EVALUATING INFRASTRUCTURE REFORMS AND REGULATION: A REVIEW OF METHODS

Working paper No. 2 / August 2010

ACCC/AER WORKING PAPER SERIES
FOREWORD

The ACCC has now been a national economic regulator for 15 years. What impact has the legislation that the ACCC (and more recently the AER) administers had on Australia’s infrastructure performance? This is an important question, but not one that can be answered without digging deep into a kit bag of evaluation tools (methods) developed over time in the social sciences – especially in economics. Understanding these methods, their strengths and weaknesses and the types of questions that can be posed using a particular method is the rationale for this report. It is therefore hoped that if evaluation methods can be better understood, researchers in universities or in other places might be encouraged to seek answers to questions about the particular impact of certain reforms and regulations on Australia’s infrastructure performance.

This is the second working paper released by the ACCC. The working paper series was established to allow staff and commissioned consultants working on research projects to make a contribution to the public policy debate. The views expressed are those of the authors and not necessarily those of the ACCC. Nevertheless, the ACCC considers that these working papers are part of its important brief to disseminate material that will inform and educate.

Joe Dimasi
Commissioner
Australian Competition and Consumer Commission
## CONTENTS

### FOREWORD

3

### CONTENTS

5

### ABOUT THE WORKING PAPER

8

1. Introduction ................................................................. 9
   1.1 Project Origins .......................................................... 9
   1.2 The Role of Evaluation in Government Today ....................... 10
   1.3 Aim and Purpose ....................................................... 11
   1.4 Overview of Findings ................................................ 11
   1.5 Conclusion .............................................................. 17

### Section I: Theoretical Foundations and Critical Issues

20

2. Elements of Infrastructure Reform ........................................... 22
   2.1 Introduction ............................................................. 22
   2.2 Corporatisation .......................................................... 22
   2.3 Competition and Structural Change ................................. 25
   2.4 Privatisation ............................................................. 28
   2.5 Regulatory Models for Economic Infrastructure ................... 29
   2.6 Conclusions ............................................................ 33

3. The Case for Government Intervention in Infrastructure Provision .................. 35
   3.1 Introduction ............................................................. 35
   3.2 The Normative Basis for Government Involvement ................... 35
   3.3 Concepts of Economic Efficiency ..................................... 37
   3.4 Market Failure and the *Prima Facie* Case for Intervention ................ 45
   3.5 Conclusions ............................................................ 49

4. The Evaluation Process ...................................................... 51
   4.1 Introduction ............................................................. 51
   4.2 The Research Questions ............................................... 51
   4.3 Types of Evaluation .................................................... 52
   4.4 Levels of Evaluation .................................................... 54
   4.5 Accounting for Aspects of Regulatory Governance .................... 55
   4.6 Types of Evaluation Design ............................................ 57
   4.7 Remaining Steps in *Ex Post* Evaluation .............................. 58
   4.8 Conclusions ............................................................ 59

5. The Counterfactual Problem ................................................ 61
   5.1 Introduction ............................................................. 61
<table>
<thead>
<tr>
<th>Section/Chapter</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>5.2</td>
<td>Essential Requirements of a Counterfactual</td>
<td>61</td>
</tr>
<tr>
<td>5.3</td>
<td>Theoretical Approaches to the Counterfactual</td>
<td>62</td>
</tr>
<tr>
<td>5.4</td>
<td>Strength and Weaknesses of Structural and Treatment Effects Approaches</td>
<td>72</td>
</tr>
<tr>
<td>5.5</td>
<td>Applied Approaches to the Counterfactual</td>
<td>73</td>
</tr>
<tr>
<td>5.6</td>
<td>Conclusions</td>
<td>81</td>
</tr>
<tr>
<td><strong>Section II: Methods and Applications</strong></td>
<td></td>
<td>82</td>
</tr>
<tr>
<td>6</td>
<td>Social Cost-benefit Analysis</td>
<td>84</td>
</tr>
<tr>
<td>6.1</td>
<td>Introduction</td>
<td>84</td>
</tr>
<tr>
<td>6.2</td>
<td>Overview of SCBA Method</td>
<td>84</td>
</tr>
<tr>
<td>6.3</td>
<td>Illustrative Case Studies</td>
<td>88</td>
</tr>
<tr>
<td>6.4</td>
<td>Strengths and Weaknesses of the SCBA Method</td>
<td>96</td>
</tr>
<tr>
<td>6.5</td>
<td>Conclusions</td>
<td>97</td>
</tr>
<tr>
<td>7</td>
<td>Computable General-equilibrium (CGE) Models</td>
<td>98</td>
</tr>
<tr>
<td>7.1</td>
<td>Introduction</td>
<td>98</td>
</tr>
<tr>
<td>7.2</td>
<td>Overview of CGE analysis</td>
<td>98</td>
</tr>
<tr>
<td>7.3</td>
<td>Development of CGE Models for use in Policy Evaluation</td>
<td>100</td>
</tr>
<tr>
<td>7.4</td>
<td>Case Studies</td>
<td>103</td>
</tr>
<tr>
<td>7.5</td>
<td>Conclusions</td>
<td>111</td>
</tr>
<tr>
<td>8</td>
<td>Econometric Methods</td>
<td>113</td>
</tr>
<tr>
<td>8.1</td>
<td>Introduction</td>
<td>113</td>
</tr>
<tr>
<td>8.2</td>
<td>Econometric Methods for Policy Evaluation</td>
<td>113</td>
</tr>
<tr>
<td>8.3</td>
<td>Key Issues to Emerge from the Literature</td>
<td>115</td>
</tr>
<tr>
<td>8.4</td>
<td>Conclusions</td>
<td>129</td>
</tr>
<tr>
<td>9</td>
<td>Productivity Studies</td>
<td>130</td>
</tr>
<tr>
<td>9.1</td>
<td>Introduction</td>
<td>130</td>
</tr>
<tr>
<td>9.2</td>
<td>Total Factor Productivity Analysis – Growth Accounting Analysis</td>
<td>130</td>
</tr>
<tr>
<td>9.3</td>
<td>Frontier Analysis</td>
<td>133</td>
</tr>
<tr>
<td>9.4</td>
<td>Data Availability</td>
<td>135</td>
</tr>
<tr>
<td>9.5</td>
<td>Attribution Problems</td>
<td>136</td>
</tr>
<tr>
<td>9.6</td>
<td>Conclusions</td>
<td>141</td>
</tr>
<tr>
<td>10</td>
<td>Qualitative Data and Analysis</td>
<td>142</td>
</tr>
<tr>
<td>10.1</td>
<td>Introduction</td>
<td>142</td>
</tr>
<tr>
<td>10.2</td>
<td>Qualitative Research Methods</td>
<td>142</td>
</tr>
<tr>
<td>10.3</td>
<td>Mixed Methods</td>
<td>148</td>
</tr>
<tr>
<td>10.4</td>
<td>Lessons from the Literature</td>
<td>149</td>
</tr>
</tbody>
</table>
10.5 Conclusions...........................................................................................................159

Section III: Conclusion .............................................................................................160

List of Abbreviations ...................................................................................................162

Appendix A: References ............................................................................................164
ABOUT THE WORKING PAPER
A major project *Developing Indicators for the Economic Evaluation of Infrastructure Reforms and Regulation* is currently being conducted by staff of the Regulatory Development Branch of the Australian Competition and Consumer Commission (ACCC). The project has a number of parts to it one of them being this methods report. The project is oversighted by an internal ACCC committee as well as an external committee. The external committee members – Professor Jeff Borland, University of Melbourne, Deborah Cope, Principal, Pirac Economics and Dr Denis Lawrence, Director, Economic Insights – have been exceptionally generous in providing insights and in working through the various drafts of this report. The final responsibility for the report rests with those who undertook the research and writing. Dr Rob Albon has led the overall project. Lin Johnson had responsibility for the methods component of the project, and the methods report also draws on contributions from Dr Chris Decker (Regulatory Policy Institute, Oxford), Craig Malam, Dr Anne Plympton, Dr Derek Ritzmann and Dr Su Wu. Katrina Evans, Kara Lamm and Adam Spencer, as student interns working with the ACCC, have also provided valuable assistance.
1. Introduction

This report aims to provide a comprehensive coverage of the issues that can arise, and the methods that can be used, in evaluating competition and regulatory reforms affecting economic infrastructure in areas such as energy, communications, water and wastewater, and transport. It focuses specifically on the process of infrastructure reforms and regulation that have been introduced in Australia since the early 1990s and have pre-occupied most other OECD countries over the last three or so decades. It draws upon research in the academic literature as well that of public bodies and other organisations which have conducted such evaluations in the past. It builds on the normative framework developed in the welfare-economics literature to consider appropriate approaches to the process of evaluation.

The report then reviews techniques that have been used in evaluating infrastructure reforms, including social cost-benefit analysis (SCBA), Computable general-equilibrium modelling (CGE), econometric analysis, productivity studies and qualitative methods. While the aim is always to capture the full range of effects flowing from reforms on an economy-wide basis, the empirical methods do vary in their degree of inclusiveness.

It is also hoped that this publication will become a useful resource for others thinking about setting-up an evaluation study of infrastructure reform and regulation, and it is intended as a guide to evaluation work within the Australian Competition and Consumer Commission and the Australian Energy Regulator (jointly referred to as the ACCC/AER).

1.1 Project Origins

National Competition Policy (NCP) is central to Australia’s microeconomic reform agenda. Over time these reforms have become embedded into the Australian economy. After more than ten years of operation it hardly seemed too soon to ask ‘what has been the impact of economic regulation on Australia’s infrastructure industries?’ and the follow-up question, ‘what can be learned from this?’

Initially there was an expectation that this question would have been posed and answered elsewhere. The ACCC through its administration of the Trade Practices Act has a major role in NCP and its expertise is in regulation rather than evaluation. Both of these reasons suggest that the ACCC should look to other organisations or independent researchers for answers about evaluation.

In keeping abreast with studies in all things regulatory, and in scouring the literature, it became clear that the work that has been undertaken in the evaluation area is not comprehensive. To further advance this area, a consultancy project was developed with NERA and the findings presented to the 2007 ACCC Regulatory Conference. NERA’s work, which was (deliberately) more narrowly focussed than the current study, outlined many of the difficulties that have subsequently become evident in the methods work. Its conclusion argued for a partial indicators approach (such as price trends, investment levels and quantity indicators) with the provision that the choice of specific partial indicators would depend on a number of critical factors identified in its report (see NERA, 2007).

This methods work attracted some attention, and while there was agreement that evaluation was important – and that the time was right – it did not stimulate further work. It became clear that, if this work was to be further advanced, the impetus had to
come from parties most interested in the impact of economic regulation on infrastructure performance, and this of course includes the ACCC/AER.

This interest in evaluation methods coincided with a longer-run concern within the ACCC/AER about the organisation’s public responsibility given the data that it holds across the regulatory areas. These data, acquired in normal regulatory processes and practices, would be fundamental to any sophisticated evaluation exercise. Given this, how should the data best be stored so that they are most useful for such purposes? Also, how could this quantity of data best be presented so that other organisations and groups could understand not only what is available, but the strengths and limitations of using certain data in particular ways (all of course within the context of the strict observance of confidentiality requirements)?

The need to advance thinking about evaluation methods relating to the economic regulation of infrastructure then came together with the recognition within the ACCC/AER of the need for longer-term data preservation. Data supplied over the years as part of ‘normal’ regulatory process have to be kept and documented in such a way that it is useful for future research, particularly for future evaluation studies. This data project was set up at the same time as the methods project, and is now well developed but still ongoing.

1.2 The Role of Evaluation in Government Today

National governments are constantly making decisions, and implementing policies, that have the potential to impact on all aspects of the economy and society. It can be difficult, however, to measure the impact those decisions have on the broader society. Government action may just have hastened a change that would have occurred anyway or change may be totally independent of government action. The change observed by some can be considered to be peripheral when observed by others. Further the change, or lack of change, may not be a consequence of the policy or program but of poor implementation resulting from an inadequate governance structure or a random unforeseeable event.

The tool of evaluation can be used by governments to gain more information for further action or by academic and interest groups to better understand change or to lobby for a different option. However, the evaluation course is not without difficulties – the first questions open a Pandora’s box. What was really intended by government/parliament? How can this be measured? What data are available to be proxies for these measures? What other changes were occurring? Do they better explain the results? When there are multiple interpretations of the data, it may not even be possible to agree on what constitutes outcomes.

Recent calls from within government for evidence-based policy-making (that is policy based on broad-ranging research) may stimulate a renewed interest in policy and program evaluation. Some past experiences may be able to be drawn-upon. Development of import tariff policy is one of the oldest and longest–standing examples

---

1 As would be the case in comparable OECD countries. See the final report of the Infrastructure Consultative Committee (2009) ‘International Insights on Design and Process for Better Regulation of Infrastructure’ on the ACCC website at: [http://intranet.accc.gov.au/content/index.phtml/itemId/1050478](http://intranet.accc.gov.au/content/index.phtml/itemId/1050478) [accessed on 30 June 2010].
of policy-based research in Australia. Gary Banks (2009, p. 107) provides this perspective:

The tariff story illustrates the crucial point that the contribution of an evidence-based approach depends on its content and the objectives to which it is directed. Evidence that is directed at supporting narrow objectives – a particular group or sector, or fostering use of a particular product or technology – will generally look quite different to that which has as its objective the best interests of the general community. Of course, this depends on having the analytical tools to enable such a broad based assessment to be undertaken. Developments in this area were also an important part of the story on tariffs, as well as in other policy areas.

Without evidence gained by research and evaluation programs, government relies on theory, or on some mixture of intuition, ideology, and ‘gut feelings’ in its policy making. However, to obtain evidence, evaluations of past policies and practices are often a key source of objective evidence, particularly when based on robust methods, quality data and insightful interpretations. As the following methods report outlines, these are not simple requirements to fulfil.

1.3 Aim and Purpose

The methods report is broadly based and aims to provide a comprehensive coverage of the issues that can arise, and the methods that can be used, in evaluating competition and regulatory reforms. The purpose is to provide a framework to guide further evaluative work. Charting diverse evaluation paths is required because nuance in the starting question can influence the approach used, data availability will govern the techniques that can be adopted, and constructing the counterfactual is as much an art as a science.

The methods report attempts to explain these relationships, the interlinking nature of the start-out questions, the methods subsequently used and the trade-offs that have to occur when data are imperfect (as they always are). Also, the report attempts to show how confidence in the counterfactual’s veracity impacts on the credibility assigned to the results from an evaluation study.

The methods report highlights the importance of using robust methods to conduct evaluations. It also highlights the fact that evaluations need to be well-formulated and holistic in scope. Caution should be exercised when considering evaluations and assessments that are based on simple and narrow comparisons between countries or other metrics.

The report has been written by economists for those with at least some understanding of economics as well those with a more specialist interest. To cater for this broad group the text has been written with a prose explanation of technical issues accompanied by some simple formulae and a few explanatory diagrams. For those who prefer a more complex mathematical exposition, where appropriate, such explanations have been inserted in boxes within the text.

A broad overview of the report’s findings is set out below.

1.4 Overview of Findings

The methods report is divided into three sections. Section I, ‘Theoretical Foundations and Critical Issues’ provides the framework. The second Section, ‘Methods and Applications,’ compares the different methods that can be used to conduct ex post evaluations of regulatory and competitive reforms and policies. Section III summarises
the key insights emerging from this report for an evaluation of the infrastructure reforms occurring in Australia under the NCP.

1.4.1 Section I

National Competition Policy (NCP) and the micro economic reforms that preceded and followed NCP marked a new direction in the long history of government involvement in infrastructure provision and oversight in Australia. This new direction was based upon strong theoretical foundations. It was also based on a new understanding that it was governments’ role to ‘steer rather than row’ and this required greater attention to incentives, to policies that exposed the provision of infrastructure to market forces and in some cases to infrastructure privatisation.

The main institutional and regulatory reforms enacted by many OECD countries in the 1980s and 1990s encompassed four key elements:

- Corporatisation gave previous ‘public service’ business operations a separate legal entity with a more commercial focus (North America is an exception here as utilities were already operating as separate entities). To prevent government-owned operations having an artificial advantage over private-sector competitors competitive neutrality requirements were also put in place.
- Competition was introduced, based on regulated access to essential facilities. This sometimes involved some form of separation of ‘upstream’ and ‘downstream’ operations.
- Regulation was made separate from both the incumbent operator and the Ministry, usually involving the creation of new, more independent, regulatory bodies.
- Many government-owned entities were subject to some level of privatisation using a wide range of methods.

The evaluation of government involvement in infrastructure requires a normative basis of ‘economic welfare’. Applied welfare economics and social cost benefit analysis (SCBA) have been built on a broadly agreed distributionally neutral welfare basis, hence the determination of the economic welfare implications of infrastructure reforms typically draws upon a well-developed understanding of allocative, cost and dynamic efficiency. The standard public economics approach to government intervention lies in finding a market failure (i.e., the free market fails to achieve one or more of the three efficiencies) and demonstrating that government intervention can effect an efficiency gain. In the case of utility industries the most common market failure is natural monopoly followed by network externality, where the aggregate value of a network to ‘society’ exceeds the sum of the individual valuations of parts of the network (particularly an issue in communications networks). It then needs to be established whether the benefits of intervention (or further intervention) outweigh the costs, and the type of intervention that would produce the greatest net benefit.

With the theoretical foundations established the focus of the report moves to the evaluation process. Comparing or selecting an evaluation method can only occur when other parameters in the evaluation process have been resolved. These include:

- Ex post evaluations of a policy or programs can focus on regulatory processes or regulatory impacts or a combination of both. The report explains this distinction and the relationship between process and impacts.
The level of focus for the evaluation can be at the system-wide, program or individual project level.

The evaluation design provides the logical framework for making conclusions about outcomes and attributing results to the policy or program of interest. The questions cannot be simply answered by the observed outcomes as other factors that are unrelated to the policy or program may have been present and contributed to the outcomes. To draw valid conclusions from the evaluation, it is necessary to compare the observed outcomes (the ‘factual’) with the outcomes that may have eventuated had the policy or program not gone ahead (the ‘counterfactual’).

The type of information required to conduct the evaluation may be quantitative or qualitative information, or a combination of both.

A method for analysing the data (discussed in Section II) must be selected.

Data has to be gathered and any problems with measurement or collection identified. Data availability and data quality has also to be thoroughly explored.

Finally, the analysis must be conducted and findings reported.

Overall, and considering all these steps, specifying the counterfactual is one of the central problems confronting *ex post* evaluation. The report devotes a chapter to the counterfactual issue.

Theoretical approaches to the counterfactual can be broadly classified under three headings:

- the ‘structural approach’
- the ‘treatment-effects approach’
- the ‘counterfactual-histories’ approach.

The report provides a detailed examination of these different approaches. While one approach (the ‘treatment-effects approach’) is considered in theory most useful for assessing the impact of historical policy interventions, in practice the literature indicates that for most evaluation studies developing the counterfactual is exceedingly difficult.

Evaluation studies are more likely to adapt the approach to the counterfactual according to practical circumstances. Much of the approach appears to depend on the specific evaluation question being posed. Understandably, there is a preference for simplicity to make an evaluation more tractable and the resulting analysis less complex. Also, no one approach to counterfactual analysis is more legitimate, or better or worse, than another, but rather that there appears to be significant diversity in the approaches that have been adopted, reflecting the circumstances of individual evaluations.

Ultimately, the choice of, and approach to the counterfactual must be consistent with the other aspects of the evaluation process, including the research question, the evaluation design and data availability.

### 1.4.2 Section II

Building on this framework the focus of the report moves to an examination of the specific methods that can be used to conduct *ex post* evaluations of regulatory and competition reforms. Individual chapters are devoted to different methods with the starting point being those methods that take account of the economy-wide effects of
reforms. These methods are particularly suited to an evaluation of policies that have large nationwide effects, or where there are substantial effects beyond the markets that are directly affected by the policy or regulation.

**Social Cost Benefit Analysis (SCBA)**

The use of SCBA in evaluations of public projects or programs focuses on benefits and costs, including those that affect the broader community. This analysis encompasses benefits and costs that are not readily measured or quantified, such as environmental impacts, changes in health and safety and externalities and takes the perspective of the costs and benefits of society as a whole. In this way SCBA differs from cost benefit analysis (CBA) which is usually used by the private sector to evaluate projects. In CBA the focus is on the private costs and benefits that accrue by measuring the cash flows associated with the project.

For a given income distribution, SCBA measures the net social benefits with the policy or program being evaluated against the net social benefits without the policy or program. A project or policy is acceptable on social cost-benefit grounds if total social benefits exceed total social costs, and thus net social benefits are positive. If SCBA is used to select from a number of options, the option with the greatest (positive) net social benefit would be the preferred policy approach.

The SCBA method has been widely used as a tool for both *ex ante* and *ex post* policy and project evaluation. The report outlines the different steps (at least seven and of varying complexity) that have to be taken to estimate the net social benefit of a policy. While SCBA as a method is conceptually easy to grasp, its application is challenging. A review of the literature provides some insight into how to respond to some of the more difficult issues including:

- specifying the counterfactual scenario
- valuing impacts that are essentially qualitative or intangible in nature
- obtaining suitable data with which to measure costs and benefits over the desired timeframe
- choosing a discount rate, and
- ensuring that results are robust and not sensitive to key assumptions, parameters and/or choice of discount rate.

Overall the report concludes that SCBA is useful for evaluating the economy-wide effects of a policy. It is also a useful method if a policy has large indirect effects that may not be captured adequately using methods that involve only partial evaluation. In addition, it has strong microeconomic foundations and can be applied on a fairly consistent basis with adjustments on a case by case basis to reflect the particular problems that are encountered. Finally, SCBA is the evaluation technique advocated by several governments for assessing *ex ante* policy changes. Nevertheless, data requirements are demanding and potentially costly, and inevitably the application of SCBA will require trade-offs between theoretical rigour and practical tractability.

**Computable General-equilibrium (CGE) models**

Computable General-equilibrium (CGE) models are potentially useful for evaluating the impacts of policies that have a widespread impact on the economy, or have economy-wide, large, indirect effects that may be overlooked if partial methods of evaluation are used. The model consists of a large number of equations that represent
the behaviour of all of the relevant economic agents – consumers, producers, governments – in the economy being modelled. The economic agents are aggregated into a limited number of ‘households’ and production sectors. A policy or reform is introduced as a ‘shock’ to the model. Through the interaction of demand and supply, CGE models compute market-clearing prices and thus determine outputs, the allocation of resources and distribution of income that are consistent with the resulting new general equilibrium.

Because of their complexity, CGE models tend to be purpose-built by a specialised multi-skilled team. However, once developed, the models can be made available to others to be tailored to a particular policy purpose, although typically retaining the theoretical structure and assumptions of the CGE model on which they are based.

In the Australian context the Productivity Commission and its predecessor the Industry Commission have been significant users of CGE models. Drawing from the case studies provided by their work, the report details some of the strengths and weaknesses of using CGE models for policy evaluation.

A general-equilibrium framework underlies the methods of both SCBA and CGE modelling and some productivity studies – that is, these techniques estimate the impact of the policy across the whole economy. Methods that rely on econometric analysis and many productivity studies, however, are based on a partial-equilibrium framework, which estimate the effects of regulation or policy in the directly affected markets only. Using a partial-equilibrium framework the effect of regulation can be found by examining its impact on supply and/or demand in the market and thus the market-clearing price and comparing this new partial equilibrium to the equilibrium without the regulation or policy. The impact of the regulation or policy on welfare can be found by examining changes in consumer and producer surplus.

Econometric methods

Econometric methods can be applied on a within-country basis or on a cross-country basis to analyse the impact of different aspects of the regulatory regime on economic outcomes in the regulated industries. Econometric methods are used to make causal inferences about the effect of a policy intervention on the variable of interest, and by assuming that all other factors are held constant (the ceteris paribus assumption), the impact of the policy or program can be isolated and identified. However, the extent to which an econometric model can be used to make causal inferences is influenced by a number of factors. These factors, and the sequential stages that are required to develop such a model, are outlined in the report.

The specification of the appropriate econometric model involves careful selection of both the dependent variable(s) and the set of independent (explanatory) variables (including the policy of interest) that are hypothesised to impact on the dependent variable. This can be challenging and while it is important that the model is not deliberately specified so as to best fit the available data, ultimately, the way in which an exogenous variable is represented in an econometric equation may be dependent on the available data. Data availability and quality is a crucial issue.

Cross-country econometric studies use econometric methods to analyse the impact of particular variables, such as aspects of the regulatory regime and/or economic and institutional factors, on economic outcomes in the regulated industries. However such studies are subject to a variety of data weaknesses – a particular variable may be defined differently in different countries and the methods used to measure that variable
may also vary across countries. Thus care must be exercised to avoid an ‘apples with oranges’ evaluation.

Also, ‘threats to validity’ arise from factors that undermine an ability to make causal inferences about the impact of the policy or program on the indicator of interest. Internal validity refers to the extent to which inferences can be made that a policy had the effect on the dependent variable suggested by the results. External validity refers to the extent to which the findings of a study can be generalised to other contexts.

Overall, while much work using econometric techniques has been undertaken, misspecification can threaten the validity of the model and undermine the ability to make causal inferences about the impact of reforms on the chosen indicators. Lack of transparency and reproducibility are also often problems with econometric methods. Further, data availability may constrain the use of econometric methods in evaluating the effectiveness of infrastructure reforms.

_Productivity studies_

Both Total Factor Productivity (TFP) Analysis and Frontier Analysis are considered in this report. TFP growth occurs where the quantity of outputs produced grow at a greater rate than the quantity of inputs used. Either an index-number approach or an econometric approach can be used to measure TFP. While both approaches have strengths and weaknesses, the index approach is typically used as it requires fewer observations and is more transparent and reproducible. TFP and TFP growth can, in principle, be measured at the economy-wide macroeconomic level, industry-wide level, the firm level, or down to production units within a firm or organisation.

Frontier Analysis is based on the notion that there is a best-practice level of technical efficiency which cannot be exceeded, a production possibilities frontier that can be reached but not surpassed – but that production can be below best-practice or within the frontier. In this framework, an observed change in productivity (e.g., measured by a change in TFP) for any one economic unit could be the result of either an increase in the best-practice technology available (which would shift the frontier) or a better and more efficient use of existing technology (which would move the economic unit closer to the frontier). As Frontier Analysis is able to distinguish between these two sources of productivity growth, it theoretically enables more detailed analysis of each firm against best practice but requires considerably more observations than index methods. It also potentially permits the origins of technical efficiency (e.g., different regulatory regimes) to be identified. The report explains both of the two significant classes of quantitative techniques for Frontier Analysis – data envelopment analysis and stochastic frontier analysis – and an assessment of their relative merits.

A review of these techniques highlights two substantive issues when moving from theory to practice. The simplest way to evaluate the impact of reforms by focusing on productivity is to obtain a time series of the relevant indicator of productivity and track its performance over time, both before and after the reform. Although conceptually simple, this approach requires a large time series of consistent data, and this is rarely available. Another critical issue is the determination of the extent to which observed changes in productivity ought to be attributed to the reforms under consideration rather than other factors occurring concurrently.

In assessing the contribution of econometric techniques and productivity studies, the report broadly concludes that, if the indirect effects of competition and regulatory reforms are widespread and large, then the results obtained from partial evaluations
alone may be misleading. In such circumstances, it would be advisable to supplement the results using additional general-equilibrium or qualitative methods. Even within the markets that are directly affected, the results of partial evaluations may also be misleading if considered in isolation. For example, to associate a reduction in prices with an improvement in welfare may be incorrect unless the impact of regulation on service quality has also been considered. Thus, an assessment of the effect of regulation on a number of key performance indicators may be necessary to give a more thorough understanding of the effect of regulatory and competition reforms.

**Qualitative data and analysis**

It is clear that there are circumstances where qualitative methods can be an appropriate technique for evaluating competition and regulatory reforms and the chapter outlines these circumstance highlighting strengths and weaknesses of using qualitative analysis. In summary, qualitative methods:

- are likely to be the main approach taken in process evaluations and reviews of regulatory governance.
- may be necessary as an alternative to quantitative methods if adequate data to support quantitative methods are not available. There will be times when insufficient time has passed to enable a large-enough dataset to be assembled for quantitative analysis.
- may be appropriate as a complement to findings based on quantitative approaches. The need to account for qualitative impacts is of particular interest to governments and a particular focus of qualitative analysis is to establish the meaning and broader policy relevance of results and findings.

An evaluation of the process, outcome and impact of competition and regulatory reforms is likely to necessitate the use of both qualitative and quantitative methods of analysis.

**1.5 Conclusion**

An *ex post* evaluation of regulatory and competition reforms typically involves a comparison of actual outcomes with those that would have occurred absent the reforms that are being evaluated – the counterfactual scenario. The report suggests that specifying this scenario is the most challenging aspect of conducting an economic evaluation. Further, the success or otherwise of infrastructure reforms can be evaluated using a number of methods (explained in section II of this report). However, the choice of method must be guided by the question that the evaluation seeks to answer.

From these two broad conclusions the report offers the following advice to those embarking on an evaluation exercise:

- Question(s) that the evaluation seeks to explain must be clearly specified and capable of being answered. Generally, a specific evaluation question is easier to answer than a more general question.
- An evaluation should be well-designed and explained and this will have different implications depending upon the approach taken to the counterfactual.
- Infrastructure reforms are intended primarily to impact on variables that lend themselves to quantification, such as prices, costs, efficiency etc and thus evaluation using a quantitative method. However, a well-designed qualitative
approach may be useful to supplement the quantitative assessment. A qualitative approach will also be the main approach taken to a process evaluation that seeks to evaluate the way in which infrastructure reforms have been implemented.

- Data used in the evaluation should be explained, including a discussion of their sources and measurement. Any adjustments made to the data should be explained. Subject to confidentiality, the data should be available to third parties so that the evaluation’s findings can be replicated.

Trade-offs will inevitably be required to be made between theoretically ideal evaluation designs and methods, and those that are achievable in practice. These trade-offs arise not only because of the difficulty of specifying a defensible counterfactual, but also because of the resources that may be available to conduct an evaluation. More generally, trade-offs are often required as the ‘ideal’ data are seldom available. To illustrate the evaluation process and its inherent features, box 1.1 contains an overview of the evaluation process followed by the Productivity Commission (PC) during its evaluation in 2005 of the impact of NCP reforms on the Australian economy (PC, 2005a).

Finally, the report makes a plea that data collection be carefully considered at the time reforms are implemented to enable defensible, credible and robust ex post evaluations to be conducted.
Box 1.1: Evaluation – An Example

In 2004 the PC was instructed to conduct an ex post evaluation of the impact of NCP reforms on the Australian economy (PC, 2005a). The terms of reference set out some of the parameters of the project, requiring the PC to consider the impacts of NCP reforms on economic indicators (such as growth and productivity) and any significant distributional impacts. The PC was also instructed to identify other areas offering opportunities for potential gains from further reform.

Project design

In accordance with the terms of reference the PC adopted an economy-wide focus in evaluating the impacts of the reforms (PC, 2005a, p. 3). Although the terms of reference did not refer to institutional or procedural arrangements, the PC noted that these elements were interrelated in the achievement of reform objectives. In addition to conducting an impact evaluation of the reforms, the PC therefore also commented on the ‘strengths and weaknesses of institutional settings in the NCP, including the role of the National Competition Council and financial transfers…’ (PC, 2005a, p. 4).

The evaluation used a combination of qualitative and quantitative methods.

Quantitative methods – CGE modelling

The PC chose to undertake its own economic modelling to evaluate the impacts of reform. Observed productivity and service price changes were used as shocks to an applied CGE model (MMRF-GR) (see chapter 7). The model was then used to generate changes for Gross Domestic Product, Gross National Product, sectoral output, employment by state, exports and imports by commodity, nationally and by state, and government revenues and expenditures (p. 50). In contrast to previous reports, no attempt was made to quantify the gains from NCP reforms, the reasoning being both that the task was considered to be scoping in nature and because of data limitations and conceptual issues.

The counterfactual

In order to quantify the impact of NCP, the PC recognised that it would first have to determine a counterfactual, in terms of what productivity and service price changes would have occurred in the absence of reforms. However, as the PC was not aware of any rigorous empirical basis for isolating the impact of NCP reforms and thus with ‘no firm basis for constructing a counterfactual’, it did not explicitly define a counterfactual but instead chose to assess the economy-wide effects of all the productivity and service price changes that occurred over the 1990s. Clearly, this would be likely to overstate the impact of the reforms. However, the PC considered that this overstatement would be counter-balanced somewhat by the likelihood that not all of the benefits of reforms that were implemented in the 1990s would be evident during that decade (PC, 2005b, p.3).

Qualitative methods

The PC chose to also include qualitative information to inform its assessment of the impacts of reform.

First, the PC released an Issues Paper (PC, 2004b), on which it received 135 submissions from interested parties. A number of workshops and public hearings were also held to obtain qualitative information from industry, government, academic, social welfare and public policy interest groups. A discussion draft was released in 2004 (PC, 2004a), upon which further written submissions were sought. This qualitative information (in the form of participant’s opinions etc) was then extensively quoted throughout the report to support the results of the quantitative CGE modelling.

Outcome

The PC concluded that the NCP had led to an increase in GDP, reduced costs and prices of goods and services and improved service quality and reliability in some sectors. These benefits were deemed to have outweighed the identified transitional costs and the failure to deliver on all promised reforms.
Section I: Theoretical Foundations and Critical Issues

Section I of this report consists of four chapters that consider issues that are central for any ex post evaluation of competition and regulatory reforms affecting economic infrastructure, regardless of the purpose of the evaluation and the method that is ultimately chosen for conducting that evaluation.

Chapter 2 reviews the types of competition and regulatory reforms undertaken in OECD countries, particularly in the 1980s and 1990s. The reforms include the corporatisation of government-owned utilities, the privatisation of some of those utilities and the introduction of competition into potentially competitive segments of the sectors in which those utilities operate. The latter reform was sometimes accompanied by structural reforms to vertically separate potentially competitive segments from natural monopoly segments. Where elements of natural monopoly remain, a key component of the competition reforms is the introduction of a regulatory regime to promote competition in potentially competitive segments by allowing third party access to essential, so-called ‘bottleneck’, facilities controlled by the natural monopolist and to regulate other aspects of the natural monopolist’s market power. Chapter 2 also considers the economic outcomes that are (or were) expected to flow from the competition and regulatory reforms.

Chapter 3 sets out the normative basis for these regulatory and competition reforms. An understanding of this welfare basis is important for several reasons:

- Welfare economics provides the rationale for the identification and articulation of the reform problem. The theories also underlie ex ante policy analysis and are used to make predictions about the economic benefits that should result from implementation of the proposed competition and regulatory reforms.
- Fundamentally, the framework of welfare economics underlies the objectives and expected outcomes that policy makers had in mind when implementing the reforms. If these expectations are not clear, the welfare framework can be used ex post to infer the policy-maker’s intentions. Understanding these objectives and expected outcomes is a crucial step in evaluating regulatory and competition policies as it provides a focus from which the evaluation can proceed.
- Welfare economics underlies the quantitative methods for ex post evaluation that are discussed later in this report, and are implicit in the qualitative evaluation methods as well. Thus an understanding of the relevant welfare economics is important to an understanding of those methods.
- The evaluation methods discussed later in this report are based on a conceptually simple comparison – of economic outcomes ‘with and without’ the reforms using counterfactual analysis. The economic theory discussed in chapter 3 can be applied to estimate economic welfare in both the ‘with’ and ‘without’ states. The difference between the two is the effect of the reform or policy being evaluated.

Chapter 4 discusses the evaluation process. It notes that formulation of the question(s) that evaluation seeks to ask is a key part of the evaluation process. Once the question(s) are specified, consideration can be given to the type of evaluation that is suited to answering that question(s) and the level at which the evaluation should be conducted. Chapter 4 also discusses the choice of evaluation design and the method of evaluation. The remaining steps that must be undertaken in performing the evaluation are also considered. Of particular relevance is the specification and collection of the
information and data that will be needed to conduct the evaluation. Sometimes, the data are inadequate for this purpose and therefore tradeoffs between the ‘ideal’ evaluation process and what can be achieved may be necessary.

The discussion in chapter 5 reveals that careful specification of the counterfactual is critical to any robust policy evaluation and is one of the central problems confronting *ex post* evaluation. This is because although conceptually straightforward, practical counterfactual analysis is difficult and inherently subjective – and hence controversial. Chapter 5 sets out the theoretical approaches to the counterfactual and the ways that these approaches can be applied in practice to derive a robust and defensible counterfactual.

The chapters in Section I provide the foundations for Section II of this report which considers the methods that may be used to perform *ex post* evaluations of competition and regulatory reforms.
2. **Elements of Infrastructure Reform**

2.1 **Introduction**

Government involvement in infrastructure provision can take a variety of forms. The infrastructure reforms considered for the purposes of this report are conceived broadly to include all of the main institutional and regulatory changes enacted by many OECD countries in the 1980s and 1990s. Five key elements have been observed in the various exercises in infrastructure reform observed around OECD countries since the 1980s, although there have been differences in detail and in the order in which the reforms have been put in place.

First, except in North America where certain entities were already separate, previous ‘public service’ commercial operations were corporatised, giving them a separate legal entity with a more commercial focus. This required competitive neutrality principles be put in place in an attempt to ensure that government-owned operations did not have an artificial advantage over private-sector competitors. Second, competition was introduced, based on regulated access to essential facilities with other firms being granted access to the natural monopolist’s infrastructure. Third, structural reform took place in the industries; this sometimes involved some form of separation of ‘upstream’ and ‘downstream’ operations. Fourth, regulation was made more independent of both the incumbent operator and the parent government department, usually involving the creation of new regulatory bodies. Fifth, government-owned entities were privatised, in part or in full, with a wide range of methods for doing so observed.

The motivations for these reforms varied across countries and even across industries and sectors, and included stated objectives such as improving productivity, increasing international competitiveness and the generation of government revenue.

In the Australian context, changes to the competitive structure of the economy have taken place through The National Competition Policy reforms. These reforms had economy-wide implications, potentially affecting economic infrastructure in energy, transport and other areas. Under this approach, competition is not valued for itself – rather, it is valued for the various efficiencies it can bring. This was made clear by the Independent Committee of Inquiry (1993, p. xvi):

> Competition policy is not about the pursuit of competition per se. Rather, it seeks to facilitate effective competition to promote efficiency and economic growth while accommodating situations where competition does not achieve efficiency or conflicts with other social objectives.

In this chapter, each reform element – corporatisation, competition and structural change, privatisation, and regulatory models – is described, followed by a discussion of the expected impact, usually as at the time the change was implemented. Current expectations as to the impact of reform elements are also presented.

2.2 **Corporatisation**

A widely observed and critical element of economic reform programs has been the corporatisation of former public-service utility operations initially embedded in government departments. In Australia, for example, (Commonwealth) Government Business Enterprise (GBE) reforms in the late 1980s affected Telecom (now ‘Telstra’), Australia Post, and six other GBES in communications and transport. Amongst the publications guiding corporatisation at the national level were Steering Committee on National Performance Monitoring of Government Trading Enterprises (1994a, 1994b and 1995); Industry Commission (1994, 1997) and Productivity Commission (1998).
There were also some State and Territory initiatives with respect to government-owned trading enterprises (GTEs) – particularly in New South Wales (see Task Force on Monitoring Performance of Government Trading Enterprises, 1991).

2.2.1 Elements of corporatisation

The key elements of corporatisation (according to current expectations) are the introduction of:

- a separate legal and administrative entity
- explicit regulation
- management accounting systems
- independent capital expenditure decision-making
- a rate of return and a regulatory asset base
- competitive neutrality, and
- clearer objectives.

The Special Premier’s Conference Coordinating Task Force on GTE Reform in Characteristics for a Fully Corporatised Government Trading Enterprise and Checklist for National Stocktake of GTE Reforms (1991), which is the basis for the competitive neutrality agreement under NCP, listed the following elements of a corporatised model:

- clear and non-conflicting objectives
- managerial responsibility, authority and autonomy
- effective performance monitoring by the owner-government
- effective rewards and sanctions related to performance
- attaining competitive neutrality in input and output markets
- effective natural monopoly regulation.

The Commonwealth’s competitive neutrality policy notes the Hilmer Committee’s view that corporatisation includes the following elements:

- clarity and consistency of objectives
- management authority and accountability
- performance monitoring
- effective rewards and sanctions, and
- competitive neutrality.

Prior to corporatisation, the details of ‘regulation’ are inherent in the public-service operation and ministerial direction. At corporatisation, regulation is typically made explicit and independent-of-operation. With respect to pricing, regulation tends to focus on cost orientation and the avoidance of monopoly pricing. Aspects of regulatory reform more generally are considered in more detail in section 2.5 below.

An integrated management accounting system typically is introduced at corporatisation. Whereas, under public-service operation there may be a disconnect between revenues (which go directly into consolidated revenue) and expenses (which are met from government appropriations following a bureaucratic process), the accounting system of
a corporatised entity makes managers more directly responsible and accountable for the revenues generated, the costs incurred and the bottom-line difference between the two.

Under public-service operation, physical investments are subject to external scrutiny, particularly, department of finance or treasury oversight. Funds for capital expenditure may be rationed. Ministers may also influence decision making. Under corporatisation, capital expenditure decisions typically became freer from political influence through the creation of independent regulation. Further, the corporation can often undertake its own borrowing to finance its capital expenditure. While the corporation has greater freedom to determine the quantity and composition of capital expenditure, it also becomes subject to more commercial disciplines; particularly rate-of-return hurdles.

Corporatisation of monopoly businesses typically involves definition and revaluation of the regulatory asset base (RAB) and application of a cost of capital. In most instances, the weighted average cost of capital (WACC) becomes the benchmark for the rate of return on a utility’s asset base. The WACC is based on the different costs associated with each of equity and borrowing, with an appropriate allowance for risk. A rate of return above the WACC is suggestive of exploitation of monopoly power while a rate of return below WACC suggests the utility is not covering its costs. Corporatisation in the late 1980s and early 1990s typically involved adoption of some form of replacement cost method of asset valuation to determine the RAB, replacing the valuation of assets at their unindexed historical cost. Asset revaluation means that management is presented with the ‘true cost’ of the capital employed. An accurate asset valuation is essential to form the appropriate basis for a rate of return requirement.

Corporatisation also necessitates the introduction of ‘competitive neutrality’ principles. Infrastructure providers are usually exempt from all taxes levied by different levels of government under public-service operation. However, the entity becomes subject to taxes at all levels of government at corporatisation. Where there is competition with private providers, this change provides for competitive neutrality between government-owned and private businesses. More generally, managers are faced with the same costs as private providers. For example, in Australia, Commonwealth GBEs were progressively exposed to taxation (at all three levels of government) beginning in the late 1980s.

Finally, corporatisation usually means that corporate objectives are made much clearer, including how trade-offs are to be approached. Under traditional infrastructure organisation, governments require providers to charge concessional prices to certain preferred users or uses (traditionally these were called ‘cross subsidies’). These remain following corporatisation, but subsidies become more transparent, and may involve more efficient funding. They usually become known as Community Service Obligations (CSOs) and Universal Service Obligations (USOs) – these are described in Box 2.1. To the extent that they require provision of goods or services at less than commercial prices, universal provision and uniform pricing both incorporate CSO

---

2 In Australia, the ‘deprival value’ methodology gained support from COAG. After revaluation the asset value usually went up substantially. This was the case of the then major Commonwealth government business enterprises (GBEs) like Qantas and Telecom in the late 1980s. In the case of Australia Post its assets were valued in the Annual Report at $512 million in 1987-88 and $1,527 million in the following year. In the water industry asset revaluation also occurred, although somewhat later than for Commonwealth GBEs. For example in 1993-94, Sydney Water’s assets had an historical value of $5.8 billion and an estimated replacement value of $14.9 billion.
elements. Further competition and regulatory reforms usually retain the CSOs and USOs imposed on infrastructure operators.

