Administrative Law 189; O'Brien, 'Statements of Reasons for Administrative Decisions: Ex Post Facto or Pars Rei Gestae, (1990) 1(3) *Public Law Review* 217.

42 The most relevant paragraphs are in bold: (1) A person who is aggrieved by a decision to which this Act applies that is made after the commencement of this Act may apply to the Court for an order of review in respect of the decision on any one or more of the following grounds: (a) that a breach of the rules of natural justice occurred in connection with the making of the decision; (b) that procedures that were required by law to be observed in connection with the making of the decision were not observed; (c) that the person who purported to make the decision did not have jurisdiction to make the decision; (d) that the decision was not authorised by the enactment in pursuance which it was purported to be made; (e) **that the making of the decision was an improper exercise of the power conferred by the enactment in pursuance of which it was purported to be made;** (f) **that the decision involved an error of law, whether or not the error appears on the record of the decision;** (g) that the decision was induced or affected by fraud; (h) **that there was no evidence or other material to justify the making of the decision;** (j) that the decision was otherwise contrary to law. (2) The reference in paragraph (1)(e) to an improper exercise of a power shall be construed as including a reference to: (a) taking an irrelevant consideration into account in the exercise of a power; (b) failing to take a relevant consideration into account in the exercise of a power; (c) an exercise of a power for a purpose other than a purpose for which the power is conferred; (d) an exercise of a discretionary power in bad faith; (e) an exercise of a personal discretionary power at the direction or behest of another person; (f) an exercise of a discretionary power in accordance with a rule or policy without regard to the merits of the particular case; (g) an exercise of a power that is so unreasonable that no reasonable person could have so exercised the power; (h) an exercise of a power in such a way that the result of the exercise of the power is uncertain; and (j) any other exercise of a power in a way that constitutes abuse of the power. (3) The ground specified in paragraph (1)(h) shall not be taken to be made out unless: (a) the person who made the decision was required by law to reach that decision only if a particular matter was established, and **there was no evidence or other material (including facts of which he or she was entitled to take notice) from which he or she could reasonably be satisfied that the matter was established**; or (b) the person who made the decision based the decision on the existence of a particular fact, and that fact did not exist.

it has a duty to accord procedural fairness, in particular a right to be heard (or make submissions) to an interested party.

It has also been suggested that reasons should set out in clear language the decision maker's understanding of the relevant law, any findings of fact on which their conclusion depends and **the reasoning process which led them to the conclusions** (emphasis added).[43] Where there is a material conflict of evidence, this should be spelt out, as should the reasons for the decision maker's preference. In addition, it is important to set out relevant parties' submissions and contentions (if not in full, then a summary) to demonstrate that you have understood and not overlooked the submissions.

It has also been suggested that the decision maker should identify the audience and keep them in mind. The primary audience is the parties. Other 'audiences' include the wider industry, policy makers, the media and the courts.

Brevity has often been stressed, but, in my experience, it is often difficult to know where brevity leaves off, and inscrutable conclusions begin. Unfortunately, every material fact is likely to rest on numerous subordinate facts. Where those subordinate facts are in contention, they should be discussed. If the facts are not in contention, they can be assumed.

Tests to be applied and statutory criteria

Some statutes are explicit as to the test to be applied by the decision maker. For example, a statute may require that something not be done unless a particular criteria is satisfied (e.g. the ACCC cannot grant an authorisation unless it is satisfied that the public benefits outweigh the public detriment).

However, quite frequently, a decision maker will be required by statute to have regard to a set of factors rather than to make a decision in accordance with a predefined test. In this situation, the decision maker may be uncertain as to the particular test to be applied or the weight to be given to each factor. There is no simple answer to this problem other than to note that courts have generally held that it is up to the decision maker, within the bounds of reasonableness, to make that assessment for itself. For example, in *Minister for Aboriginal Affairs v Peko-Wallsend Ltd*, Mason J stated:

> [I]n the absence of any statutory indication of the weight to be given to various considerations, it is generally for the decision maker and not the court to determine the appropriate weight to be given to the matters which are required to be taken into account in exercising the statutory power.[44]

---

43 Ansett Transport Industries v Wraith (1983) 48 ALR 500 at 507.

44 (1985–86) 162 CLR 24 (1985 at 41. See Aronson & Dyer, Judicial Review of Administrative Action (LBC Information Services, Sydney, 1996) p. 297.

However, according too much or too little weight to a relevant factor may result in the decision being 'so unreasonable that no reasonable person could have come to it'.[45] In this case, the decision is liable to be overturned on judicial review.

What is the scope for review?

There are two basic forms of review. A 'merits review' is where the review body effectively looks at the facts afresh and makes its own judgment on those facts (see, for example, reviews to the Australian Competition Tribunal). The right to a merits review will be expressly provided for in the regulatory statute or some other more general administrative review statute.

'Legal review' refers to a review on some legal ground. The Commonwealth AD(JR) Act codifies the legal grounds of review and sets out the types of decisions to which the Act applies. Most final administrative decisions by federal agencies will be subject to judicial review under the AD(JR) Act. Some states have also enacted legislation in this area. However, even in the absence of specific legislation, courts have a broad power under the common law, to review decisions of administrative bodies on legal grounds.[46]

Conclusion

At the beginning of this paper I asked five questions: What is good information? How do you get it? How do you use it? What are good reasons? And what is the link between evidence and reasons? The questions are fairly obvious. So too have been my observations in respect of the questions. However, the questions are critically important, and, while they may be obvious to a decision maker who takes time to think about them, not all do. For those that do, while there are no easy answers, these questions provide the key to good, credible and defensible decision making.

---

45 Minister for Aboriginal Affairs v Peko-Wallsend Ltd (1985–86) 162 CLR 24 at 41–42 per Mason J.

46 See Aronson & Dyer, Judicial Review of Administrative Action (LBC Information Services, Sydney, 1996) Ch. 2.