**Box 2.1: Community Service Obligations and Universal Service Obligations**

CSOs relate to government requirements of infrastructure operators to supply specified services at non-commercial prices, such as concessional fares on public transport and utility access rebates for pensioners and other low-income customers. Obligations may apply generally or in specific circumstances.

USOs are requirements to provide specified standards of service ‘universally’; that is, to all customers seeking supply, and are usually accompanied by uniform pricing provisions. Examples include uniform local telephone call charges and a single standard-letter postage rate.

### 2.2.2 Impact of corporatisation

The anticipated impacts of corporatisation include that prices will become more reflective of costs (rather than being related to ‘political’ and ‘social’ considerations) and that capital expenditures will be made on a more commercial basis. Reductions in production costs more generally are also expected. There may also be improvements in product quality, although the effects on product quality can be difficult to predict. On the one hand, the previous public-service approach with its cost of service underpinning and politicised decision making might tend towards higher quality. It may also be easier to achieve quality through an administrative relation inside the firm, compared with having to specify quality in a contract with an external supplier. On the other hand, corporatised (and privatised) suppliers are likely to be more attuned to customers’ needs, thus improving service quality. However, corporatised suppliers may have an incentive to reduce costs by reducing quality, especially where they do not face competition in the provision of infrastructure services. For this reason, quality-of-service regulations are likely to be carried over into the new regulatory bodies. Oum, Adler and Yu (2006) provide an overview of the literature on the effect of various forms of ownership of airports on aspects of performance, including quality.

It is widely accepted however, that corporatisation is only one phase of the transition towards a more socially optimal outcome. For example, meeting rate-of-return targets might still be easy because of the protection from competition given by statutory monopoly. The statutory protection from competition may also allow some room for ‘cost-padding’ (incurring unnecessary costs), *inter alia*, to enjoy the consumption of perquisites and to allow management the indulgence of Hicks’s ‘easy life’ (Hicks, 1935). Further reforms such as the gradual removal of monopoly rights would be needed to realise the potential gains fully.

### 2.3 Competition and Structural Change

The approach to competition in infrastructure industries in OECD countries has commonly been through the ‘essential-facility’ model, where natural monopoly infrastructure that it is inefficient for rivals to duplicate, but essential for them to enter upstream or downstream markets, is made subject to regulation-based access. The approach taken in practice has been one of permissiveness with respect to entry into any part of an incumbent’s operation. However, in telecommunications entrants have tended to enter downstream markets relying on access to incumbent facilities for the natural monopoly part of the network. Entry into the upstream market would involve the replication of the current telecommunications network. Duplication of such a network would be uneconomical due to its required size and capital intensity. Therefore it is more profitable for firms to enter into the downstream market given the large initial outlays and maintenance costs for competing in the upstream market.
In Australia the introduction of competition to previously monopoly markets was fundamental and central to the National Competition Policy. Even prior to implementation of the Hilmer reforms, the introduction of competition was a focus of infrastructure reform in Australia. In 1992, the telecommunications market was opened to competition when Optus was granted a telecommunications licence in competition with Telecom (now Telstra). The telecommunications market was opened completely to competition in 1997.

The traditional structure of government-owned infrastructure providers in OECD countries, including Australia, was one of vertical integration, combining upstream and downstream operations in a single entity. Structural separation was an integral part of the reform agenda in OECD countries, and was closely related to the introduction of competition. Like several other countries, Australia adopted an ‘essential facility’ model, of regulation-based access to monopoly infrastructure into potentially competitive parts of the market.

As specified in the Hilmer Report (p. xxxi), structural separation of three kinds needed to be considered:

- **Separation of commercial operation from regulation**: The anticipated benefit of this is clearer objectives, less politicised decision-making and a clearer focus on commercial operations. In addition, when competition is introduced, regulation located independently of the access provider is important to prevent the owner of the monopoly infrastructure setting price and non-price terms and conditions that unduly favour its own commercial interests in upstream and downstream markets.

- **Separation of natural monopoly elements from potentially competitive parts**: This is central to the classic ‘essential facilities’ approach taken by Hilmer and is central to NCP. There is a strong emphasis on the need for regulation of the essential facility.

- **Creation of smaller independent business units**: The anticipated benefit of this is an improved managerial focus; closer linking of effort and reward; and the forcing of a coherent internal pricing structure for transactions between business units.

Telecommunications was treated outside of (and before) the main competition reforms and the approach taken to vertical structure was not consistent with that prescribed in Hilmer. The elements of the approach applied to telecommunications are outlined in other sections of this chapter.

### 2.3.1 The trade-off between vertical separation and economies of integration

There is often debate about whether the natural monopoly element of infrastructure should be separated from potentially competitive services. In rail, for example, Australian governments have taken different approaches. The national network is separated such that the owner of the network is not involved in the provision of above-rail service. In contrast, in Queensland, Queensland Rail owns the network and competes with private train operators. These differences in approach result from governments having different views about the trade-off between competition benefits of structural separation compared with the economies of integration.

The joint production of the upstream and downstream services in a single entity may involve economies of integration (coordination, sharing common facilities, etc.) that is eliminated if structural separation occurs. This presents a trade-off or dilemma between improved allocative efficiency (closer alignment of prices to underlying costs from
more effective competition – see chapter 3) on the one hand, and reduced cost efficiency (chapter 3) on the other. This issue has been particularly important in regard to electricity, where mergers have led to significant post-reform reconsolidation, and telecommunications. The issue was emphasised by King and Maddock (1996) in their book on ‘Unlocking the Infrastructure’.

2.3.2 The impact of competition

At the time of implementing reforms, expectations about the outcomes from greater competition were diverse, with there being little detail on the way competition would work. At one extreme, the theory of contestability (Baumol, Panzar, and Willig (1982)) seemed to offer the attainment of both the main objectives, simultaneously preserving the advantages of natural monopoly (avoiding cost inefficiency from fragmentation) and gaining the advantages of competition (lower prices, greater dynamism and higher quality). However, at the other extreme of expectations, there was close interrogation of the key underlying assumptions of the theory of contestability; particularly those supporting its ‘perfect’ form. In particular, ‘perfect contestability’ assumed:

- absence of sunk costs
- no physical or legal tying of customers to incumbents, and
- no restrictions on access by potential competitors to suitable technologies.

At least one – and often all three – of these assumptions were not met in infrastructure industries, raising the question of just what degree of efficiency gain could be expected in typical real-world utility industries?

Apropos of this debate, the terms ‘effective competition’ and ‘workable competition’, are often used interchangeably, and are often not defined carefully. These ideas had been the subject of active debate in the 1930s, 1940s and 1950s, including the involvement of economists such as Robinson (1933) and Chamberlin (1933). Shepherd (1984) reintroduced these terms into the debate in his critique of contestability. Shepherd closely examined the underlying assumptions of contestability and highlighted some potential problems with Baumol et al’s (1982) discussion of the process of competition.

Notwithstanding the inherent imperfections, the benefits expected from introducing competition ‘across the board’ include the following:

- Reduction in prices and an increase in quantities in end-user markets away from monopoly levels.
- Rebalancing of prices from the break-down of cross subsidies. Competition will occur in areas where profits are being made, putting downward pressure on prices and tending to reduce those profits. This undermines the taxation base for funding of implicit subsidies inherent in uniform and concessional pricing, meaning that prices in the subsidy areas will move upwards towards costs.
- Stimulation of improved work practices and improved managerial incentives for cost minimisation more generally, resulting in the production of a given level of output with a reduced amount of inputs.
- New entrants in upstream and/or downstream markets introduce new production methods and new services into those markets, trying to differentiate themselves from incumbents, thus promoting greater ‘dynamic efficiency’ (chapter 3).
2.4 Privatisation

Except in those jurisdictions where private ownership was already present – principally the United States and Canada – privatisation was another important element in infrastructure reform across the OECD. In Australia, the context of the National Competition Policy was primarily one of ‘government monopoly’, and the emphasis was on structure, competition and regulation. Privatisation was not required under NCP, but NCP was concerned with issues of competitive neutrality between government and private operators, and at the same time some governments were considering privatisation of government businesses – at the Commonwealth level (e.g., privatisation of AUSSAT in 1992) and at the State level (Victoria’s privatisation of electricity suppliers). As a result, privatisation rapidly became an integral part of the reforms to infrastructure governance of the 1990s, particularly in energy, telecommunications and airports.

2.4.1 Approaches to privatisation

Privatisation can be pursued through a variety of different approaches. These include:

- **Trade Sale:** This is where the entire operation, having been corporatised into a singular legal entity is sold to a single buying entity.
- **Merger:** This is where the government-owned corporation is merged with a privately owned operation, thus ‘diluting’ the extent of government ownership and injecting private-sector methods into the management and organisation.
- **Stock Exchange Float:** This can involve either part or all of the shares in the utility being made available through public listing on a stock exchange. In the case of large government-owned corporations, often transfer of the utility to private ownership occurs in stages or ‘tranches’ – as in the case of Telstra.
- **Management Buyout:** In this case, ownership and control passes to the previous government employees and management, often at a nominal price.
- **Structural Division and Sale:** Vertical or horizontal disaggregation resulting in the formation of separate business entities followed by sale of some or all of the separate entities thus established. For example, reform of the postal industry in Britain involved the previously vertically and horizontally integrated Post Office being broken up into separate business units and then selling some of those units to non-government buyers.

2.4.2 The effects of privatisation

Identification and quantification of the implications of privatisation are difficult for a number of reasons. For instance, the impacts will differ according to the actual privatisation approach taken, with, for example, a management buyout being different from a stock-market float. Further, the impacts of privatisation per se might be difficult to separate from the effects of other changes being made – seldom is privatisation enacted separately from other infrastructure reforms. Broadly, there are three main issues that need to be considered (see Vickers and Yarrow, 1988, chapter 2 for a more detailed account).

First, where government ownership is replaced by more commercially focused shareholder ownership, owners might be expected to monitor management more closely because it is in their direct pecuniary interest to exert stronger discipline on management. On the other hand, government owners may also have regard to
performance, because it will be a key determinant of the amount of dividends that can be extracted from the utility.

Second, a privately owned entity (particularly where publicly listed) is always on the market – in effect, a ‘market for corporate control’ is established. Where current owners are not achieving enhanced performance, alternative owners will see an opportunity to benefit from taking the firm over and improving performance. Again, however, government owners also extract dividends from government business enterprises and are likely to have an interest in the sale value in the event of privatisation. Discipline on management will not be completely absent under government ownership, and again the extent of any impact is not completely clear.

Third, where corporate-based owners from the same industry area assume control, they may bring operational and market expertise; greater incentives for cost efficiencies (chapter 3); and improved dynamic efficiencies (chapter 3). The question arises as to what would have happened in the absence of this form of privatisation. Further, these types of benefits are inherently difficult to conceptualise and to measure.

As in the case with corporatisation, the possibility does arise that product quality may suffer in an attempt to lower costs and increase profits. The establishment of independent regulators – discussed directly below – may be required to monitor and ensure the quality of the product.

2.5 Regulatory Models for Economic Infrastructure

Corporatisation in its various guises in many jurisdictions was usually accompanied by independent regulation, separate from both the minister and from the incumbent. Competition and privatisation also usually necessitated further changes to regulatory governance. In the United States, which already had an independent telecommunications regulator, the dominant forms of utility regulation were rate-of-return regulation and (the similar but different) ‘cost-of-service’ regulation. However it emerged that more incentive-compatible regulatory models were required in both the United States and Europe.

2.5.1 The critique of rate-of-return regulation

Broadly speaking, rate-of-return regulation is a form of regulation that allows a supplier to recover the aggregated costs associated with providing a set of regulated services, including an allowable rate of return on a regulated asset base. The supplier submits to the regulator information about consumer demand and its estimate of the total operating expenses and capital costs (including a post-tax return on investment) of supplying the relevant services in a given test year, which amount may be passed through to consumers in the prices charged. Traditionally, this type of assessment is conducted relatively frequently on the basis of an estimate of the costs and demand associated with a recent historic ‘test’ year. Conceptually this could be either a historic test year or a forecast test year. However, in most implementations, the assessment is based on a historical annual ‘test year’ cost of service. These estimates may take into account future changes in inflation or any expected and measurable future changes in costs.

Averch and Johnson (1962) analysed the effects of rate-of-return regulation. The model is driven by the assumption that the regulated firm is allowed a rate of return above the rate of return consistent with zero economic profit but below the monopoly rate of return. This leads the regulated firm to choose an inefficient mix (chapter 3) of capital equipment and other inputs, with an excessive amount of capital equipment
‘over capitalisation’) to produce a given level of output. The marginal product of capital (MPPK) is driven down but remains positive (now known as the ‘A-J effect’).

However, this over-capitalisation is not the only inefficiency from rate-of-return regulation, because excessive operational expenditure may also be allowed under the ‘cost-of-service’ approach. This is where regulators allow actual costs incurred to be passed on without scrutiny of their efficiency or reasonableness. Having some discretion with respect to costs can encourage the consumption of perquisites and the enjoyment by management of an easy life (Albon and Kirby, 1983, and references there-in).

On the other hand, Spence (1975) identified a possible advantage of rate-of-return regulation where a monopolist is able to choose both price and service quality. Spence (pp. 428–429) argued:

The unregulated monopolist’s selection of producer characteristics is likely to be biased away from the social optimum … Rate of return regulation has some merit in these circumstances as a second-best strategy … Simple price constraints may cause quality to be set … below optimal levels.

However, this analysis is in the context of regulation of a previously unregulated private monopoly, and this was not the usual circumstance in the Australian context of government-owned monopolies.

2.5.2 Incentive regulation

In general terms, price cap regulation involves the regulator setting a maximum allowable average price (or revenue) path for a set of relevant services for a specified period, which, to some degree, is independent of the actual costs associated with the provision of those services. The utility is allowed to increase the weighted average of the prices of a basket of its services by no more than the increase in the CPI less a percentage amount, X. Physical quantities of each service are the weights used in determining the weighted average. Usually these are previous-period weights. It is possible that not all services are included in the basket. In particular, services in competitive areas are often left out. Further, there may be ‘sub-baskets’ of services subject to specific restrictions.

The key generalised aspect of price-cap regulation, and one that distinguishes it from traditional rate-of-return regulation discussed above, is that average prices are set so as to be independent of the *controllable* costs of the supplier for a significant period of time.

Price capping has both practical and theoretical underpinnings. CPI–X (RPI–X in the UK reflecting that country’s use of a retail price index or ‘RPI’) was initially developed as a temporary control mechanism in the transition to full competition for use in UK telecommunications (Littlechild, 1983). It has since applied in a range of industries in a number of countries. Vogelsang and Finsinger (1979) had earlier developed the theoretical underpinnings of price-capping regulation.

There are four main objectives of CPI–X price capping:

- **Achieving Greater Productivity**: The utility has an incentive to pursue productivity improvements. Where it fails to achieve cost reductions consistent with X its profits will fall. Further, as it can keep any cost savings in addition to those reflected in X – at least in the regulatory period – it has an incentive to aim for greater cost reductions than are provided by productivity growth of X per cent.
• **Passing on Productivity Growth to Customers:** CPI–X forces the utility to pass on the cost reductions (reflected in the set value of X) in lower prices to customers rather than allowing them to result in higher profits.

• **Whittling Away Monopoly Profit or Existing Cost-Inefficiency:** Where the utility commences regulation with above-normal profits and/or existing cost inefficiency, X can be set above TFP growth in order to whittle these away. Indeed, the gradual elimination of above-normal profits was the emphasis of Vogelsang and Finsinger in establishing the theoretical basis for this form of regulation.

• **Restructuring Prices:** CPI–X allows the utility to restructure its pricing towards greater efficiency. As the cap applies to the weighted average of the utility’s prices and not to specific prices, the utility is able to raise (at least relative to the CPI change) one or more of its prices if other prices are reduced sufficiently to satisfy the cap. When freed in this way, the pursuit of profitability will lead the utility to change its pricing structure towards a ‘Ramsey-Boiteux’ configuration of prices; exploiting the more inelastic demands in keeping with the ‘inverse-elasticity rule’ (Albon, 2000).

**Implementation problems**

Implementation problems can arise in determining the value of X. At the time of the implementation of CPI–X there are many possible ‘starting points’ that have relevance to designing a regulatory regime that optimises the achievement of its objectives. The two elements of the utility’s costs are operating costs and capital cost. Determining the capital cost is the subject of considerable controversy. It requires establishing the weighted average cost of capital (WACC) and the asset base to which the WACC is to be applied. Armstrong, Cowan and Vickers (1994, p. 183) note that:

> [a]t first sight it might seem strange to emphasize the role of the cost of capital and the asset base when one of the objectives of the RPI–X system is to escape the well-known inefficiencies of rate-of-return regulation. But each regulator has the duty to ensure the firm can finance its operations, and it is clear that regulators pay close attention to these issues when setting X …

Further, a utility may be operating at less than world’s best practice, using too many inputs to produce its outputs after correcting for influences on its costs beyond the control of management. Attempts at determining whether there is cost inefficiency are described as ‘benchmarking’ studies because they relate the utility’s performance to some best-practice benchmark or ‘yardstick’. Where there is evidence of cost inefficiency relative to world’s best practice, one of the objectives of regulation is to provide an incentive for the elimination of this excess cost.

The roles of X are to ensure that productivity improvements are passed on and (in some interpretations) that existing above-normal profits and cost inefficiencies are removed. Usually X is set to reflect expected growth in total factor productivity (TFP) based on past TFP growth, but perhaps with an eye to possible future developments affecting costs. It may also include an amount to reduce existing monopoly profits and/or existing cost inefficiency (chapter 3). The value of X is re-assessed towards the end of each discrete regulatory period, usually of three to five years’ duration. At this time, in addition to resetting X, the regulator may also make an adjustment to initial prices for the next period to reflect any changes in productivity (a so-called Po adjustment).
The greater is X, the tighter is the constraint. Obviously the regulated utility would prefer a low X (allowing higher prices and profits) while customers would prefer a higher X (lower prices). But the setting of X also has to make reference to, on the one hand, the incentive for the utility to reduce its costs and, on the other, its need to cover its full costs.

*Expectations from improved regulation*

Cost reductions resulting from productivity improvements with the same service quality unambiguously reflect cost efficiency gains. However, where cost reductions come at the expense of service quality there can be a counteracting negative effect on welfare. Therefore productivity change and the overall financial result also have to be considered in the light of service quality. This gives rise to two issues. First, how much do consumers value quality of service? Are they prepared to trade off lower service quality for a lower price or are they prepared to pay more to ensure higher quality levels? This issue requires careful analysis but there are severe empirical difficulties in determining how much customers value service attributes. Second, what is the best way of regulating service quality – the ‘stick’ or the ‘carrot’?

CPI–X belongs to the group of ‘incentive-compatible’ regulations and has clear advantages over direct regulatory structures previously used in Australia and US-style rate-of-return regulation with respect to the incentives it provides the utility to pursue pricing and operational efficiency. It requires careful investigation of the ‘starting point’ for the regulation with respect to the existing degree of cost recovery (are there above-normal profits?), the extent of cost inefficiency (by how much do the utility’s costs deviate from international best-practice cost levels after taking into account factors in its operating environment beyond the control of management?) and the extent of inefficiency of the existing pricing structure. It is also necessary to determine the likely course of productivity growth, usually based on past performance but with a ‘forward-looking’ perspective as well.

Just as with other aspects of infrastructure reforms, the impact of the changes with respect to regulation will depend, in part, on the other changes being made and on the system of governance put in place to oversee the implementation of the reforms. Having said that, the types of reforms discussed here would be expected to result in allocative efficiency gains (more efficient pricing structures) and greater cost efficiency (less regulatory-induced distortion of input choice). Issues surrounding quality are complicating factors in relation to the evaluation of the impact of changes to regulatory arrangements.

### 2.5.3 Impact of regulation on capital expenditure

The impact of economic regulation of infrastructure on the quantity and composition of capital expenditure (capex) has been one of the most intense debates since the infrastructure reforms began. Under the previous government-monopoly circumstances, capex decisions often appeared to be made on ‘engineering’, ‘political’ and ‘social’ grounds, with little explicit regard for commercial criteria. Corporatisation led to the introduction of specific commercial criteria for capex decision-making in GBEs, and competition meant that capex responsibility broadened to include competitors entering previously closed markets as well as incumbents supplying access. The ‘economist’s view’ is that an unregulated monopolist will produce too little, meaning lower than socially-optimal physical investment. Under this view, increases in production driven by the lower prices brought by competitive pressures tend to increase
the amount of capex as against the monopoly counterfactual, including in downstream markets. On the other hand, others have argued that regulated prices – particularly for access – have been set too low to be sufficient to compensate incumbents adequately for the costs and risks they incur, including the loss of real options. The conclusion on the impact of regulation on the quantity and composition of capex is not settled, and increasingly sophisticated arguments for a higher WACC continue to appear. Perhaps the ultimate resolution – if there is one – will be empirically based? On the issue of stimulation versus stifling, it may only be possible to appeal to the data on what has happened.

2.6 Conclusions

Corporatisation, structural change and competition, privatisation of former public-service utility operations, and the introduction of independent and more incentive-compatible regulation are the five key elements of infrastructure reform observed in OECD countries since the 1980s. Corporatisation forces managers of the organisation to become accountable for the balance between costs and revenues. It also provides the firm with clearer objectives and removes politics from the decision making process with regard to capital expenditure and operation.

Structural change involves alterations to the industrial organisation. These changes come about through three different separations. First, separation of commercial and regulatory operations ensures a distinction between the two objectives and prevents any overlap in decision making. It also prevents any unfair access arrangements for competitors. Second, separation of the natural monopoly and potentially competitive elements allows for the possibility of entry by other firms and greater efficiency in contestable areas. Third, breaking the incumbent up into smaller business units leads to clearer objectives for each unit.

The introduction of competition takes place through the separation of the natural monopoly and constable elements. New firms are able to enter the industry and compete with the incumbent in the contestable market, with regulated access to the necessary infrastructure. A notable critique of this separation is that it can lead to higher cost inefficiencies due to loss of ‘scale of integration’. The generated benefits include reductions in monopoly distortions and the possibility of greater dynamic efficiency.

Privatisation is a step on from corporatisation which involves a greater degree of accountability for management of the firm. Managers become subjected to the usual pressures of private industry. Particularly for publicly listed companies, takeovers become a possibility and the demands of the owners for profit force the adaptation of efficient practices and cost structures. Private owners can also bring expertise and knowledge to the organisation which can lead to greater cost efficiencies.

CPI−X price capping is a method used whereby the increase in price charged by the utility is bounded to a level that is less than the increase in the CPI. This form of regulation provides incentives to the firm to increase productivity in order to boost profits and to pass on these productivity gains to the consumers. However it should be noted that there is some subjectivity in determining the appropriate bound on price increases.

Rate-of-return regulation involves allowing the firm to recover the operating and capital costs associated with providing services; they are allowed to pass on these costs to the consumer through higher prices. This model can result in over-capitalisation and the
adoption of inefficient cost structures should these expenditures not be scrutinised effectively be the regulator.

Each element was implemented with the expectation that it would lead to desirable economic outcomes, such as closer alignment of prices with underlying costs (usually expected to lower prices) and increased cost efficiency. Evaluating the extent to which these expectations have been met is the main purpose of this project.

Regardless of the actual motivation at the time of particular reforms, in this report infrastructure reforms will be judged on the basis of standard economic criteria. That is, they are assessed on the basis of whether or not they effect improvements in economic welfare, as compared with the welfare consequences of these changes not being made. The interpretation of each of ‘economic welfare’ (chapter 3) and of the ‘counterfactual’ (chapter 5) are key concerns of the methods report. It is also concerned with the governance put in place to effect and administer the reforms, as this governance structure can have a bearing on the outcome of the reforms (chapter 4).
3. The Case for Government Intervention in Infrastructure Provision

3.1 Introduction

For centuries governments have intervened in infrastructure provision including lighthouses, ports, roads and bridges, and postal services; and have become involved in new areas like rail, electricity, gas and telecommunications as these have been introduced. The approach taken to infrastructure intervention changed markedly in the 1980s and 1990s, particularly towards greater exposure of infrastructure sectors to market forces, privatisation and attention to incentives. The infrastructure reforms over this period were outlined in chapter 2.

The purpose of this chapter is to set out the theoretical foundations for government intervention in infrastructure provision. The theoretical approach requires the use of positive and normative economic analysis, drawing on the standard microeconomics of markets, industrial organisation and game theory, production economics, utility economics and theoretical and applied welfare economics.

Government involvement in the economy in general, including in infrastructure provision, is premised on a normative concept of ‘economic welfare’. This concept involves value judgements, and such judgements can involve differences of opinion. Nonetheless, applied welfare economics and social cost benefit analysis (SCBA) have been built on an agreed distributionally neutral welfare basis following the work of – in particular – Pareto (1906), Kaldor (1939) and Hicks (1939) in the first half of the twentieth century, and what has become a widely accepted appeal for uniformity from Harberger in 1971.

An important lesson from microeconomic theory and welfare economics is that the interaction of buyers and sellers in unrestricted markets is – under certain assumptions – the best way to ensure that scarce resources are allocated in a way that produces the goods and services that consumers want at the lowest possible cost. In other words, markets usually deliver efficient outcomes. There are, however, a number of reasons why markets might fail to deliver efficient outcomes and some form of government intervention may be justified to address this market failure.

This chapter explores the reasons why markets for infrastructure services may fail to deliver efficient outcomes and thus explains the theoretical justification for government intervention in infrastructure provision to improve economic efficiency.

3.2 The Normative Basis for Government Involvement

Modern welfare economics provides the framework for considering where and how government intervention is justified. The idea of economic efficiency – interpreted in terms of allocative, cost and dynamic dimensions – is a normative concept, built on a number of key value judgements:

- **Individual Basis.** Social welfare is based solely on the welfares of individuals that make up society. The ‘state’ *per se* does not have welfare that is independent of its constituents.
- **Individual Sovereignty:** Individuals are the best judges of their own welfare.
- **Pareto Criterion:** Following Pareto (1906), society is deemed to be better off as a consequence of some change in economic arrangements when that change results in at least one individual being made better off, and no individual being made worse off. This is usually known as the ‘Pareto criterion’ and a change satisfying
it is known as a ‘Pareto improvement’. In figure 3.1, there is a utility possibility frontier representing possible combinations of utility that are achievable when the conditions for Pareto efficiency are satisfied (explained below). The utility possibility frontier slopes downwards left to right. Pareto efficiency is achieved when it is impossible to make one individual better off without making another individual worse off, thus it is impossible to effect a Pareto improvement when on the frontier. All points inside the frontier are inefficient, and starting from the inside a Pareto improvement is possible through a reallocation of resources. For example, starting at A, any move to the north (individual 1 better off; individual 2’s utility unchanged); to the east (individual 2 better off; individual 1’s utility unchanged) or to the northeast (both better off) are Pareto improvements.

• ‘Potential Pareto Criterion’: A number of economists debated the Pareto criterion in the 1930s and 1940s, and expressed concern about its restrictiveness. With the example of the Repeal of the English Corn Laws in the 1840s in mind, they observed that few economic policies actually achieved a no-loser outcome, so that the Pareto criterion would seldom be satisfied. Most moves are like A to D or A to E in figure 3.1. Repealing the Corn Laws removed an import barrier that caused the price of wheat to be higher than it would be under free trade – consumers gained (cheaper bread) and landowners lost (lower price for their wheat). Kaldor (1939) and Hicks (1939) both argued for adoption of the value judgement that a change was desirable if the gainers from the change could compensate the losers from the change, and still remain better off themselves. This effectively means that following a change taking the economy from A to D, it would be possible to effect a redistribution taking it in to the area north-east of A. This is sometimes known as the ‘Kaldor-Hicks Criterion’, the ‘potential Pareto criterion’, the ‘compensation principle’ or ‘Harberger’s third basic postulate’ (see immediately below).

---

3 Hicks’s version was that the potential losers could not bribe the potential winners not to want the change. At a more technical level, Scitovsky (1941) demonstrated that the Kaldor criterion alone is not symmetric so that it is possible for outcome A to be an improvement over outcome B, while B is also an improvement over outcome A. While the combined K-H criterion does not have this problem, it can be non-transitive so that A may be an improvement over B, and B over C, but A may not be an improvement over C.
Harberger was interested in the operationalisation of these underlying value judgments into guidelines for applied welfare economics. In Harberger (1971) he specified a set of three basic postulates for applied welfare economics, namely that:

- The competitive demand price for a given unit measures the value of that unit to the demander.
- The competitive supply price for a given unit measures the value of that unit to the supplier.
- When evaluating the net benefits or costs of a given action (project, program, or policy), the costs and benefits accruing to each member of the relevant group (e.g. a nation) should normally be added without regard to the individual(s) to whom they accrue.

These postulates have been broadly accepted by economists and form the basis of work in applied welfare economics and social cost benefit analysis (SCBA) analysing issues in areas such as taxation reform, international trade policy, environmental policy, and – as in this report – infrastructure reform.

3.3 Concepts of Economic Efficiency

The economic welfare justifications for infrastructure interventions typically draw on the trilogy of efficiencies – allocative, cost and dynamic. While many older microeconomic texts grappled towards enunciation of each of these three meanings, they did not achieve an integrated discussion. For example, Reynolds (1985) discusses all three efficiency concepts at different places throughout the book; without drawing them together. Now, these three aspects of economic efficiency are becoming common place – for example, the texts by Cabral (2000, pp. 26-27) and Motta (2004) explicitly adopt this three-part classification. In Australia, for example, these three concepts are specified in the Independent Committee of Inquiry (1993, Hilmer Report, p. 4) and are used consistently by the ACCC/AER. Therefore, they form a natural basis of evaluation for the purposes of this report.

3.3.1 Allocative efficiency

Allocative efficiency relates to the allocation of finite resources to their most valuable uses. Intuitively, this means that goods and services be produced at levels where the
value of the marginal unit to consumers matches the addition to the cost to the economy of producing that unit – ‘price = marginal cost’. The idea can also be illustrated through a ‘utility possibility frontier’ which maps the utilities available to the individuals (1 and 2) in the economy (figure 3.1).

A more sophisticated idea of allocative efficiency can be gained by exploring the three underlying conditions for its satisfaction.

Efficiency in production: This occurs where available productive resources (like ‘labour’ and ‘capital’) are allocated in a way that the output of any one good can only be increased by reducing the output of some other good. As illustrated by figure 3.2 in box 3.1 below, this puts the economy on the production possibility frontier.

Box 3.1: Efficiency and Inefficiency in Production

Figure 3.2: Efficiency in Production

In this diagram, it is assumed that only two goods are produced in the economy – goods A and B. The horizontal and vertical axes represent the labour (L) and capital (K) inputs respectively devoted to each good in production – the origin for production of good A is in the lower left hand corner of the box, and the origin for production of good B is in the upper left right corner of the box. It is also assumed that there are fixed amounts of L and K and these are used for a constant production technology. The curves Q_{1A} and Q_{2A} are isoquant curves for the production of Q_1 and Q_2 unit of good A, where Q_2 is greater than Q_1. The curves Q_{1B}, Q_{2B} and Q_{3B} are isoquant curves for the production of good B, where Q_1 < Q_2 < Q_3. Isoquants show the various combinations of factors of production (in this case L and K) that can produce the same level of output.

Efficiency in production is achieved at any point along the dotted line labelled ‘efficiency locus’. The efficiency locus maps the points of tangency between the isoquants for goods A and B. At all of the points along the ‘efficiency locus’, the allocation of inputs between the two goods is Pareto efficient. For example, if inputs are allocated as shown at point E_0, then Q3 units of B and Q1 units of A will be produced. In order to get to E_1, for instance, the amount of labour and capital allocated to good B, and hence the quantity of good B produced would have to fall so that the amount of labour and capital allocated to production of good A, and thus the amount of good A produced, can rise. When the production point is on the ‘efficiency locus’, the only way to produce more of one good is to reduce the amount of another good.

When the production point is off the ‘efficiency locus’, there is room for the possibility of a Pareto improvement. If the economy was at the point D, then any re-allocation of inputs in the area M+N would represent a Pareto improvement. Notice that only a re-allocation inside the area but on the ‘efficiency locus’ would represent Pareto efficiency in production.
Efficiency in consumption: This occurs where the bundle of goods produced is allocated such that no one individual can be made better off without somebody else being made worse off. This can be represented in an Edgeworth box diagram similar to that used to exposit the concepts of efficiency and inefficiency in production, but with the amounts of each good available to each consumer on the axes, and indifference curves (not isoquants) in the box. Efficiency in consumption is where one consumer can only be made better off at the expense of the welfare of the other consumer, along a consumption-efficiency locus of tangencies of indifference curves.

Overall efficiency condition: The third condition is an ‘overall’ one, meaning that the economy trades off between goods in production as individuals trade them off in consumption. This can be represented in terms of a ‘utility possibility frontier’ (Figure 3.1) that slopes downwards, again with the fundamental characteristic that one individual can only be made better off through a reallocation that results in another individual being worse off (Boadway and Bruce, 1984, ch. 3).

Partial-equilibrium interpretation

In applied work, allocative efficiency is normally interpreted in terms of the change in net economic surplus from a microeconomic application. These analyses are usually undertaken in a partial-equilibrium framework. A microeconomic application like a tax, subsidy or price control will have implications for consumer surplus, producer surplus and the government’s net economic position.

Allocative efficiency is maximised when the benefit that an individual derives from the last unit of consumption of a product is equal to the marginal cost (MC) of that unit. The benefit from consumption of a unit of product is measured by a consumer’s maximum willingness to pay for that unit (WTP). Consumer surplus represents the difference between the maximum amount that the individual would have been willing to pay for the unit and the actual price paid for the unit as given by market price. The marginal cost of a unit of production is measured by the *opportunity* cost of the inputs used to produce that unit, or the value of the inputs in their next best use. The difference between the opportunity cost (or the minimum amount of revenue that a supplier is willing to receive in order to supply a unit of output) and actual market price of the output gives an indication of ‘producer surplus’.
If consumers are willing to pay more for an additional unit of production than the opportunity cost of producing that unit, then allocative efficiency is improved if inputs are shifted into production of that unit. In this way, allocative efficiency is maximised when willingness to pay equals the opportunity cost of the marginal unit of the product, as is illustrated by \( WTP^* = MC^* \), in figure 3.3. At this point, production is equal to \( A^* \). If a product is supplied in a competitive market, then the market price provides an indication of both the maximum willingness to pay and the opportunity cost of the marginal unit. At this market price, no further gains can be made to either consumer surplus or producer surplus.

**Figure 3.3: Partial-equilibrium Efficiency Analysis**

Now consider the imposition of a new tax. This policy will reduce consumer surplus, reduce producer surplus and generate taxation revenue. With reference to the diagram above, the effect will be similar to moving from a point of efficiency to a point where activity is below the optimal level \( A^\text{Low} \). A wedge will be created between willingness to pay \( WTP^\text{Low} \) and the cost of production, \( MC^\text{Low} \). Since the revenues generated by the government will be exceeded by the loss of consumer surplus (shown by the triangle D) and producer surplus (shown by the triangle E), a deadweight loss will be generated. The change in the deadweight loss (\( \Delta \text{DWL} \)) is the efficiency gain or loss from the microeconomic application.

**General-equilibrium interpretation**

While the standard partial-equilibrium approach of \( \Delta \text{DWL} \) is integral to understanding the reforms considered in this report, it may not tell the full story – there might be repercussions in related markets, such as markets for substitutes or markets with which there is a production relationship. These broader effects give rise to the need for a ‘general-equilibrium’ analysis taking into account the impact of these broader effects on economic welfare.

General-equilibrium welfare analysis was established by Hotelling (1938) and by Harberger (1964, 1971). Given the presence of substitutes and complements for infrastructure services, it is often necessary to take a general-equilibrium approach to economic efficiency. For example, suppose that a reform results in a reduction in the price of a service, increasing the demand for that service (with a consequent direct efficiency gain) and also reducing the demand for a substitute service. The efficiency analysis must account for the efficiency effect in the market for the substitute, which
may be positive or negative. For example, it will be negative if price is increased above long-run marginal costs (LRMC) (there will be a reduction in consumption of something which is valued at more than its cost to the economy) and positive if price is held below LRMC (so that reduced units cost the economy more than their value).

A corollary of the general-equilibrium approach, and another issue of importance to the normative framework, is the theory of second-best, developed simultaneously in a number of areas (taxation, trade policy, etc.) in the 1950s and formalised by Lipsey and Lancaster (1956). There are two main messages from the theory of second-best:

First, where one condition for Pareto efficiency is not satisfied, it does not necessarily improve welfare to satisfy another of the conditions. For example, if price exceeds long-run marginal cost in both industry \(i\) and industry \(j\), it does not necessarily improve efficiency to set price equal to long-run marginal cost in industry \(j\) alone.

Second, where one or more conditions for Pareto efficiency are not met then there is a 'second-best efficient' divergence of price from LRMC in the area that can be controlled (first-best would involve all conditions being satisfied). That is, it is possible to ‘make the best of a bad lot’.

This has obvious implications for developing methods for the evaluation of a set of reforms in one part of the economy, when there are other areas of the economy subject to departures from the optimality conditions.

*Intertemporal dimension*

Allocative efficiency also has an intertemporal dimension, with the need to take into account the margin between present and future consumption. With respect to this intertemporal margin, inappropriate policy can affect this margin, perhaps leading to too much present consumption and too little future consumption – essentially, too little capital expenditure.

### 3.3.2 Cost efficiency

The second category of economic efficiency is associated with terms such as ‘cost efficiency’, ‘productive efficiency’, ‘technical efficiency’ and ‘X-efficiency’. In this report the generic term ‘cost inefficiency’ is used to describe a situation where a given level of output is produced at a higher-than-minimum cost, where the source of the inefficiency could be any one or more of inefficiency in production (also an allocative inefficiency); technical inefficiency (an engineering concept) and ‘X-inefficiency’ (perhaps including pursuit of an easy life).

Inefficiency in production is a source of cost inefficiency and also an element of allocative inefficiency. A firm faced with a distorted high price of an input – such as a high wage rate or a high rate of return on capital under rate-of-return regulation – could be forced to distort its input choice resulting in excessive production cost. A profit-maximising firm producing inefficiently in this way would still be efficient in a technical sense (see immediately below), but would be led away from the cost-minimising input combination by distortion of an input price. The consequent higher cost would not totally represent an increased cost to the economy as it would have a transfer element. For example, an input tax would generate tax revenue that was a transfer from the input supplier to the government. For another case, a requirement to use a minimum amount of labour in a production process will force a substitution of labour for capital and increase production costs (see Figure 3.4 in box 3.2).
The efficient point of production is at the point of tangency between the isoquant and the isocost line whereby the firm hires $K^*$ units of capital and $L^*$ units of labour to produce the quantity $Q^*$. The introduction of a minimum labour requirement, (to use at least $L'$ units), results in a substitution to a new input mix. The new input mix employs an inefficiently high amount of labour ($L'$) and inefficiently low amount of capital ($K'$).

The source of inefficiency stems from the new production mix resulting in the same quantity of output as the optimal input mix but at a higher cost. The relative return paid to capital and labour is unchanged as a result of the policy and therefore the slopes of the isocost lines are likewise unchanged. The increase in outlay to reach $Q^*$ can be seen diagrammatically by the difference between the intercepts of the two isocost lines, as represented by the distance A-B.

Yet another case – considered in chapter 2 – is the impact of rate-of-return regulation. In this instance, regulation of an excessive rate of return on the capital asset base can result in over-expansion of that base, in order to increase profits. While this ‘Averch-Johnson effect’ results in a distortion of input choice, it does not cause technical inefficiency.

**Technical inefficiency**

Technical efficiency refers to the efficiency of turning inputs into outputs, where an improvement means using less inputs to produce a given output – or, alternatively, producing a greater output from a given bundle of inputs. This requires the concept of an ‘efficiency frontier’ and measurement of the distance of the actual input bundle from that frontier. In figure 3.4, the isoquant curve showing the use of a combination of capital and labour in the production of a given level of output $Q^*$ represents the best-practice frontier. Any production point above the curve (i.e., using more of labour and/or capital to produce $Q^*$) is technically inefficient. Any production point below the curve (i.e., using less of labour and/or capital to produce $Q^*$) is infeasible with the given technology.

Techniques that are used in assessing cost inefficiencies include TFP analysis and frontier analysis. TFP analysis involves comparing the growth rates in outputs and inputs with an improvement in efficiency being represented by greater growth in the
latter than the former. Frontier analysis attributes the change in observed TFP to either a shift in the positioning of the frontier curve or a movement towards it from an inefficient point inside the feasible region. These techniques are discussed in detail in later chapters.

**X-inefficiency**

Leibenstein (1987) introduced and then elaborated the concept of ‘X-inefficiency’ which is cost inefficiency arising from a lack of competitive stimulus. He contended that ‘X-inefficiency is similar to technical inefficiency’, but that ‘there is nothing technical about the most substantial sources of non-allocative inefficiencies in organisations’ (p. 934). Leibenstein identified relaxation of maximising behaviour; inertia; incomplete contracts and discretion as causes of X-inefficiency. Additional examples include excessive executive perquisites, and rent seeking behaviour through political lobbying for protection and favourable regulations. This has similarities with Hicks’s (1935) belief that for some monopolists, ‘the best of all monopoly profits is a quiet life’.

**Allocative and cost inefficiency combined**

The concepts of allocative inefficiency and cost inefficiency can perhaps be better understood from the following simple case (Figure 3.5) where the two types of inefficiency are present at the same time. This case begins with monopoly pricing at $P_{mon}$ where the monopolist has determined price by setting marginal revenue (MR) to the inefficient long-run marginal cost, $C_{ineff}$. Were both allocative efficiency and cost efficiency to be present, the market would be at $P_{eff}$, $C_{eff}$ and $Q_{eff}$. Instead, there is allocative inefficiency of the red area, and cost inefficiency of area B.

Suppose now a policy can be implemented that causes the monopolist simultaneously to price at long-run marginal cost and to adopt efficient costs ($C_{eff}$). The producer now sets price equal to the new (lower) long-run marginal cost ($C_{eff}$) at $P_{eff}$, and quantity increases to $Q_{eff}$. 
Considering first allocative efficiency, there will be a gain of area $A + C + D$ (shaded in red) representing the excess of total consumer valuation (willingness to pay) of the new units ($Q^{mo}$ to $Q^{eff}$) over the cost to the economy of producing the new units ($Q^{mo}$ to $Q^{eff}$). At the quantity $Q^{eff}$, there is no consumer surplus unrealised since there are none who value any extra units at more than they cost to produce.

Considering now cost efficiency, total output is now produced at $C^{eff}$ per unit rather than $C^{ineff}$ per unit. The gain is in the form of the savings to the firm for producing a given quantity when cost is reduced to the efficient level. This is a benefit of area $B$ where considered as occurring with the original level of production maintained, and a gain of area $B + C + D + E$ when considered over the new level of output.

The total gain from the removal of both cost inefficiency (area $B$) and allocative inefficiency (area $A + C + D$) is equal to area $A + B + C + D$.

### 3.3.3 Dynamic efficiency

The third efficiency, dynamic efficiency, relates to the innovation (or ‘introduction’) of new or improved production processes, organisational structures and services. While the other two concepts are known as ‘neoclassical’, dynamic efficiency is known as ‘Austrian’, based particularly on the work of Schumpeter (1942). This interprets the competitive process as one of both ‘constructive competition’ (applying new techniques and organisation) and ‘creative destruction’ (replacing existing less productive methods). These processes are driven by industry participants striving for monopoly power, but, as Schumpeter put it, monopoly power is constantly subject to ‘the gale of creative destruction’.

The dynamic efficiency approach presents some interesting issues and challenges for designing methods to evaluate the impacts of policies of infrastructure reform considered in this report. In particular:

- **Market Power**: In the neoclassical view, market power (particularly monopoly) is something to be removed or regulated wherever possible, particularly in pursuit of...
allocative efficiency gains. However, under the Schumpeterian view, the longer-run benefits of market power in terms of lower costs, more product variety, etc. must be kept in sight when assessing the impact of market power, leading to a trade-off situation. According to this view, the transitory gains from monopoly profits act as an important part of the dynamic process of innovation and discovery.

- **Long-run competitive equilibrium**: Australian regulatory authorities often set the efficiency benchmark as the long-run equilibrium of a competitive industry. In contrast, the Austrian approach does not see the long-run equilibrium of perfect competition as a suitable *desideratum*. In fact, any sort of ‘equilibrium’ is not a particularly compatible concept for the Austrian approach, as markets are seen as being (as Schumpeter described it) in constant turmoil, never actually reaching equilibrium.

- **Bias to allocative and cost efficiency**: Schumpeter has provided evocative terminology for describing what is happening in markets, but has not suggested ways of measuring these effects. Accordingly, Cabral (2000) emphasises that economists pay less attention to dynamic efficiency than to allocative and cost efficiency because these can be measured more readily than dynamic efficiency.

### 3.4 Market Failure and the *Prima Facie* Case for Intervention

The public-economics literature considers circumstances where the free operation of the market can and cannot result in an efficient outcome. The idea stretches back over more than two-hundred years to Adam Smith’s notion that the ‘invisible hand’ can result in what Smith described as an outcome that is ‘the most agreeable to the whole of society’. In more contemporary terms, the contention is that pervasive perfect competition under assumed conditions can produce a state of Pareto efficiency. The standard public-economics approach to government intervention lies in finding a market failure (i.e., the free market fails to achieve one or more of the three efficiencies) and demonstrating that government intervention can effect an efficiency gain. Significant potential market failures in infrastructure provision are natural monopoly and network externality.

There are two main steps in establishing the case for government intervention to improve the efficiency of a market.

First, it is necessary to identify where the market fails to achieve one or more of the trilogy of economic efficiencies. In the case of utility industries the most common market failure is natural monopoly, where cost efficiency dictates that at least some production components should be carried out by a single producer only, but where an unregulated monopoly supplier would restrict output and charge a ‘monopoly price’. The other main market failure associated with utilities is network externality, for example where private action will result in too few subscribers to ICT networks, as individuals will only account for their private benefits and not the benefits of their actions for others. The finding of a market failure gives rise to a *prima facie* case for government intervention.

Second, it then needs to be established whether intervention can actually improve the situation (effect an efficiency gain) over *laissez-faire*. Do the benefits of intervention outweigh the costs of intervention? Which of alternative interventions would produce the greatest net benefit? In the context of infrastructure reforms in the late 1980s and early 1990s, the prevailing approach was one of statutory monopoly, government...
ownership and non-separation of operations from ‘regulation’ with consequent ‘ politicisation’ of pricing and capital expenditure. So rather than a consideration of de novo government intervention, the focus was on alternative interventions.

However, Demsetz (1969) highlighted four possible fallacies related to government intervention in markets:

- The ‘free-lunch fallacy’ – nothing is ‘free’ so there needs to be a focus on the costs – as well as the benefits – from proposed interventions.
- The ‘Nirvana fallacy’ – markets and intervention are never perfect; so it is always necessary to compare like with like (i.e., imperfect intervention with imperfect market).
- The ‘grass–is-always-greener fallacy’ – don’t jump from market failure to intervention without investigating the costs and the benefits. It cannot be assumed that government intervention will necessarily lead to an efficiency improvement.
- The ‘people-can-be-different fallacy’ – the suggestion that intervention could induce people not to exploit asymmetries of information or could remove moral hazard altogether. Realistically people will remain self seeking in the interventionist state of the world.

The message here is that the step from market failure to government intervention is not automatic – proposals to intervene need to be subjected to a rigorous cost-benefit analysis. The caution required can also be emphasised by observing that comparisons between the market and the interventionist state must be on a like-with-like basis; comparing imperfect market outcomes with imperfect interventions. The ideas of ‘workable competition’ and ‘effective competition’ (considered briefly in chapter 2) reflect this caution by suggesting ‘real-world’ benchmarks for evaluation of comparative performance, rather than the textbook notion of ‘perfect competition’ with its long list of practically unattainable attributes.

3.4.1 Natural monopoly

A defining feature of ‘utility’ is the existence of ‘natural monopoly’ where a function or group of functions can be carried out at a lower overall cost by a single producer than by more than one producer. Natural monopoly stems from the underlying physical characteristics of the industry and has a number of dimensions. First there are those economies from a single producer being responsible for a given quantity of a single good (‘economies of scale’). Second, there are those economies arising from producing given quantities of two or more goods together (economies of scope). Third, there are economies arising from the avoidance of duplication of distribution or collection networks (network economies or ‘economies of density’).

Single product ‘natural monopoly’

Natural monopoly in the single-product case occurs where a given quantity of a particular good can be produced at a lower cost by one producer than if the quantity were produced by more than one producer. The given quantity of particular relevance is the quantity necessary to supply the entire market. Traditionally single-product natural monopoly was thought of in terms of ‘economies of scale’ – decreasing long-run average costs as output increases.
However, natural monopoly can occur without economies of scale. ‘Sub-additivity’ is the crucial concept. Global sub-additivity requires that every conceivable way of dividing the output among two or more producers must lead to a higher total cost than if the output were produced by the single firm. Box 3.3 provides more detail about natural monopolies and illustrates strong natural monopoly in Figure 3.6 and weak natural monopoly in Figure 3.7.

**Box 3.3: Natural Monopoly**

Consider the cost of producing some quantity, $Q$, of a particular good. Each producer has the same cost function relating cost to input use. This output level can be produced by a single producer with cost $C(Q)$ or by a number ($i = 1, \ldots, n$) of producers each with a cost of $C(q_i)$, where their total output is $\sum q_i = Q$. The cost function is defined as ‘sub-additive’ where

$$C(Q) < C(q_1) + C(q_2) + C(q_3) + \ldots + C(q_n). \quad (3.1)$$

While it is usually associated with decreasing (long-run) average cost (LRAC), sub-additivity can occur where average cost is rising. Single product natural monopoly with falling LRAC is known as ‘strong natural monopoly’ as compared with ‘weak natural monopoly’ where LRAC is rising at the level of total demand.

Strong and weak natural monopoly are shown respectively in the following two diagrams:

**Figure 3.6: Strong Natural Monopoly**

![Diagram of strong natural monopoly]
Multi-product natural monopoly

In the case of producing some bundle of two or more goods, the traditional concept of natural monopoly has been that of ‘economies of scope’ — the bundle can be produced at a lower cost by a single producer than by two or more producers. Intuitively, the economies arise through the sharing of utilities used in producing the different goods.

In the multi-product case the concept of ‘economies of scope’ is central:

\[ C(q^1, q^2) < C(q^1, 0) + C(0, q^2) \]  \hspace{1cm} (3.2)

Here the superscripts refer to different products. This means that a given ‘bundle’ of the two products could be produced at lower cost by a single producer than with two separate producers each producing the same quantity of one of the products.

Economies of scope and multi-product sub-additivity are not exactly the same thing. Economies of scope and declining average incremental cost for each service are sufficient for sub-additivity. However, while economies of scope are necessary for sub-additivity, declining average incremental cost is not (See Berg and Tschirhart, 1988, pp. 34-44, for a detailed treatment of multi-product natural monopoly).

Networks

Another distinguishing feature of utilities giving rise to natural monopoly is the existence of networks. Utility operations involve one or more of the following three types of network. First, there are distribution networks for the retail distribution of the product to final consumers. Typical examples are: the customer access network (CAN) in telecommunications; the system of wires; ducts, poles, substations etc. in the local reticulation of electricity; and the postal distribution network. Second there are collection networks, particularly those for the removal of wastewater and sewage;
garbage collection; and the gathering of recyclable materials. Third there are long-distance transmission networks, most notably in electricity, gas, water, telecommunications and posts.

The natural monopoly associated with networks is difficult to categorise because of the diversity of networks across different infrastructure areas and because of the different ways that network economies arise. However, what is fundamental is that natural monopoly always seems to involve the sharing of common infrastructure. Perhaps as a combination of economies of scale and economies of scope, it is ‘economies of density’ that captures some of the essence of network economies.

3.4.2 Network externalities

Another possible market failure involved with utility operations is network externality, where the aggregate value of a network to ‘society’ exceeds the sum of the individual valuations. This can arise because of interdependence between subscribers who value the ability to call, and to be called by, additional subscribers. An individual’s marginal private valuation of subscription may be less than the marginal social valuation which includes the externality amount.

Network externality is particularly an issue in communications networks (telecommunications, internet and posts). The presence of a network externality can provide a *prima facie* justification for interventions such as universal service obligations (USOs) and customer service guarantees.

As noted, the existence of natural monopoly provides only a *prima facie* case for government intervention to improve efficiency, and hence social welfare. In order to intervene it is necessary to show that intervention would actually improve outcomes. In practice there are a number of forms that intervention can take.

3.5 Conclusions

Social welfare is conceived in terms of notions of economic efficiency – allocative efficiency, cost efficiency and dynamic efficiency – and welfare change is defined by the ‘Kaldor-Hicks’ criterion. Following a strong call for consistency by Harberger, these underlying ideas of applied welfare economics have found remarkable acceptance in key areas such as taxation analysis, international trade policy, environmental economics and SCBA. These ideas tend to be reflected in the relevant legislation and guidelines operating in relation to infrastructure regulation.

The virtues of *laissez-faire* operation of markets must be tempered by the presence of market failures. The economic justification for *prima facie* government intervention in the provision of infrastructure services is the notion that such intervention will improve economic efficiency and social welfare. Intervention may be required because of the natural monopoly characteristics of infrastructure provision which mean that it is not possible to improve social welfare by strengthening competition alone.

However, the existence of natural monopoly alone, and the potential for efficiency gains, is not sufficient justification for intervention. The normative basis for government intervention in infrastructure provision is encapsulated in the Kaldor-Hicks (potential-Pareto) criterion, where economic change is desirable if the gainers could compensate the losers while still remaining better off. This is interpreted in terms of the trilogy of economic-efficiency concepts covering allocative efficiency, cost efficiency and dynamic efficiency. Intervention should only be considered if it is likely
to actually improve outcomes, having regard to the costs of intervention as well as the benefits.

The primary purpose of this project is the development of indicators for the evaluation of infrastructure reforms and regulation. The various indicators or methods discussed in later chapters rely on the welfare notions set out in this chapter.
4. The Evaluation Process

4.1 Introduction

The previous chapters of this report have considered the types of infrastructure reforms and regulation introduced across OECD countries since the 1980s and the economic justifications for reforms of infrastructure areas. In highlighting the value of assessing the impact of such reforms after a specific period, this report also recognises the need for evaluative tools to ascertain the causal relationship between reforms and observed outcomes. This chapter introduces some key aspects and issues associated with the evaluation process, and in doing so, it provides a framework for the remainder of the report. It focuses particularly on explaining the types of evaluation that may be done and the steps required to do these.

Before undertaking an ex post evaluation of competition and regulatory reforms it is necessary to decide what question(s) the evaluation is seeking to answer, and thus whether a process and/or impact evaluation should be conducted.

Having decided the research question and what type of evaluation to conduct, there are a number of additional steps that must be completed in the evaluation process. This chapter discusses each of these steps in turn:

- At what level will the evaluation focus? This chapter draws a distinction between three levels of analysis – system-wide, program or individual project levels – and notes that these are not necessarily exhaustive.

- What evaluation design will be used? The evaluation design provides the logical framework for making conclusions about outcomes and attributing results to the policy or program of interest. For the purposes of evaluating competition and regulatory reforms where it is not possible to conduct a controlled experiment, the choice of design is one of the three approaches to the counterfactual – a quasi-experimental ‘treatment-effects’ approach, a causal-structural model, and a ‘counterfactual-histories approach’.

- Decide a method for analysing the data – see the chapters in Section II for further details of methods that can be applied.

- Gather the data and identify any problems with measurement or collection.

- Conduct the analysis and report the findings.

4.2 The Research Questions

The starting point for any ex post evaluation is to determine why the evaluation is being conducted. This requires an understanding of the questions that the evaluation seeks to address. Rossi et al. (2004) note that specification of the evaluation question(s) is sometimes done in a perfunctory manner. They advocate, however, that the task of specifying the question be given detailed attention. This is because the evaluation questions give structure to the evaluation and guide the remainder of its process including the choice of evaluation design, the evaluation method, and the information that is required to conduct the evaluation. According to Rossi et al. (2004), good evaluation questions:

- are reasonable and appropriate by addressing issues that are meaningful in relation to the policy or program

- address issues that are of concern to key stakeholders
• are answerable with available research techniques, and
• are formulated so that the criteria by which success will be judged are explicit or
can be determined in a straightforward way.

For the purpose of ex post evaluation of competition and regulatory reforms, the trilogy
of economic-efficiency concepts lays the theoretical foundation for evaluation and may
also be empirically estimated. The questions for which answers are sought often focus
on the effects of the reforms on key economic variables. The variables of interest may
be high-level indicators such as economy-wide GDP, sector or industry level variables
such as concentration indices and industry performance, or firm-level variables such as
prices and costs. The evaluation may also seek to determine whether society is ‘better
off’ with the reforms than it would otherwise be. How to address these types of
questions in the so-called ‘impact evaluation’ is the main focus of the report.

There is, however, recognition that the way in which competition and regulatory
reforms are implemented can also influence the effectiveness of those reforms and their
impacts on key economic variables. Thus, an ex post evaluation may focus on the way
in which the reforms were implemented, including the performance of the regulator
itself. This type of evaluation is known as ‘process evaluation’.

Sometimes, an ex post evaluation includes an assessment of both the impact of the
regulation or policy on key economic variables as well as the manner in which the
reform or policy was implemented.

4.3 Types of Evaluation

There is an extensive literature on the conduct of evaluations in the social sciences.
Rossi, Lipsey and Freeman (2004, p. 54) identify a variety of types of evaluation.
Depending on the evaluation questions, an ex post evaluation of a policy or program
can focus on regulatory processes, regulatory outcomes and impacts or a combination
of each. This section describes each of these types of evaluations.

4.3.1 Impact evaluations

The impact evaluation, or the so-called ‘summative evaluation’ in the literature,
examines the effect of a policy or program by considering the extent to which the
policy or program produces the expected outcomes from its implementation.

In the context of this report, impact evaluation considers the effect of a policy or
program on economic outcomes, such as prices, costs, service quality, employment or
efficiency. It effectively involves a comparison of what have been observed as the
outcomes and what the outcomes would have been had the policy or program not been
implemented. The difference is the overall or net effect of the policy or program. The
assessment also considers unintended outcomes, whether positive or negative. For
example, Heckman, Lalonde and Smith (1999) examine the impacts of US and
European active labour market policies, such as job training, job search assistance, and
job subsidies on participants in those policies, the net social benefits (or costs)
associated with those policies and the methods used for the evaluation. Heckman et al
find that such policies are often ineffective. For most participants, the benefits are
modest and at worst, participation is harmful. Furthermore, many programs and
initiatives do not pass a cost-benefit test.

Impact evaluations address questions such as:

• What is the overall or net effect of the policy or program?
• Were the desired outcomes achieved?
• Did the policy or program have any unintended effects?

There may be a need to make a distinction between impact evaluations and outcome evaluations, the latter being the more narrowly defined ‘impact assessment’ sometimes used in the literature of social science. The outcome evaluation does not encompass the unintended consequences on other industries or sectors, or the economy as a whole. Despite this distinction, the methods for conducting ‘outcome’ and ‘impact’ evaluations are generally similar. Most of these methods are quantitative in nature, but may be supplemented with qualitative appraisals.

There may be a further need to distinguish impact evaluation from outcome monitoring, where the latter focuses on reporting of observed outcomes for the policy and program implemented. As an example, the Department of Employment and Workplace Relations regularly reports on the proportion of participants who find employment after completing a labour market program (see, for example, Department of Employment and Workplace Relations, 2004). The report would not consider the employment outcomes that might have occurred in the absence of the program participation – the counterfactual – such as the impacts of the program can be evaluated. This distinction is important, given the research findings of labour market programs that suggest that outcomes for program participants are often poorly correlated with program impacts (see for example, Heckman et al. 2002).

4.3.2 Process evaluations

The conduct of impact evaluations is the main focus of this report. Nevertheless, an evaluation of regulatory processes (also known as ‘formative’ evaluation) can also be an important aspect of ex post evaluation of regulatory and competition reforms. Hall (2009) defines process evaluation as focusing on how a program (the implementation of a policy) is being delivered, whether this is according to its charter, what kinds of service participants are being given, and what facilities are available to the program. Process evaluations generally seek to evaluate regulatory effectiveness. Nevertheless, there is a high degree of consensus that the development of objective outcome based measures of regulatory effectiveness is difficult to devise. Accordingly, some evaluative studies focus on the extent to which processes are designed in a way that is likely to foster good outcomes. Typically, the approach in that type of study is to compare the attributes of a specific regulatory system with a hierarchy of principles and standards which are said to reflect ‘best practice’.

Generally speaking, processes and outcomes are linked and better processes tend to lead to better (and more policy consistent) outcomes. However, the linkages are neither mechanistic nor one-to-one, and findings of poor or sub-standard processes do not automatically sustain an inference that outcomes have been sub-standard or deficient. Nor can inferences be made from observed sub-standard outcomes as to whether the process of implementation has been satisfactory or not.

4 Note that many studies in the literature, including Rossi et al. (2004), do not specifically make a distinction between the two terminologies. Others, such as European Monitoring Centre for Drugs and Drug Addiction (EMCDDA), have treated the two terms differently as to whether unintended consequences are accounted for. See the EMCDDA’s website at http://www.emcdda.europa.eu/html.cfm/index9932EN.html [accessed on 2 June 2010] for more information.
4.3.3 Combined process and impact evaluations

Both the process and impact evaluations are informative to the policy-makers and other key stakeholders in the competition and regulatory reform. Each of the two types of evaluations is supported by a number of methods extensively researched in the literature.

It is possible to conduct either a process evaluation or an impact evaluation in isolation. However, if an impact evaluation is conducted in isolation, it is important to realise that information about why the regulatory reform has had the impact that could be obtained from a process evaluation may be lost. Implicit assumptions may have been made about the impact of the regulatory framework and the competency and resources of regulatory institutions and actors. Assumptions, such as no casual effect of these factors, may be subject to considerable debate.

Alternatively, an ex post evaluation of regulatory or competition policy may include both. This combined approach may be optimal if resources and data permit as it recognises the links between processes and outcomes and takes account of the influence of both regulatory governance and the substance of regulatory policies. If the combined approach is adopted, it does not mean that there needs to be an excessive focus on process and regulatory governance evaluation. Rather, it is important to be aware of the link between regulatory governance and the assessment of performance.

Evaluation may focus on examining specific aspects of the regulatory system for consistency with the ‘best practice governance principles’ and making an assessment of whether those aspects are likely to help or hinder in the achievement of specific regulatory goals. This is particularly the case when (and only when) it is not possible to look directly at the outcomes and therefore a proxy for measuring the effectiveness needs to be found. Alternatively, a process assessment may be warranted if the assessment of outcomes has found deficiencies and the implementation process needs to be examined to identify potential reasons for those deficiencies, noting such reasons could lie in the design of the regulation or in its implementation.

The focus of this report will be the assessment of the impacts of the competition and regulatory reform. Section II considers the various methods that may be suitable for impact evaluations. As for process evaluation, section 4.5 will further consider the issue in terms of the relevance of regulatory governance to the evaluation.

4.4 Levels of Evaluation

Evaluations of the impacts of competition and regulatory policies can be carried-out at different levels of aggregation, from highest to lowest as follows:

- **System-wide level.** This would involve an assessment of the total impact of all regulatory policies introduced or implemented during the time period under consideration across all relevant sectors.
- **Program level.** This level of evaluation assesses the impact of all regulatory policies introduced/implemented in a particular infrastructure area, such as telecommunications, energy or post.
- **Project level.** This level focuses on evaluation of regulatory policies introduced/implemented in particular areas of regulated activity, such as mobile telephony or gas transmission.

The choice of the appropriate level of an evaluation principally depends on the objectives of the exercise – that is, the research question of interest. Accordingly, the
Theoretical literature does not provide a definitive ‘answer’ as to an appropriate level of evaluation, nor is there typically any explicit consideration in the literature of the advantages or disadvantages of each possible evaluative level. In short, the approach taken in theoretical work is to adopt a level of evaluation that is fit for the purpose of addressing the research question of interest. The choice as to the appropriate level of evaluation will also be determined, in part, by considerations regarding the choice and availability of data at each level, and the precision and complexity of the counterfactual developed.

As discussed in the next chapter, the question as to the level of evaluation directly affects the choice of the counterfactual (and indeed whether or not one or more counterfactual(s) is required).

4.5 Accounting for Aspects of Regulatory Governance

Consideration of aspects of regulatory governance is now commonplace in studies of the effects of major public policy interventions. Regulatory governance matters because it is one possible cause of the economic outcomes that are observed following the implementation of a regulatory policy or program. In effect, regulatory governance assessment is a major application of process evaluation.

Figure 4.1 depicts the interrelationships among the regulatory framework, reform policies, regulatory governance and economic outcomes graphically. Figure 4.1 indicates that economic outcomes will also depend on exogenous factors that are unrelated to the policy or program that is being evaluated. These exogenous factors must be considered when defining the counterfactual scenario(s) as explained in more detail in chapter 5.

Figure 4.1: The Determinants of Economic Outcomes

There is a vast and expanding literature on regulatory governance, now comprising a significant part of the literature on regulation in general, and on privatisation and post-privatisation regulatory reform in particular. Despite this, there does not appear to be a generally accepted definition of regulatory governance. Some definitions focus on narrow notions which refer to the mechanisms employed to constrain regulatory
discretion. For example, Levy and Spiller (1996, p. 205) define governance structures as:

The mechanisms that societies use to constrain regulatory discretion and to resolve conflicts that arise in relation to these constraints.

In contrast, other researchers adopt broader notions that encompass the frameworks within which a suite of political and social reforms and institutional developments are implemented. For example, Minogue and Carino (2006, p. 4) note:

An evaluation of regulatory governance goes beyond an examination of the formal rules that govern the relationship between public and private sectors to the broader framework of state-public relationships and drawing on disciplinary contributions that range across economics, law, politics and public policy and management.

An important and useful distinction is frequently drawn in the literature between regulatory governance and regulatory substance or content; between how regulation operates within a given framework, and what is regulated (or the substantive regulatory rules). This distinction is captured by Brown et al. (2006, p. 5), on behalf of the World Bank, who note:

Regulatory governance refers to the institutional and legal design of the regulatory system and the framework within which decisions are made. Regulatory governance is the ‘how’ of regulation. It involves decisions about the independence and accountability of the regulator, the relationship between the regulator and policymakers; the process – formal and informal – by which decisions are made; the transparency of decision-making; the predictability of decision-making; and the organizational structure and resources of the regulator.

Regulatory substance is the content of regulation. It is the actual decisions, whether explicit or implicit, made by the specified regulatory entity or other entities within the government, along with the rationale for the decisions. Regulatory substance is the ‘what’ of regulation.

At a conceptual level, the distinction between the substance of a policy and its implementation is sometimes referred to in the context of distinguishing between ‘implementation-neutral’ and ‘implementation-specific’ policies. It follows that a complete evaluation of the effectiveness or performance of past regulatory policies must take account of both regulatory governance and the substance of regulatory policies. Taking no account of regulatory governance is to implicitly assume that the regulatory framework, and the competency and resources of regulatory institutions and actors, do not affect the effectiveness of regulations. In effect, regulatory governance is to conceive of a formal or legalistic – command and control type – framework of how rules are made and then implemented.

More specifically, regulatory governance is an important aspect of evaluation of competition and regulatory reforms as there is likely to be a link between the quality of the regulatory framework within which regulation occurs (the process, or conduct of regulation) and the outcomes of regulation (the focus of outcome and impact evaluations – see above). For instance, while a given policy intervention may have been optimal in terms of its substance or content, its implementation may have been flawed or overly influenced by specific political or stakeholder interests so that the outcomes of the policy intervention may not be as expected. In such circumstances, the cause of poor performance of the policy intervention is arguably correctly attributed to the regulatory authority. Alternatively, it is possible that the policy intervention is itself flawed from the outset, and the implementation is adequate in the circumstances. In this case it would be inappropriate to attribute responsibility for adverse effects of the policy solely on the regulatory agency or on wider attributes of the governance
framework. Responsibility for the poor performance of the policy intervention in this instance arguably rests with the government. In other cases, both the policy intervention and governance framework may be optimal but the effects of the policy intervention are not as expected. This situation may arise if factors that neither the government nor the regulator can control (i.e., uncontrollable factors) have substantially impacted on outcomes.

The recognition that there is a link between how regulation is implemented and economic outcomes is important if the evaluation seeks to attribute outcomes to particular causes. In this respect, it has been suggested that the most useful method of accounting for regulatory governance in such evaluations is to examine specific aspects of the regulatory system and assess whether they are likely to help or hinder in the achievement of specific objectives and goals of public policy in that sector (such as the relative independence of the regulator, the propensity to consult with stakeholders, etc.). Those aspects of regulatory governance commonly identified in the literature as being linked to regulatory outcomes are discussed in detail in chapter 10.

4.6 Types of Evaluation Design

The evaluation design provides the logical framework for making conclusions about outcomes and attributing results to the policy or program or interest. The questions cannot be simply answered by the observed outcomes as other factors that are unrelated to the policy or program may have been present and contributed to the outcomes. To draw valid conclusions from the evaluation, it is necessary to compare the observed outcomes (the ‘factual’) with the outcomes that may eventuate had the policy or program not gone ahead (the ‘counterfactual’).

Defining and identifying the plausible counterfactual(s) are the essential element of evaluation design. The task of determining the unobservable outcomes is complex, but can be accomplished using appropriate methods. For the purposes of evaluating competition and regulatory reforms where it is not possible to conduct a controlled experiment, the choice of design is effectively from one of three broad approaches to the counterfactual:

- a ‘structural approach’ in which the counter-factual is a base case set of assumptions chosen by economic modelling.
- a ‘treatment-effects approach’ in which a quasi-experimental exercise is conducted to determine the counter factual, or
- a ‘counterfactual-histories approach’, under which a qualitative analysis is conducted to specify a counterfactual.

The structural (or structural-econometric) approach has its origins in early econometric work and is based on economic modelling involving the specification of a system of equations that seek to capture behavioural relationships and the causal links among variables. It models a set of outcomes (Y) in terms of its determinants, which are specified in economic theory.

The treatment-effects approach has its origins in statistics and is based on quasi-experimental designs. The traditional treatment-effects approach postulates a counterfactual relationship without explicitly modeling the factors that determine the possible outcomes. It uses observational data from quasi-experiments to gauge the hypothetical outcome.
A third approach, the counterfactual-histories approach, is originated from economic history studies and seeks to apply qualitative methods to ascertain the sensitivity of historical developments to specific events.

The alternative approaches taken to the counterfactual will be reviewed in greater detail in chapter 5. The structural (or structural-econometric) approach is the principal approach in economics to the modelling of a counterfactual and to making inferences about causality. It determines the counterfactual by making *ceteris paribus* assumptions to the modelled determinants. In the treatment-effects approach, the counterfactual is defined by the outcomes of the ‘control’ group – or those units that have not been exposed to the ‘treatment’. The counterfactual-histories approach is primarily concerned with highlighting contingency in history, and in particular, the sensitivity of historical developments to specific events.

### 4.7 Remaining Steps in *Ex Post* Evaluation

Once the research question, the level of evaluation and the evaluation design have been decided, a number of additional steps must be completed. Data needs must be determined, appropriate data gathered and a method for analysing the data must be decided upon.

Consistent with the chosen research design, it is necessary to decide what type of data will be needed to conduct the evaluation. This might include data on chosen indicators of performance as well as data on factors that might explain that performance. The types of performance indicators will depend, in part, on the preceding steps. For instance, if a project-level evaluation is to be conducted, indicators of performance in particular areas of regulated activity will be necessary. In contrast, for a system-wide comparison, higher-level, aggregated indicators of performance across the entire system may be more appropriate. The required data may be quantitative, qualitative, or a combination of both.

Having determined the required data, it is necessary to identify any problems with measurement or collection of the data. Data can be gathered from a variety of primary and secondary sources. Sometimes measures of the preferred indicator are not available and the use of proxy indicators may be necessary. For example, if the indicator of interest is unmeasurable, then it may be necessary to use a dummy variable to represent the effect of that indicator or use a more qualitative approach to evaluate the indicator of interest. In addition, data problems might arise if data-collection methods change over time, or the definition of a particular data item is changed.

Next (or perhaps simultaneously) it is necessary to decide on a method for analysing the data. The chosen method may be quantitative, qualitative or a combination of both. Quantitative research and analysis use numerical descriptions of information. The interpretation and presentation of information is more likely to utilise established statistical theory which can be employed if requirements on quantitative data have been met. The availability, and particularly the suitability, of quantitative information or data can therefore have a large bearing on the decision to employ quantitative methods.

The final steps involve conducting the analysis using the chosen method and available data and reporting the findings. The chapters in Section II examine the main quantitative and qualitative methods for *ex post* evaluation of competition and regulatory reforms. These chapters describe how the analysis may be performed using a particular method, as well as the problems that may be encountered, and the solutions (if any exist) that might be adopted.
Finally, the way in which the evaluation results are presented and reported can be as important as the analysis itself. In this regard, a key requirement is that the process of applying the method and using the data, including any underlying assumptions or adjustments, is transparent. In addition, evaluation reports should:

- be drafted in plain-English where possible. If technical language must be used, it should be explained clearly.
- provide sufficient evidence to support the report’s conclusions and recommendations
- provide an easy ‘audit trail’ for the reader to check calculations, supporting evidence and assumptions
- provide the results of any sensitivity analysis.

The process of undertaking an evaluation is unlikely to be as neatly sequential as set out above. Rather, it is likely to involve lots of feedback loops – for example, knowledge of data sources will feedback to the choice of evaluation method and the type of evaluation that it will be possible to undertake.

4.8 Conclusions

A number of points can be drawn from this chapter to provide guidance for the economic evaluation of infrastructure reforms. Briefly, these points relate to:

- the importance of specifying the purpose of evaluation and the question(s) that the evaluation seeks to answer
- the need to choose the type of evaluation (process, outcome/impact or combined) that is best suited to address the research questions
- the level at which the evaluation will be conducted to best address the research question
- the choice of research design – one of a quasi-experimental treatment effects approach, a structural approach or a counterfactual-histories approach to the determination of the counterfactual
- the choice of data and the importance of having adequate and consistent data with which to conduct the evaluation.

Conceptually, the conduct of ex post evaluations follows a number of discrete steps. In practice, there is likely to be considerable overlap between the stages. In addition, tradeoffs may have to be made. For example, the evaluator may decide that a system-wide evaluation is desirable only to discover that the available data are inadequate for that task. In such circumstances, it may be necessary to focus the evaluation on a different level, or to use the available data but recognise its limitations when drawing conclusions about the success or otherwise of competition and regulatory reforms.

An additional point to take from this chapter is that good policy can be jeopardised by poor implementation. Governance is another explanatory variable in determining the relationship between economic outcomes and economic reforms. For example, if access to essential facilities is controlled by the incumbent infrastructure provider, a less desirable outcome with respect to economic welfare would be expected. This must be kept in mind when evaluating which factors drive the outcomes from reforms.
The remainder of this report provides more information about the choices that must be made when conducting *ex post* evaluations and the trade-offs that may be required. The literature provides some guidance as to how to deal with the tradeoffs. Therefore the following chapters also examine the lessons learned from the literature.
5. The Counterfactual Problem

5.1 Introduction

Evaluations of the economic impact of policy reforms involve a comparison of observed effects (the ‘factual’) with the outcomes that may plausibly have happened if the policy had not gone ahead (the ‘counterfactual’). Such a comparison is necessary because the impact of a policy or reform is always incremental to what would have happened in its absence (Commonwealth Government, 2006).

A counterfactual query comprises three elements – what is to be explained (the consequent); the variables or factors that account for this outcome (the antecedent); and the principle(s) assumed to link the two such that a change in the antecedent might be expected to cause a change in the consequent. The degree of confidence about the results of any evaluation employing a counterfactual will reflect the confidence in the underlying causal model. Counterfactuals are thought experiments, representing alternatives to actual history or observable outcomes. In other words, a counterfactual inquiry seeks to answer questions about what might have happened had a particular event (a cause) not occurred. By carefully considering possible alternative outcomes for the indicators of interest, a counterfactual analysis considers the range of factors that may also ‘cause’ particular outcomes in addition to the relevant policy or program, including factors that are not related to the policy or program that is being evaluated. A comparison of actual outcomes with a carefully specified counterfactual enables the ‘effect’ of the particular ‘cause’ to be isolated and identified.

Perhaps because of the practical difficulties of defining a counterfactual, ex post evaluations of competition and regulatory reforms have not always been based on explicit comparisons of actual outcomes against a robust and defensible counterfactual. Clearly, in this situation the conclusions that are drawn from these evaluations about the success or otherwise of the reforms or policies being evaluated are substantially weakened.

The purpose of this chapter is to identify and discuss some of the general theoretical and practical issues associated with the development of a robust and defensible counterfactual for the purposes of evaluating competition and regulatory reforms. The chapter begins by explaining the term ‘counterfactual’ and its links to causal reasoning. An outline of some of the challenges associated with the use of counterfactuals is then presented. Three theoretical approaches to counterfactual reasoning – the ‘structural approach’, the ‘treatment-effects approach’ and the ‘counter-factual histories approach’ – are discussed. The chapter concludes by considering the practical application of counterfactuals to the analysis of competition and regulatory reforms and policies.

5.2 Essential Requirements of a Counterfactual

There are four crucial requirements of an appropriate counterfactual.

First, precision and clarity in the definition and formulation of the counterfactual and of the evaluation framework in which it is employed are essential (McCloskey, 1987). The more precisely the counterfactual and framework are specified, the more robust and defensible the results of that counterfactual analysis are likely to be. The counterfactual’s precision and clarity will be enhanced if it is developed in such a way that key assumptions, parameters and variables, and relevant data can be clearly identified and the results are replicable. The assumptions underlying the counterfactual will be critical to the conclusions drawn about the effect of a policy or program.
Although, in counterfactual reasoning it is not necessary that either assumptions or predictions are ‘true’ or observable, it is crucial that the assumptions and predictions are linked in some way, and that there is at least an implicit underlying causal model.

Second, the counterfactual should be as simple as possible (Heckman and Vytlacil, 2007). An evaluation exercise designed to address a specific economic question is likely to be more powerful and convincing in a model with a smaller number of assumptions or parameters. Put differently, addressing one question will typically require fewer assumptions, and place less demands on the data, than answering a wide array of questions. However this has also been referred to as the basic paradox of the counterfactual – simpler models may be harder to believe than more complex models, but they are more likely to be representative of the actual past occurrences because they place fewer constraints on history.

Third, the level at which the evaluation will occur should be defined (see chapter 4 for details) as this will determine the detail with which the counterfactual query should be posed. Specifically, when seeking to undertake a system-wide comparison of the impact of policies, the appropriate counterfactual will be one focused on what might have been expected to occur if an entire program of reforms were not introduced. Alternatively, when undertaking a series of project-level evaluations, it will be necessary to develop a number of more specific (‘tailored’) counterfactuals.

Finally, the temporal aspects associated with the counterfactual should be considered. This involves consideration of whether a single event is hypothesised to lead to a counterfactual path or is the counterfactual assumed to comprise a number of discrete stages, each of which cumulatively influences the hypothetical outcome.

5.3 Theoretical Approaches to the Counterfactual

A central problem of counterfactual reasoning is how to estimate the outcomes that might have occurred if a policy or program had not been implemented. This is because the only outcomes that are observable are those that have occurred with the policy or program in place. As the counterfactual is unobservable, it is important that it is constructed in a manner that is as theoretically rigorous as possible so that the counterfactual is both defensible and reasonable.

The counterfactual scenario can be specified using quantitative or qualitative techniques. The theoretical literature has developed two main quantitative methods for addressing the counterfactual problem for the purpose of economic evaluation. The first approach, known as the ‘structural approach’, has its origins in early econometric work and is based on economic modelling that seeks to uncover a structure underlying the data generating process. The second approach, known as the ‘treatment-effects approach’, has its origins in statistics and is based on quasi-experimental evaluation designs. These approaches are based on the evaluation designs that were discussed in Chapter 4, and, as mentioned there, the choice of evaluation design will have implications for the approach to the counterfactual.

A key difference between the treatment-effect approach and the structural approach is the detail with which they specify counterfactual outcomes. The structural approach models the counterfactuals more explicitly than the statistical treatment-effect literature (Heckman, 2005). The ‘structural approach’ models a set of outcomes (Y) in terms of its determinants or causes, where those determinants include the policy or program being evaluated as well as other determinants that are selected on the basis of economic theory. The counterfactual is found by varying a particular determinant, or cause, while
holding the other possible determinants or causes constant. In contrast, the ‘treatment-effects approach’ does not rely primarily on specifying a theoretical model, but rather compares observed outcomes between a ‘treatment’ group that has been exposed to the policy or program, and a ‘control’ group that has not. In this approach, the ‘control’ group acts as the counterfactual. Heckman and Vytlacil (2007) explain the distinction as being between understanding the ‘effects of causes’ (the goal of the treatment-effects literature) and understanding the ‘causes of effects’ (the goal of the structural economic modeling approach).

There is a third approach to specifying the counterfactual which is sometimes used for evaluative purposes. In economic history, the counterfactual is sometimes considered in some academic work employing what might be loosely termed the ‘counterfactual-histories approach’. This approach seeks to apply qualitative methods to ascertain the sensitivity of historical developments to specific events. As considered above, the new economic history revolution beginning in the mid-1960s saw the application of economic theory, econometric techniques and other quantitative methods to the study of history. In effect, ‘climetric’ or ‘new economic history’ studies apply quantitative techniques to historical information in the determination of the counterfactual.

5.3.1 The ‘structural approach’ to the counterfactual

The principal approach in economics to the modelling of a counterfactual, and to making inferences about causality, is the structural (or structural-econometric) approach. This involves the specification of a system of equations which contain parameters and variables that attempt to capture behavioural relationships and specify the causal links between variables. Implicit in this approach is an assumption that economic theory permits structure to be placed on the various parameters of the model to derive robust empirical results.

The term ‘structural’ conveys the idea of a representation of the world that is stable under a large class of interventions or of modifications of the environment. In very general terms, this approach involves the specification of a system of equations which contain parameters and variables that attempt to capture behavioural relationships and specify the causal links between variables (Goldberger, 1972). As conventionally applied, the structural approach to modelling causality involves the representation of an endogenous variable as a function of other exogenous (and endogenous) variables and error terms. Economic theory is used to guide the construction of these structural models and to provide guidance about which variables to include or exclude from the model (Heckman and Vytlacil, 2005).

Mathematically, conventional structural models typically have the following common functional form:

\[ Y = f(X, U) \]  

(5.1)

where \( Y \) is an effect, \( f \) is a deterministic function, \( X \) represents a cause and \( U \) represents some further extraneous random variable. Structural models can consist of hundreds of such functional relationships.

The theoretical literature sometimes distinguishes between the ‘all causes’ structural approach and the ‘deep structural’ approach to modelling causality (Heckman, 2005). The theory of these two approaches is considered in more detail in box 5.1. Both models employ a form of causality that enables the causal effect of a particular variable to be estimated by holding other variables constant (a \textit{ceteris paribus} assumption). In a structural approach, the notion of causality is linked to the model itself, rather than to
the data, and it is possible that the data are consistent with more than one structural, or causal, model. A counterfactual can be constructed by varying the ‘cause’ in question, holding all other variables constant.

**Box 5.1 Structural Approaches to Modelling Causality**

In the ‘all causes’ structural model, a range of outcomes \( y \) in different states \( s \) are derived from the deterministic mapping of various causal input variables to outputs. Following Heckman (2005), the functional relationship can be written as:

\[
y(s) = g_s(x, u_s) \quad (5.2)
\]

where \( x \) is the set of observable variables and \( u_s \) is the set of unobservable variables. In this case, \( x \) and \( u_s \) are fixed variables and are specified by the relevant economic theory. The function \( g_s \) maps the observable and unobservable inputs into possible ex post outcomes. Consequently, all of the outcomes \( y_s \) in equation (5.2) are explained in a functional sense by the arguments of \( g_s \). Multiple outcome equations may be observed in such a model.

A key property of a causal model that comprises multiple outcome equations as defined in (5.2) above is so-called ‘invariance’. This refers to the requirement that the function \( g_s \) be ‘stable’ or ‘invariant’ to changes in the values of either \( x \) or \( u_s \). Put differently, the structure of the model itself must not change in response to changes in the proposed economic environment. So, for example, in evaluating the effect of introducing a price cap on a monopolist’s price, the model would need to maintain the assumption that the monopolist’s cost function remains unchanged in both the price-cap and non-price-cap scenarios.

The ‘deep structural’ approach effectively unbundles the different components of \( g_s \) and maps each component into a range of outcome values \( y \). More specifically, some aspects of the policy regimes, characterised by the components, are explicitly modelled in terms of component distribution over the policies evaluated. In these models \( s \) represents a map of possible variations in treatments which are applied to the generating inputs \( q_s, x \) and \( u_s \):

\[
y(s) = g(q_s, x, u_s) \quad (5.3)
\]

where \( q_s \) is the observed characteristics of the treatments, \( x \) is the set of observable variables, and \( u_s \) is the set of unobservable variables.

An advantage of the deep structural approach, such as presented in equation (5.3) is that it allows for the generation of counterfactuals across treatments from a base set of characteristics \( (x, u_s) \). In this way, these types of deep structural models are seen as providing the basis for solving policy problems that seek to examine the impact of new policies (treatments). So, for example, if a new policy intervention is characterised by known transformations of variables \( (q_s, x, u_s) \) that can be mapped and lie within the domain of \( g \), then it is conceptually possible to estimate the outcome(s) of that policy \( y(s) \).

The main criticism of the structural approach, in its traditional form, is that it requires strong assumptions, particularly in relation to functional form and exogeneity. While efforts have been made in recent years to address these concerns, for example by weakening the parametric structures of some of these structural models, the complexity and multi-dimensionality of many of the structural models have led some researchers to focus on simpler methods of conducting counterfactual analyses (such as the treatment-effects approach).

As conventionally applied, the structural approach to modelling causality involves the representation of an endogenous variable(s) as a function of other exogenous (and endogenous) variables and error terms. The structural approach is further discussed in chapter 8.

**5.3.2 Treatment-effects approach**

The conventional treatment-effects approach to causal inference draws upon statistical work in relation to randomised experiments. It seeks to apply these methods to more complex settings, including observational settings where random experiments are not
possible and thus quasi-experimental techniques must be used. When quasi-experimental techniques are used the treatment group is usually selected on the basis of particular criteria (for example, regulation may be applied to all firms in an industry) rather than randomly. A control group must then be chosen (again non-randomly) to match the treatment group as closely as possible. As the groups are not randomly selected and assigned, the results of quasi-experiments may be subject to selection bias. In other words, the observed differences between the treatment and control groups may be influenced by initial, unobservable differences between the two groups which are not cancelled out by the randomisation process.

It is assumed that the relevant units of analysis (individuals, firms, etc.) can be exposed to only one of two alternative states, but that each individual could, a priori, have been exposed to either state. Critically, each state is characterised by a distinct set of conditions – such as different policy environments – exposure to which potentially affects the outcome of $Y$. In the literature these two states are sometimes referred to as ‘treatment’ and ‘control’, and the units of analysis are ‘randomly’ assigned to be in one of the two states.

The key assumption of the counterfactual framework underlying the treatment-effects approach is that individuals assigned to these treatment and control groups have potential outcomes in both states; the one in which they are actually observed, and the one in which they are not observed (the counterfactual state). In formal terms, let $Y'_i$ and $Y^c_i$ equal the potential outcomes for each unit of analysis $i$ that would result from the exposure to the treatment (t) and control (c) conditions respectively. The causal effect of the treatment on the outcome for each unit of analysis (labelled as $\delta_i$) is defined as the difference between the two potential outcomes in each of the states (Winship and Morgan, 1999):

$$\delta_i = Y'_i - Y^c_i \quad (5.5)$$

Table 5.1 summarises several quasi-experimental designs for conducting ex post evaluations. These may form the basis of a treatment-effects approach. The reduced-form approach evaluates the causal effect of a treatment on a unit of observation by estimating the difference between the value of the outcome if the unit is treated, and the value of the outcome if the unit is not treated. For example, Design A and Design C both use before-after observations from the treated population only. Design A requires cross-sectional data from a single point before the treatment and cross-sectional data from a single point after the treatment. In contrast, Design C uses time-series data of the performance indicator before and after the treatment, and thus takes account of trends in the performance in estimating the impact. In these designs, the observations ‘before’ the treatment act as the counterfactual (see chapter 5). In Design A, the control group (counterfactual) is represented by the value of the dependent variable before the treatment. In contrast, in Design C, the counterfactual is based on the assumption that the rate of change of the dependent variable before the treatment would continue.
Table 5.1: Selected designs for quantitative evaluation – based on Venetoklis (2002)

<table>
<thead>
<tr>
<th>Design</th>
<th>Description</th>
<th>Treatment group</th>
<th>Control group</th>
<th>Value of indicator before treatment</th>
<th>Receipt of treatment</th>
<th>Value of indicator after treatment</th>
<th>Estimated impact $\Delta Y$</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>Before/after treatment comparison group design – one group, non-random assignment of treatment, cross-sectional observations</td>
<td>$Y^T_{t-1}$</td>
<td>$Y^C_{t-1}$</td>
<td>$P^T_1$</td>
<td>$Y^T_{t+1}$</td>
<td>$Y^T_{t-1} - Y^T_{t+1}$</td>
<td></td>
</tr>
<tr>
<td>B</td>
<td>Before/after treatment comparison group design – two groups – treatment and control, non-random assignment of treatment, cross-sectional observations</td>
<td>$Y^T_{t-1}$</td>
<td>$Y^C_{t-1}$</td>
<td>$P^T_1$</td>
<td>$Y^T_{t+1}$</td>
<td>$(Y^T_{t+1} - Y^T_{t-1}) - (Y^C_{t+1} - Y^C_{t-1})$</td>
<td></td>
</tr>
<tr>
<td>C</td>
<td>Before/after treatment comparison group design – one group, all treated, longitudinal observations.</td>
<td>$Y^T_{t-3}, Y^T_{t-2}, Y^T_{t-1}$</td>
<td>$Y^C_{t-3}, Y^C_{t-2}, Y^C_{t-1}$</td>
<td>$P^T_1$</td>
<td>$Y^T_{t+1}, Y^T_{t+2}, Y^T_{t+3}, \ldots$</td>
<td>$A - B$</td>
<td></td>
</tr>
<tr>
<td>D</td>
<td>Before/after treatment comparison group design – two groups, non-random assignment of treatment, longitudinal observations.</td>
<td>$Y^T_{t-3}, Y^T_{t-2}, Y^T_{t-1}$</td>
<td>$Y^C_{t-3}, Y^C_{t-2}, Y^C_{t-1}$</td>
<td>$P^T_1$</td>
<td>$Y^T_{t+1}, Y^T_{t+2}, Y^T_{t+3}, \ldots$</td>
<td>$[(A - C) - (B - D)]$</td>
<td></td>
</tr>
</tbody>
</table>

Note: a. For simplicity, the treatment is assumed to be administered at a single point in time.

Notation:
- R is random assignment to treatment and control groups
- NR is non-random assignment to treatment and control groups
- T is the treatment group receiving the policy intervention
- C is the control group not receiving the policy intervention
- t–1 is the time period just before the intervention.
- t+1 is the time period just after the intervention
- P is the intervention/treatment
- Y is the value of the indicator of interest
- A is rate of change of $Y^T$ after the treatment
- B is rate of change of $Y^T$ before the treatment
- C is rate of change of $Y^C$ after the treatment
- D is rate of change of $Y^C$ before the treatment
Design A and Design C both assume that the ‘before’ and ‘after’ groups would be identical in the absence of the treatment. Thus it is assumed that no other exogenous factors than the treatment itself have influenced the value(s) of the dependent variable. This is a very strong assumption which is unlikely to apply in practice. Thus the results of evaluation studies based on this design may give misleading results to which significant qualification may be required. For example, in reviewing road safety studies, the Institute of Transportation Engineers (2009) considers that ‘naïve’ before-and after studies that measure the change in number of accidents before and after the implementation of safety treatment should not be recommended. Nevertheless, it may be necessary to use Design A or Design C if suitable data on a control group are unavailable. For example, a policy may affect all members of the population (e.g., it is country-wide). In another instance, data on unregulated firms that may have formed a control group may not be available. Alternatively, data on a control group that was not exposed to the treatment may be available. Members of the control group should have key relevant characteristics as similar as possible to those of the treatment group. The treatment and control groups may be compared both before and after the treatment. Where data are limited, the two groups may only be compared after the treatment. However, as the treatment and control groups have not been randomly assigned, there will inevitably be additional explanations – other than the treatment – for observed differences between the groups. These possible explanations must be kept in mind when drawing conclusions about the effect of a policy on observed outcomes. Designs B and D are examples of this approach.

Using Design B as an example, the net effect of the policy intervention could be calculated as:

$$\Delta Y = (Y_{t+1}^T - Y_{t+1}^C) - (Y_{t+1}^C - Y_{t+1}^C)$$  \hspace{1cm} (5.4)

Equation (5.4) is known as a difference-in-difference impact estimator. It measures the difference between the treatment and control groups of their average change in terms of the dependent variable, before and after the treatment. Design B is illustrated by Card and Krueger (1994)’s cross-state difference-in-difference analysis of the effect of the change to minimum wages in New Jersey.

The choice of quasi-experimental design may depend on the type of policy being evaluated, but will also often be dependent on the availability of data. Where a policy only affects a portion of the population (for example it may be a regional policy) and if data on a control group are available, the difference-in-difference designs, Design B and Design D, are preferable to Design A and Design C. Design B is based on before-after cross-sectional observations of the performance variable in both the control and treatment group. In contrast, Design D is based on before-after longitudinal observations of the dependent variable in the control and treatment groups. In Design B, the counterfactual is represented by the change of the dependent variable for the control group. It assumes that, absent the intervention, the treatment group should have the same level of change as the control group over time. In Design D, the counterfactual is based on the assumption that absent the intervention, the rate of change of the dependent variable for the treatment group after period $t$, would be same as the observed rate of change of the dependent variable for the control group after period $t$.

In policy evaluations designed to examine the impacts of past interventions, the relatively parsimonious approach of the treatment-effects method may be better suited,
and more convincing, than the explicit structural models which require a large number of assumptions.

However, the approach has sometimes been criticised for not clearly stating the economic question that the estimated parameters are seeking to answer. In particular, it is noted that the estimators for ‘causal effects’ in the treatment-effects approach tend to make implicit behavioral assumptions that are not fully exposited. Consequently, this approach may be of limited usefulness when examining how the estimated ‘effects’ in one environment will translate to new environments, or in estimating the ‘effects’ of programs never previously implemented.

Furthermore, this approach may provide little information about the causes of outcomes. For example, if there are differences between the group of interest (that is subject to regulation) and the control group (that is unregulated), a range of factors, including the approach of the regulator, could drive the observed differences.

In its simplest form, the treatment effects approach evaluates the causal effect of a treatment on a unit of observation by estimating the difference between the value of the outcome if the unit is treated, and the value of the outcome if the unit is not treated (Heckman, 2001). In most applications, the average causal effect on a population, which is the average difference between treated and untreated outcomes across all units in a population or some relevant sub-group, is estimated. Alternatively, the approach may focus on the average or the marginal impact for those treated. The average treatment effect on those treated measures the average difference between the observed outcomes for those receiving the treatment and the outcomes that they would obtain if they have not been treated. The marginal treatment effect on the treated is the marginal effect of the treatment on the outcome, and can be measured in a parametric non-linear model like probit or Tobit model.

As conventionally applied the treatment-effects approach to policy evaluation begins by identifying an outcome of interest (Y). Examples of Y might be the level of GDP, factor productivity, prices, investment, etc.
The structural approach derives hypothetical outcomes on the basis of economic theory and structural economic models. In contrast, the treatment-effects approach overcomes the problem of not being able to observe the outcomes of individuals in states to which they have not been exposed by using observations of outcomes of individuals who have been exposed to a particular state. These observed outcomes act as the counterfactual of the outcomes that would have been observed had the individual been exposed to the alternative state. By comparing these outcomes, the effect of the policy or program can be identified.

The simplicity of the treatment-effects approach to estimating causal effects and counterfactual outcomes is intuitively appealing. However, the approach makes a number of implicit assumptions which may impact on its usefulness. Important among these is the so-called ‘stable unit treatment value’ assumption – that the change in

---

**Box 5.2: A Method for Estimating Causal Effects**

One method for estimating causal effects in longitudinal models is the so-called ‘interrupted time series’, which employs standard time-series methods on multiple observations over time for a single unit in order to estimate a causal effect of a variable. This approach assumes that the treatment is introduced at a specific time (k) and has an immediate impact. The purpose of the analysis is then to estimate how the dependent variable would evolve over time in both the presence and absence of the treatment effect.

Formally, this type of model can be represented as follows:

\[ Y_t = b_{0t} + T_k b_{1t} + e_t \]  \hspace{1cm} (5.6)

In this case note that both the intercept \( b_{0t} \) and the treatment effect \( b_{1t} \) can potentially vary over time. More typically it is assumed that both the treatment and control conditions grow linearly over time, implying that the differences \( (b_{0t+1} - b_{0t}) \) and \( (b_{1t+1} - b_{1t}) \) are constants. One drawback of this approach is the assumption that the future is sufficiently like the past and that the past can therefore be used to estimate how \( Y_t \) would have evolved in the absence of treatment.

The relationship modelled in equation (5.6) is depicted graphically in Figure 5.1. Note that in this example the treatment has resulted in both a shift in the intercept and a change in the slope of \( Y_t \). The long dashed line presents the predicted evolution of \( Y_t \) in the absence of the treatment.

**Figure 5.1: Graphical Representation of the Treatment Effect**

The structural approach derives hypothetical outcomes on the basis of economic theory and structural economic models. In contrast, the treatment-effects approach overcomes the problem of not being able to observe the outcomes of individuals in states to which they have not been exposed by using observations of outcomes of individuals who have been exposed to a particular state. These observed outcomes act as the counterfactual of the outcomes that would have been observed had the individual been exposed to the alternative state. By comparing these outcomes, the effect of the policy or program can be identified.

The simplicity of the treatment-effects approach to estimating causal effects and counterfactual outcomes is intuitively appealing. However, the approach makes a number of implicit assumptions which may impact on its usefulness. Important among these is the so-called ‘stable unit treatment value’ assumption – that the change in
treatment status of one unit/individual does not affect the treatment status of another unit/individual – that is, there is no interference across treatments. When such complex effects are present, the powerful simplicity of the counterfactual framework lessens. The stable unit treatment value rests on two additional assumptions – that outcomes for the treatment and control groups would be the same in the absence of the treatment, and that the potential effect of the treatment is the same for both the treatment and control groups. These assumptions are crucial to enable differences between the treatment and control groups to be attributed to the treatment rather than to other ‘causes’ which are not explicitly considered.

An equally important implicit assumption relates to the method by which the units (or individuals) are assigned to the treatment and control group populations. Ideally, individuals should be randomly assigned to either group. However, this is not the case with observational data, where, by definition, assignment to a treatment or control group is not based on an explicit randomisation scheme (like in experimental studies). In most cases, treatment assignment will be correlated with a potential outcome variable, and consequently the standard estimator will usually yield an inconsistent estimate of the true treatment effect when it is applied to observational data. This is a particular problem for the evaluation of regulatory reforms where the application of the ‘treatment’ – the regulation – is explicitly linked to characteristics of particular firms or sectors (e.g., the presence of natural monopoly, market power, etc – see chapter 3). Several methods have been developed to address the problems that arise when the assumption that outcomes in the treatment and control groups would be the same in the absence of the treatment no longer holds. These include analysis of covariance and matching, regression discontinuity design, propensity scores and instrumental variables. Winship and Morgan (1999), Blundell and Costa Dias (2000), Cobb-Clark and Crossley (2003), and Heckman et al. (1999), provide a more detailed discussion of those techniques.

5.3.3 The counterfactual-histories approach

A final approach to counterfactual analysis is what might be loosely termed the ‘counterfactual-histories’ approach. In simple terms, this approach to the counterfactual is concerned with highlighting contingency in history, and in particular, the sensitivity of historical developments to specific events.

This approach to counterfactual reasoning is an established, albeit sometimes controversial, method among historians and political scientists (Ferguson, 1997). There are generally two common attributes of the method used in this approach. The first is identifying the so-called ‘point of divergence’, with a focus on examining the mechanisms by which the counterfactual, or alternative, world might be ‘tipped away’ from the actual world. The second is examining how the alternative counterfactual world path unfolds following this initial breakpoint. In all cases, the overarching purpose of the exercise is to ascertain the importance of this incident or breakpoint on subsequent events or on specific outcomes of interest.

The counterfactual-histories approach has been criticised by some commentators for being ‘unduly speculative’ and not meeting the standards of scholarly research (see for example, Carr, 1990). However, in the view of proponents, good counterfactual historical reasoning is not an act of unconstrained imagination, but involves plausible reasoning grounded in indirect evidence. Such indirect evidence that can include: ‘laws of nature’ or theory, rationality (such as game theory), causal analysis, or in some cases, even generally accepted ‘folk theorems’ (Bunzl, 2004). In economics it has been
suggested that there are additional methodological constraints which can ground the reasoning (McCloskey, 1978).

Proponents of counterfactual history also argue that the difference between factual and counterfactual reasoning can be exaggerated, and that counterfactual arguments, like any ‘factual’ historical argument, are only as compelling as the logic and evidence offered by the researcher to substantiate the links between the hypothesised antecedent and its expected consequences (Lebow, 2000). Additionally, in both factual and counterfactual work, when the level of analysis moves from the individual level to the ‘systems’ level, the balance between evidence and inference shifts decisively in the direction of the latter. In both cases, this typically requires the introduction of the so-called ‘structural arguments’ that are built on a chain of inference using assumed behavioural principles as anchor points to connect the reasoning.

The new economic history revolution since in the mid-1960s has seen the application of economic theory, econometric techniques and other quantitative methods to the study of history, known as the ‘cliometric’ or ‘new economic history’ literature. Generally speaking, this involves the application of economic theory and quantitative methods to try and explain specific historical instances of economic or institutional change. In this way it has been contrasted with the traditional approaches of economic history which were typically qualitative in nature, and based largely upon the established methods of historical research.

Cliometrics is seen by some economists to aid the construction of ‘solid economic theory’, by giving theoreticians access to large historical datasets which can allow for trends to be revealed over time. In historical work its contributions are seen as two-fold. First, it provides a method of bringing out new quantitative facts through the application of economic theory to historical events and observations which have previously been considered in a different way. Second, it permits the re-interpretation of some important aspects of economic history on the basis of economic theory. In this way it can give history a greater coherence (McCloskey, 1978).

In terms of scope, the cliometric approach is sufficiently broad to allow the method to be applied to a wide range of economic contexts. It has been applied to address issues of macroeconomic interest, such as the history of capitalism and slavery (Conrad and Meyer, 1958; Deane and Cole, 1964), as well as more specific issues such as the effect of the convergence of interest rates in the United States, the economic effects of the removal of import tariffs (Smiley, 1975), and the effects of the establishment of the US Interstate Commerce Commission on railroad competition (Ulen, 1977).

Most famous among cliometric studies are those of Nobel laureates Robert Fogel and Douglass North, both of whom focused on the effects of changes in transportation pricing. Fogel (1964) measured the contribution of the railways to American economic growth. The method of this work involved estimating the impact on national income of the development of a counterfactual transport system without railroads. Fogel found that, without railroads, national income would have declined by at most 2 per cent in 1890, and that the use of canals would have resulted in a similar change to growth. North (1968) focussed on the impact of institutional changes on productivity (particularly in ocean shipping), and specifically on how institutional changes could have a larger impact on productivity than technological changes.

The new economic history has been criticised on the basis of the quality and reliability of the historical data to which it is applied, as well as the implicit assumptions that are
made about the market and contextual environment in which certain historical break-
points occurred (Stone, 1977). Nevertheless, cliometrics should not be considered as a
separate approach to counterfactual as it essentially employs econometric techniques to
answer economic history questions (Angrist and Pischke, 2008).

5.4 Strength and Weaknesses of Structural and Treatment Effects Approaches

In applied work, the approach to the counterfactual is interdependent with the research
question, decisions regarding the evaluation design and method of analysis (see chapter
4), as well as the availability of data and the resources available for the evaluation
exercise. For example:

- If data on a control group are unavailable then it may not be possible to use the
treatment-effects approach that applies differences-in-differences analysis. On
the other hand, the treatment-effects approach (for example a simple before-and-
after comparison) may be necessary if insufficient data are available to construct a
robust structural model and counterfactual.
- Development of a full structural model may not be warranted by the research
question or by the resources available to the evaluation team, in which case a
treatment-effects approach may be preferable.

For the purposes of evaluating recent and ongoing competition and regulatory reforms,
the choice of approach to the counterfactual will typically be between the structural and
treatment-effects approaches. The cliometric and counterfactual histories approaches
are usually associated with evaluations of superseded reforms or policies. Regardless
of the approach to the counterfactual that is ultimately adopted, an understanding of the
strengths and weaknesses of each of the two main quantitative approaches is important
to understanding the robustness of the results that are derived from each approach.

The main apparent advantages of the structural approach to the counterfactual are
similar to the advantages associated with the structural approach to evaluation:

- The approach explicitly models the relationship between the policy and the
outcome in accordance with the underlying economic theory. How different
components of the policy can affect the outcomes can be examined under the
structure models.
- The approach explicitly accounts for the impact of unobservable economic and
behavioural parameters ($u$) on outcomes, and therefore allows for the sources of
potential variability among observationally identical coordinates to be
incorporated into the modelling. This can be contrasted with the conventional
treatment-effects approach where the failure to explicitly account for
unobservables gives rise to concerns about selection bias and problems of
inference.
- Structural models for counterfactuals are based on defined economic parameters
that are policy-invariant. Thus the functions can be varied hypothetically to
examine a range of potential outcomes. In contrast, the treatment-effects
approach focuses primarily on one, or a narrow range, of effects (Heckman,
2001).

On the other hand, there are a number of well-recognised limitations of the traditional
structural approach to the counterfactual. Chief among these is the problem of
functional form, and in particular, the need for strong and often deterministic
assumptions to be made about the functional form of outcome equations and the joint
distribution of the unobservables (Reiss and Wolak, 2007). For example, it is noted
that estimates of changes in consumer surplus may be strongly affected by whether it is assumed that demand is linear as compared to an assumption of constant elasticity. In addition, concerns are often raised in that structural models seek to address a large number of questions through their models and require a large number of simplifying assumptions. The models use many equations and thus place strong demand on data. Finally, there may be difficulties in translating complex economic theories into estimable relations. This can complicate an evaluation exercise and limit what can be done with the available data, and may involve the introduction of conditioning variables which are not part of the economic theory.

It is the complexity and multi-dimensionality of many structural models that have led some researchers to focus on simpler methods of conducting counterfactual analyses (such as the treatment-effects approach). Indeed, the principal benefit of the treatment-effects approach is its relative simplicity and tractability when compared to the structural approach. Its focus on only one outcome parameter, instead of many, allows the identification of that outcome under weaker conditions and with fewer potentially deterministic assumptions (Reiss and Wolak, 2007). By focusing only on a narrow question, the treatment-effects approach can potentially avoid many of the problems confronted in the structural approach which requires explicit economic models to be developed. However, the limited focus of the treatment-effects approach makes it less suitable for use in evaluating the extension of a policy into new areas, or the introduction of an entirely new policy.

On the other hand, there are a number of perceived limitations of the treatment-effects approach. The problems associated with using observational data to make causal inferences are of particular concern to some economists. In addition, the treatment effects approach does not seek to understand the factors underlying outcomes, with differences between the ‘control’ and ‘treatment’ groups usually attributed solely to the treatment. Heckman and Vytlacil (2007) provide a detailed discussion of this topic.

Recent work on treatment effects in economics seeks to address some of the limitations of the approach for economic evaluation. For example, some recent treatment-effects work has sought to modify and extend the conventional approach to incorporate additional information from outside the observed data and to address more general forecasting problems. At the same time, advances in the traditional structural model emphasise the non-parametric identification of certain relationships, allowing for different responses to treatments, robustness to new environments, and simplicity (Heckman and Vytlacil, 2007). Thus, while the two approaches are sometimes presented as polar opposites in the theoretical literature, the differences between the approaches in practice are a matter of degree rather than kind.

5.5 Applied Approaches to the Counterfactual

As noted, the approach to the counterfactual typically flows from decisions regarding evaluation design and the method of analysis as well as data and resource availability. The following discussion of the approaches that have been taken to the counterfactual in ex post evaluations of competition and regulatory reforms seeks to illuminate the preceding theoretical discussion of approaches to the counterfactual and the strengths and weakness of the varying approaches, and to provide practical guidance as to how the approaches can be applied.

In the literature, there are two general types of applied studies that have employed counterfactual reasoning to assess the effect of competition and regulatory policies. The
first has been undertaken by applied academic economists and has focused on evaluating the impact of various policy initiatives – for example, the welfare impacts of privatisation of telecommunications and/or electricity. The purpose of these studies is typically to test economic theories by applying and extending existing empirical techniques. Criticisms of these academic studies focus on the extent to which a study has extended the existing literature, rather than on the adequacy of the underlying data and simplifying assumptions made to deal with those problems. Thus, some of the approaches of these academic economists may not be well suited to evaluations undertaken by governments or their agencies. For instance, the academic researchers typically simplify the analysis by assuming that the counterfactual scenario follows from a single step/antecedent, even if the policy has actually proceeded in multiple steps. While this simplification makes the analysis more tractable; it can influence the results of the analysis and researchers may wish to explore whether it is possible to match the path of the counterfactual to that of the policy being evaluated. Nevertheless, the academic literature can provide some very useful insights, particularly about the application of rigorous research methods.

The second type of study has been undertaken by government bodies, regulatory agencies, or by economic consultants commissioned by those bodies. These studies include evaluations of the performance of both regulatory policies and agencies. The studies are often used directly in the policy-making process, and thus may be less concerned with academic rigor than with due policy processes, including the need for consultation with key stakeholders. Criticisms of these studies tend to focus more on the assumptions made in conducting the analysis, the limitations of the data and the robustness of the findings and thus the validity of any policy recommendations that are made on the basis of those findings. These studies thus potentially provide insights into how to define defensible counterfactuals that may be subject to political scrutiny and debate.

5.5.1 **Key messages from the applied academic literature**

A number of key messages are clear from the applied academic literature that has been surveyed for this report. The first is that the use of counterfactual analysis in practice does not correspond directly with any of the theoretical approaches described above. Theoretical notions of counterfactual analysis tend to be stringent in terms of the approach undertaken, seeking to accord closely with the more general principles of econometric or statistical modelling. The theoretical approach to the counterfactual is largely quantitative in nature and therefore requires a robust understanding and treatment of various econometric or statistical techniques. In contrast, many applied examples of counterfactual analysis used to evaluate competition and regulatory reforms do not appear to accord strictly with the stringent requirements presented in the theoretical work. Although there is a wide spectrum of approaches, counterfactuals in such cases appear to be more pragmatically developed. For example, some academic studies that have attempted to quantitatively measure the effects of various major policy initiatives (such as privatisation) appear to be hybrids, combining elements of the treatment-effects and the structural approaches. Similarly, while there is much diversity in the approaches to the counterfactual in ex post evaluations of the impact or effects of regulatory policies (typically undertaken by government bodies or regulators), some of these studies appear to adopt an approach that is more qualitative in nature and more closely resembles a counterfactual-histories approach.
Furthermore, in the academic work examined for this report, a single antecedent was invariably used for developing the counterfactual scenarios. Such an approach agrees with general considerations of simplicity, as sequential stages would typically make an evaluation less tractable, and the analysis more complex. However, as will be discussed below, this approach is not always appropriate for evaluations conducted by, or on behalf of, government bodies.

An additional message is that the practical approaches typically represent trade-offs between what is theoretically desirable and that which can be practically achieved, rather than a rejection of the theoretical ideal. For example, Morrison and Winston (1999) examined the economic effects of deregulation of US transportation industries. They argue (p. 473):

> the most accurate way to measure the economic effects of deregulation in an industry is a counterfactual analysis that estimates the price, cost and service quality changes that are solely attributable to deregulation and thus would not have occurred had an industry still been regulated. Such an analysis can be complicated. … To sidestep such complications, a simpler approach is to perform a ‘time-series’ counterfactual by evaluating changes in real costs and prices over time, thus capturing the change in regulatory environment and providing current comparisons. … Although this simpler approach does not constitute a rigorous counterfactual, the conclusions we draw from it are strongly suggestive and qualitatively consistent with the few rigorous counterfactuals that have been performed for the transportation industries using less recent data.

Similarly, Galal, Jones, Tandon and Vogelsang (1994, p. 19), who examined the welfare consequences of privatisation of public enterprises in four countries across twelve industries by applying a social cost benefit analysis (SCBA) approach (see chapter 6) argue:

> It would be desirable to have what statisticians call a ‘treatment group’, in this case the enterprises that have been divested, and a ‘control group’, in this case the enterprises that have not been divested but are otherwise similar to the members of the treatment group. … This option, unfortunately, is not open to us. … It is difficult, if not impossible to find a control group against which the performance of divested enterprises could be compared.

Although Galal, Jones, Tandon and Vogelsang (1994) were not able to adopt the theoretically ideal approach, their study explicitly adopts a ‘counterfactual approach’ in contrast to other studies, such as Morrison et al. (1999), which adopt simple pre- and post-privatisation comparisons of key indicators. The approach adopted in Galal, Jones, Tandon and Vogelsang (1994) appears in many ways to be a hybrid of the treatment-effects and the structural approaches to the counterfactual. On the one hand, the study adopts a treatment-effects like approach to welfare analysis by comparing welfare under the counterfactual scenario to the estimated welfare under the actual scenario. On the other hand, in building up each of the scenarios the analysis has aspects of the structural approach, albeit a fairly pragmatic example of that approach. Their approach is considered in more detail in box 5.3.

**Box 5.3: Approach to the Counterfactual – Galal, Jones, Tandon and Vogelsang (1994)**

Specifically, the approach adopted is to posit four factors that are considered to affect welfare: changes in consumer surplus, changes in profitability, changes in inputs such as labour costs; and changes in the welfare of competitors. These parameters have been selected *a priori* in a manner similar to a structural approach, although the study is not explicit in discussing the underlying reasoning for the selection of these parameters and not others. Nor is the study explicit about the underlying causal relations, although the authors note that the general approach is to estimate causal relationships for each parameter which is then ‘built-up’.
Although the model used by Galal *et al.* (1994) was structural in nature, the approach to estimating the model’s parameters differed from the standard structural approach which typically consists of a number of simultaneous equations which collectively define the various parameters and values of the variables in the counterfactual equation of interest. In contrast, the approach adopted by Galal *et al.* (1994) was much less structured and involved the authors acquiring a detailed knowledge of each parameter in each case – based on interviews etc – and then applying judgement as to which parameters should be changed, and by how much, in the counterfactual scenario. Effectively, the changes that could be attributed to the divestiture were ‘subtracted out’ of the factual scenario to yield the counterfactual.

Built on the SCBA methodology applied by Galal, Jones, Tandon and Vogelsang (1994), a number of academic case studies, as summarised in box 5.4, have taken instructive approaches to the counterfactual problem. Each of these approaches is pragmatic and a hybrid of a number of theoretical approaches, responding to data and other practical constraints, including the resources available to conduct the evaluation. There are, however, a number of additional notable insights.

First, it is sometimes necessary and appropriate to make simplifying assumptions depending on the purpose of the evaluation and the nature of the constraints confronting the evaluation. Nevertheless, the nature of these assumptions and their likely effect on the evaluation results should be clearly explained.

Each of the SCBA case studies specified at least two counterfactuals in recognition of the considerable uncertainty surrounding the likely evolution of chosen indicators absent the policy being evaluated. Ultimately, a particular counterfactual may have been preferred but the results obtained from alternative counterfactual scenarios provided an indication of the sensitivity of the findings to key assumptions about the counterfactual. The practice of specifying a number of counterfactuals improves the transparency of the evaluation and thus enhances the robustness and defensibility of the preferred counterfactual and the results obtained.

The preferred counterfactual in each of the case studies was typically based on conservative and defensible assumptions so that net social benefits would be understated, rather than overstated. A conservative approach to the counterfactual helps to deflect potential criticism that only known ‘successes’ will be evaluated, or that the evaluator will make biased assumptions in order to achieve particular results. The latter is a particular concern if the evaluation is not conducted by a body that is independent of the policy or program being evaluated. Finally, the choice of a conservative counterfactual helps to offset the effect of any unobserved (positive) factors that have contributed to actual outcomes but have not been explicitly accounted for in the counterfactual.

Finally, to the extent possible, researchers used benchmarks to inform and provide ‘reality checks’ on their posited counterfactual. These benchmarks may have been similar firms or industries in overseas countries, or firms or industries within the country that had not been subject to reform.

Alongside such evaluations of the impact of privatisation policies on welfare, another set of studies has taken as its evaluative focus specific indicators of welfare such as prices, and therefore consumer surplus. This approach has been particularly prominent in studies that have sought to assess the effects of privatisation on electricity prices in Great Britain (see for example, Yarrow 1992; Branston and Wilson, 2009).

The method typically adopted in these studies is to compare the electricity prices observed following privatisation with a counterfactual set of prices that might have existed in the event that the electricity companies were not privatised.
The counterfactual scenario is developed in two stages. First, a pricing identity for electricity is specified which splits the electricity price into its various components. Second, each of these components is varied under a counterfactual scenario to assess what would have been likely to happen if privatisation did not occur. The components are varied on the basis of specific assumptions as to how government policy would have been in the absence of privatisation – for example, government policy in relation to coal, the most significant component of electricity prices.

Once again it is clear that this approach does not directly accord with either of the two general approaches discussed above – it has aspects of a structural approach in specifying a causal relationship between costs and prices \textit{a priori}, while at the same time it is similar to the treatment-effects approach in not being specific about the treatment of unobservables. In this regard, it could be also considered as a pragmatic and hybrid approach developed in response to the evaluation question of interest.
Box 5.4: Approaches to the Counterfactual

<table>
<thead>
<tr>
<th>Author(s) and Year</th>
<th>Approach to the counterfactual</th>
</tr>
</thead>
<tbody>
<tr>
<td>Boardman et al., 2009</td>
<td>This study adopted a variant of the ‘treatment effects’ approach to specifying the counterfactual which was facilitated by the existence of the privately owned Canadian Pacific Railway (CP) both before and after the privatisation of Canadian National Railway (CN). The factual scenario (CN under private ownership) was CN’s actual annual observed costs from 1993 to 2003. Two counterfactual scenarios were estimated – a ‘base-case’ and a ‘conservative’ scenario. The results were found to be sensitive to the assumptions underlying the counterfactual.</td>
</tr>
<tr>
<td>Wolf and Pollitt, 2009</td>
<td>This study was able to use a private-sector competitor for the purpose of specifying the counterfactual scenarios. Two counterfactuals were constructed based on different assumptions about operating costs compared with the factual scenario. The assumptions used to derive the counterfactual scenarios were deliberately conservative.</td>
</tr>
<tr>
<td>Barmack et al., 2007</td>
<td>This study specified the capital and operating costs under a Competition (factual) scenario and a Regulation (counterfactual) scenario. The Competition scenario was based on the actual evolution of the electricity system in New England as at the time of the study and projections out to 2008. The Counterfactual regulation case was constructed by adding capacity to the system in order to meet reserve margin targets using previous regulatory experience as a guidance of likely behaviour absent the policy. The capacity expansion plan under the counterfactual was based on actual expansion. The authors applied a series of logical steps to derive the counterfactual expansion plan from the actual expansion based on their assumptions about the target reserve margin and how the regulator would have behaved in the counterfactual scenario and the implications of this behaviour for capacity and type of plant. The authors acknowledged that their scenarios simplified the electricity system but considered the simplification should be sufficient for the purpose of their high-level assessment.</td>
</tr>
<tr>
<td>Pollitt and Smith, 2002</td>
<td>This study estimated three counterfactual scenarios based on different assumptions about cost efficiency gains. It also made some simplifying assumptions about the time path of efficiency gains.</td>
</tr>
<tr>
<td>Domah and Pollitt, 2001</td>
<td>This study constructed four counterfactual scenarios based on 0 per cent, 1 per cent (the pro-privatisation scenarios), 2 per cent (the central-case scenario) and 3 per cent (the pro-public scenario) reduction in operating costs under continued public ownership.</td>
</tr>
<tr>
<td>Newbery and Pollitt, 1997</td>
<td>This study constructed a counterfactual model using actual post-privatisation data and ‘best estimate’ of the likely future evolution of the industry. The predictions were based on the regulatory accounts of the restructured companies, consolidated to form a post-privatisation ‘Central Electricity Generating Board (CEGB)’ – inverted commas indicating the metaphysical nature of the construct. Two counterfactual scenarios – a pro-privatisation scenario and a pro-CEGB (government ownership) scenario based on different assumptions about the types of plant that would have been built in the counterfactual and forecast fuel prices – were considered.</td>
</tr>
<tr>
<td>Galal et al., 1994</td>
<td>This study developed detailed knowledge of history and outcomes of each particular case and applied judgement as to causes of observed changes. Information used include: comparable enterprises in home country and other countries, particularly for evidence of the impact of technological change and to provide guidance on counterfactual behaviour; qualitative information from industry participants to supplement quantitative data. It adopted a conservative approach to counterfactual specification by assuming that all observed changes in performance were exogenous unless there was strong argument to the contrary.</td>
</tr>
</tbody>
</table>
5.5.2 Lessons from the applied Government/Institutional literature

As noted above, there are additional challenges in specifying the counterfactual for evaluations conducted on behalf of government bodies because of the political context in which those evaluations are conducted. For example, in its 1995 ex ante study of the Hilmer reforms, the Industry Commission (IC) found it difficult to distinguish between reform programs that would have been implemented in the absence of a national competition policy and those that would not have been implemented. To overcome these problems, the IC included all reforms that were implemented or announced in the years immediately prior to its review and which fell within the scope of the draft national competition policy principles that existed at the time of the review. Further, it assumed that all reforms were implemented at the same time and to the same (maximum) extent and that all productivity growth observed after implementation would be caused by the reforms. The IC referred to this as the ‘outer envelope’ of reforms or an estimate of an ‘upper limit’ of the expected benefits from the reforms (IC, 1995, p. 5). This ‘optimistic’ approach contrasts with the conservative approach to the counterfactual that is typically adopted by academic researchers.

At the time of the IC review, the potential introduction of the Hilmer reforms generated considerable public and political interest, so that the IC’s approach to the counterfactual, and thus its findings, were subject to close scrutiny. In this context, Quiggin (1997) criticised the IC’s approach to the counterfactual, suggesting that it would be more appropriate to assume that, in the absence of reform, rates of productivity growth would be unchanged from their historic trend level. Quiggin’s approach is similar to the conservative approach suggested by the academic literature.

The Productivity Commission (PC, 2005b) also encountered problems specifying a counterfactual with which to evaluate the ex post effect of the NCP reforms. Its estimate of the impact of current NCP arrangements included the influence of non-NCP reforms and other factors such as technological change. This was because the PC found it difficult to specify a counterfactual that took into account the effect of those reforms relative to NCP reforms. According to the PC, this would require judgements that would be ‘contentious’. As it was not aware of any rigorous empirical basis for isolating the impact of NCP reforms, its modelling did not ‘purport to quantify the economy-wide impacts of NCP reforms in the sectors concerned’ (PC, 2005a, p. 49). Instead, the PC drew on information about productivity and service-price changes over the 1990s and assessed the economy-wide effects of those changes and their associated regional impacts. The PC considered that this overstatement of the impact of NCP reforms would at least partly be offset by the fact that not all of the benefits of NCP reforms undertaken in the 1990s would have been realised during that time.

Similar difficulties were encountered by the PC (2006) in specifying the counterfactual in its CGE modelling exercise of the potential benefits of the National Reform Agenda (NRA). The difficulties partially related to the ex ante nature of the evaluation; for example the lack of specificity about the nature and extent of the policy program, and its likely subsequent impact on productivity, prices and workforce participation. In addition, the PC encountered difficulties in disentangling the potential effects of the proposed NRA reforms from the potential effects of the NCP reforms that were already underway, and thus deciding which reforms should be included in the ‘base-case’ or counterfactual scenario. Ultimately, the base-case
mainly reflected the economic settings prevailing in the different areas associated with the NRA in the early 2000s.

Government evaluations often have to deal with how to define a counterfactual scenario when it is not obvious how much impact other key drivers of industry/sector performance would have had in the absence of the reform or policy being evaluated. For example, the ACG (2004) was commissioned by the Australian Communications Authority to evaluate the consumer benefits from regulatory reforms in Australian telecommunications services. The ACG considered that specifying a counterfactual in telecommunications services was complicated by the fact that rapid technological progress played a greater role in influencing productivity than in some other infrastructure areas. Therefore assumptions made regarding the pace of technological change in the counterfactual could have a significant impact on modelling results.

For the purposes of specifying the counterfactual, the ACG (2004) assumed that, had the 1997 telecommunications reforms not been introduced, the annual percentage rates of change in the values for the structural variables of the economy would have been the same as their historical values as calculated over the 1986-87 to 1993-94 period (ACG, 2004, p. 11). In other words, the ACG attributed all of the post-1997 technological changes that are above the historical growth rate to the telecommunications reforms. This approach is similar to that taken by the IC (1995) and is different to the approach taken in the academic-applied literature.

In contrast, the ACMA (2005) addressed a similar question as that addressed by ACG (2004) but took a different approach to specifying the counterfactual by assuming that, in the absence of telecommunications reforms, the telecommunications industry would not experience any additional effects from the reforms to those already embedded in the economy at the end of 2003-04.

Drawing on the applied academic literature, an alternative approach may be to:

- define a number of counterfactuals based on different, clearly defined, assumptions about the effect of technological progress absent the reform of interest
- present the evaluation findings based on each counterfactual
- assess the sensitivity of the findings to the assumptions regarding the counterfactual, and
- choose the most robust and defensible counterfactual on the basis of available information regarding past outcomes and outcomes in similar environments.
5.6 Conclusions

The purpose of this chapter is to identify and discuss theoretical and practical approaches to developing a counterfactual in the context of an evaluative exercise. Theoretical approaches to the counterfactual can be broadly classified under three headings – the ‘structural approach’, the ‘treatment-effects approach’ and the ‘counterfactual-histories’ approach. Broadly speaking, the ‘treatment-effects approach’ is considered most useful for assessing the impact of historical policy interventions on observed outcomes. In practice, however, the use of counterfactuals in evaluative exercises typically does not correspond directly with theoretical approaches, but often involves pragmatic adaptations in response to contextual considerations. In this respect, the exact relationship between much of the work at the applied level and the theoretical approaches is sometimes unclear.

Despite these uncertainties, there is widespread agreement on certain basic attributes that are necessary for the development of a robust, defensible counterfactual upon which the results of an evaluation can be asserted. Foremost among these are clarity and precision, required not only in relation to the elements of the counterfactual reasoning (in defining the antecedent, the consequent, and the linking causal principles), but also in structuring the evaluative exercise (relevant variables and parameters, unobservables, outcomes of interest, and the interpretative framework). The counterfactual query will also necessarily define the level at which an evaluation will occur, and will also have a temporal aspect – namely, whether the counterfactual scenario follows from a single step/antecedent or occurs over a number of stages.

The key message from the applied work is that no single approach to counterfactual analysis is more legitimate, or better or worse, than another, but rather that there appears to be significant diversity in the approaches that have been adopted in theoretical and applied economic work. Ultimately, the approach to the counterfactual must be consistent with the other aspects of the evaluation process, including the research question, the evaluation design and data availability.
Section II: Methods and Applications

This section of the report contains five chapters that discuss the specific methods that can be used to conduct \textit{ex post} evaluations of regulatory and competition reforms and policies. Four of these chapters (6 to 9) cover quantitative methods and one chapter (10) reviews qualitative methods. Most researchers rely on a mix of quantitative and qualitative methods.

The first four chapters discuss a variety of quantitative methods for \textit{ex post} evaluation:

- Chapters 6 and 7 respectively consider the methods of social cost-benefit analysis (SCBA) and computable general-equilibrium (CGE) modelling. Both of these methods are capable of taking account of the economy-wide effects of the policy or regulation under review. Therefore, these methods are particularly suited to an evaluation of policies that have large nationwide effects, or where there are substantial effects beyond the markets that are directly affected by the policy or regulation.

- Chapters 8 and 9 contain a discussion of the quantitative methods that take a more focused approach to evaluation by concentrating primarily on the direct impact of the policy or regulation under review. Each of the methods discussed in these chapters are based on a more partial-equilibrium analysis, and typically invoke a \textit{ceteris paribus} assumption. Thus these methods are well-suited to evaluations at the program level and where the indirect effects of a policy or regulation are small. Chapter 8 surveys various econometric methods that can be used for policy evaluation on a within-country or cross-country basis, while chapter 9 considers productivity-analysis methods — total factor productivity (TFP) analysis and the frontier methods, data envelopment analysis (DEA) and stochastic frontier analysis (SFA).

The section concludes with chapter 10 which considers qualitative methods for \textit{ex post} policy evaluation. These methods focus on the use of qualitative information collected from a variety of sources including face-to-face interviews and group discussion, observation, submissions and other documents. Qualitative evaluation methods are often used for process evaluations and studies of regulatory governance, including on a cross-country basis. The distinction between qualitative and quantitative methods can be blurred, and it is common, particularly in official government reviews of the outcomes of policies and regulation, for qualitative methods to supplement quantitative approaches. Further, quantitative techniques may sometimes be applied to qualitative information.

Each of the methods discussed in this section of the report has various strengths and weaknesses which are discussed in the relevant chapter. Thus, no method is universally more appropriate than any other method for the purpose of \textit{ex post} evaluation of competition and regulatory policies. Instead the choice of method, or indeed combination of methods, will depend on the circumstances of the case, may involve a number of trade-offs and be guided, among other things, by the:

- nature of the question being asked and the level at which the evaluation is being conducted
- relative magnitude of the direct and indirect effects of the policy or regulation
- resources available to conduct the evaluation, including time, human capital and financial resources
- data available for conducting the review.
6. Social Cost-benefit Analysis

6.1 Introduction

This chapter examines Social Cost-benefit Analysis (SCBA) as a method for conducting evaluations of competition and regulatory reforms, usually on an economy-wide basis. As a generic approach, cost-benefit analysis (CBA) is a method used for policy or project evaluation in both the private and public sector. It compares the net benefits with a project, policy or program with the net benefits without the project, policy or program.\(^5\) As such, CBA relies on counterfactual analysis (chapter 5) and thus the issues raised in that chapter are relevant here. The focus in private sector evaluations that use CBA is primarily on the economic costs and benefits that accrue to particular individuals or groups; these are generally measured by the cash flows associated with the project. In contrast, the use of CBA in evaluations of public projects or programs focuses on the broader concept of social benefits and costs. Hence, when used for public sector evaluations, the method is usually termed ‘social’ cost-benefit analysis to distinguish it from the use of CBA for private project evaluation. The term ‘social’ implies that the costs and benefits accrue to society as a whole, not just private transactors. This broader focus encompasses benefits and costs that are not readily measured or quantified, such as environmental impacts, changes in health and safety and externalities.

The chapter begins with an overview of SCBA and its theoretical basis as well as an overview of how SCBA has been applied to both ex ante and ex post policy evaluation. A more detailed examination of the various strengths and weaknesses of SCBA for use in policy evaluation is then presented, drawing on the lessons learned from case studies that have used it as the main method of evaluation. The chapter concludes with an assessment of when the method may be an appropriate tool for policy evaluation.

6.2 Overview of SCBA Method

SCBA measures, for a given income distribution, the net social benefits with the policy or program being evaluated and the net social benefits without the policy or program. Net social benefit is the difference between social benefits and social costs. Conceptually, the SCBA method expresses all individual costs and benefits in dollar terms (the process of ‘monetising’ costs and benefits is discussed later). Individual costs and benefits are then added to give total costs and benefits. Total costs are subtracted from total benefits to give an estimate of net social benefit (or cost if the resulting figure is negative). Costs and benefits that occur at different times can be directly compared by using a discount rate that reflects the social opportunity cost of the capital needs of a regulatory measure (Australian Government, 2007).

A project or policy is acceptable on social cost-benefit grounds if total social benefits exceed total social costs, and thus net social benefits are positive. If a choice must be made between competing options, the option that maximises net social benefits would generally be selected. An option that maximises net social benefit also maximises allocative efficiency (chapter 3). In the context of ex post evaluation, it can be said that allocative efficiency has been enhanced if a policy has generated net social

\(^5\) For a more detailed discussion, there are numerous textbooks, including Sugden and Williams (1978); Pearce and Nash (1981) and Mishan and Quah (2007).
benefits, although it is not generally possible to conclude that the policy has maximised allocative efficiency – an alternative policy or different implementation strategy may have delivered greater efficiency gains.

SCBA is widely used by governments and their agencies for *ex ante* appraisal of policy options to assess whether a proposed policy or program should be undertaken and to choose between alternative policy options. SCBA is also increasingly being used for *ex post* evaluation of the outcomes achieved by a project or policy and to assess whether an existing policy or program should be continued or modified.

### 6.2.1 Ex ante policy evaluation in Australia

The use of *ex ante* SCBA is implicit in the principles of good regulatory process that were endorsed by the Australian Government in 2006 as a way of addressing the underlying causes of poor regulation. For example, principles two and three (Australian Government, 2007, p. 2):

2. A range of feasible policy options (including self-regulatory and co-regulatory approaches) need to be identified and their *benefits and costs (including compliance costs)* assessed within an appropriate framework [emphasis added].

3. Only the option that generates the greatest net benefit for the community, taking into account all the impacts, should be adopted.

A regulatory impact statement (RIS) is a document prepared by the body responsible for a regulatory proposal, following consultation with affected parties. It formalises and provides evidence of the key steps taken as part of a good policy development process and includes an assessment of the costs and benefits of each option, and a recommendation supporting the most effective and efficient option. A RIS must be prepared for any federal regulatory proposals that are likely to have significant impacts on business or individuals, or the economy. This includes proposals to restrict competition. In short, a RIS can only recommend a restriction on competition where both the benefits to the community as a whole outweigh the costs, and the government’s objectives can be achieved only by restricting competition. For major proposals, a full quantitative cost-benefit analysis will often be appropriate (Australian Government, 2007). The Council of Australian Governments (COAG) has also agreed that governments will improve the quality of regulatory impact analysis through the use of SCBA (COAG, 2007).

To facilitate the use of SCBA in public decision making nationally, a CBA Unit has been established within the Office of Best Practice Regulation (OBPR), providing training in the use of CBA and offering technical assistance in the conduct of CBA. Information about CBA is also contained in an appendix of the *Best Practice Regulation Handbook* (Australian Government, 2007). In addition, the Commonwealth Department of Finance and Deregulation has published a *Handbook of Cost-Benefit Analysis* (Commonwealth Government, 2006). These documents provide relevant information and more detail for practitioners about the application of CBA for *ex ante* public decision making and are a useful starting point for readers who are interested in a more detailed discussion of the practical application of the SCBA method than is given in this chapter. The information contained in these publications is also applicable to the use of SCBA for *ex post* evaluation of policies and programs.
6.2.2 *Ex post policy evaluation*

SCBA can also be used to conduct *ex post* evaluations of public policies and programs. In an *ex post* context, the counterfactual query asks what net social benefits would have been if the policy or program being evaluated had not been implemented. Once costs and benefits in the counterfactual scenario have been identified, they can be compared with actual costs and benefits to determine the incremental impact of the policy (Frontier Economics, 2006).

In the Australian setting, under the legislation review and reform program, Australian governments were obliged to review all legislation that restricts competition and have an ongoing obligation to review new legislation that restricts competition. The guiding principles (clause 5(1) of the Competition Principles Agreement) are that legislation should not restrict competition unless the benefits of the restriction to the community as a whole outweighs the costs, and the objectives of the legislation can only be achieved by restricting competition.

To assist governments in carrying out the legislation review and reform program, the National Competition Council (Centre for International Economics (CIE), 1999) published review guidelines which set out a comprehensive framework for conducting NCP reviews. The NCC’s review guidelines do not explicitly mention SCBA. However, the approach is implicit in the guiding principles and framework set out in the guidelines. This included the use of SCBA.

6.2.3 *Steps in SCBA*

Regardless of the policy being evaluated and whether the evaluation is conducted *ex ante* or *ex post*, there are a number of common steps in SCBA:

**Step 1 – specify the policy options to be evaluated**

If SCBA is applied *ex ante*, there is likely to be a number of alternative policy options including the ‘do nothing’ or base-case scenario (which is the counterfactual). In an *ex post* setting, an SCBA evaluation is conducted by comparing net social benefits under at least two scenarios: the ‘factual’ scenario, or what actual outcomes are observed with the policy in place; and the ‘counterfactual’ scenario, or what outcomes would have been observed had the policy not been in place. In an *ex post* evaluation, the counterfactual scenario seeks to attribute causation, or to evaluate the extent to which *ex post* outcomes have been caused by the policy under consideration rather than by changes in other exogenous factors. As foreshadowed in chapter 5, specifying the counterfactual scenario(s) is one of the greatest challenges in SCBA.

**Step 2 – determine the boundaries of the relevant costs and benefits**

This step identifies the boundaries (geographical or otherwise) of the costs and benefits as well as the economic agents that will be included in the SCBA assessment. Often, the regulation of economic infrastructure has a national dimension or impact that extends beyond those directly affected by the regulation to encompass all residents. These indirect costs and benefits should be included in any SCBA of significant competition and regulatory policies. In contrast, if the social costs and benefits of regulatory policies fall mainly on those directly affected by the policy, a SCBA can be confined to an assessment of those direct effects. This may be the case if policy or program is confined to a state or region and the associated costs and benefits are also fairly localised. However, if regional policies have large effects that
spread beyond a region, it would be appropriate to widen the geographic scope of the SCBA.

**Step 3 – list the costs and benefits (impacts) and select indicators to measure the costs and benefits in the counterfactual and factual scenarios**

In many instances, the identification of impacts and measurement indicators is relatively straightforward, while in other instances the choice of measurement indicator will be less obvious. For example, this may be the case where a policy could lead to an improvement in quality of service. It is not immediately clear what is the best indicator to use to measure improvements in quality. In energy, for example, alternative indicators of quality of service include; the number of system interruptions, average outage time, and the number of customer complaints. The choice of measurement indicator in such instances may depend on the availability of data.

**Step 4 – estimate the effect of the policy on the chosen measurement indicators over the life of the regulatory proposal**

In the context of *ex ante* policy evaluation, the focus of SCBA is necessarily forward-looking, requiring forecasts of the expected benefits and costs of a proposal for each year of the life of the proposal, as well as the likely path of the indicators in the absence of the policy. For *ex post* evaluation, it is also necessary to ‘forecast’ the historic path of the measurement indicators in the counterfactual scenario where the policy had not been implemented. Forecasting expected future benefits and costs is complex and is likely to require increasing levels of speculation and assumption the further into the future the forecast is projected.

**Step 5 – monetise (attach dollar values to) the impacts identified in step 4**

The dollar values of the net benefits from a policy provide an indication of its welfare effects. Costs and benefits are monetised by assessing the maximum dollar amount that individuals would be willing to pay to obtain (or avoid) the impact. This amount represents the individual’s valuation of the impact, and may be positive or negative depending on whether the impact makes the individual better or worse off (a benefit or cost respectively). Market prices often provide an indication of willingness to pay as a market transaction implies that the purchaser values the item at least as highly as the price paid for the item – if the purchaser’s valuation was lower than the market price, the transaction would not occur. However, practitioners often confront problems in valuing willingness to pay or opportunity cost as the product/service is either not supplied in a competitive market, or there is not a market price at all. For instance, environmental, health and safety impacts are not usually traded in a market or readily measurable in monetary terms. There are various methods that seek to measure and value costs and benefits where market prices are not a reliable measure. These are described in more detail in Australian Government (2007). Each method seeks to estimate the so called ‘shadow price’ of an impact for which market price is an unreliable indicator. Regardless, the valuation of intangibles which are not readily measured using market prices means that subjective assumptions and judgements may have to be made, each of which may inadvertently bias the findings of the evaluation. This limitation can be minimised, however, by carefully setting out the nature of the assumptions and judgements and their rationale, and conducting sensitivity analysis on the key assumptions and judgements as set out in Step 7 below.
A measure of the impact of a proposal on society as a whole can be made by summing the individual valuations. This simple addition of costs and benefits assumes that a dollar gain to one individual cancels out a dollar loss to another. This ‘dollar is a dollar’ assumption (chapter 3) means that in SCBA evaluation, efficiency effects are separated from distributional (or equity) effects and are estimated on the basis of the prevailing distribution of income. If this focus on efficiency leads to distributional outcomes that are politically or socially undesirable, some have suggested that these undesirable outcomes should be addressed through taxes and direct transfer payments in order to preserve the theoretical basis of SCBA (Musgrave, 1969).

Step 6 – discount costs and benefits to derive estimate of net present value

The costs and benefits of a policy typically occur at different times over the life of the policy. Because of the time value of money, whereby a dollar today is valued more highly than a dollar tomorrow, it is necessary to discount the dollar value of future costs and benefits to their present value (PV) before summing the PV of costs and benefits to derive an estimate of the net present value (NPV) of a policy. The NPV, in turn, is a measure of net social benefit. The choice of discount rate impacts the size of the PV of costs and benefits and therefore the estimation of NPV and net social benefit. The Commonwealth Government’s preferred approach is to base the discount rate on market-determined interest rates, which indicate the value to the current population of future net benefits. There are, however, a number of different market-determined interest rates that may be chosen and each is likely to be controversial. Therefore, it is important to test the sensitivity of the results of a SCBA to the choice of discount rate and to take any such sensitivity into account when drawing conclusions about the effectiveness of a policy.

Step 7 – conduct sensitivity testing of the results

In order to complete the preceding steps, it will be necessary to make a number of assumptions and choices that will invariably affect the magnitude, and possibly the sign, of the net social benefits derived from the SCBA evaluation. Thus, to ensure that the findings of the SCBA are robust, it is necessary to conduct sensitivity analysis by varying the nature of the assumptions and choices. For example, the sensitivity of the results to the choice of discount rate can be assessed by estimating net social benefits using a number of plausible discount rates and comparing the results in each case. Similarly, if an economic model is used to estimate costs or benefits it may be appropriate to re-estimate net social benefits using alternative specifications or assumptions. Of course, it may be necessary to weigh the benefits of sensitivity analysis with the costs of conducting that analysis. For instance, while it may be relatively straightforward to conduct sensitivity analysis of the choice of discount rate, it may be less straightforward to conduct sensitivity analysis of the choice of economic model. Ultimately, the extent of sensitivity analysis that is conducted will be driven by the size of the economic impacts of the project/policy and the constraints under which the evaluation is conducted.

6.3 Illustrative Case Studies

The preceding discussion indicates that SCBA is conceptually straightforward and can be undertaken by following a number of clearly defined steps. In practice, there is considerable variation in the scope of SCBA, the types of policies that are evaluated using this method, and the nature of, and solutions to, the difficulties of specifying the counterfactual scenario (chapter 5); valuing impacts that are essentially qualitative or
intangible in nature; obtaining suitable data with which to measure costs and benefits over the desired time frame; choosing a discount rate, and finding that results are sensitive to key assumptions, parameters and/or choice of discount rate.

The SCBA technique has been applied on a number of occasions by academic researchers to evaluate the *ex post* impact of regulatory policies. In addition, SCBA has been used to assess whether legislation that restricts competition should be retained as part of the NCP legislation review program. These case studies provide some useful insights into the ways in which the practical difficulties that may be encountered in applying the SCBA method to *ex post* policy evaluation might be addressed, as well as the strengths and weaknesses of the method.

### 6.3.1 Academic case studies

A brief summary of selected academic case studies that have used SCBA to assess the *ex post* impact of competition and/or regulatory reforms in a number of countries and sectors is given in box 6.1. Each of these case studies is based on the method for *ex ante* evaluation of privatisation proposals using cost-benefit analysis as detailed in Jones, Tandon and Vogelsang (1990). Jones *et al* (1990) envisaged that this *ex ante* method could also be used for *ex post* evaluation of privatisation decisions, and as can be seen from box 6.1, the method has also been applied to a number of other types of competition reforms. The approach of Jones *et al* (1990) to SCBA recognises that, although the SCBA method is soundly based in economic theory, practitioners will inevitably encounter practical difficulties in applying the method. Thus their approach seeks to strike a balance between theoretical rigour and practical applicability.

The method calculates the benefits or costs of privatisation as the difference between a ‘factual’ and ‘counterfactual’ scenario(s). For the case of privatisation (as envisaged by Jones *et al* (1990)), the overall change in welfare, or the efficiency gains, can be written as:

\[
\Delta W = V_{sp} - V_{sg} + (\lambda_g - \lambda_p)Z \tag{6.1}
\]

where:

- \(V_{sp}\) is the social value of the firm under private operation (the factual scenario) as given by the present value of expected net benefits accruing to society as a whole from the private operation of the firm.
- \(V_{sg}\) is the social value of the firm under continued government ownership (the counterfactual scenario) as given by the present value of expected net benefits accruing to society as a whole in the absence of privatisation.
- \(\lambda_g\) is the shadow multiplier on government revenue. \(\lambda_g\) measures the value of $1 dollar of government revenue in terms of its consumption equivalent. In other words, a $1 increase in government revenue yields the same increase in social welfare as \(\$\lambda_g\) of consumer welfare. The shadow multiplier on government revenue recognises the possible need to adjust the various components of the welfare change if an economy is highly distorted.
- \(\lambda_p\) is the shadow multiplier on private funds. \(\lambda_p\) measures the value of $1 dollar of private funds in terms of its consumption equivalent. In other words, a $1 increase in private funds yields the same increase in social welfare as \(\$\lambda_g\) of consumer welfare. The shadow multiplier on private funds recognises the possible need to adjust the various components of the welfare change if an economy is highly distorted.
Z is the sale price.

Equation (6.1) can also be written as the change in welfare of each of the economic agents that may be affected by the privatisation and written as:

$$ \Delta W = \Delta CS + \Delta PS + \Delta Gov $$

(6.2)

where:

$ \Delta CS $ is the change in consumer surplus

$ \Delta PS $ is the change in producer surplus, and

$ \Delta Gov $ is the change in government surplus (Government revenue minus taxation).

Equation (6.2) can be expanded to allow for the welfare impact on other stakeholders such as employees and competitors to be modelled if appropriate.

The different components of welfare are estimated by constructing a simple model of costs and demand for each of the markets in which the privatised firm operates. The total change in welfare is found by aggregating the results across markets.

Further analysis of the case studies listed in box 6.1 provides some insights into how the SCBA can be applied to ex post evaluation. For example, a feature of each case study is the amount of detail that is presented about the history of the sector and firms that are directly affected by the reform or program that is being evaluated. Detail is also provided about the nature of the reform or program, including its purpose, expected outcomes and timing. A discussion of observed outcomes of key economic variables that may have been affected by the reforms is also provided, as is consideration of other factors that may have influenced the observed outcomes. All of this detail has an important role in understanding the need for intervention, specifying the counterfactual and attributing causation.

The approaches taken to the counterfactual scenario by the researchers summarised in box 6.1 were considered in chapter 5. To reiterate, the case studies suggest that specification of a credible counterfactual is of central importance to policy evaluation based on the SCBA method, but as this is a complicated task, researchers rarely adopt one of the theoretical approaches. Instead, the approach to the counterfactual is more pragmatic and often combines elements of both the treatment-effects and structural approaches. Sometimes the researchers specified a number of counterfactual scenarios to test the sensitivity of their results to the key assumptions underlying the scenario. Where the results were found to be sensitive to key assumptions, and thus there were a range of results, the preferred counterfactual, which formed the basis of the conclusions reached in the various studies, was typically chosen to be conservative.
### Box 6.1 Selected Academic Case Studies – SCBA Method

<table>
<thead>
<tr>
<th>Author(s) and Year</th>
<th>Country, Industry or Sector</th>
<th>Reforms</th>
<th>Findings</th>
</tr>
</thead>
<tbody>
<tr>
<td>Brunekreeft, (2008)</td>
<td>Germany, Electricity transmission</td>
<td>Ownership unbundling</td>
<td>Net social benefits were likely to be positive, but small.</td>
</tr>
<tr>
<td>Pollitt and Smith, (2002)</td>
<td>UK, Railways</td>
<td>Restructuring and privatisation of British Rail.</td>
<td>Estimated net social benefits of £1.1 billion although the efficiency gains from privatisation and restructuring in the first four years were more than offset by restructuring costs.</td>
</tr>
<tr>
<td>Newbery and Pollitt, (1997)</td>
<td>UK, Electricity</td>
<td>Privatisation and restructuring of the Central Electricity Generating Board in the UK in 1990.</td>
<td>Estimated a permanent cost reduction of 5 per cent per annum, or an additional 40 per cent return on assets. Consumers and the government lost from the privatisation and restructuring</td>
</tr>
<tr>
<td>Galal, Jones, Tandon and Vogelsang, (1994)</td>
<td>UK – the telecommunications, airline and trucking, Chile – electricity generation, electricity distribution and telecommunications, Malaysia – the airline, container ports and lottery industries and Mexico – airlines</td>
<td>Privatisations</td>
<td></td>
</tr>
</tbody>
</table>
Further analysis of the case studies shown in box 6.1 also indicates that the results of a SCBA can be sensitive to assumptions that affect both the counterfactual and factual scenario, such as the choice of discount rate, and assumptions about parameter values and functional forms of any models used to estimate costs and benefits. Indeed, variations in these assumptions can influence the size and sign of results obtained from the SCBA method. For instance, Wolf and Pollitt (2009) performed sensitivity analysis on seven of the key assumptions underlying their estimate of net social benefit and found that the incremental impact varied by up to 172 per cent. Similarly, Barmack, Kahn and Tierney (2007, p. 173) noted that the choice of discount rate was critical to their assessment because of differences in the timing of investment patterns between their counterfactual and factual scenarios. Thus the sensitivity of SCBA results to key assumptions, parameter values, discount rates and model specifications (if used) is centrally important. As with the approach to the counterfactual (chapter 5) such potential sensitivities generally lead academic researchers to base their preferred estimates of net social benefits on conservative, and hence more defensible, assumptions about the values of key drivers of outcomes, the parameters of economic models and the choice of discount rate used in an evaluation using the SCBA method.

Data issues

Further analysis of the academic case studies shown in box 6.1 also provides some practical insights into the use of data in SCBA evaluations. These insights are helpful as the availability of data is likely to impact all stages of SCBA, and may ultimately determine the approach taken. For example, these studies have primarily relied on publicly available data, frequently accounting data taken from published financial accounts. The researchers have recognised that data issues can constrain their evaluations and have thus taken a number of approaches to improve the adequacy of their data, and thus the rigour of their evaluation. This has been approached in three ways.

First, researchers have supplemented quantitative information with qualitative information obtained from industry participants, including views of the likely evolution of the industry or sector absent the policy. These data were used to supplement rather than replace quantitative data and to inform, challenge or confirm the authors’ observations and assumptions (Wolf and Pollitt, 2009). This approach is particularly useful where there are measurement problems associated with the available data.

Second, missing data have been estimated from available data. For instance, using a number of assumptions, Pollitt and Smith (2002) estimated total industry costs and unit costs from available data on total industry revenue and industry profits. Where data are estimated, it would be appropriate to fully document the nature of the estimation, including any assumptions made, and to conduct sensitivity analysis to assess the robustness of the results to changes in estimated values.

Third, data have been adjusted to correct for inconsistencies. For instance, Domah and Pollitt (2001) obtained data on key indicators from a range of sources and noted inconsistencies in the way in which data were reported and published. In order to use these data, it was necessary to make a number of adjustments to deal with the inconsistencies. Where data are adjusted, it would be appropriate to fully document the nature of the adjustment, including any assumptions made, and to conduct sensitivity analysis to assess the robustness of the results to the choice of adjustment method.
6.3.2 Use of SCBA in NCP legislation review and reform program

SCBA was a main method used in the Australian NCP legislation review and reform program. The process for conducting a SCBA under the NCP legislation review and reform program consisted of three steps: establish a review; conduct a review, and implement a review’s findings. The NCC’s guidelines (see above) recognised that the first step (establish a review) may be hindered by a number of factors including lack of data for analysis, limited resources to conduct a review, and measurement difficulties. In contrast to the academic case studies, the NCC’s guidelines envisaged a review being directed by a Terms of Reference and overseen by a steering committee that provided quality control and ensured the Terms of Reference were followed.

The NCC’s guidelines set out a number of steps to conduct an evaluation. These are similar to those discussed above. However, the guidelines make a number of additional useful points about the practical application of SCBA:

- The theoretical impact of restrictions on competition should be identified from ‘first principles’ using a consistent economic framework that provides guidance on the impact of removing legislative restrictions on key economic variables.

- As quantification of costs and benefits can be expensive and theoretically demanding, it may be appropriate to only quantify key costs and benefits with less important, or uncertain, parameters considered in a probabilistic manner and through sensitivity analysis.

- If quantification is not possible, qualitative and subject assessments may be necessary so that benefits and costs can be at least listed and ranked. All assumptions used to derive the list and rankings should be backed up with as much evidence as possible and the nature of that evidence made clear. Qualitative methods are discussed in more detail in chapter 10.

- The NCC’s guidelines recognised that measurement of costs and benefits would be influenced by the available data and suggest a number of data sources and methods for estimating the counterfactual including; subjective assessments of key stakeholders; international benchmarks; accounting models; engineering models; and economic models.

- Net social benefits can be found using economic models that consider changes in consumer and producer surplus. Alternatively, a simple cost-benefit accounting framework (using a spreadsheet) can be used. Regardless of the method, the results should be subject to sensitivity analysis.

This last point suggests that there can be considerable variation in the sophistication with which SCBA is applied. For evaluations of policies that have small direct effects and limited indirect effects, or where data are limited, it may be appropriate to conduct a simple ‘desk-top’ SCBA using available data set out in a spreadsheet.

In contrast, where a policy has considerable direct and indirect effects, a more sophisticated approach is likely to be appropriate. This approach may include the collection of additional qualitative and quantitative data, consideration of the views of key stakeholders and the use of economic models to provide a framework for evaluating the data and estimating net social benefit.
In contrast to the approaches taken by academic researchers, a key aspect of the conduct of a review under the NCC’s guidelines was an appropriate level of public consultation with key stakeholders and the public to provide information about costs and benefits. The NCP Review of the *Wheat Marketing Act 1989* (the WMA) in 2000 provides some insight into the potential complexity and cost involved in applying the SCBA to *ex post* evaluation of a policy that has large direct and indirect effects, as well as significant social, or distributional outcomes. The review of the WMA was conducted by an independent Committee over eight months in 2000. At the time of the review, the WMA dealt with the relationship between the regulatory authority, the Wheat Export Authority (WEA) and AWB (International) Limited (AWB). Australia was the world’s fourth largest wheat exporter and sales of export wheat contributed significantly to the Australian economy (Irving et al, 2000).

Box 6.2 contains a summary of the key aspects of the WMA review, including the approaches taken and the findings. Of particular note is the length of the review process, the extent of consultation and the sophistication of the techniques used by the review’s commissioned economic consultants, the Allen Consulting Group (ACG) to evaluate the economic impacts of the single-desk arrangements. This evaluation was an input to the broader SCBA undertaken by the Committee, which incorporated the qualitative and quantitative information supplied by stakeholders. The ACG used a range of methods, including qualitative, econometric and CGE modelling for their evaluation. Despite the application of these expensive economic tools, the Review committee found itself unable to determine conclusively the impact of the existing arrangements on social welfare. This outcome highlights the potential complexity of *ex post* policy evaluation and suggests that the use of sophisticated economic models does not always yield clearer results.
Box 6.2 NCP Review of Wheat Marketing Act 1989

The Review Committee’s terms of reference included a requirement to conduct a ‘public benefits test’ as follows:

- Analyse and quantify the benefits, costs and overall effects of the existing WMA compared to the relevant possible alternatives, including non legislative approaches.
- Identify the impact on different groups likely to be affected by either the current arrangements or the implementation of viable alternatives.

The review process was comprehensive, involving the receipt of over 3,300 submissions from interested parties and extensive consultation in Australia and overseas. Additional domestic consultation followed the release of a draft report.

The WMA’s objectives were not explicitly stated in the legislation. Thus the review committee had to infer the objectives that were sought to be achieved by the so-called ‘single-desk’ arrangements that required granting an effective monopoly to AWBI for the export of bulk wheat and restricting the choice of growers who did not wish to market their wheat through AWBI.

The Review Committee followed the process set out in the NCC’s guidelines (Centre for International Economics, 1999). It considered that the counterfactual scenario was the supply of Australian export wheat through a competitive market. There were conflicting views about the nature of the costs and benefits of existing arrangements and likely outcomes under the counterfactual. Given the breadth of the likely impact of the arrangements, the committee found it necessary to restrict the detailed assessment of costs and benefits to those associated with a number of key propositions and three key effects – a single desk price premium, an impact on market development and innovation, and an effect on grain supply chain costs.

The Committee found it difficult to measure the impact of deregulation on growers and the broader community and confronted data problems which meant that proxy indicators had to be used in some instances. In addition, there were a range of methods that could be used to evaluate the key issue of whether the existing arrangements generated price premiums. On balance, the committee considered that the introduction of more competition into export wheat marketing would likely deliver net social benefits but recommended the continuation of the single desk until 2004 so that better evidence about net benefits of the arrangements could be gathered.

To assist its deliberations, the Committee appointed the Allen Consulting Group (ACG 2000a and 2000b) to undertake a technical assessment of the economic and social effects of the WMA. The ACG presented the committee with two separate reports although the results of each were found by following a similar process. Each study was conducted using a number of analytical tools:

- for both studies, a literature review of single-desk arrangements and wheat industry trends
- for the assessment of economic impacts, a qualitative ‘first principles’ assessment of the economic impact of reform.
- for the assessment of the social impacts, the qualitative first principles assessment was based on evidence from the available literature, including existing evidence of rural consolidation.
- econometric modelling to estimate the maximum premium that AWBI could expect to earn from price discrimination based on the assumption that AWBI is a profit maximising, price discriminating monopolist.
- general-equilibrium modelling to estimate the second order impacts of removal of AWBI’s single desk using estimated first-order impacts derived from the qualitative assessment and econometric modelling. The ACG modelled alternative scenarios, including a worse case scenario, reflecting differing assumptions about the impact of the removal of single desk arrangements on the price premium and domestic efficiency impacts.

The ACG found it difficult to quantify and assess many of the claimed social impacts. Thus its modelling of these social impacts focussed predominantly on employment effects.
6.4 Strengths and Weaknesses of the SCBA Method

The SCBA method is widely used for policy evaluation and follows a fairly standardised process. It is a particularly suitable method where the policy being evaluated has large indirect effects that may not be adequately measured using partial-evaluation methods. Nevertheless, an understanding of the strengths and weakness of the SCBA method is appropriate in order to assess whether it is the best method of ex post evaluation in a particular circumstance.

The Australian Government (2007) lists the strengths of the SCBA method as follows:

- It provides quantitative information about the effects of a policy, including the positive and negative effects by identifying and measuring all social costs and benefits.
- It quantifies the impact of a policy in a standard manner, which promotes comparability and encourages a consistent approach to evaluation.
- It captures the various linkages between the policy and other sectors of the economy (for example, increased safety has or has not reduced health care costs).
- In addition, although changes in social welfare can be assessed within the context of an economic model, this does not have to be the case. Thus, in contrast to some other techniques, the application of the SCBA method does not require high-level economic knowledge.
- There is an inherent attraction in using SCBA to evaluate a policy or program ex post if SCBA was used to justify the implementation of that policy or program.
- The technique has a proven track record in ex post evaluation as indicated by its use in the NCP legislation review and reform program. This program and the case studies examined in this section indicated that SCBA is a flexible method that can be applied to a range of sectors to address a range of questions.
- The SCBA method produces a single dollar value estimate of the impact of a particular policy or program. If this is positive, then the policy or program might be labelled a success although it is not possible to conclude whether alternative options that were not adopted would have generated a larger positive number. On the other hand, if the outcome of the SCBA is a negative dollar figure, then it can be unambiguously concluded that the policy or program has failed to improve efficiency and welfare.

Despite its advantages, there are also a number of weaknesses associated with the use of the SCBA method:

- A SCBA can be costly to apply if an attempt is made to quantify all costs and benefits. However, this weakness can be partly overcome by focussing only on the key costs and benefits. Nevertheless, a trade-off may be necessary between theoretical rigour and practical necessity.
- As with all evaluation methods, SCBA can be subject to data limitations which make the measurement and valuation of costs and benefits difficult.
- There may be problems associated with valuing intangible costs and benefits. The subjective assumptions and judgements that must be used to value these costs and benefits may inadvertently bias the findings of the evaluation. This limitation is less important if a SCBA approach is used to assess a policy or
program where the key associated costs and benefits can be readily quantified. Furthermore, as the SCBA method explicitly seeks to incorporate intangible costs and benefits into the evaluation framework it may result in a more comprehensive evaluation than methods that do not explicitly consider intangible outcomes.

- It may be difficult to interpret what constitutes effective or successful reforms under this framework. For example, it is hard to say whether the net social benefit estimated from a SCBA should have been higher or lower in the circumstances and therefore difficult to determine if the value is evidence of an effective policy.

- Although SCBA has a strong micro-theoretic foundation, the theoretical welfare economics on which it is based is subject to the ‘public choice critique’ which argues that the theory ignores the political aspect of policy determination and thus makes government intervention appear more desirable than it actually is (Buchanan, 1962).

6.5 Conclusions

SCBA is a widely used method for evaluating the economy-wide effects of a policy. It is also a useful method if a policy has large indirect effects that may not be captured adequately using methods that involve only more focused evaluation. It has strong microeconomic and welfare foundations, and can be applied on a consistent basis with adjustments on a case-by-case basis to reflect the particular problems that are encountered. However, the application of SCBA will often require trade-offs between theoretical rigour and practical applicability. As with other methods where such trade-offs are necessary, it is essential that the trade-offs are carefully documented and subject to sensitivity analysis. Finally, the cost of the SCBA method is likely to vary widely depending on the nature of the policy being evaluated, the data that are available and the purpose of evaluation.
7. Computable General-equilibrium (CGE) Models

7.1 Introduction

This chapter examines the use of computable general-equilibrium (CGE) models – CGE analysis – to evaluate competition and regulatory policies. The chapter starts by describing CGE analysis and its theoretical basis. A discussion of the use of CGE analysis for policy evaluation is then provided. This discussion is informed by a number of case studies. The chapter concludes with an overview of the strengths and weaknesses of CGE analysis for policy evaluation purposes.

7.2 Overview of CGE analysis

CGE analysis – also known as applied general-equilibrium (AGE) analysis – is a quantitative method for evaluating and forecasting the economy-wide effects of policies. Specifically, CGE analysis is the use of a numerically specified general-equilibrium model for policy evaluation and forecasting.

CGE analysis is based on the theoretical Walrasian general-equilibrium framework (as formalised by Arrow and Debreu (1954)). It converts this framework, which is based on the interactions of a large number of individual consumers and producers who supply and demand inputs and final outputs, into operational models of real market economies. In the Walrasian general-equilibrium framework, prices of inputs and outputs are determined by the interaction of demand and supply. General equilibrium exists if prices are such that demand equals supply in each market (excess demand is zero), and economic profits are zero across markets. This notion of general equilibrium contrasts with that of partial equilibrium which focuses on the equality of demand and supply in a single market. As noted in the previous section, a general-equilibrium outcome arising from competitive markets will be Pareto optimal in the absence of market failure.

CGE models are a complete specification of economic interactions in an economy or region and may be built for sub-national (regional), single-country, multi-country or global analysis. Sub-national or single-country models are generally used for analyses of country-specific policy issues and proposals. Multi-country and global models tend to have less sectoral detail than single-country models and are primarily designed for analyses of multi-lateral policies such as free-trade agreements as well as trade-flows.

A CGE model may be either static or dynamic. In a static CGE model, the adjustment process from one state of general equilibrium (without the policy shock) to another (with the policy shock) is not explicitly modelled. In contrast, in a dynamic CGE model the adjustment path is explicitly modelled. Dynamic models are more realistic, but this benefit must be traded off against the additional cost and complexity of constructing such models.

CGE models vary in terms of the number of sectors, factors and the specification of international trade. Nevertheless, any CGE model consists of a large number of equations (typically in the hundreds) that represent the behaviour of all of the relevant economic agents – consumers, producers, governments – in the economy of region that is being modelled. To keep the models tractable, the economic agents are aggregated into a limited number of types of ‘households’ and production sectors. The degree of aggregation depends in part of the type of analysis and the degree of precision required. For example, in multi-country models, the level of aggregation
can be quite high, both to contain the number of equations and because only a high-level analysis of likely impacts is needed. In contrast, for sub-national or single country analysis, an existing model of the national economy might be disaggregated to provide a finer level of analysis. For example, the ORANI CGE model of the Australian economy has been disaggregated in a number of applications to more closely represent the industries and/or regions that are particularly relevant to the analysis.

Regardless of the degree of aggregation, each of the agents represented in the CGE model is assumed to demand or supply goods, services and factors of production, as a function of their respective price. Through the interaction of demand and supply, CGE models compute market clearing prices and thus determine outputs, the allocation of resources and distribution of income that is consistent with this equilibrium (Borges, 1986). Unlike partial-equilibrium approaches, CGE models capture the indirect effects of policy changes as well as the effects in markets that are directly affected by the policy.

In order to build a CGE model, it is necessary to specify production functions for each of the goods and services represented in the model as well as demand functions for each of the factors of production (inputs) and final goods and services in the model. The assumed nature of the production and demand functions can impact on the results obtained from CGE analysis. In applied CGE models it is often assumed for simplicity that production technology exhibits constant returns to scale and that consumer preferences are homothetic, although there is no particular basis for such assumptions. These assumptions help to make the underlying complex mathematical structure of a CGE model more tractable. However, if it is evident that these assumptions do not accord with real-life behaviour, the results of a CGE modelled based on such assumptions may not reflect actual outcomes.

A fully operational CGE model always contains more variables than equations. Thus, in order to solve the model, some variables must be set exogenously, or outside the model. The choice of exogenous variables is called ‘model closure’ and may be controversial.

Once the exogenous variables and the parameters of the production and demand functions are specified (Shoven and Whalley, 1984, pp. 1009-1010), the CGE model can be solved for the equilibrium prices of the endogenous variables using data contained in a social accounting matrix (SAM) or input-output table. A SAM contains information about income, expenditure and financial flows in an economy. The matrix is arranged so that entries in rows represent sellers and hence receipts, whereas entries in columns represent buyers and hence payments. There is a corresponding column for every row – in other words all economic agents are both buyers and sellers – and the totals of each corresponding row and column are equal. A SAM represents a single year and is thus a static ‘snap shot’ of an economy.

---

6 Constant returns to scale implies that if all inputs are increased by a factor ‘m’, output will also increase by the same factor ‘m’. Preferences are homothetic if a consumer is indifferent between bundles hx and hy (h>0) whenever the consumer is indifferent between x and y. The assumption of homothetic preferences means that the marginal rate of substitution between x and y is unaffected by equal proportionate changes in all quantities. The assumption also means that the income elasticity of demand is equal to 1 for all goods, or in other words, as income increases, consumption increases by a proportionate amount.
Once the equilibrium values of the endogenous variables have been determined, equilibrium quantities for the factors of production and outputs of final goods can then be determined simultaneously. A number of algorithms can be used for the numerical determination of the general-equilibrium solution (Shoven and Whalley, 1984). The resulting solution is called the ‘base-line’ solution.

For policy-evaluation purposes, the CGE model could then be modified to include the relevant policy. This involves modelling a ‘counterfactual equilibrium’ by ‘shocking’ the model with the relevant policy under consideration, and comparing this with the base-line equilibrium. A policy shock is introduced by changing the value of an exogenous variable(s).

A simplified CGE model is shown in box 7.1.

### 7.3 Development of CGE Models for use in Policy Evaluation

CGE models are computationally complex. Thus the development and application of these models has been influenced by advances in mathematics and computing technologies which enabled larger, more complex models to be solved. Key points in the history of CGE modelling include the development of the first CGE model by Johansen (1960) and the development by Scarf (1967) of an algorithm for computing a Walrasian general equilibrium. Scarf’s algorithm enabled the subsequent development of a series of models designed to evaluate tax and trade policy issues in a number of countries, including multi-country trade models (Shoven and Whalley, 1984).

Because of their complexity, CGE models tend to be purpose-built by a specialised multi-skilled team, typically within a university or large government agency (Powell and Snape, 1992). Once developed, however, the models can be fully documented and made available to external economists who are familiar with their usage. For example, the Centre of Policy Studies’ MONASH Model is fully documented in Dixon and Rimmer (2002). The Monash Multi-Regional Forecasting Model (MMRF) model is also documented. A version of the MMRF model was used by the PC to analyse the potential benefits of the national reform agenda (PC, 2006). The model can then be tailored to the particular policy issue under consideration, although tailored models typically retain the theoretical structure and assumptions of the CGE model on which they are based.

Sometimes, a national model is tailored so that sub-national or regional analysis can be performed using the national model. There are two different ways to add a regional dimension to a CGE model: top-down and bottom-up. Under the top-down approach, aggregate national results for variables are disaggregated into regional components. The top-down approach does not require new data and can accommodate a large number of regions, but can not easily model region-specific behaviour.

In contrast, the bottom-up approach to regional CGE modelling links a series of independent regional CGE models which interact through trade and primary factor flows. This type of model is very data intensive and costly to solve. Hence, in practice a more aggregated model must be used which may sacrifice some sectoral or regional detail.

Hybrid models combine top-down and bottom-up approaches. For example, MMRF-GREEN is a bottom-up model of the eight Australian states (see box 7.3).
The CGE models that have been used to evaluate policies vary widely in size, complexity and the types of policies that are assessed. Typically, however, the models have been used to evaluate policy changes that affect a large share of economic activity and have large indirect effects, or when it is important to consider changes in the sectoral structure of output, trade, demand, employment and/or prices (Devarajan and Robinson, 2002, Fossati and Weigard, 2001).

In order to use a CGE model for evaluative purposes, it is necessary to conceptualise the effect of a policy as a quantitative impact that can be applied to the model, thus generating a counterfactual scenario which can be compared to the ‘base-line’ general-equilibrium. The policy instruments traditionally incorporated into CGE models are typically tax-instruments and productivity improvements. Productivity improvements are usually simulated by a reduction in the amount of labour and/or capital used to produce a given level of output.

Once the shock has been conceptualised, it is necessary to determine its magnitude, and this must be done outside the model. One approach is to benchmark the prices or productivity of firms that had undergone reform against those that had not (Productivity Commission, 1999). The difference between the two is the predicted amount by which price or productivity performance would improve if reform was introduced elsewhere. This ‘benchmarking approach’ has been criticised, however, for comparing ‘apples with oranges’, as it does not control for all the other factors that might account for the difference in the price or productivity performance of firms that had been subject to regulatory reform compared with those that had not (Dee, 2005).

The magnitude of the improvement might also be estimated from partial studies of a specific reform (see chapter 9 for more information). As the model’s results are highly dependent on the ‘shocks’ that are imposed, the ‘quality’ of the partial studies used to estimate the magnitude of the ‘shock’ plays a key role in determining the ultimate impact of various modelled competition and regulatory reforms. Sometimes, the partial studies involve extensive research. However, sometimes modellers need to make ‘best guesses’ of the probable change in specifying a particular shock.

The policy shocks most typically analysed in a CGE framework during the 1970s and 1980s, such as expected changes in import tariffs, were relatively easy to quantify. However, since the 1990s, CGE modelling has increasingly been used to evaluate the impact (both ex ante and ex post) of other policies, including regulatory policies, which are more difficult to conceptualise and quantify as an exogenous shock. For instance, the various NCP reforms potentially influence the behaviour of economic agents by changing incentives. The extent to which it is possible adequately to represent a particular policy as a quantitative shock will influence the usefulness of CGE modelling for policy evaluation.
Box 7.1: A Simplified CGE Model

This simplified CGE model is based on Shoven and Whalley (1984). It consists of two final goods (X₁ and X₂), two factors of production (capital (K) and labour (L)) and two classes of consumers. General equilibrium is represented as a set of prices and levels of production in each industry so that market demand equals market supply for both goods, and all economic profits are zero.

In order to determine these equilibrium prices and quantities, production functions for the two final goods and demand functions for the two factors of production and the final goods are specified. It is also necessary to make a number of additional assumptions about consumer and producer behaviour and the nature of markets. These assumptions are based on the underlying general-equilibrium theory and ensure that the resulting outcome is Pareto efficient. In particular, the following assumptions are made: First, consumers and producers maximise utility and profit respectively. Second; excess demand functions are homogeneous of degree zero in prices meaning all independent variables can be increased by a constant value without affecting the value of the dependent variable (for demand, this means that if income and prices are both changed proportionately, demand will be unchanged). Third, Walras’s Law is satisfied meaning that, if all other markets in an economy are in equilibrium, then the remaining market must also be in equilibrium. (Walras’s Law also implies that the sum of excess demand across the economy must equal zero). Fourth, all product and factor markets are competitive. Fifth, relative prices simultaneously clear all markets.

In this model, closure is achieved by assuming that the fixed endowment of labour and capital for each of the two consumers is determined exogenously (represented by L and K respectively). This means that there are 12 endogenous variables to be determined by the model -- the prices of the two final goods (P₁, P₂), the prices of labour and capital (Pₐ and Pₖ), and eight quantities (the allocation of each final good between the two consumer groups (X₁, X₂) and allocation of capital and labour between the two consumer groups (K₁, K₂, L₁, L₂)).

The equilibrium equations for this simplified model can be written as:

(a) Demand equals supply for factors (K and L)

\[ K₁(P₁, P₂, Q₁) + K₂(Pₐ, Pₖ, Q₂) = K \]  
(7.1)

\[ L₁(P₁, P₂, Q₁) + L₂(Pₐ, Pₖ, Q₂) = L \]  
(7.2)

Where: K₁(P₁, P₂, Q₁), K₂(Pₐ, Pₖ, Q₂), L₁(P₁, P₂, Q₁), L₂(Pₐ, Pₖ, Q₂) are factor demands derived from cost minimising the production functions for each commodity.

K, L are fixed endowments (supply) of capital and labour respectively.

(b) Demand equals supply of goods X₁ and X₂

\[ X₁₁(P₁, P₂, Pₐ, Pₖ) + X₁₂(P₁, P₂, Pₐ, Pₖ) = Q₁ \]  
(7.3)

\[ X₂₁(P₁, P₂, Pₐ, Pₖ) + X₂₂(P₁, P₂, Pₐ, Pₖ) = Q₂ \]  
(7.4)

Where: X₁₁, X₁₂, X₂₁, X₂₂ are demands for goods X₁ and X₂ by each consumer group (1 and 2) respectively found by maximising the consumer utility function subject to consumer budget constraints.

Q₁ and Q₂ are total supply of each product, i, as given by the production function.

(c) Profits are zero in each industry

\[ PₖK₁(Pₐ, Pₖ, Q₁) + PₐL₁(Pₐ, Pₖ, Q₁) = P₁Q₁ \]  
(7.5)

\[ PₖK₂(Pₐ, Pₖ, Q₂) + PₐL₂(Pₐ, Pₖ, Q₂) = P₂Q₂ \]  
(7.6)

Equilibrium in this model is characterised by four prices, Pₐ, Pₖ, P₁ and P₂ so that equations (7.1) to (7.6) hold. Once the exogenous variables and the parameters of the production and demand functions are specified (Shoven and Whalley, 1984, pp. 1009-1010), the model can be solved for these equilibrium prices using data contained in a social accounting matrix (SAM) or input-output table (see below). Equilibrium quantities for the factors of production and outputs of final goods can then be determined simultaneously to give the ‘base-line’ solution. The model can then be ‘shocked’ to include the relevant policy. The resulting ‘counterfactual’ general-equilibrium outcome can then be determined and compared with the base-line equilibrium outcomes.
7.4 Case Studies

This section examines the lessons that can be learned from analysis of a selection of Australian case studies that have used CGE models to evaluate competition and regulatory reforms. These are summarised in box 7.2. CGE models have also been used extensively in other countries to evaluate various types of policies.

Most of the case studies summarised in box 7.2 were undertaken by the Productivity Commission (PC), or its predecessor, the Industry Commission (IC). These studies used variants of models descended from the ORANI model of the Australian economy which was purpose-built as part of the IMPACT project for ex ante policy evaluation, particularly those related to trade and industry policy (Powell and Snape, 1992). Most of these studies have involved assessment of NCP reforms. The models are developed and documented by the Centre of Policy Studies (CoPS) at Monash University. Box 7.3 contains a general description of the features of the main CGE models used in the policy evaluations summarised in box 7.2.

As noted above, in order to solve a CGE model, data in the form of a social accounting matrix or input output table is required. Input-output tables developed by the Australian Bureau of Statistics form the basis of the data-bases underlying the Australian CGE models. The various CGE models can then be solved using GEMPACK (General Equilibrium Modelling PACKage).

7.4.1 Lessons from the case studies

Analysis of the case studies in box 7.2 shows how CGE models have been used to evaluate competition and regulatory policies in Australia, and demonstrates some of the strengths and potential pitfalls of those models when used for that purpose. Each of these is discussed in greater detail below but can be summarised as:

• Existing CGE models may not be appropriate for the evaluation task at hand.
• CGE modelling is unlikely to be an appropriate method of evaluation if the evaluation must be completed within a short time frame.
• The choice of parameter values and exogenous variables influence the results of CGE modelling.
• Specification of a defensible ‘counterfactual scenario’ is crucial to derive useful results from a CGE model. This is discussed in Chapter 5.
• The task of determining an appropriate policy shock with which to develop the ‘counterfactual’ general equilibrium is a key aspect of a CGE modelling exercise.
• CGE modelling is only an adequate tool for policy evaluation if it is possible to adequately represent the policies being evaluated as a quantifiable shock to the model.
• If quantification of the policy shock is possible, the chosen quantification method can influence the results obtained from a CGE model.
• Careful checking of the modelling approach is important.
• CGE modelling may be one component of a wider policy-evaluation method such as SCBA (see chapter 6) or an approach that combines quantitative and qualitative methods (see chapter 10).

Existing CGE models may not be appropriate for the modelling task at hand.

Each of the case studies in box 7.2 started with an existing ‘off-the-shelf’ CGE model of the Australian economy. However, in each case the modellers considered that some modification of the existing model was appropriate to better represent the likely impact of the policy being evaluated and the particular focus of the evaluation.

In many instances, despite the modifications, the modellers still considered the ‘modified’ model to be inadequate for the purpose of evaluating the particular competition or regulatory reform under consideration. The developers of the ‘off-the-shelf’ models sometimes responded to these inadequacies by developing new ‘off-the-shelf’ models that were intended to be better suited to these types of policy evaluations. For example, the dynamic MONASH CGE model of the Australian economy has evolved from the comparative static ORANI model. Similarly, ‘top-down’ regional versions of the MONASH model were developed to model the impacts of NCP reforms on regions, and ‘bottom-up’ regional models evolved to deal with issues that could not be adequately addressed with a ‘top-down’ regional model.

The PC (1999a) cautioned that its modelling results should be interpreted with care as simplifying assumptions were used to incorporate the NCP reforms into the modelling framework. In addition, the modelling included estimates of the effects of output and employment of some reforms which were already implemented. The PC (1999a) also noted that its choice of model was not ideally suited to the task at hand and, thus, the regional estimates of region-specific reforms needed to be interpreted with care. In particular, its CGE model took a ‘top-down’ approach to regional analysis which, as noted above, is not well suited to tracing the regional effects of region-specific reforms. Where such effects are important, a ‘bottom-up’ modelling approach is preferable.

CGE modelling is unlikely to be an appropriate method of evaluation if the evaluation must be completed within a short time frame.

The Industry Commission (IC) (1995) was asked by the COAG in August 1994 to report in March 1995 on the potential gains from the Hilmer reforms. The IC found it necessary to exercise ‘considerable judgement’ in reaching a conclusion because of imperfections in its modelling exercise. Some of these imperfections arose from the time constraints under which the IC was operating, which meant that it had to be selective in both the choice of reforms evaluated and the way in which the reforms were evaluated. The IC also reported that it faced a number of modelling difficulties which were exacerbated by the time constraints under which it operated. For instance, it was not able to disaggregate the model to the ideal level but had to ‘make do’ with a level of industry disaggregation for which data were readily at hand using an existing disaggregated version of ORANI and data and techniques developed in some of its previous inquiries. This resulted in the HILORANI model.

Additionally, the IC did not have time to incorporate into the HILORANI database the data from an updated input-output table (for 1989-90) that had recently become available. This was of particular concern for those industries that had undergone substantial microeconomic reform since the previous input-output table was published for 1986-87.
**Box 7.2: Summary of Selected Australian Studies using CGE Modelling**

<table>
<thead>
<tr>
<th>Author and year</th>
<th>Subject</th>
<th>Model Used</th>
<th>Counterfactual/shock</th>
<th>Main Findings</th>
</tr>
</thead>
<tbody>
<tr>
<td>Industry Commission (1995)</td>
<td><em>Ex ante</em> evaluation of the growth and revenue implications of Hilmer and Related reforms</td>
<td>HILORANI (a modified version of ORANI)</td>
<td>For each sector where reform was expected to occur, estimates were made of likely changes in labour and capital productivity, output prices and other parameters. These estimates were applied to the HILORANI model to estimate the change in the general-equilibrium levels of GDP, consumption and other endogenous variables arising from the reforms relative to the counterfactual where no reform takes place.</td>
<td>In the long run, the Hilmer and related reforms were expected to lead to an annual gain in real GDP of 5.5 per cent or $23 billion per year (1993-94 dollars). Sectoral results were mixed, but all broad sectors were expected to gain.</td>
</tr>
<tr>
<td>Productivity Commission (1999a)</td>
<td>Transitional evaluation of the impact of NCP reforms on rural and regional Australia.</td>
<td>MONASH-RR, a regional version of the MONASH model</td>
<td>Examined direct initial impact of NCP reforms on rural and regional Australia by looking at a range of cost, price, service quality and employment indicators in sectors directly affected by NCP reforms using partial evaluation tools and available data. Used these partial impacts to shock the MONASH-RR model to estimate longer term economy-wide impacts.</td>
<td>At the national level, NCP reforms were expected to lead to an annual gain in real GDP of about 2.5 per cent. At the regional level, implementing NCP reform was estimated to increase gross regional product in all Australian statistical divisions except Gippsland in Victoria.</td>
</tr>
<tr>
<td>Productivity Commission (2005)</td>
<td><em>Ex post</em> evaluation of current NCP arrangements and report on future competition related reform priorities. The sectors modelled were electricity, gas, urban water, telecommunications, urban transport, ports and rail freight, with the results relating to the changes observed over the period from 1989-90 to 1999-00.</td>
<td>MONASH Multi Regional Forecasting Competition Policy Reform Model (MMRF-CR)</td>
<td>Conducted partial evaluation of productivity and service-price changes over the 1990s and used this information to model the economy-wide effects of those changes.</td>
<td>Once observed productivity improvements and service price rebalancing in the relevant infrastructure sectors had fully worked their way through the economy, Australia’s real GDP was projected to be around 2.5 per cent ($20 billion) higher than it would otherwise be.</td>
</tr>
<tr>
<td>Source</td>
<td>Description</td>
<td>Methodology</td>
<td>Findings</td>
<td></td>
</tr>
<tr>
<td>---------------------------------------------</td>
<td>-----------------------------------------------------------------------------</td>
<td>-----------------------------------------------------------------------------</td>
<td>--------------------------------------------------------------------------------------------</td>
<td></td>
</tr>
<tr>
<td>Victorian Government (2005) (Submission to PC 2005)</td>
<td>Estimate of impact of NCP reforms for Victoria in 11 areas: electricity and gas; water; rail transport; road transport; telecommunications; statutory marketing arrangements; unincorporated enterprises; postal services; federal airports/Airservices Australia; ports; and an overall review of anti-competitive legislation.</td>
<td>MMRF-GREEN Used estimates from IC (1995) and PC(1999) of the impact of NCP reforms on productivity and other economic variables to shock the MMRF-GREEN model.</td>
<td>Estimated that Australia’s GDP would be 1.5 per cent, or $11 billion higher in the long run as a result of NCP reforms. Victoria’s gross state product would be 2.0 per cent or $3.8 billion higher as a result.</td>
<td></td>
</tr>
<tr>
<td>Productivity Commission (2006)</td>
<td>Ex ante estimate of the potential economic and fiscal impacts of COAG’s proposed National Reform Agenda (NRA).</td>
<td>MMRF-National Reform Agenda (MMRF-NRA), an updated version of MMRF-GREEN.</td>
<td>The NRA and what is known about its implementation plans and detailed reform objectives guided the quantification of its potential direct impacts to form the policy ‘shocks’. These shocks were then used to model the potential maximum (‘outer-envelope’) longer-run effects of the NRA, assuming full implementation of the NRA and complete adjustment to the effects of reform.</td>
<td>It is not possible to aggregate results into a single measure of the ‘impact of the NRA’, but the NRA is expected to significantly raise activity levels and incomes in all jurisdictions in addition to benefits from ongoing reform programs. Reforms in energy, transport and infrastructure could provide resource savings of around $10 billion and after a period of adjustment, GDP could be increased by nearly 2 per cent. Governments’ combined net revenues could rise by up to around $5 billion.</td>
</tr>
<tr>
<td>Allen Consulting Group (2004)</td>
<td>Ex post estimate of benefits to consumers and small business of Telecommunications Act 1997 and Part XIB and Part XIC of Trade Practices Act 1974.</td>
<td>MONASH-TELCO which disaggregates the ‘communications industry’ in the MONASH model into two industries, ‘telecommunications services’ and ‘other’.</td>
<td>To model the ‘actual’ or base case path of the economy with the telecommunications reforms in place, ACG used available information about outputs, prices, growth rates of employment and capital and sales revenue over the period 1986-87 to 1993-94 to endogenously determine the values of structural variables that are not directly observable, such as changes in tastes and technologies. The values of these endogenously determined variables then formed the basis of</td>
<td>In 1997-98, national output was 0.16 per cent higher than it would otherwise had been if the telecommunications reforms had not been implemented. By 2003-04, this had increased to 1.25 per cent, or $10.4 billion per annum. Real benefits per household were $924 by 2003-04, or $7.1 billion in total. Total benefits to small business were</td>
</tr>
</tbody>
</table>
communications’ the shocks to the model for the period 1997-98 to 2001-02. The modelling results were the ‘actual’ case of the economy.

To specify the counterfactual path of the economy, the ACG assumed that had the 1997 telecommunications reforms not been introduced, the annual percentage rates of change in the values of the structural variables of the economy would have been the same as their historical values over the period 1986-87 to 1993-94. Thus all gains were attributed to telecommunications reforms.

ACIL Tasman for ACMA (2005)

<table>
<thead>
<tr>
<th>Estimate of economic welfare from developments in the telecommunications industry during 2004-05.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tasman-Global Telco, based on Tasman-Global model with an enhanced ‘technology’ bundle that represents developments in the telecommunication s sector.</td>
</tr>
<tr>
<td>Compared actual developments in the telecommunications sector in 2004-05 with the base-case in which the telecommunications sector did not experience any additional effects from reforms over and above those already embedded in the economy at the end of 2003-04.</td>
</tr>
</tbody>
</table>

Estimated that the Australian economy had grown by $1.97 billion more than it would have absent the reforms. But ACIL Tasman stressed that this was an ‘order of magnitude’ not a precise number.

Combined with a GDP estimate of $10.4 billion in 2003-04, ACIL Tasman estimated economy-wide benefits from telecommunications reform and subsequent market developments of around $12.4 billion in 2004-05.

Consumer benefits had increased by 0.24 per cent relative to the base case in 2004-05 or $1.3 billion. Total consumer benefits were estimated to be around $8.4 billion in 2004-05.

Small business benefitted by $216 million in 2003-04 relative to the base case, or $2.4 billion in total.
Box 7.3: Key Features of Selected Australian CGE Models used for Policy Evaluation

<table>
<thead>
<tr>
<th>Model</th>
<th>Number of Sectors/regions</th>
<th>Key Features</th>
<th>Documentation</th>
</tr>
</thead>
<tbody>
<tr>
<td>MONASH</td>
<td>113 industries and 115 commodities.</td>
<td>The model is a dynamic extension of the comparative-static ORANI model of the Australian economy.</td>
<td>The MONASH Model is fully documented in Dixon and Rimmer (2002).</td>
</tr>
<tr>
<td>MONASH-RR,</td>
<td>113 industries and 115 commodities, The states and territories are separately represented. The states are further divided into 55 statistical subdivisions.</td>
<td>MONASH-RR is a regional version of the dynamic MONASH model. Regional detail is obtained by a 'top down' approach whereby national results are disaggregated to give regional results. The 'top-down' approach requires less data and is computationally simpler than a ‘bottom-up’ approach. However, region-specific supply behaviour is not easily modelled and propinquity effects (when growth in one region benefits neighbouring regions) are ignored.</td>
<td></td>
</tr>
<tr>
<td>MMRF-GREEN</td>
<td>Up to 144 commodities/industries. The states/territories are disaggregated into 57 sub-state regions. Each region is modelled as a mini-economy.</td>
<td>A dynamic regional model developed from the comparative static MMRF model used in PC (2005). Regional detail is obtained using a ‘bottom up’ approach where results are found by linking a number of independent regional CGE models through trade and primary factor flows.</td>
<td>A technical description of MMRF-GREEN is provided in Adams, Horridge and Parmenter (2000).</td>
</tr>
<tr>
<td>MMRF3</td>
<td>MMRF3 models the Australian economy as six separate States and two Territories In each region there are 58 industries, each producing a single commodity, one representative household consumer, and one government in each state.</td>
<td>MMRF3 is an updated version of MMRF-GREEN and is based on the MMRF-NRA model used by PC (2006). MMRF3 allows for inter-fuel substitution in the generation of electricity, explicit modelling of the national electricity market, inter-modal substitution in road and rail transport. It has an updated database and improved treatment of the GST.</td>
<td>Documentation of MMRF3 can be found on the CoPS website.</td>
</tr>
</tbody>
</table>
The task of determining an appropriate policy ‘shock’ is a key aspect of a CGE modelling exercise.

In contrast to previous applications of CGE modelling to trade and tax policy issues, the IC (1995) found it difficult to adequately quantify some of the proposed policy reforms to ‘shock’ the CGE model. It noted (IC 1995, p. 1):

Some of the reforms being considered are broad strategies [for fostering a climate of improved economic performance] rather than specific policy changes; or may even have the important but intangible effect of locking in gains from changes that have already been introduced. Moreover, some of the big gains from reform are likely to be of the dynamic kind that are difficult to predict, let alone measure.

In addition, the IC (1995) was not able to model the size of the productivity gains that would be expected to flow from a better regulatory or legislative government structure for competition policy (such as improved managerial incentives for GBEs) nor the speed with which those productivity improvements would be realised. Thus, the IC had to make these judgments outside a model framework (IC, 1992, p. 31).

A comprehensive approach to determining the policy shock would take account of all available external information about the likely effects of a policy on the variable of interest and the other factors that would be expected to impact on the variable with or without the policy. In this regard, the PC’s 1999 inquiry into the impact of NCP reforms on rural and regional Australia generated considerable public interest. A key aspect of the inquiry process was the gathering of information about the views and experiences of people living and working in rural and regional communities. In addition to its standard public inquiry procedures in larger cities, information gathering was supplemented by an extensive round of visits to a variety of localities in all States and Territories during which informal discussions were held with a wide range of interested parties.

The PC (1999) examined the direct initial impact of NCP reforms on rural and regional Australia (which formed the basis of the policy shocks) by looking at a range of cost, price, service quality and employment indicators in the sectors directly impacted by NCP reforms. It also considered the factors other than NCP reforms that were likely to influence those indicators and found (PC, 1999, p. 308):

… the effects on most, but not all, regions of the NCP reforms are likely to be less significant than those resulting from the broad economic forces which are continually reshaping economic and social conditions in Australia’.

The method chosen to quantify a policy shock can influence the results obtained from a CGE model

Where quantification was possible, the IC (1995) chose ‘world’s best practice’ as benchmarks of the productivity gains that might be expected to result from the NCP reforms. The difference between existing Australian productivity and ‘world’s best practice’ was the policy shock applied to the model.

The IC’s (1995) modelling work was subject to a range of criticisms, including its specification of the counterfactual, as noted in Chapter 5. Quiggin (1997) also criticised the IC’s estimates of the direct benefits of microeconomic reform in 19 areas of the economy and contended that the use of ‘world’s best practice’ as benchmarks over-estimated most of the likely productivity gains and represented upper bounds to possible achievement rather than likely outcomes.
The PC (2006) found it difficult to quantify the likely direct impact of future regulatory and competition reforms that were not clearly specified or where the impact was difficult to determine from first principles. Thus it cautioned (PC, 2006, p. xxviii) that:

There has been limited information on which to base such estimates [of potential maximum gains from NRA reforms] and the results should be viewed as exploratory or, at best, broadly indicative.

*The choice of parameter values and exogenous variables can influence the results of CGE modelling*

The results of CGE modelling can be sensitive to the underlying assumptions about parameter values and functional forms (for production functions and consumer utility functions) and the results can be criticised on the basis of those assumptions. Thus modellers tend to use parameter values and functional forms that are simple and non-controversial (such as Cobb-Douglas and Constant Elasticity of Substitution (CES)), although this sometimes equates to the modeller’s ‘best guess’ or judgement (van der Mensbrugghe et al., 1989).

In this regard, the Victorian Government recognised the limitations of its modelling results and cautioned that these are ‘indicative only’ as (Victorian Government, 2004, p. 42):

… general-equilibrium modelling involves various simplifying assumptions and approximations, including estimations of the direct impact of the reforms in the affected industries, assumptions inherent in the structure of the model (such as substitution elasticities and industry shares) and the extent to which other features of the economy do not change in response to the reforms.

Most recently, the PC has provided the following up-front caveat (PC, 2006, p. v):

Like all model-based exercises … the results in this study reflect a range of assumptions and caveats. And they need to be seen as projections of what is potentially achievable, not forecasts of what necessarily will be achieved.

*Careful checking of the modelling approach is important*

The PC’s enabling legislation contains a number of provisions that implicitly recognise the difficulties inherent in using CGE models for policy evaluation. Under s 58(3) of the *Productivity Commission Act 1998*, if a PC report relies on formal mathematical economic modelling, the PC must either utilise at least two different economic models and make explicit reference to the assumptions and result of those models in its report, or appoint and report on the views of an independent reference panel on the modelling.

*CGE modelling may be a component of a wider policy evaluation method*

In certain circumstances, CGE modelling can have a role in identifying orders of magnitude of policy impacts and identifying broader impacts that may not be captured by partial evaluation methods.

The various caveats placed by modellers on the results of their CGE modelling, as discussed above, suggest that the usefulness of CGE modelling for policy evaluation may be enhanced if it is combined with other evaluation methods.

For example, ACIL Tasman (2005) supplemented its CGE modelling results with regional data and analysis based on detailed, sub-state data provided by carriers in assessing the impact of telecommunications industry developments. It also used
spatial data from a range of sources, including databases on demographics, industrial structure, business numbers, employment and telecommunications availability (especially relevant to terrestrial broadband). The regional analysis was used to provide a breakdown of benefits for each region by industry.

ACIL Tasman also used qualitative analysis to shed light on the quantitative results and their interpretation. Its qualitative analysis was informed by industry experts’ knowledge of recent developments in the telecommunications sector that were likely to have impacted on consumers – in particular, the internet and mobile phones.

As noted, in the previous section, CGE modelling may be a useful input to a full SCBA, if the impact of the policy is sufficiently large to justify the time and expense of CGE modelling.

Other general strengths and weaknesses

Paradoxically, some of the apparent strengths of CGE models are often closely linked to some of their perceived weaknesses. For instance, the results of CGE simulations can be unpredictable and unexpected because of the complexity of the models. This unpredictability underlies some of the criticisms of CGE models, in general, as ‘black boxes’ that generate solutions that lack transparency and hence credibility. NERA (2007) says that ‘it is consequently easy to imagine situations in which results are driven by errors in the data or in the construction of the model itself, rather than the explanatory variables’. Borges (1986) responds to such criticisms by noting the strong theoretical foundations of CGE models and the systematic manner in which the behaviour of all economic agents is modelled. In this regard, Borges (1986, p. 16) contends that:

The theoretical foundation of such models makes it possible to trace back, in every case, the simulation results and determine which factors are crucial in explaining them. … it is impossible – except for errors – that the model will lead to results which are contrary to what the underlying theory predicts.

If this defence is accepted, then a CGE model that generates unexpected results may be useful to identify general-equilibrium effects of policy changes that initially were not obvious and thus may be overlooked by partial-evaluation methods.

Nevertheless, modelling results are sensitive to the magnitude of the ‘shock’ that is applied to the base-line results, as well as the choice of which variables to treat as exogenous (the closure method). Both aspects can be controversial. The usefulness of the results derived from a CGE model may also be impacted by the level of sector and household aggregation.

Thus sensitivity analysis in relation to key assumptions is an important element of any CGE modelling exercise.

7.5 Conclusions

This chapter has examined the CGE method for policy evaluation and shown that CGE models are potentially useful for evaluating the impacts of policies that have economy-wide implications and large indirect effects that may be overlooked if partial methods of evaluation are used. Borges (1986, p. 8) suggests that CGE modelling is capable of providing ‘coherent’ answers to complicated questions in a systematic way and argues that ‘its usefulness is unquestionable for certain types of issues’. In this regard, CGE models are a ‘powerful and informative tool’ to deal with important practical issues. However, Borges (1986) cautions that the models ‘should
be developed with great care and used with prudence’, and notes that CGE modelling requires substantial resources so that it is neither a fast nor ready instrument, and is unsuited to handling ‘trivial’ problems. Powell and Snape (1992) provide an insight into the resources required to apply CGE modelling to a policy evaluation question and describe the substantial data gathering, analysis and adjustment that CGE modelling requires.

According to NERA (2007), the principal advantage of using CGE models is that they potentially provide the most comprehensive picture of all of the costs and benefits associated with a policy at a macroeconomic level. However, there appears to be general acceptance that such models should only be used when the impact on the overall economy of the policy under consideration is widespread or when the indirect effects of such policies are large (Borges 1986, p. 30). If CGE modelling is to be used for policy evaluation, it may be appropriate to supplement the results with other quantitative and qualitative information to strengthen confidence in those findings.

The review of the use of CGE models to evaluate NCP reforms highlights the complexities involved in trying to isolate the impact of a particular reform(s) from the impact of other reforms as well as the other macroeconomic and social factors that influence actual outcomes. Previous studies also highlight the complexity of separating the dynamic effects of reforms that start at different times and whose effects are not realised uniformly through time. These difficulties are inherent to the policy evaluation process. However, it is important that the choice of evaluation method does not exacerbate the problem.

CGE relies on the specification of a counterfactual and the credibility of results will be impacted by the defensibility of the assumptions underlying the counterfactual scenario. The CGE method also requires a number of other assumptions to be made which can influence the results. However, the potential impact of these assumptions can be gauged by conducting sensitivity analysis.
8. Econometric Methods

8.1 Introduction

The quantitative methods considered in this and the following chapter can be classified into two broad groups, although there is a degree of overlap between the two. The first group (this chapter) consists of those methods that use econometric techniques to make inferences about the effect of competition and regulatory reforms, either on a within-country basis or on a cross-country basis. The second group (chapter 9) is loosely termed ‘productivity and efficiency studies’ and includes the techniques of total factor productivity (TFP) analysis, and the frontier analysis methods – stochastic frontier analysis (SFA) and data envelopment analysis (DEA).

The discussion of each of the methods starts with a general overview of the method and then proceeds to consider the various strengths and weaknesses of the approach when used to evaluate competition and regulatory reforms, with reference, where appropriate, to the literature. Each chapter concludes with an appraisal of the usefulness of the relevant method for evaluating competition and regulatory reforms.

8.2 Econometric Methods for Policy Evaluation

Gujarati (2003, p. 2) describes econometrics as ‘an amalgam of economic theory, mathematical economics, economic statistics and mathematical statistics’ and provides a variety of quotes that seek to define the term ‘econometrics’ and the role of the econometrician (a practitioner of econometrics). Although the definitions vary, there is broad agreement that econometrics involves the development and application of statistical methods to explain or test economic principles and theory using data on a range of economic variables.

Econometrics is a sub-discipline within the economics sphere on which a large theoretical and empirical literature has developed (see Gujarati (2003) for an introductory explanation; and Greene (2000) and Amemiya (1985) for more advanced discussions). Econometrics uses observational data to estimate structural equations that attempt to find the statistical relationship between an independent variable(s) (X) and the dependent variable of interest (Y). Observational data are drawn from real-world situations where the assignment of individuals into treatment or control groups is beyond the control of the researcher. When the independent variable is a particular policy intervention, econometric methods can be used to make causal inferences about the effect of that policy intervention (the cause – X) on the dependent variable of interest (Y – the effect). By assuming that all other factors are held constant (the ceteris paribus assumption), the impact of the policy or program can be isolated and identified. However, the extent to which an econometric model can be used to make causal inferences is influenced by a number of factors, including the model’s specification. These issues are discussed later in this chapter.

Several sequential steps should be taken if an econometric method is used for policy evaluation.

Step 1 – formulate the question

Primarily, it is necessary to formulate the question that the evaluation seeks to answer. The answer to this question will guide the second step.
Step 2 – specify the model

Model specification is the process of identifying the dependent variable, the independent variables that are hypothesised to impact on the dependent variable and the form of the mathematical relationship between them.

A stylised cross-sectional regression model that can be used for evaluation purposes is shown in equation (8.1) (Venetoklis, 2002):

\[ Y_i = \beta_0 + \beta_1 TREAT_i + (\beta_2 \ldots \beta_n) C_i + \varepsilon_i \]  

(8.1)

where \( Y_i \) is the dependent variable of interest, or the performance indicator

\( TREAT_i \) is the policy intervention (the cause)

\( C_i \) is the set of other exogenous variables that are hypothesised to impact on \( Y_i \)

\( \beta_2 \ldots \beta_n \) is a vector of regression coefficients and \( \varepsilon_i \) is the error term

\( \beta_0 \) is the intercept term.

\( \beta_1 \) is the estimate of the average causal effect of the policy intervention (\( TREAT_i \)) on the dependent variable (\( Y_i \)). Assuming that all other relevant factors are held constant, if \( TREAT \) is binary (equal to 0 or 1), then \( \beta_1 \) explains the average difference in outcome \( Y \), between the treatment and no treatment group. If \( TREAT \) is continuous, then \( \beta_1 \) measures the partial effect of \( TREAT \) on the average value of \( Y \) given other variables are held constant.

Equation (8.1) is a stylised, single-equation linear regression model. In practice, a variety of model specifications involving multiple equations and non-linear assumptions could also be used. The interested reader should consult an econometrics textbook (such as Angrist and Pischke, 2008) for further information on the various types of model specifications.

Step 3 – identify the model

It must then be determined if it is possible to find unique values for the model’s parameters (the \( \beta \)s). There are several approaches to identifying the model.

Step 4 – gather the data

Once the model is specified, datasets are constructed for the dependent variable and the various independent variables. If suitable data are not available, it may be necessary to use appropriate proxy data. If any of the variables are not measurable with sufficient accuracy, then it may be necessary to take account of the impact of these unmeasurable variables by using dummy variables. The use of dummy variables is often necessary to represent the effect of regulation or competition policies. For example, in equation (8.1) above, the independent variable, \( TREAT_i \) may be a dummy variable that equals 1 if the unit \( i \) is exposed to the treatment (policy) and 0 if it is not. Alternatively, \( TREAT_i \) can be a continuous variable that varies according to the ‘dose’ of the treatment.

The datasets to which econometric analyses are applied can be classified as time-series data, cross-sectional data, or panel data. Time-series datasets contain observations on a particular variable over time; for example, prices observed over several years. For the purposes of evaluation, the data should ideally include measurements both before and after the policy. Cross-sectional datasets contain observations on many agents taken at a single point in time; for example, an
individual household’s expenditure on energy in a given month. Panel datasets contain both time-series and cross-sectional observations.

**Step 5 – estimate the econometric model**

The econometric model is then estimated using the appropriate econometric estimator. Single-equation regression methods – such as ordinary least squares for estimation of a linear relationship – model a single variable (the dependent variable) as a function of one or more explanatory (or independent) variables. In some contexts, single-equation methods may have poor statistical properties. Therefore, simultaneous-equation methods, such as two-stage least squares, three-stage least squares and instrumental variables may be more appropriate. Instrumental variables methods, in particular, are often used to estimate causal effects where it is necessary to use quasi-experimental designs. It is a method of regression analysis that seeks to simulate a randomisation process by addressing the problems that arise when the data do not include all factors that influence the dependent variable. In the first step, the treatment (policy intervention) variable is used as a dependent variable in a regression equation that has two types of independent variables – variables taken from the observed data and an additional variable called an ‘instrument’. The chosen instrument variable should be one that is strongly correlated with the treatment variable but is not directly correlated with the impact variable or the error term. In the second step, the predicted value of the policy intervention (the treatment variable) generated from the first regression is used as the main independent variable in another regression model which uses the policy impact variable as a dependent variable. The instrument affects the dependent variable only through the instrument’s impact on the policy-treatment variable. The resulting estimator is consistent.

**Step 6 – test the goodness of the fit**

The goodness of the fit of the estimated equation should then be tested to determine how consistent it is with the data. There are a number of statistics that can be used to determine goodness of fit, including the coefficient of determination, or $R^2$, and the chi-squared ($\chi^2$) statistic. Again, interested readers should consult an econometrics text book for further information.

**Step 7 – respecify the model**

The equation can be respecified, if necessary, to improve the fit. If re-specification is necessary, then steps 2 to 6 would need to be repeated.

**8.3 Key Issues to Emerge from the Literature**

There is a large body of literature that uses econometric methods to evaluate aspects of competition and regulatory reform on various indicators of performance (impact variables). The literature can be broadly classified as single-country case studies or cross-country analysis. The case-study literature is more directly relevant for the purpose of evaluating particular infrastructure reforms and regulation. A selection of this case-study literature is shown in box 8.1. The studies shown in box 8.1 provide an indication of the types of evaluation questions that might be answered by using econometric methods. The studies focus particularly on the:

---

7 For a more extensive review of some of the relevant literature, see, for example, Megginson and Netter (2001) and Sappington (2003).
• effect of incentive regulation on the quality of retail telephone service in the United States
• impact of regulation and competition on cable television rates, and
• impact of privatisation on operating performance in a variety of sectors.

The breadth of evaluation questions that can be examined using econometric techniques is greater than the selection of case studies shown in box 8.1. Nevertheless, the method-related issues that arise when using econometric methods are essentially the same regardless of the questions to which econometric methods are applied. The key issues are model specification, threats to validity, transparency and reproducibility, and the availability of data.

The purpose of closer analysis of the case studies in box 8.1 is to highlight the approaches to these key issues and the lessons that can be learned from the literature. This is done in the following sub-sections.

8.3.1 Model specification and threats to validity

The specification of the appropriate model is a critical aspect of using econometric methods for any purpose, including for policy evaluation. Model specification involves careful selection of both the dependent variable(s) and the set of independent variables (including the policy of interest) that are hypothesised to impact on the dependent variable.

The dependent variables – performance indicators

When using econometric methods for evaluating the effect of regulatory and competition reforms, the focus of evaluation is often on the performance of the regulated firms, sector or industry. The choice of a performance indicator(s), or dependent variables, will depend on the question(s) that the evaluation seeks to answer as well as the objectives of reform (to promote competition or to regulate monopoly power), the nature of the industry, the stage of market development, and the characteristics of the regulatory regime (NERA, 2007).

Some of the performance indicators (based on NERA, 2007) that might be used as dependent variables in an econometric assessment of the impact of competition and regulatory reforms include:

• the price of output
• production costs
• quality of service
• investment
• demand-side indicators and
• supply-side indicators.

These are outlined in more detail below. Box 8.1 contains examples of some actual indicators that have been used in evaluations of aspects of competition and regulatory reforms. In practice, the definition and measurement of a particular performance indicator may depend on the availability of data.

When competition is weak, the prices of outputs are often misaligned with efficient costs of production – they may be higher than efficient costs if output is supplied by
an unregulated monopolist, or lower than efficient costs if there are cross-subsidies. An important objective of most competition and regulatory reform is to promote prices that are aligned with efficient costs. For most regulated outputs, this means that prices are expected to fall. However, if regulation leads to the unwinding of cross-subsidies, the prices of previously subsidised output may rise. This is particularly the case in the Australian water sector. Thus an evaluation may seek to assess the impact that regulation has had on the price of output.

If a price indicator is used as a dependent variable, it is important that the underlying definition and measurement of ‘output’ and quality of output is consistent over time. If consistency is not possible, adjustments to the price dataset may be necessary. The level at which prices are measured can also be important. For example, if there are regional differences or differences across consumer groups it would be appropriate to use a price indicator that captures these differences and changes to them. An evaluation of output prices should generally be accompanied by an evaluation of production costs, quality of service, investment and innovation to ensure that the partial evaluation is as robust as possible.

When competition is weak, production costs tend to be above efficient levels. Thus, most regulatory and competition reforms seek to promote cost efficiency. An evaluation may therefore seek to assess the impact of regulatory or competition reforms on production costs at the sector, industry or firm level. Cost indicators need to be consistently defined and measured over time. Interactions between costs, quality of output and investment levels should be carefully considered.

In markets where competition is weak, an apparent reduction in costs and/or prices might be achieved by lowering the quality of the product or service being supplied. Thus, an evaluation may also seek to assess the effect of regulation on the quality of a regulated service as observed cost and/or price reductions may be achieved by a reduction in service quality. If this is the case then the welfare effects of regulatory reform are ambiguous. It will generally be necessary to use proxy variables to measure service quality.

Regulatory reforms often include a requirement to promote efficient investment in infrastructure. Nevertheless, there is frequently a debate about whether regulation actually stifles, rather than promotes, efficient investment (chapter 3). Thus an examination of the effect of regulation on investment would be a useful evaluation question. However, it is difficult to assess the efficiency of investment, requiring, *inter alia*, detailed information about the timing, location and nature of investments. The level of investment is not necessarily a suitable proxy for efficient investment as it says nothing about those issues. Therefore an assessment of investment effects that was based solely on an assessment of investment levels would need to be supplemented with additional information.

In sectors or industries where regulation seeks to promote competition in potentially competitive upstream and/or downstream markets, it may be appropriate to assess the effect that regulation has had on demand-side indicators of customer behaviour in those markets such as customer switching. It may also be fruitful to assess the effect that regulation has had on supply-side indicators of:

- market structure such as market shares, concentration ratios and barriers to entry
- supplier behaviour, such as evidence of anticompetitive conduct or misuse of market power in possible contravention of the *Trade Practices Act 1974*. 

It may be necessary to use proxy or dummy variables for some of these indicators which are more difficult to quantify and measure.

Choice of Exogenous Variables

The choice of exogenous, or control variables, is a critical issue for econometric evaluation methods as different model specifications can lead to different conclusions about the nature of the causal relationship between the dependent variable and the policy intervention. To avoid the potential for the model’s specification to be guided by the data (by constantly respecifying the model in response to poor goodness of fit), there should be an *a priori* belief that a particular independent variable would have a direct impact on the dependent variable. Economic theory has a central role to play in establishing these *a priori* expectations and in facilitating the choice of control variables. Prior empirical results can also provide useful guidance when selecting independent variables.

Econometric techniques themselves can also help to identify the relevant independent variables. For example, a multiple-regression analysis of a number of potential factors that may influence the dependent variable may be performed. This seeks to assess the *causes of effect*, or to explain the potential factors that may influence a certain outcome. The regression results help to identify those variables that are statistically significant and therefore likely to actually be an important factor in influencing actual effects/outcomes. Those statistically significant variables may then be selected as control variables, assuming this choice is consistent with *a priori* expectations. Kennedy (2003, Chapter 5) provides more information about, and a critique of, this approach.

Although it is important the model is not deliberately specified so as to best fit the available data, ultimately, the way in which exogenous variables are represented in an econometric equation may be dependent on the available data. For example, if data on the variable of interest are unavailable, it may be necessary to use a proxy variable to approximate the effect of the variable. Alternatively, if the variable of interest is unmeasurable, it may be necessary to use a dummy variable to approximate its impact.

 Threats to Validity

The term ‘threats to validity’ refers to factors that undermine an ability to make causal inferences about the impact of the policy or program on the indicator of interest (Campbell and Stanley, 1963). Threats to validity can arise in any evaluation method that is not based on an experimental design. When using econometric methods, the significance of these threats will vary on a case-by-case basis having regard, among other things, to data issues and model specification.

Threats to validity may be either internal or external. In this context, internal validity refers to the extent to which valid inferences can be made that the policy or program of interest has had the effect on the dependent variable that is suggested by the results of econometric estimation (the $\beta_1$ in equation (8.1) above) and that there are not alternative explanations for the outcome that have not been taken into account.

Threats to internal validity may arise from a number of sources. These include:

- **Omitted variables in the model’s specification** – if a relevant independent variable has been omitted from the econometric model, and there is correlation between the omitted variable and an included explanatory variable, then $\beta_1$ will
be a biased estimate of the true relationship between the outcome and the regulatory or competition reform and it is likely that there are alternative explanations for the observed changes in the dependent variables. There are a number of tests of omitted variables bias that could be performed.

- Trends in outcomes – where changes in outcomes are a function of the passage of time *per se*. This type of threat may be reduced by including a ‘time’ variable in the econometric model.

- Mismeasurement of data caused, for example, by changes in definition of or measurement of variables.

- Endogeneity of the policy that is being evaluated. Endogeneity arises when a variable that is theoretically exogenous, is in fact endogenous to the system being modelled. This means that variables that affect a particular outcome are themselves affected by that outcome. Endogeneity is a particular issue for the evaluation of competition and regulatory policies where the process of reform can be influenced by the performance of the industry that is subject to reform. In the case of regulation, it is often the case that the *method* and *application* of regulation is influenced by the performance of regulated firms. Several econometric techniques such as the use of instrumental and lagged variables and dynamic modelling have been developed to address endogeneity. However, these techniques require more and better data which may not be available. Thus trade-offs may be necessary. In this regard Jamasb, Mota, Newbery and Pollitt (2004, p.10) caution that:

  ... poor specification, or indeed single equation analyses which do not attempt to address the issue of endogeneity will give rise to indefensible empirical estimates and non-robust estimates.

The studies by Boardman and Vining (1989), Martin and Parker (1995), Hazlett (1997) and Fischer, Guitierrrez and Serra (2003) provide examples of how threats to internal validity have been addressed in the literature.

Boardman and Vining (1989) use a control (or non-treated) group of firms to draw comparisons with a treated group. The authors attempt to evaluate the impact of government ownership on performance by comparing the performance of state owned enterprises (SOEs) (the control group) with privately owned firms (the treatment group). Boardman and Vining (1989) find that after controlling for size, market share and other factors, private firms are significantly more profitable and efficient than either mixed-ownership firms or state-owned enterprises.

In contrast, the studies by Martin and Parker (1995), Hazlett (1997) and Fischer, Guitierrrez and Serra (2003) compare the performance of a single group before and after a particular treatment – privatisation in Martin and Parker (1995) and Fischer *et al* (2003), and regulation in Hazlett (1997). A weakness of these designs is the likelihood that factors other than the treatment may account for changes in the performance indicator of interest. In this regard, Hazlett (1997) finds that, despite expectations, subscriber growth to cable television services slows down during the two years of regulation. However, Hazlett acknowledges that factors other than regulation may also account for the changes in the pre- and post-regulation performance of cable markets. In addition, Fischer, Guitierrrez and Serra (2003) acknowledged that counterfactual analysis would provide a more complete evaluation of the privatisation process. In these studies, it is likely that there are alternative
explanations for the outcomes that have not been taken account of. This weakens the extent to which causal inferences can be made about the effect of the regulatory or competition policy of interest on the observed outcomes.

In contrast to internal validity, external validity refers to the extent to which the findings of a study can be generalised to other contexts. The studies by Ai and Sappington (2002), Roycroft and Garcia-Murillo (2000), Banerjee (2003), Clements (2004) and Sappington (2003) all use a comprehensive, common dataset and, thus, are directly comparable. Although addressing similar questions about the effects of incentive regulation on service quality, the studies do not provide unequivocal findings. Instead, the findings vary according to how service quality is defined and measured, how the regulatory regimes are classified, the choice of exogenous variables and the econometric estimation methods used (Sappington, 2003). These studies indicate that careful attention to model specification is vital to using econometric methods for policy evaluation particularly if the results are likely to be generalised to other contexts. Indeed it may be inappropriate to make assumptions that the findings of the evaluation are relevant to other contexts.

**Transparency and reproducibility**

Many potential threats to validity are dealt with during the modelling phase. However, the extent to which the results of an evaluation have been influenced by the modellers’ treatment of these threats may not be readily apparent to third parties. Thus it is important that an evaluation that uses econometric methods is transparent in terms of the model’s design, the data used and approaches taken to threats to validity. This transparency should enable third parties to replicate the results of the evaluation and test the extent to which these are both defensible and robust.
<table>
<thead>
<tr>
<th>Authors</th>
<th>Evaluation Question</th>
<th>Dependent Variable(s)</th>
<th>Independent Variables</th>
</tr>
</thead>
</table>
• Aggregate Investment  
• Operating revenue and cost  
• Profit  
• Basic local service rates | • Demographic characteristics  
• Political characteristics  
• General economic conditions in the state  
• Industry characteristics  
• Characteristics of state regulators |
| Roycroft and Garcia-Murrilo (2000) | The impact of local exchange competition on telephone service quality for Regional Bell Operating Companies. The impact of state level regulation, automation, merger and deployment of fibre optic cable and digital switches on service quality. | ‘Trouble reports’ – reports of initial and repeated troubles – provide measure of both network reliability and service quality. | • Competition (number of competitors per business line)  
• Regulation  
• Downsizing/automation (number of lines/employee)  
• Technology – percentage of carrier cable that is fibre optic and percentage of switches that are digital  
• Mergers |
| Clements (2004)         | The impact of regulation, competition and employees per line on quality of service in US local telephone industry. | Estimated econometric equations for a range of ‘people and process oriented (PPO) quality variables’ and ‘equipment and system oriented (ESO) quality variables’  
PPO quality variables included:  
• The percent of installation orders for local telephone service that were completed by the commitment date  
• The number of complaints filed with state and federal regulators about service quality per 1000 access lines.  
ESO service quality variables included:  
• The percentage of access lines served with digital switches  
• The percent of cable that is fibre. | • Price of local service  
• Employees per number of access lines  
• Competition  
• Regulation  
• Proportion of business access lines  
• Growth in number of access lines |
<table>
<thead>
<tr>
<th>Authors</th>
<th>Study Title</th>
<th>Methodology</th>
<th>Findings</th>
</tr>
</thead>
<tbody>
<tr>
<td>Boardman &amp; Vining (1989)</td>
<td>The impact of ownership on performance of firms in a variety of industries. Used data on 500 largest Canadian firms in 1986, including 12 State owned enterprises (SOE) and 93 mixed (part SOE) enterprises.</td>
<td>Performance as measured by profitability and technical efficiency Three measures of profitability • Return on assets • Return on sales • Net income Two indicators of technical efficiency • Sales per employee • Sales per asset.</td>
<td>Several variables relating to market structure, including indicator or market concentration.</td>
</tr>
<tr>
<td>Martin and Parker (1995)</td>
<td>The impact of privatisation on performance in various industries. Evaluated the performance of 11 UK firms that were privatised in the 1980s using data from five survey periods – nationalisation period, pre-privatisation period, post-announcement period, post-privatisation period, recession period.</td>
<td>Rate of return on capital Productivity as measured by value-added per employee hour.</td>
<td>Did not explicitly consider impact of other factors on performance. Authors considered the limitations of this method in their concluding comments.</td>
</tr>
<tr>
<td>Kelly and Ying (2009)</td>
<td>The effect of competition and regulation of cable television rates in the US using firm-level data from 1993 to 2001.</td>
<td>Average customer rate.</td>
<td>• Wages (a proxy for marginal cost) • Total number of channels • Number of basic subscribers • Multiple System Operator (MSO) affiliation and size • Rates for basic and upper tier programming • Regulation • Competition</td>
</tr>
<tr>
<td>Hazlett (1997)</td>
<td>Did regulation of cable television rates lead to an increase in subscription to, and viewing of, cable television (alternatively, was regulation successful in constraining prices and thus increasing subscriptions)? Compared trends in levels of subscription before and after regulation.</td>
<td>Various measures of the level of subscription to cable television services.</td>
<td>• Regulation • Time</td>
</tr>
</tbody>
</table>
8.3.2 Data issues

The availability of data is a fundamental issue for the use of econometric methods for policy evaluation and can dictate the way in which an econometric model is specified and the types of questions that can be examined using this approach. For example, Fischer, Guitierrez and Serra (2003) were restricted to considering privatisations for which adequate data on employment, production and profitability were available. In addition, their choice of performance indicators (dependent variables) was sometimes guided by the availability of data rather than the suitability of the variable for the task. In this regard, they used sales per worker as an indicator of productivity despite recognising that this indicator may reflect market power as well as underlying changes in productivity.

As discussed in chapter 4, econometric methods that attempt to evaluate the impact of regulation must model regulation itself. Most of the case studies that explicitly seek to evaluate the impact of regulation use a dummy variable to indicate the presence (variable equals ‘1’) or absence (variable equals ‘0’) of regulation. The main difficulty with this approach is that the regulatory regime is represented at a high level only—it is either present or absent. Finer details of regulatory regimes, including the effects of various incentives for performance and penalties for non-compliance cannot be readily captured by this approach. As noted by Sappington (2003, p. 371), all relevant details of regulation must be accounted for before definitive links can be drawn between observed behaviour and the prevailing regulatory regime.

Several case studies also use a dummy variable (takes the value 0 or 1 to indicate the absence or presence of a factor that is expected to have a significant impact on the dependent variable) or a proxy variable (in place of a variable that cannot be measured) to capture the effects of variables that cannot be represented more directly. These techniques can, however, introduce additional problems that require trade-offs to be made. For example, Ai and Sappington (2002) and Clements (2004) used a dynamic fixed-effects method that included year-specific and state-specific dummy variables to capture the effects of the variety of economic, social and political determinants of telephone service quality that could not be readily measured. However, in both studies, the dummy variables were highly correlated with several demographic variables. This multicollinearity problem was overcome by omitting the demographic variables. However, this solution potentially introduced omitted-variable bias (Sappington, 2003).

8.3.3 Cross-country econometric studies

A further set of studies uses cross-country time series data (panel data) to examine the effects of regulation across a number of countries. The panel data consists of both different economic units (such as countries) and different periods of time. These cross-country studies are mainly econometric studies that analyse the impact of particular variables, such as aspects of the regulatory regime and/or economic and institutional factors, on economic outcomes in the regulated industries. The aim of a cross-country econometric analysis is to understand whether these variables have a positive or negative impact on the economic performance of the sector by testing general propositions about the effects of competition and regulatory reforms, rather than providing detailed evaluations of specific reforms in a particular country.
A typical statistical cross-country analysis

The goal of an econometric cross-country study is generally to analyse differences in an economic variable among countries and across time. An analysis of regulatory regimes might focus on differences in productivity or efficiency growth between economies and/or across time, and the extent to which explanatory variables such as features of the regulatory regime, broader economic factors, and perhaps other non-economic factors (e.g. the political system) are associated with differences in the variables (e.g. productivity growth, etc.) of interest.

The analysis generally considers a ‘panel’ of countries, over a particular number of years. In technical terms, a panel dataset has both cross-section and time-series dimensions, so that variables $x_{it}$ are generally indexed both in respect of the economic unit ($i = 1,\ldots,I$) and period of time ($t = 1,\ldots,T$) (for instance, see Greene, 2007, and Hsiao, 2003).

This type of analysis begins by collecting a dataset, comprising both dependent and explanatory variables, for each country and time period.

The dependent variables typically consist of one or more variables that measure some sort of outcome or level of performance of the regulatory regime, such as the growth rate of subscriber density, or operating efficiency of the regulated operators.

The explanatory variables are those variables that it is thought might explain any differences across countries and/or across time in the performance of the regulatory regime. Explanatory variables typically include both:

- Variables describing the characteristics of the regulatory regime, including both formal characteristics (such as aspects of the legislative framework) and informal characteristics (such as the degree of independence of the regulator in practice). These also include characteristics such as what type of regime it is (cost of service, price cap, etc.), whether the regulator is independent, and so on.
- Variables describing the economic and institutional environment of the wider economy, such as GDP per capita, the government share in GDP, the type of legal system, and many other potential variables of this type.

The analysis consists of applying statistical and econometric techniques to this dataset, to try to evaluate the impact of the explanatory variables on regulatory performance and other dependent variables. A great wealth of statistical and econometric techniques have been (and continue to be) developed for the analysis of panel data (among others, see Greene, 2007, Hsiao, 2003, and Arellano, 2003).

A general panel data model is written as:

$$y_{it} = \alpha + \beta X_{it} + u_{it}$$  \hspace{1cm} (8.2)

meaning that a dependent variable $y_{it}$ indexed both in terms of the economic unit $i$ and time $t$ is regressed against a set of explanatory variables $X_{it}$, and an error term $u_{it}$. Different specifications of the error term result in different specifications of the panel data model. For instance, in a ‘fixed effects model’, the error term is specified as:

$$u_{it} = \mu_i + v_{it} ,$$  \hspace{1cm} (8.3)
where $\mu_i$ is a time-invariant component specific to each economic unit $i$, and $\nu_{it}$ is an ‘orthodox’ random error term, $\nu_{it} \sim iidN(0, \sigma^2_\nu)$ (randomly drawn from a normal distribution). This model is known as a fixed effects model because the $\mu_i$ component in this specification is assumed fixed over time. Other common specifications include the random effects model, where the combined error term is specified as above, but the time-invariant component $\mu_i$ is a random term drawn from a distribution $\mu_i \sim iidN(0, \sigma^2_\mu)$.

The aim of the analysis is usually to understand whether these characteristics have a positive or negative impact on the economic performance of the sector. The primary aim is to test general propositions regarding the potential economic effect of regulation, rather than providing detailed recommendations on specific reforms. For instance, they might test the proposition that some characteristic such as what type of regulatory regime (cost of service, price cap) exists, or whether or not it is codified in a formal statute, might affect levels of investment, or industry productivity.

A selection of cross-country statistical analyses of regulation is shown in box 8.2. The studies use a variety of statistical and econometric techniques that have been (and continue to be) developed for the analysis of panel data – see for example Greene (2007), Hsiao (2003) and Arellano (2003).

Cross-country studies of competition and regulatory reforms use a form of counterfactual analysis whereby countries that have not implemented a particular reform or policy act as the counterfactual for those that have. The panel data will typically include countries that have implemented a particular reform or policy and those that have not. The countries that have not implemented the reform or policy of interest act as the counterfactual to those that have. The effect of the reform or policy is determined by applying cross-country econometric and statistical techniques to the panel data.

Cross-country econometric studies are subject to a variety of potential method-related and data weaknesses that should be kept in mind when interpreting their findings. As it is difficult to assemble a robust cross-country panel dataset, most of the data used in the studies shown in box 8.2 are panel data compiled by multi-national organisations such as the OECD and the World Bank. Nevertheless, the data for many countries in cross-country panel datasets may be unreliable for a variety of reasons (Maddala, 1999). For example, a particular variable may be defined differently in different countries, and the methods used to measure that variable may also vary across countries. Thus care must be exercised to avoid an ‘apples with oranges’ evaluation.

Levine and Renelt (1991, p. 3), writing about cross-country regressions of economic growth on policy parameters, examine some of the issues that can influence the reliability of findings based on cross-country growth studies. One issue relates to aggregation and sampling of data, and whether countries are the appropriate unit of study. It might be more appropriate to conduct the analysis at a more disaggregated level to enable the influence of country-specific factors on performance to be isolated. Levine and Renelt (1991, p. 5) caution that cross-country regressions that do not account for each country’s sectoral composition and worldwide sectoral trends may capture spurious relationships or miss significant relationships.
Regression analysis also assumes that data are sampled from a single population. However, it is not clear whether cross-country data meet this criterion. As noted by Harberger (1987, p. 256):

What in the world do Thailand, the Dominican Republic, Zimbabwe, Greece, and Bolivia have in common that merits their being put in the same regression analysis? Answer: For most purposes, nothing at all. He who puts them in the same regression should have a very good reason for doing so.

A variety of methods can be used to choose the countries to be included in the sample (based for example on GDP per capita). However, Levine and Renelt (1991) note that there is not a simple resolution to this issue and it is important that reasons for choosing a particular sample are clearly explained and an assessment made of how the findings change if the sample changes.

In many types of econometric studies, coefficients represent estimates of elasticity or behavioural relationships that provide an indication of how much the dependent variable will change if an independent variable changes by one per cent. However, this is not the case with coefficients derived from cross-country regressions, whose equations generally do not represent behavioural equations. Care must therefore be taken in interpreting these coefficients.

Although measurement and data issues arise in all empirical work, the problems tend to be magnified in cross-country studies for a number of reasons including:

- cross-country differences in data collection processes
- difficulties in measuring the relevant policy variable across countries
- international comparisons of monetised variables are difficult because of issues associated with the need to convert these variables to a common currency
- measurement error, index problems and data availability, and data accuracy (whether the data are measured accurately and appropriately represent the concepts for which they are measures or proxies).

Levine and Renelt (1991, p. 19) consider that sensitivity analysis of the findings of cross-country studies is important given the ‘serious methodological, conceptual and statistical problems with cross-country regressions’. 
**Box 8.2: Cross-country Econometric Studies**

<table>
<thead>
<tr>
<th>Author(s) and Year</th>
<th>Country, Industry or Sector</th>
<th>Reforms</th>
<th>Findings and Techniques</th>
</tr>
</thead>
</table>
• Use: (i) the World Bank Performance Database and the Electricity Regulatory Governance Database, which contains data on several aspects of the governance of electricity agencies in the region.  
• Run various models to explain the impacts of change in ownership and different characteristics of the regulatory agency on performance.  
• Private ownership increases performance (other than coverage), especially in the transition from public to private ownership.  
• Both the existence and the experience of a regulatory agency have a material impact on performance.  
• The quality of regulatory governance improves utilities’ performance. |
| Ros (1999) | Telecommunications industry. Uses data from the International Telecommunications Union (ITU). | The impact of privatisation and competition on providers’ performance:  
(1) the rate of network expansion  
(2) the degree of network operating efficiency, as measured by main lines per employee as a proxy. | • Panel data analysis.  
• Competition indicated by whether country officially permits competition in local, long distance or international.  
• Privatisation associated with more mainlines per capita.  
• Competition is not associated with network expansion, but is associated with greater operating efficiency. |
• Private ownership positively associated both with levels |
<table>
<thead>
<tr>
<th>Study</th>
<th>Industry</th>
<th>Sample</th>
<th>Findings</th>
</tr>
</thead>
</table>
- Private ownership positively associated with levels and growth rates of operating efficiency (as measured by the lines/employee proxy).  
- Competition positively affects efficiency, but does not affect network expansion. |
- Measures competition as the market share not owned by incumbent.  
- Shows that competition is positively correlated with increases in performance indicators, and decreases in prices.  
- Performance increased by privatisation plus an independent regulator.  
- Privatisation alone does not bring similar benefits, and is negatively correlated with connection capacity. |
8.4 Conclusions

Econometric methods are often used for policy evaluation and there is a large body of case-study literature, covering both within-country and cross-country analyses, from which to draw. Analysis of the existing literature suggests that two key issues are likely to arise in undertaking an econometric study. First, care must be taken in specifying the model, in terms of the dependent and exogenous variables chosen. Misspecification can threaten the validity of the model and undermine the ability to make causal inferences about the impact of reforms on the chosen indicators. Second, the availability of data is a fundamental issue, which may dictate the choice of model specification. Ultimately, data issues may place a constraint on the use of econometric methods in evaluating the effectiveness of infrastructure reforms. If econometric methods are used for evaluations of competition and regulatory reforms it is very important that the model and data used are transparent to third parties. Unless it is possible for others to replicate the findings of an econometric study, such findings are unlikely to be defensible.
9. Productivity Studies

9.1 Introduction

This chapter examines the use of productivity studies for the evaluation of competition and regulatory reforms. Productivity studies include the techniques of total factor productivity (TFP) analysis, and the frontier analysis methods – stochastic frontier analysis (SFA) and data envelopment analysis (DEA).

These techniques may be used as an evaluation method on their own. For instance, typically one of the high-level objectives of competition and regulatory reforms is to improve ‘productivity’ – either of regulated firms, an industry or more broadly an entire economy. In Australia, for example, the NCP reforms were broadly motivated by Australia’s poor on-going productivity performance relative to other OECD countries. The NCP was implemented as a series of micro-economic reforms which targeted inefficiencies at the industry or sector level with the intention of boosting national productivity. A productivity study can help to evaluate the extent to which productivity has changed since the introduction of NCP reforms. As is discussed later in the chapter, however, one of the key issues with productivity studies is the extent to which it is possible to attribute observed changes in productivity to a particular policy or program rather than to other factors.

Productivity studies may also be used as inputs into other quantitative methods. For instance, they may be used to derive productivity data which are used in forming shocks for a CGE model, or used as inputs into an econometric study.

The chapter starts with an overview of each method of TFP analysis, SFA and DEA. It then discusses the data requirements for conducting a productivity study using these techniques. A discussion of the attribution problem, or how to account for other factors which might influence productivity, is then presented before the chapter concludes.

9.2 Total Factor Productivity Analysis – Growth Accounting Analysis

The roots of total factor productivity (TFP) analysis go back to the works of Tinbergen (1942), Abramovitz (1956), Kendrick (1956), Solow (1957), Stigler (1961), Denison (1962), Griliches and Jorgenson (1966), and Jorgenson and Griliches (1967).

TFP measures the effectiveness with which a production process turns all the productive inputs used into the outputs. TFP growth occurs where the output quantities produced grow at a greater rate than the input quantities used. TFP is firmly rooted in the concept of economic production functions as it analyses the transformation of productive inputs into outputs. TFP and TFP growth can in principle be measured at any level of aggregation where a production process can be defined, from the economy-wide macroeconomic level, to the industry-wide or firm level, down to production units within a firm or organisation.

In practice, there are two main methods used to estimate TFP growth: the index number approach and the econometric approach. The index number approach applies an index number formula to construct input and output quantity indices. The TFP growth rate is then defined as the difference between the rate of output quantity growth and input quantity growth. In contrast, the econometric approach estimates TFP growth directly by estimating a cost or production function for all firms under consideration.
The Australian Energy Market Commission (AEMC) (2008 and 2009), which is currently reviewing the potential use of TFP as an alternative to the building-blocks method for the *ex ante* regulation of electricity and gas networks, identifies a number of advantages and disadvantages to each approach. For example, the econometric approach requires assumptions to be made about the nature of the production process and firm behaviour. Such assumptions can be contestable and may influence the findings of studies which use the econometric approach. In contrast, it is not necessary to make assumptions about production processes and firm behaviour if the index approach is used.

The econometric approach also generally requires more data than the index approach but has the advantage that it is able to incorporate and analyse production features such as economies of scale. A practical constraint on the use of the econometric approach is a limitation on the number of outputs and inputs than can feasibly be considered. This constraint does not apply to the index method. Balancing these strengths and weakness, in practice, the index approach is typically used as it requires fewer observations and is more transparent and reproducible. Box 9.1 provides more information about the index approach to TFP analysis.

The AEMC has recently released a detailed spreadsheet model of network TFP which compares processes and outcomes under TFP–based regulation and building blocks regulation (Economic Insights 2010).

Traditional TFP indices only measure TFP growth and not TFP levels. However, information on TFP growth rates alone across firms or countries does not provide the context necessary to interpret differences in growth rates. All else equal, firms starting from lower TFP levels could be expected to achieve higher TFP growth rates as they can make easy ‘catch-up’ gains compared to more efficient performers who, all else equal, will have lower TFP growth rates despite their higher efficiency levels. Multilateral TFP indices provide a means of comparing both TFP levels and growth rates in panel data. For instance, the Bureau of Industry Economics (1996) compared productivity levels and growth rates between the Australian and US electricity supply industries spanning the period of corporatisation of the Australian industry and found that the gap between Australian and US TFP levels narrowed markedly during the corporatisation period.
Box 9.1: Index of TFP Growth

Berndt (1980) summarised an index of TFP growth for period \( t \), \( TFP(t) \), in continuous time as being

\[
TFP(t) = \frac{\dot{Q}}{\dot{O}(t)} - \frac{i}{I(t)} \tag{9.1}
\]

where: \( \dot{Q} \) and \( \dot{I} \) are consistent indices of total outputs (aggregating the different outputs \( Q_1, Q_2, \) etc in some appropriate manner) and total inputs (aggregating the different inputs \( I_1, I_2, \) etc in some appropriate manner); the production function is ‘well-behaved’ in that it exhibits certain orthodox differentiability and curvature characteristics; and where \( \dot{Q} \) and \( \dot{I} \) are the time derivatives of \( Q(t) \) and \( I(t) \), respectively.

This type of approach is commonly known as ‘growth accounting’; the difference in output growth and input growth yields the rate of productivity growth defined in this way. To see how growth accounting works in the context of a simple production function, consider a simple production function in two factors, capital (\( K \)) and labour (\( L \)):

\[
Y(t) = A(t) \cdot F(K(t), L(t)) \tag{9.2}
\]

The variable \( A \) measures the shift in the production function at given levels of capital and labour. This variable is commonly thought of as measuring the level of technology, so that an increase in \( A \) represents technological advance, represented by way of an outward shift in the production function (although there are some potential issues with this interpretation: see Hulten, 2000, p.6). If this function is differentiated with respect to time, and it is additionally assumed that the price of each input \( K \) and \( L \) is the value of its marginal product, the terms can then be re-arranged as follows:

\[
\frac{\dot{Y}}{Y(t)} = \frac{\dot{A}}{A(t)} + s_K \frac{\dot{K}}{K(t)} + s_L \frac{\dot{L}}{L(t)} \tag{9.3}
\]

Where \( s_K \) and \( s_L \) are the observable shares in income of the factors of production \( K \) and \( L \), respectively.

The result is a simple and elegant framework for productivity measurement in continuous time. However, most data are available in discrete-time rather than in continuous-time format, and so a discrete-time equivalent measure is needed. One such measure is the Tornqvist index (following Tornqvist (1936), Diewert (1976) and others) although the Fisher index has superior properties and is now favoured by statistical agencies (Diewert 1993).

A Tornqvist index for inputs is calculated as a geometric weighted average of the relative changes in each of the inputs. In our simple two-factor production function, the Tornqvist index of discrete-time input change from period \( t-1 \) to period \( t \) is therefore calculated as:

\[
\frac{I_t}{I_{t-1}} = \left( \frac{K_t}{K_{t-1}} \right)^{\overline{s}_K} \cdot \left( \frac{L_t}{L_{t-1}} \right)^{\overline{s}_L} \tag{9.4}
\]

where \( \overline{s}_K \) is the average income share of capital over periods \( t-1 \) and \( t \) i.e. \( \overline{s}_K = (s_{K,t-1} + s_{K,t})/2 \), and similarly for \( \overline{s}_L \). An output index can similarly be derived.

Diewert (1976) showed that a Tornqvist index is: (1) an exact index number for its continuous-time equivalent (Divisia) index if the underlying production function follows a certain functional form known as trans-log, developed by Christensen, Jorgenson and Lau (1973), and (2) is a good second-order approximation if it is not. Therefore, discrete-time indices thus derived are valuable proxies for the continuous-time analysis, if continuous-time data are not available (as they usually are not).

Once reliable indices of total input growth and total output growth are derived, measures of TFP growth can then be derived as the difference between the two, as outlined earlier:

\[
\frac{TFP_{growth_{t-1,t}}}{TFP_{t-1}} = \frac{Outputs_{t}}{Outputs_{t-1}} - \frac{Inputs_{t}}{Inputs_{t-1}} \tag{9.5}
\]
9.3 Frontier Analysis

Frontier analysis is based on the notion that there is a best-practice level of technical efficiency. This level of efficiency is represented by a production possibilities frontier that can be reached but cannot be exceeded. Production can lie on or within the frontier. Production within the frontier would represent production at less than the best-practice level of technical efficiency. An organisation or economy might be below its best-practice level for a variety of reasons including organisational inefficiency or poor management and constraints on adoption of new technology or resources use.

A consequence is that, based on this notion and in this framework, an observed change in productivity (e.g. by TFP measurement) for any one economic unit could be the result of (1) an improvement in the best-practice technology available, shown by a shift outwards of the technology frontier; or (2) better and more efficient use of existing technology, shown as a ‘catch-up’ of the economic unit to the frontier rather than of a movement of the frontier itself.

In conducting productivity analysis, it can often be important to be able to distinguish between these two sources of productivity growth. Frontier analysis assists in doing this by analysing productivity in two sequential steps (conceptually, in any event):

- First, define the production frontier as being the observed maximum output for given input combinations (e.g. for given capital-labour combinations, in a simple two-factor production function).
- Second, analyse each economic unit in relation to the production frontier, and where a unit lies within the frontier. The extent of the distance from the frontier is attributed to inefficiency.

The principal advantage of frontier techniques is that they allow the separation of an observed change in TFP into two quite different phenomena – a change in the best-practice frontier, and a shift in relation to (towards or away from) the efficiency represented by the frontier:

\[ TFP_{growth} = FrontierShift + EfficiencyChange \]  

A significant analytical consequence of this is that it theoretically enables analysis (‘benchmarking’) of each firm against best practice. Secondly, it potentially also permits the origins of technical efficiency (e.g. different regulatory regimes), to be identified – cross-country studies are potentially useful here. For instance, the question might be what differences in economic, institutional and other factors influence differences in technical efficiency across countries and across time?

Two main alternative classes of quantitative techniques for frontier analysis are used: (1) data envelopment analysis; and (2) stochastic frontier analysis.

9.3.1 Frontier analysis – DEA

DEA is a linear programming technique used to conduct frontier analysis. The data are ‘enveloped’, with the envelope defining the best practice frontier. The technical efficiency of individual economic units is then calculated as the ratio of the distance (technically, the distance-function) of the economic unit from the origin, to the distance of the frontier from the origin at the same relative proportions of the use of inputs into production. Efficiency, being a ratio, is commonly expressed as a number in the range from 0 to 1. An ‘efficiency score’ of 1 would indicate production on the
frontier, with an efficiency score of (for instance) 0.6 indicating that only 0.6 of the possible output is produced for that set of factor proportions, given the available frontier technology.

A potential advantage of DEA is that it is a non-parametric technique and does not require econometric or statistical estimation of the parameters of a posited relationship. However, a commensurate drawback is that results can be skewed by outliers (e.g. resulting from measurement errors in respect of individual units). If an outlier lies beyond the true frontier, that outlier will mistakenly be enveloped by the DEA techniques and will be shown to ‘define’ the frontier – with the result that the frontier is mis-specified due to the outlier and the resulting measure of technical efficiency is under-stated. Another disadvantage of DEA is that, like econometric methods, it requires a relatively large number of observations to implement.

9.3.2 Frontier analysis – stochastic frontier analysis

Stochastic frontier analysis is a parametric, econometric technique used to conduct frontier analysis. Rather than the frontier being constructed by the data being enveloped as in DEA, the frontier is estimated on the basis of the data. Typically, the basis for estimation is an econometric specification of a production function, such as a trans-log production function. The production function is transformed into a production frontier by way of the error term. Whereas an ‘ordinary’ production function is typically estimated using an orthodox random error term, a frontier production function is estimated using a composite error term, with one component being a true random error term (random over a symmetric distribution), but the second ‘frontier’ component being drawn from a one-sided distribution, so that this component can cause the observed output value to lie within the frontier but not outside of it. One such SFA model is described in box 9.2.

Box 9.2: SFA Model

In the SFA model due to Aigner, Lovell and Schmidt (1977):

\[ y_i = \alpha + \beta' x_i + \epsilon_i \]  

(9.7)

\[ \epsilon_i = v_i - |u_i| \]  

(9.8)

where \( y \) represents the output, \( x \) is the set of inputs into production, and the error term of the regression, \( \epsilon \), is a composite error term made up of the underlying error terms \( v \) and \( u \). Both the underlying random errors \( v \) and \( u \) are normally distributed with zero means and constant variances, but the component term \( -|u| \) is strictly non-positive, as it represents the ‘inefficiency’ estimate – a unit can lie within the frontier (\( |u| > 0 \)) or on the frontier (\( |u| = 0 \)) but cannot lie outside the frontier (which would be given \( |u| < 0 \), which cannot be).

This or similar models can be estimated, if one makes certain assumptions regarding the functional form for the frontier error component, for degrees of freedom reasons. Common approaches include the assumptions that technical efficiency is: (1) time invariant; (2) varies systematically over time according to some deterministic process such as an exponential functional form (Battese and Coelli, 1992) or other process (Kumbhakar, 1990); or (3) are drawn from some distribution to be estimated (Battese and Coelli, 1995).

The results that can be obtained from SFA are broadly equivalent to those from DEA. First, the TFP growth of each economic unit can be estimated. Second, the position
and the rate of the shift of the frontier, for the units (the sector, the economy) as a whole, and also at any one given set of factor proportions, can be estimated. Third, the inefficiency of each economic unit in respect of the frontier at any given time period, and the shift towards (catch-up) or away from the frontier of that unit over time, can be estimated. Finally, using these results, the sources of TFP growth of each unit can be separated into (A) frontier shift, or technological growth, and (B) catch-up towards the frontier, or changes in technical efficiency.

The advantages and disadvantages of SFA mirror those of DEA. A primary advantage is that, as SFA involves econometric estimation incorporating a ‘true’ random error component, SFA is not as vulnerable as DEA to measurement errors in respect of individual economic units – a single outlier has less influence on the measured frontier using SFA than it would using DEA. A potential disadvantage is that SFA requires estimation of a number of parameters, which gives rise to the standard issues concerning standard errors of estimates, and of results calculated using estimated parameters. A second potential disadvantage is that, in order to make estimation possible and operable, SFA typically requires that some (usually quite restrictive) functional form be forced on to the inefficiency term, which might typically be much more restrictive than the movement over time of the ‘true’ underlying inefficiency term. As discussed in the following section, a further potential disadvantage is that SFA can in some circumstances be more susceptible to data deficiencies, and, like other econometric methods and DEA, it requires a relatively large number of observations to implement.

9.4 Data Availability

The simplest way to evaluate the impact of reforms by focussing on productivity is to obtain a time series of the relevant indicator of productivity and track its performance over time both before and after the reform. Although conceptually simple, this approach requires a large time series of consistent data. Newbery and Pollitt (1997, p. 273) highlight some of the measurement difficulties:

[I]t is hard to measure capital inputs accurately, particularly when technological progress renders old capacity less valuable …. and it is hard to measure inputs and outputs consistently with changes in technology and services.

NERA (2007, pp. 19-23) discusses some of the data challenges that Australian empirical TFP studies have encountered. NERA notes that the quantity and quality of Australian sector-level data are often insufficiently robust for use in sector-level studies of productivity. Specifically:

The data are often not uniform, are prone to inadequate and often highly aggregated reporting, and frequently do not measure the key inputs required to calculate robust sectoral productivity trends.

Similarly, the AEMC (2009, p. 47) reiterates that, in relation to the energy sector, the existing data are not sufficiently consistent, reliable or robust enough to support TFP analysis of the standard required to base entire price cap decisions on. Thus, a dataset must be created to allow a robust TFP measure to be formed (for the application of incentive regulation). This would require specification of the minimum data required with consistent definitions established and reported on.

More broadly, Lawrence (1998) highlights the potential implications of these data deficiencies for evaluating the impact of Australian microeconomic reforms, and by implication, economic regulation of infrastructure sectors:
A major omission in Australia’s reform process is that we have not implemented data supply and collection mechanisms to ensure that we can evaluate the success of those reforms later. ... In future, we need to build supplying data in as a prerequisite for the restructured industry.

However, the international literature suggests that if data are available, a study of efficiency can be used to evaluate the impact of competition and regulatory reforms in a particular sector. Box 9.3 contains a summary of a selection of such studies in the water, telecommunications and electricity sectors. It should be noted that data issues were also relevant to these studies. For example, von Hirschhausen et al (2006) note that the choice of variables for measuring inputs and output in their efficiency study are constrained by data availability.

9.5 Attribution Problems

As mentioned, the simplest approach is to compare the sector’s productivity growth before and after the policy or event of interest as in Saal and Parker (2001), Arocena and Waddams Price (2002) and de Boer and Evans (1996). A key issue with this approach, however, is to determine the extent to which observed changes in productivity ought to be attributed to the reforms or regulation under consideration rather than any of the other factors that may have contributed to changes in productivity. The Productivity Commission (PC, 1999b) classifies these influences into three broad groups. The first group consists of those effects that have a direct and tangible link to the relationship between inputs and output. According to the PC, these include technological change, accumulation of physical and human capital, economies of scale, scope and specialisation, management practices and work arrangements. The second group contains underlying factors that influence productivity indirectly, and includes the extent of competition and other demand and supply factors. The third group contains the policy, institutional and social factors that have a general and indirect influence, particularly over the longer term.

The attribution problem is not specific to efficiency studies and the practical solutions that are adopted in these studies are similar to those that arise for other methods. For example, a common approach is to assume that all other influences have remained constant over time so that all observed changes in productivity are attributable to the policy or regulation of interest. This approach, however, is likely to substantially overstate the effectiveness of reform and/or regulation. For instance, De Boer and Evans (1996) recognise that the changes in TFP derived in their study are also likely to reflect changes in quality and technology, as well as the regulatory regime which is of interest. Thus De Boer and Evans (1996) supplement their calculation of TFP with information about changes in the quality of service, particularly changes in the penetration rate of electronic exchanges and the number of customers waiting at the last day of the month for services such as the installation of a phone, total faults and repeat faults. On the basis of this supplementation, they conclude that quality of service has improved and thus their estimate of TFP growth is a ‘lower limit of gains in x-efficiency’. In contrast, they do not make any adjustments for the impact of technical change but acknowledge that (De Boer and Evans, 1996, p. 30):

the numerical estimates of productivity change must be viewed cautiously because of data measurement issues ... they will reflect technical change as well as organisational change.

A preferable approach is to compare productivity growth with that of a comparable industry or sector in another jurisdiction, either in Australia or in another country. This comparable group is effectively the control group so that differences between
this group and the treatment group (the industry or sector that has been exposed to the regulation or policy) may give a broad indication of the effect of reforms or regulation. For example, comparing the TFP of a sector that is subject to economic regulation with the TFP of a broadly comparable sector in another jurisdiction that is not subject to economic regulation, may give an indication of the effect (positive or negative) of economic regulation on the sector – this is the basic approach taken in cross-country econometric studies, as outlined in Chapter 8.

At the firm level, this is essentially the approach taken by Hattori (2002) who compares the productivity performance of electric utilities in Japan and the US using the SFA method. Similarly, Australia’s aggregate productivity performance is frequently compared with other OECD countries. Both OECD (2003) and OECD (2004) suggest that Australia’s microeconomic reform agenda has played a significant role in its improved relative productivity performance over the last decade or so.

In practice, the extent to which comparison of productivity performance between a ‘treatment’ and ‘control’ group can occur depends critically on the availability of data. NERA (2007) suggests that statistical offices in most countries publish suitable sector-level data, however, Australian data are often inadequate for this purpose. In contrast international benchmarking at the industry-level may be precluded by the shortage of international, comparable industry-level data. Where limited data are available, it may be necessary to qualify the results that are obtained from the benchmarking exercise. Where a ‘like-with-like’ cross-jurisdictional comparison is not possible, measured differences in productivity may be the result of influences that are not comparable and constant across jurisdictions (also see chapter 8 for a discussion of data issues in such cross-country studies). In addition, differences in productivity may be the result of factors that are not easy to measure and compare including cultural, political, educational and institutional influences (NERA, 2007).

For example, De Boer and Evans (1996, p. 33) adopt a form of international benchmarking when attempting to assess the relative influence on efficiency of deregulation and competition compared with the relative influence of privatisation. The authors use previous studies on the relative performance of monopoly telecommunications incumbents in New Zealand and the UK to conclude (p. 34):

> The competitive environment may have contributed to productivity gains. … The comparison between Telecom and British Telecom is suggestive that the absence of both regulatory barriers to entry and concomitant price regulation may have stimulated productivity growth over that of a more regulated industry environment.
### Box 9.3: Selected Productivity Studies

<table>
<thead>
<tr>
<th>Authors and Year</th>
<th>Evaluation Question</th>
<th>Type of Efficiency Study</th>
<th>Output and input indicators</th>
</tr>
</thead>
</table>
| Aubert and Reynaud (2005) | Water industry in Wisconsin – where there were 211 different regulated water utilities, and can be regulated under different regimes (price cap or rate-of-return). | The type of regulatory regime – price cap or rate of return. | • Stochastic frontiers analysis.  
• The impact of the different regulatory regimes on differences in the firms’ levels of cost efficiency – in firms that are (regulatory regime aside) operating in a broadly similar environment.  
• Rate of return regulation could be the most or least efficient, depending on whether the regulator collected much or little data. |
| Saal and Parker (2001) | What is the impact on productivity and price performance of privatisation in the water and sewerage industry in England and Wales? | Before-after study of labour and TFP | TFP growth was estimated as output growth less a Tornqvist weighted index of input usage growth, using a quality-adjusted measure of output:  
\[ Y_t = S_{water,t} \text{Water}_t \times \text{Quality}_w + S_{sewerage,t} \text{Sewerage}_t \times \text{Quality}_{sewage, t} \]  
where \( S_{water,t} \) and \( S_{sewerage,t} \) are the respective shares of water and sewerage in total turnover at time \( t \), \( \text{Water}_t \) and \( \text{Sewerage}_t \) are base levels of water and sewerage treatment demand indexed to equal 100 in 1990, and \( \text{Quality}_w \) and \( \text{Quality}_{sewage, t} \) are indices of the quality of water and sewerage treatment indexed to equal 1 in 1990. |
| De Boer and Evans (1996) | The impact of regulation on efficiency in the New Zealand telecommunications market. | Before-after study of TFP growth | \[ g(t) = g(q) - r_s g(i) \] where \( g(.) \) is rate of growth, \( q \) is output, as measured by ‘total minutes of use of the network’, \( r_s \) is returns to scale, \( g(i) = \alpha_1 g(m) + \alpha_2 g(l) + \alpha_3 g(k) \), \( \alpha_i \) is cost share of factor i; m, l, and k are material, labour and capital input indices.  
De Boer & Evans supplement the TFP calculation with information about changes in quality of service.  
They do not make any adjustments for the impact of technical change but caution that their numerical estimate of TFP change must be viewed cautiously as it will reflect both technical and organisation change.  
Performed benchmarking of their results with those of previous studies of productivity growth of British Telecom following privatisation. |
| **Von Hirschhausen, Cullman and Kappeler (2006)** | Efficiency of electricity distribution companies in Germany | Estimates efficiency using DEA of data on 307 companies. Results are verified using SFA | The choice of variables for measuring inputs and outputs was constrained by data availability. The method used is to estimate a base model that includes labour, grid size and peak load (as proxies for capacity) as inputs, and units sold and number of customers as outputs. Variations to this base model are then made by splitting total customers into industrial and residential categories, using turnover as a measure of output, and disaggregating the network into aerial cables and underground lines. |
| **Delmas and Tokat (2003)** | Short run impact of retail deregulation on productive efficiency of electric utilities in the United States. | Uses DEA to estimate the productive efficiency of the utility. This measure then becomes the dependent variable in a regression model that is used to test the impact of retail deregulation on efficiency. | To estimate productive efficiency using DEA, Delmas and Tokat use regulatory data from 1998 to 2001 for 177 US electric utilities. Inputs are labour cost, plant value, production expenses, transmission expenses, distribution expenses, sales, administrative and general expenses, and electricity purchased from other sources. Measures of output are: quantities of low-voltage sales (residential and commercial), high-voltage sales (industrial, interchanges out, and wheeling delivered), and sales for resale to other utilities in megawatt hours. The productive efficiency estimates from the DEA study are used as the dependent variable in a regression model. The independent variables in that model are divided into several categories related to the level of deregulation that utilities face, the nature of the competitive environment, the level of vertical integration of utilities, the size of utilities, whether firms are involved in mergers with other utilities, the amount of power generated from nuclear energy, renewable energy and the power grid to which the utility belongs. |
| **Arocena and Waddams Price (2003)** | Effect of switch from rate-of return regulation to incentive regulation on efficiency of both publicly and privately-owned Spanish electricity generators. | Estimate TFP change in Spanish coal-fired generating plants between 1984-1997 using the Malmquist index approach to assess whether there was a statistically significant change in TFP growth after the introduction of incentive regulation. | Analysed regulatory data from 33 generators, 19 of which were privately owned and 14 of which were publicly owned. Variables are measured in physical units – output is annual generation per unit in megawatt hours, capital is megawatts of capacity, labour is average number of employees at each plant, fuel is measured in millions of therms (coal-fired). |
| Hattori (2002) | Potential efficiency gains available to Japanese utilities. | Decomposes the TFP change into technical efficiency change, differences in scale efficiency and technological change. | Estimates and compares the efficiency of 12 US and nine Japanese electric utilities during the period 1982-1997 using SFA. Uses translog input distance function to represent technology of electricity distribution. | Used annual data on major Japanese and US vertically integrated electric utilities over the period 1982 to 1997. The US data are regulatory data provided to the FERC. The Japanese data are from publicly available sources, including company annual reports and industry statistical yearbooks. Labour is measured as the number of full time employees. Capital input is measured as transformer capacity. Output is measured as total electricity sales (GWh) to household customers, and total electricity sales (GWh) to business users. |
9.6 Conclusions

The overarching objective of much regulatory and competition reform is to improve the efficiency of the directly affected sector, and thus, indirectly, the performance of the entire economy. There are several ways that regulatory and competition reform can promote efficiency – for example by encouraging cost efficiency and the closer alignment of prices with efficient costs, improvements in service quality and innovation, promoting efficient investment, and promoting new entry by lowering entry barriers. Partial evaluation methods can be a particularly useful way to evaluate whether or not the regulatory and competition reforms have actually led to these expected outcomes being realised.

However, as stressed above, these partial-evaluation methods only consider the effects of regulatory and competition reforms in markets that are directly affected. If the indirect effect of reforms is widespread and large, then the results obtained from partial evaluations alone may be misleading. In such circumstances, it would be advisable to supplement the results using additional general-equilibrium or qualitative methods.

Even within the markets that are directly affected, the results of partial evaluations may also be misleading if considered in isolation. For example, it may be misleading to associate a reduction in prices with an improvement in welfare unless the impact of regulation on service quality has also been considered. Thus an assessment of the effect of regulation on a number of key performance indicators may be necessary to give a more thorough understanding of the effect of regulatory reform. Similarly, the popularity of index-based TFP studies indicate they have a role to play in providing a tractable method where the number of observations is relatively limited.

Each of the methods of partial evaluation discussed in the two preceding chapters is a well known economic tool that has various strengths and weaknesses. Econometric methods, for example, may be complicated by data issues and problems caused by misspecification. However, there are a range of tests to detect these problems and the techniques for addressing the various limitations are constantly evolving. The popularity of the econometric method suggests that it should be given careful consideration when deciding how to conduct an evaluation of economic regulation of infrastructure sectors, particularly at the sector level.

Nevertheless, data issues may be a constraint on the extent to which partial evaluation methods can be applied in practice, particularly the methods for evaluating efficiency.
10. Qualitative Data and Analysis

10.1 Introduction

The nature of economic regulation, with its focus on prices, costs, investment, efficiency and other numerical variables, suggests that an evaluation of its impact would normally be based primarily on an analysis of quantitative data using the quantitative methods discussed in chapters 6 to 9. Nevertheless, there may be instances where collection and analysis of non-numeric qualitative data is also appropriate or necessary (particularly where a process evaluation is being conducted) or as a complement to quantitative methods (when there are deficiencies in the data, or where impacts are not readily quantifiable). This chapter reviews qualitative data collection and evaluation methods, their strengths and weaknesses, and the types of qualitative methods that may be used to conduct evaluations of the effects of infrastructure reform and regulation. The chapter also considers a number of studies that have used a qualitative approach to evaluation. These studies show that the distinction between qualitative and quantitative approaches is often blurred. The studies also show that the qualitative methods that are applied are not always the same as the methods described in this chapter.

10.2 Qualitative Research Methods

Qualitative research has a long history in the social sciences such as sociology, anthropology, philosophy, linguistics and psychology. Denzin and Lincoln (2000) provide a summary of this history. Its use in program evaluation is more recent, having emerged during the 1970s.

10.2.1 Elements of qualitative research methods

Qualitative research focuses on the analysis of qualitative information which typically consists of words rather than numbers. Nevertheless, the dividing line between qualitative and quantitative methods is typically blurred and Bamberger (1999) notes that it is difficult to draw clear distinctions between quantitative and qualitative approaches. Instead, he considers it useful to consider methods of sample selection, research design, data collection and recording and data analysis as being arranged along a quantitative/qualitative continuum. As Bamberger (1999, p. 9) notes

> While some studies rely exclusively on quantitative methods for sampling, data collection and analysis, and others rely exclusively on qualitative methods, many studies mix and match statistical sampling techniques, qualitative data collection and statistical analysis from the quantitative and qualitative traditions.

It is possible, however, to draw some broad distinctions between qualitative and quantitative approaches to sampling, data gathering and analysis. By their nature, qualitative methods are often not based on the experimental or quasi-experimental research designs or structural models that form the basis of quantitative research. Nevertheless, as with quantitative methods, an evaluation based on qualitative methods should be well designed and based on specific evaluation questions (see chapter 4). The design should specify the populations to be studied, the subgroups to be included and excluded, the type of data to be gathered and how best to obtain those data, in order to best answer the evaluation questions.

Sampling in qualitative research

If it is not possible to study the entire population of relevance to an evaluation, it is necessary to choose a method of choosing a sample from within that population. In
this regard, there is typically a distinction between the way in which samples are selected for quantitative and qualitative research. Quantitative sampling approaches typically select a representative sample from the population based on random or probability sampling methods. The size of the sample is determined by the theory of probability. This approach enables hypotheses to be tested and the results based on analysis of the sample to be generalised to the entire population.

In contrast, qualitative research is generally not based on data gathered from a non-random sample of a population (although it can be). Instead, the size and composition of the sample is determined by the study’s objectives. There are three main non-random approaches to sampling in qualitative research:

- Purposive sampling groups participants according to pre-specified criteria that are relevant to the research question (e.g. age, location, gender). Sample sizes depend on the study’s objectives and the resources and time available. Participants are only included in the study if they meet the pre-specified criteria.

- Quota sampling is similar to purposive sampling, but the number of participants with particular characteristics (e.g. age, location, gender) to include in the sampling is decided while designing the study. The decision enables the researcher to focus on those people who are thought most likely to have experience, knowledge of, or insights into the evaluation question. Participants are included until the prescribed quotas are met.

- In the snowball sampling method, participants with whom contact has already been made refer the researcher to other people who could potentially participate in the study.

**Methods of gathering qualitative data**

Once a sample has been chosen, there are a number of ways in which qualitative data can be obtained from participants.

In-depth interviews, based on a questionnaire, enable data to be collected based on personal spoken accounts on a topic. The questions are usually open-ended and invite participants to answer in detail. Questions are used flexibly to allow the interviewee to explore the topic using their own unique insights, experiences, perceptions, opinions, feelings and knowledge. The data obtained from in-depth interviews consist of verbatim quotations with sufficient context to be interpretable (Patton, 2003).

Focus groups can also be used to obtain qualitative data. Focus groups are a type of group interview, facilitated and moderated by a researcher, where participants are brought together to discuss a particular issue or set of issues. A focus group usually consists of four to eight people and meets for 60 to 90 minutes. During a focus-group session, data are collected simultaneously using several different methods. Generally, the session is recorded (audio, video or both). The researcher also takes notes and additional observers may take notes from behind a one-way window outside of the room.

Documentary analysis uses public and private documents as data. Documents can be written documents, (such as parliamentary papers, administrative historical records, public reports, private letters, diaries and notebooks), non-text documents (such as photographs, maps and plans, videos and DVDs) and internet-based documents such as websites and web discussion groups. Data consist of excerpts from documents captured in a way that records and preserves context.
Participant observation and ethnography involves the researcher participating in and/or observing people's lives for an extended period of time. Observations of activities, behaviours, actions, conversations, interpersonal interactions, organisational or community processes and any other aspect of observable human experience can be recorded in field notes, memos, and tape-recordings. These observations, and descriptions of the context in which the observations were made, can then be used as data in an evaluation.

**Analysis of qualitative data**

There are numerous approaches to interpreting and analysing qualitative data, all reflecting different schools and traditions in qualitative research. The type of approach will reflect the research questions, the researcher’s ‘hunches’ (*a priori* reasoning) and other research or theoretical frameworks (Cabinet Office, 2003). There are three main approaches to analysing qualitative data.

First, descriptive analysis identifies key dimensions of experiences, attitudes and behaviours. The aim of a descriptive analysis is to display the richness and depth of these dimensions by describing them in detail. This might involve, for example, a detailed description of the way in which a policy or program is organised and delivered, the circumstances of service users, or the range of (intended and unintended) outcomes experienced.

Secondly, associative analysis identifies patterns, repetition and association in the qualitative data to enrich understanding of the research issues. Associative analysis also seeks understanding of why the patterns, repetition and associations occur. The purpose is to use the associations or patterns found in the data to enrich understanding of the phenomenon in question, and to prompt further questioning and analysis to understand why the association or pattern exists.

Thirdly, interpretative and explanatory analysis builds or finds explanations in the data for the views, behaviours and accounts described and for the patterns and associations found. An interpretative and explanatory analysis might seek to explain, for example, why some users experience positive outcomes and others do not, or how and why unintended consequences arise.

**Validation**

Well-designed qualitative research will involve a process of validation of the research findings. The most common forms of validation include:

- **Participant/member validation**, which involves taking findings back to research participants to gain their views on the interpretation, fill in gaps, and check on bias. It should be noted, however, that research participants will have their own particular perspective on the issues.
- **Peer or expert validation**, which involves taking the findings to a wider group who are knowledgeable of the issues being researched, to see how findings relate to their expertise and knowledge.
- **Validation through corroboration with other research** which is used to assess the credibility of findings.

**Triangulation**

A qualitative research design may also include elements of triangulation, bringing together different types of data, or different ways of looking at data, to answer the
research question. There are two main reasons for using triangulation (Downward and Mearman, 2007). First, triangulation is used to enhance understanding or to confirm the accuracy of data by increasing the ‘persuasiveness’ of evidence, enhancing the ‘validity’ of insights and adding ‘completeness’ to accounts. Second, it aims to combine research methods on pragmatic grounds (see for example, Cresswell (1995) and Tashakkori and Teddlie (1998)) where pragmatists are concerned with ‘what works’ and solutions to problems (Patton, 1990).

There are four main types of triangulation (Downward and Mearman, 2007):

- Methodological triangulation combines qualitative with quantitative data
- Data triangulation combines data from more than one method within the qualitative research approaches, or from more than one source or population.
- Investigator or analyst triangulation involves more than one researcher looking at the data to check and challenge interpretations and analysis and add a further perspective to the data analysis.
- Theory triangulation involves exploring data from different theoretical positions and assumptions. It can also involve testing the fit of the data with different theories.

### 10.2.2 Strengths and weaknesses of qualitative methods

The issue of whether a quantitative or qualitative approach to program evaluation is appropriate will depend on the nature of the evaluation and the data available to conduct the evaluation. There are a number of situations in which a qualitative approach to policy evaluation may be an appropriate method. These include when it is necessary to:

- understand the processes, impacts and outcomes of the policies and program
- address the complexity of social intervention and draw out and understand the multiple layers of policies
- address the personal, structural and environmental contexts of policy
- draw out processes, barriers, and facilitators of policy interventions
- highlight the intended and unintended consequences of an intervention
- identify people’s values, judgements and choices, and the impact policy has on their perceptions.

Qualitative research is particularly useful to gain an understanding of why, how, and under what conditions policies, programs and projects succeed or fail. It is thus particularly well suited to process evaluations (see chapter 4) and is likely to be the main method used in such evaluations. In this regard, Patton (2003, p. 2) notes:

Qualitative methods are often used in evaluations because they tell the program’s story by capturing and communicating the participants’ stories. Evaluation case studies have all the elements of a good story. They tell what happened when, to whom and with what consequences.

One of the great appeals of qualitative evaluation is the ability to render experiences in detail. In the context of program evaluations, it is often easier to communicate key findings using cases studies.
Qualitative methods can also be used for impact evaluations. As the Cabinet Office (2003, p. 8:6) explains:

Qualitative research also plays a key (although sometimes neglected) role in understanding impacts and outcomes. Rather than providing quantitative measurements of gross or net impact, it can answer more detailed questions which might be summarised as ‘how, under what circumstances, in what ways and for which types of people is the policy working … and what do we mean by “working” anyway?’ It can tell us about the range and types of impacts a policy has, giving a voice to outcomes that were not anticipated or intended and which an evaluator might not have thought to consider.

Sigsgaard (2002) summarises a qualitative approach to performance measurement called the Most Significant Change (MSC) approach. The MSC was originally designed for project evaluation in developing countries and is used as an alternative quantitative method to assess performance. The MSC approach involves an interviewer asking persons who have been involved in the project to identify positive or negative changes they have observed over a fixed time for one or more domains of interest. These same persons are then asked to indicate which change is the most significant and why. By interviewing different stakeholders, a series of change-related stories are recorded. These stories are reviewed and the most significant are selected along with the reasons for the choices. These performance stories are verified by additional investigation. The qualitative MSC method may be appropriate in settings where adequate quantitative data are not available or cannot be readily collected.

Governments frequently wish to understand an issue in detail and ensure as much information as possible is taken into consideration, when making policy decisions. Further, governments normally require that consultation occurs as part of either explicit or perceived mandates to be open and transparent, particularly in the context of regulatory review. Qualitative techniques – which examine issues in relatively more detail than quantitative approaches and seek to draw out and understand the multiple layers of policies and their impacts – may be an appropriate way to ensure that these requirements are met.

On the other hand, qualitative methods suffer from a number of apparent weaknesses (Rao and Woolcock, 2003, p. 167).

First, it is hard to make compelling claims about causality because it is not possible to control for other mitigating factors or to establish the counterfactual (Mohr, 1999). Nevertheless, consideration and documentation of rival explanations can add analytical rigor (Patton, 2002). Negative case analysis can also strengthen claims about causality. This involves searching for and discussing elements of the data that do not support, or appear to contradict, patterns or explanations that are emerging from data analysis – that is, elements that challenge the researcher’s own view of reality. Alternative explanations are considered and the findings are based on the explanation that best fits the majority of cases.

Second, it is difficult to make generalisations about the applicability of the findings to the wider population. This is because the individuals or groups being studied are usually small in number or have not been randomly selected. In this regard, there is often a trade-off between the depth (increasing the validity of the data) and breadth (increasing the representativeness of the data). As Patton (2002, p.14) notes:

qualitative methods typically produce a wealth of detailed information about a much smaller number of people and cases. This increases the depth of understanding of the cases and situations studied, but reduces the generalisability.
Qualitative methods focus on fewer cases but proponents argue that the quality and completeness of qualitative information outweighs any disadvantages due to a lack of representativeness. Furthermore, the Cabinet Office (2003) notes that whether or not the findings of qualitative research can be generalised depends on the type of investigation and the type of generalisation to be made.

A third weakness of the qualitative approach is that it is difficult to replicate the research findings and thus independently verify the results as the individuals or groups being studied are often selected idiosyncratically (for example on the basis of a judgement call by the lead investigator) or on the recommendation of other participants.

Finally, the analysis of qualitative information frequently involves interpretative judgement on the part of the researcher. Thus two researchers looking at the same information may arrive at different interpretations. This criticism may also be applied on a relatively weaker basis to quantitative methods as McDavid and Hawthorn (2006, pp. 9 & 10) note:

There are no program evaluations that can be done without the evaluators’ own experiences, values, beliefs and expectations playing an important role …

Key to understanding all evaluation practice is accepting that no matter how sophisticated your designs, measures and other methods are, you will exercise professional judgement in your work.

Nevertheless, McDavid and Hawthorn (2006) reject the notion that the need for interpretative judgement implies an ‘anything goes approach’ to qualitative analysis. Instead they advocate a ‘structured’ approach to evaluations that relies on understanding the tools that have been developed and applying them in ways to maximise the defensibility of findings. A similar position is taken by Patton (2002, p. 553) who notes:

being able to report that you engaged in a systematic search for alternative themes, divergent patterns, and rival explanations enhances credibility. This can be done both inductively and logically. Inductively it involves looking for other ways of organising the data that might lead to different findings. Logically it means thinking about other logical possibilities and then seeing if those possibilities can be supported by the data.

As discussed above, the potential for the qualitative researcher’s judgement to influence the findings can also be mitigated using ‘triangulation’. Such an approach strengthens the credibility of qualitative evaluations (Patton, 2002). Credibility can also be strengthened by seeking comments from industry experts and stakeholders as to whether the findings accord with their own views, taking account that those views are also unavoidably subjective (Miles and Huberman, 1994).

Given its various strengths and weaknesses, if it is to be robust and credible – and hence high quality – qualitative analysis should be systematic, rigorous, and comprehensive, and involve very close attention to detail (Cabinet Office, 2003). Patton (2002, p.5) identifies some of the factors that influence the quality of qualitative analysis:

The quality of qualitative data depends to a great extent on the methodologies, skill, sensitivity and integrity of the researcher. .. Generating useful and credible qualitative findings through observation, interviewing and content analyses requires discipline, knowledge, training, practice, creativity and hard work.
The final evaluation report is one of the key ways in which the quality of a qualitative evaluation can be demonstrated. The final report of a well designed and conducted qualitative evaluation should:

- provide a clear explanation of the research design and how the design was implemented
- reflect the complexity and subtlety of the research issue
- convey the depth and richness of the qualitative research data
- explain clearly what criteria have been used in the appraisal of the policy or service and the basis on which those criteria were selected
- demonstrate a clear link between the data and the conclusions based on analysis of that data
- provide an assessment of the relevance of the findings for areas beyond the studied sample and setting.

### 10.3 Mixed Methods

Historically, research in program evaluation has been heavily polarised along quantitative and qualitative lines (Rao and Woolcock, 2003, p. 168). However, quantitative and qualitative methods each have various strengths and weaknesses that have been highlighted in this report, and the distinction between the two is not always sharp. Thus, potentially, the methods may be complements to, rather than substitutes for, each other. The choice of an appropriate evaluation method should be based on a balancing of the relative strengths and weaknesses of the various methods having regard to the research question, the resources available to conduct the evaluation, and the audience for whom the report is intended (Cresswell, 1995).

Recognising the potential complementarities, Patton (2002, p. xxii) notes that a ‘mixed-methods’ approach may be preferable:

> The classic qualitative-quantitative debate [about the relative value of different methods and alternative paradigms] has been largely resolved with recognition that a variety of methodological approaches are needed and credible, that mixed methods can be especially valuable and that the challenge is to appropriately match methods to questions rather than adhering to some narrow methodological orthodoxy.

A mixed-methods approach, based on a combination of quantitative and qualitative data and methods, recognises that the strengths of one approach potentially complement the weaknesses of the other and vice versa. Papaconstantinou and Polt (1997, p.12) explain the synergies between quantitative and qualitative approaches to evaluations of economic policies as:

> Many participants felt that evaluation schemes should … combine quantitative and qualitative indicators of the impact of the policy measure and of the private and estimated social benefits of the induced change in behaviour. Ideally, they should be combined with the qualitative information from user-surveys, in-depth case studies and interviews to produce the variety of information needed by different users of evaluations. Single-approach evaluations might in-effect be downright misleading and putting too much emphasis on single quantitative estimation, while useful as a measure of cross-checking, might miss the essential qualitative effects of new initiatives.

Rao and Woolcock (2003) note that mixed-methods can yield insights that neither a quantitative or qualitative approach would produce on its own. They note (pp.165-166):
many important issues cannot be meaningfully reduced to numbers or adequately understood without reference to the immediate context in which they occur. Qualitative methods are particularly effective in delving deep into issues of process; [thus] a judicious mix of qualitative and quantitative methods can provide a more comprehensive evaluation of an intervention.

There are several examples of when a mixed-method approach may be useful in economic evaluation. For example, SCBA requires (among other things) that costs and benefits be monetised where possible, for the purposes of comparison. However where there are no observable market prices, the evaluation must contain a qualitative dimension in order to account for all impacts. Using qualitative methods, it may be possible to infer ‘willingness to pay’ by observing consumer behaviour. Alternatively, ‘willingness to pay’ might be estimated by asking people what they would be willing to pay for a particular benefit, or in the case of a cost, identifying the amount of compensation consumers would demand in order to accept the cost (HM Treasury, 2008, Green Book).

Additionally, a mixed-method approach may be a pragmatic solution to the time and resource constraints under which most evaluations must be conducted. In this regard, a well designed mixed-methods approach can maximise the range and validity of information that can be used given the time and resources available (Bamberger et al p. 11). Datta (1994, p. 67) also highlights the practical reasons for adopting a mixed methods approach:

[T]he differences [between the qualitative and quantitative paradigms] are less sharp in practice than in theoretical statements. The best examples of both paradigms seem actually to be mixed models. …. Perhaps this is not surprising. …. Most evaluations are conducted under many constraints. These include relatively short time frames, relatively little money, often intractable measurement challenges. … In most circumstances, evaluators have to do the best they can and need more, not fewer, approaches on which they can draw.

The following section examines some of the literature that has used qualitative methods to evaluate economic policies and the insights that can be drawn from these case studies. In most instances, a mixed-methods approach has been used, and this approach is particularly common in evaluations that are conducted for, or by, governments. This likely reflects the imperative for governments to take account of the widest available information, including the views of stakeholders.

10.4 Lessons from the Literature

A selection of the literature that has used qualitative or mixed-methods for evaluation of economic policies is shown in box 10.1. An analysis of the case studies indicate that qualitative methods are frequently the main method used in process evaluations and reviews of regulatory governance. In addition, qualitative methods and data are also used to complement or supplement quantitative methods that are used to conduct impact evaluations.

10.4.1 Use of qualitative methods in process evaluations and reviews of regulatory governance

As noted, qualitative methods are well suited to the conduct of process evaluations. Thus, there is a large body of literature of regulatory governance which seeks to answer questions about how policies are implemented. Some of this is summarised in box 10.1. The literature is divided into single-country assessments and studies that use the cross-country methods that are discussed above. A key finding of much of this literature is that the design and governance of regulation is likely to impact on
industry outcomes (Rodrik, 2003). The methods used to gather the information upon which these findings are based include a range of qualitative techniques such as the use of questionnaires. Some studies make extensive use of interviews with various interested parties, including: regulatory and government officials, executives of regulated firms, and consumers. Incorporating such information into the evaluation requires a qualitative approach in many circumstances, which may supplement other quantitative analysis.

Stern and Holder (1999) provide a systematic approach to evaluating qualitative or descriptive aspects of a jurisdiction’s regulatory regime. The authors review the literature on regulatory governance and suggest six broad criteria that might be used to evaluate regulatory governance in any one particular country (and at a particular moment in time). Each broad criterion in turn is to be evaluated on the basis of a series of questions regarding the regulatory structure and regime. These six criteria, and the associated questions, are outlined in box 10.2 below. The authors then apply their proposed criteria to the governance structures of 12 infrastructure industries across six developing Asian economies, based on comprehensive case studies analysed and discussed in NERA (1998). Each industry is then scored on a scale of A to E for each of the six criteria, resulting in 72 individual scores. The results are then tabulated and used to make a series of observations and suggest some conclusions.
**Box 10.1: Case Studies of Qualitative and Mixed-Methods**

<table>
<thead>
<tr>
<th>Authors and Year</th>
<th>Type of Study</th>
<th>Research Question</th>
<th>Methods Used</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wolf and Pollitt (2009)</td>
<td>Academic</td>
<td>Effect on welfare of partial privatisation of Norway’s Statoil.</td>
<td>Used social cost benefit analysis and mainly quantitative methods and data. However, informal interviews were conducted with a former member of the Norwegian state administration, a former board member of Statoil, and an external adviser to the government, all of which were closely involved in Statoil’s privatisation. These informal interviews were used ‘to inform, challenge or confirm’ Wolf and Pollitt’s principal observations and assumptions</td>
</tr>
<tr>
<td>ACIL (2005)</td>
<td>Consultant to Government Regulatory Authority</td>
<td>Estimate of the consumer benefits from Australia’s telecommunications reform</td>
<td>Quantitative CGE modelling was main method used. However, additional qualitative information, such as views of industry experts and participants, provided information about the potential effect of product innovation on end-users in that sector. ACIL considered that the qualitative information supplemented its quantitative analysis</td>
</tr>
<tr>
<td>Productivity Commission (1999a), (1999c), (2005a), (2005b), (2006)</td>
<td>Reports by Government Statutory Authority</td>
<td>Ex ante or ex post evaluation of impacts of aspects from micro-economic reforms</td>
<td>The Productivity Commission uses a variety of quantitative and qualitative methods and data in the conduct of its inquiries. Information is obtained from written and oral submissions to public hearings, workshops, forums and in response to draft reports and preliminary findings. The PC’s enabling legislation requires it to take account of all interests in the community generally, which normally includes a range of qualitative impacts</td>
</tr>
<tr>
<td>Dixit et al (2007)</td>
<td>A publication of the Electricity Governance Initiative (EGI). EGI is a joint undertaking of the World Resources Institute and Prayas Energy Group (India).</td>
<td>Provides a framework for assessing and promote good governance in the electricity sector. The framework assesses a number of qualitative indicators of regulatory governance, including the extent to which decision making processes are transparent, allow for public participation, remain accountable to the public interest and include dispute resolution processes</td>
<td>The toolkit consists of a set of qualitative research questions which are used to gather information with which to derive indicators that are used to assess decision-making processes. For each research question, indicator values of (i) Low (ii) Low-Medium (iii) Medium (iv) Medium-High or (v) High are possible. Each value is based on a documented explanation of the extent to which particular attributes of electricity governance have been met. The design of the indicators seeks to minimise the scope for arbitrary or inconsistent value judgments. The choice of, and number of, indicators seeks to balance the need to be comprehensive and capture the full range of governance considerations against the need for simplicity and manageability</td>
</tr>
<tr>
<td>Source</td>
<td>Format</td>
<td>Purpose</td>
<td>Methodology</td>
</tr>
<tr>
<td>-----------------------------</td>
<td>---------------------------------</td>
<td>----------------------------------------------------------------------------------------------------------------------------------------</td>
<td>------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------</td>
</tr>
<tr>
<td>OECD (2005)</td>
<td>Report of multi-national,</td>
<td>This is one of the Economic Surveys is published every 1.5 to two years for each OECD country. Each survey identifies the main economic challenges faced by the particular country and analyses policy options to meet these challenges.</td>
<td>Each economic survey follows a consistent process and reports in a consistent manner. Each survey is led by two members of the OECD’s Economic and Development Review Committee and is often informed by experts in the home-country. The examined country is represented by a delegation of high-level officials from across government departments. The analysis is conducted within the OECD secretariat which, at an early stage of the project, visits the country and meets with a wide range of government officials, academics, social partners and other experts to collect information – both qualitative and quantitative. The secretariat’s tentative conclusions are also discussed at a later stage with key policy makers, such as the Minister of Finance, top government officials, the central bank and also labour unions and business confederations. Although a range of quantitative information is presented, and the results of other quantitative studies may be considered, the main analytical approach in the surveys is qualitative.</td>
</tr>
<tr>
<td>Access Economics (2006)</td>
<td>Report of Private Consultant</td>
<td>Uses a ‘score-card’ approach which is intended to focus on the extent to which the design of a particular regulatory regime is likely to foster good decisions and outcomes.</td>
<td>The scorecard for each regulatory regime was constructed by rating four broad areas of regulatory design – independence, the extent to which the regime is focussed on promoting efficiency; the transparency, predictability and consistency of the regime; and the accountability of the regulator. Indicators were identified for each of these areas. Each indicator was then assigned a rating on a scale of ‘Very Poor’, ‘Poor’, ‘Fair’, ‘Good’ and ‘Very Good compared with best-practice design principles.</td>
</tr>
<tr>
<td>European Competitive</td>
<td>Report of an industry association</td>
<td>Compiles an annual regulatory scorecard that compares the regulatory environment in 18 EU Member States, Norway and Turkey in the electronic communications sector and the effectiveness of this environment in promoting the objectives of the EU regulatory framework. The scope of the survey includes the wider institutional and legislative environment affecting</td>
<td>The Scorecard is based on responses submitted by National Regulatory Authorities (NRAs) and ECTA members to a detailed questionnaire consisting of 104 questions. The questionnaire was compiled following consultation with NRAs, ECTA members, the European Commission, and the European Regulators Group (ERG). On the basis of consolidated responses received for each country from the various stakeholders, a comparative quantitative analysis was carried out, resulting in an overall score for the effectiveness of the regulatory environment in each country.</td>
</tr>
</tbody>
</table>
the sector as well as the application of regulation by national telecommunications regulators.

<table>
<thead>
<tr>
<th>Cross-country Descriptive Analyses</th>
</tr>
</thead>
<tbody>
<tr>
<td>Stern and Holder (1999)</td>
</tr>
<tr>
<td>Ocaña (2002)</td>
</tr>
<tr>
<td>CEER (2004)</td>
</tr>
<tr>
<td>European Commission (2005)</td>
</tr>
</tbody>
</table>
### Box 10.2: The Stern and Holder (1999) Regulatory Criteria

| 1. Clarity of Roles and Objectives | • Does the primary legislation have a clear definition of the regulator’s functions and duties? If not, are these functions and duties formally set out in legal instruments and/or other documents?  
| • Is it clear where the regulator has an advisory role rather than a decision making role?  
| • Are there any functions carried out jointly or any that are ambiguous, between the regulator and those of the relevant Minister(s)?  
| • Does the legislation establish unambiguously which entity is responsible for what regulatory functions?  
| • Does the regulator have any responsibility for commercial activities? |
| 2. Autonomy | • What is the relationship between the regulatory body and the government  
| • How are members of the regulatory body appointed and dismissed?  
| • How is the regulatory body financed? Who is to approve the funding?  
| • In what cases does the regulator have only an advisory role? |
| 3. Participation | • Does the regulator formally involve stakeholders in: (1) major decisions? (2) the proposed approach to taking major decisions?  
| • Are consultation responses made public (either in full or in a summary of responses?)  
| • Does the regulator comment publicly on points made in consultation responses, and how these have affected the final decision?  
| • Are there any other indications of whether or not consultation responses influence the final decision?  
| • In what other ways are stakeholders involved in regulatory decision making and processes? |
| 4. Accountability | • Is there a formal mechanism for the regulator’s decision to be challenged? If so, how does it work and has it been used?  
| • Is there also a legal right of redress?  
| • Do firms or other parties comment on decisions through informal channels (such as direct representations to Ministers, or through the media)?  
| • To whom is the regulatory body accountable (government, parliament, courts)?  
| • Does the regulator have to answer questions before parliament?  
| • Can the regulator be dismissed for failing to fulfil his duties?  
| • Is there under primary law a facility for judicial review? |
| 5. Transparency | • Are major regulatory documents (e.g. licences) in the public domain?  
| • Does the regulator publish major decisions (or advice)?  
| • Does the regulator publish the reasoning behind major decisions?  
| • If decisions/reasons are published, is this voluntary or compulsory (e.g. a legal requirement to justify decisions)?  
| • If decisions/reasons are not published, are any participants (such as the firms themselves) told of the reasons for major decisions? |
| 6. Predictability | • How easily can the regulator’s functions and duties be changed and what is involved?  
| • How easily can key regulatory documents be changed and what is involved?  
| • To what extent are regulatory principles (e.g. on the procedural approach to tariff reviews, the definition of the base rate or the rate of return which a firm should be allowed to earn) set out formally?  
| • Have the regulator’s decisions demonstrated a consistent approach?  
| • Is there a published timetable of regulatory events every year?  
| • To what extent are regulatory arrangements in the sector part of a coherent approach (i.e. echoed in other infrastructure sectors) and to what extent are they ad hoc (i.e. difference in institutional structure from other infrastructure sectors in the country)? |
An example of a systematic approach to single-country case studies is provided by Brown, Stern and Tenenbaum (2006) in the World Bank’s *Handbook for Evaluating Infrastructure Regulatory Systems*. This is essentially an evaluation of regulatory governance (chapter 4). The purpose of the Handbook is to provide a guide for World Bank and other professionals to evaluate the effectiveness of regulatory regimes. It proposes an evaluation method for the conduct of extensive, structured, single-country case studies. It proposes that the best approach to regulatory evaluation is through well written, comprehensive case studies prepared by recognised and respected individuals who demonstrate a clear understanding of a country’s political, economic, and legal realities (Brown et al, 2006, p. 37). It proposes this type of study in preference to cross-country analyses (statistical or qualitative).

This publication provides extensive sets of broad principles that it considers make for good regulatory governance. In respect of each it provides criteria against which all case studies of regulatory regimes should be conducted. In turn, the publication provides a detailed set of questions and points of evaluation in respect of each of these broad standards. In consequence, if case studies are conducted faithfully according to this World Bank guide, there ought to be a high degree of uniformity and inter-compatibility among different single-country case studies.

The UK’s experience with implementation of the recommendations of the Hampton Report (Hampton, 2005), also provides some useful insights into how well-designed and applied qualitative methods might be used in process reviews. Further detail about the implementation and review of the Hampton Report is contained in box 10.3. As can be seen from that box, a range of qualitative methods were used to generate evidence on which to base the implementation reviews’ findings. Each review took the same approach based on guidance documents prepared for the purpose of conducting the reviews. A range of documented qualitative methods were used. Consistent with the qualitative approach to program evaluation, it was explicitly recognised that the review’s findings reflected, in part, each team member’s own judgement.
The Hampton Report reviewed the UK’s regulatory and enforcement regimes and made a number of recommendations to reduce the burden on business created by regulatory systems, including through the adoption by regulators of a number of principles (known as the Hampton principles). In 2006, the UK’s National Audit Office (NAO) and the Better Regulation Executive (BRE) were asked to jointly review the five largest UK regulators’ progress in implementing the Hampton principles (Environment Agency, Financial Services Authority, Food Standards Agency, Health and Safety Executive and the Office of Fair Trading). The results of the reviews were published in separate reports in March 2008. The reviews were conducted over a two week period in 2007 by teams of four senior individuals representing the NAO, the BRE, a regulator and an external body. Each review was preceded by a longer period of intensive research and preparation by the NAO and the BRE.

The Review teams used guidance from the NAO and the BRE in relation to the high level questions for the review and the framework in which the review should be conducted. Guidance was also provided on the types of evidence sought. The principal methods used were:

- interviews with regulatory staff and senior managers
- interviews with other stakeholders including individual businesses and trade associations
- focus groups of inspectors and businesses
- observational visits, including inspections
- document review, particularly of strategies and plans, risk method and data on effectiveness and outcomes.

Conclusions were based on the evidence available to the review team, and their own judgement. The assessment reports represented the agreed views of individual teams members.

Based on National Audit Office (2008).

Although used in a slightly different context, the European Commission’s Merger Remedies Study (DG Competition, 2005) also provides a good example of how qualitative methods might be used to conduct *ex post* evaluations. Further information about this document is shown in box 10.4. Of particular note, is the manner in which the study carefully explained its research design, the methods used to collect and analyse the data, and to validate the findings drawn from the data analysis. In addition, the study set out the limitations in its own research methods. The quality of the research report enables the reader to readily draw their own conclusions about the robustness and credibility of the study’s findings.
The European Commission Merger Regulation (ECMR) allows parties to European mergers that raise competition concerns to offer commitments to the European Commission (EC) that address those concerns. The objective of the EC’s Merger Remedies Study (DG Competition, 2005) was to conduct an ex post review of the design and implementation of commitments offered and accepted by the EC so as to identify areas where further improvements to its merger remedies policy and procedures may be necessary. The focus of the Study was to identify the factors and/or processes that may have positively or negatively influenced the effective design and implementation of merger remedies.

The Study analysed 40 decisions adopted by the EC in the period 1996-2000, covering 96 remedies. The selection of decisions was aimed at creating a balanced sample in relation to the types of remedies, the stage at which remedies were accepted, and the type of industry.

After the cases were selected, interview teams reviewed the case files and prepared sample questionnaires of around 120 questions, which were evaluated in a pilot test of nine cases and modified as required. The questionnaires were tailored according to the type of interviewee (the buyers, sellers, trustees).

Using these questionnaires, interviews were conducted with 145 company officials who had been personally involved in the negotiation and/or implementation of merger remedies. Each interview followed a structured but open interview format. The open format allowed the interviewees to raise issues on their own initiative and enable the interview team to discuss either case and/or remedy specific issues in depth. The responses were tape-recorded and detailed minutes prepared. The answers were cross-checked with answers from other interviewees who had been involved in the same remedy. The extent of the robustness and consistency of the replies were discussed both within the interview teams and with other members of the study team. Interview responses were also compared with information from other sources, such as the case file, written answers to specific follow-up questionnaires, and other publicly available sources.

On the basis of the collected information and the case file, interview teams drafted remedy reports for each remedy in accordance to a standard format, and held case discussions with the interview team and with other members of the study group. The remedy reports were also given for comment to DG COMP officials who had conducted the merger matter at the time.

Limitations of the method

The evaluation team recognised that their method had a number of limitations. In particular, it was difficult to determine market outcomes in the absence of a fully fledged new market investigation and the study relied to a large extent on the interviews and data provided by key participants in the process. Secondly, the team recognised that exogenous factors likely contributed to observed market outcomes. It was not easy to attribute causality using their qualitative method although the team recognised that this would still be difficult even if a ‘full-blown’ quantitative market investigation had been conducted.

Finally, the study sought to compare actual market developments with the results that would have been likely to occur, absent the remedy – the counter-factual scenario. The study team considered that the interview method was a useful way to get industry participants’ views on these issues but there was limited opportunity to interview companies that had complained of possible anti-competitive effects during the initial investigation. The study team considered that these limitations circumscribe the robustness of their results which should therefore be treated as mainly indicative.


Several of the studies shown in box 10.1 are based on what might be termed mixed-methods for program evaluation. For example, a number of reports by the Productivity Commission are based on a range of quantitative and qualitative data and quantitative and qualitative methods for analysing those data. Examination of many of the PC’s reports indicates that the views of inquiry participants are frequently quoted in its analyses and findings. This qualitative information is obtained in a number of ways including public hearings, submissions and workshops. The views of many participants are normally quoted directly in the inquiry’s draft and final reports, along with the results of its own quantitative and qualitative analysis. The issue with
this general approach is not the use of quotes per se; rather it is one of ensuring the weighting of the views is done objectively.

10.5 Conclusions

Qualitative analysis involves the use of a wider range of information and often a higher level of detail than quantitative analysis. This level of detail can generate better understanding of observed outcomes, particularly in the context of measuring impacts among complex economic relationships. However this value must be balanced against the often limited ability to draw wider implications from specific circumstances.

In summary, the preceding discussion suggests a number of instances where qualitative methods may be an appropriate technique for evaluating competition and regulatory reforms.

First, qualitative methods are likely to be the main approach taken in process evaluations and reviews of regulatory governance (see chapter 4 for a discussion of these). Sometimes, quantitative techniques can be used to code qualitative information gathered as part of those evaluations. Once coded, the data may be interpreted using quantitative methods. For example, the frequency of certain words appearing in documents or testimony can indicate overall themes and patterns.

Qualitative methods may be necessary as an alternative to quantitative methods if adequate data to support quantitative methods are not available. Cubbin et al (2008) note this was a reason that qualitative methods were often used in early case studies of competition and regulatory reforms, such as Bishop, Kay and Mayer (1995). In those early studies, insufficient time had passed to enable a large-enough dataset to be assembled for quantitative analysis. In spite of such trade-offs, these early case-studies had the advantage of enabling more detailed examination and understanding of the nature of relationships at play. Nevertheless, these early studies are good examples of the trade-offs that might be necessary when conducting real-world evaluations.

Finally, qualitative methods may be appropriate as a complement to findings based on quantitative approaches. For instance, the need to account for qualitative impacts is of particular interest to governments, due in part to an interest in varying sources of information, but also because of the requirement for policy evaluation to generate prescriptive content. A particular focus of qualitative analysis is to establish the meaning and broader policy relevance of results and findings. In the context of evaluation, qualitative analysis emphasises the ‘how’ and ‘why’ associated with the phenomena of interest, in addition to measuring ‘what’ actually happened. An evaluation of the process, outcome and impact of competition and regulatory reforms is likely to necessitate the use of both qualitative and quantitative methods of analysis.
Section III: Conclusion

This report has considered the issues that might arise, and the methods that may be used, when conducting an *ex post* evaluation of institutional, competition and regulatory reforms to economic infrastructure, as occurred throughout the OECD, during the 1980s and 1990s. The five key elements of these infrastructure reforms were corporatisation, structural change and competition, privatisation of former public-service utility operations, and the introduction of independent and more incentive compatible regulation.

Guided by the predictions of economic theory, a range of reforms were implemented in the expectation that desirable economic outcomes would be achieved. It was hoped that these changes would be reflected at the macroeconomic level by a more productive and efficient economy as indicated, for example, by increases in real GDP per capita and increased productivity. At the micro-economic level, the reforms were expected to lead to closer alignment of prices with underlying costs and increased cost efficiency.

In Australia’s case the period after the implementation of the initial reforms was followed by a large increase in real GDP and real GDP per capita. Should we then conclude that the reforms have been a success? This report demonstrates that any such conclusion based on such limited evidence would not be justified. Evaluating the success or otherwise of infrastructure reforms is much more difficult. The literature extensively cited in this report documents the struggles and pitfalls of different evaluation processes.

An *ex post* evaluation of regulatory and competition reforms typically involves a comparison of actual outcomes with those that would have occurred absent the reforms that are being evaluated – the counterfactual scenario. The report suggests that specifying this scenario is the most challenging aspect of conducting an economic evaluation. Further, the success or otherwise of infrastructure reforms can be evaluated using a number of methods (explained in section II of this report). However, the choice of method must be guided by the question that the evaluation seeks to answer.

From these two broad conclusions the report offers the following advice to those embarking on an evaluation exercise:

First, the question(s) that the evaluation seeks to explain must be clearly specified and capable of being answered. Generally, a specific evaluation question is easier to answer than a more general question.

Second, the evaluation should be well-designed and explained and this will have different implications depending upon the approach taken to the counterfactual:

- If a treatment-effects approach is taken, the evaluation should consider the likelihood and implications of any initial differences between the treatment and control groups, and factors other than the treatment which might explain observed differences between those groups after the policy intervention.

- If a structural approach is taken, the assumptions underlying the specification of the counterfactual scenario and the rationale for those assumptions should be carefully set out. Similarly, the assumptions underlying any economic models
used to conduct the evaluation (such as functional forms and/or parameter values) should also be clearly explained.

- Sensitivity analysis of the findings to changes in key assumptions or parameters should be conducted and the outcomes and implications of the sensitivity analysis explained in the written evaluation report.

Third, by their nature, infrastructure reforms are intended primarily to impact on variables that lend themselves to quantification, such as prices, costs, efficiency etc. Thus, normally a quantitative approach to evaluation would be preferable to a qualitative approach. However, a well-designed qualitative approach may be useful to supplement the quantitative assessment. A qualitative approach will also be the main approach taken to a process evaluation that seeks to evaluate the way in which infrastructure reforms have been implemented.

Fourth, the data used in the evaluation should be explained, including a discussion of their sources and measurement. Any adjustments made to the data should be explained. Subject to confidentiality, the data should be available to third parties so that the evaluation’s findings can be replicated.

A recurring theme in the empirical literature that has been reviewed in preparing this report is that trade-offs will inevitably be required between theoretically ideal evaluation designs and methods, and those that are achievable in practice. These trade-offs arise not only because of the difficulty of specifying a defensible counterfactual, but also because of the resources that may be available to conduct an evaluation. For example, a CGE structural-modelling method may be the preferred approach in principle, but a suitable model may not be available because of time and financial constraints. Thus, a practical approach to the evaluation may require a trade-off to be made in the selection of evaluation method in order to accommodate the resources available to conduct the evaluation.

More generally, trade-offs are often required as the ‘ideal’ data are seldom available for the conduct of an evaluation. An absence of reliable and consistent data may mean that the preferred approach to evaluation cannot be undertaken and that compromise approaches must be used.

Given the significance of competition and regulatory reforms relating to economic infrastructure, and the expectations for positive improvements in economic performance that accompany them, it is anomalous that data to enable the impact and effectiveness of these reforms to be evaluated may not be readily available. It is a final suggestion of this study that data collection be carefully considered with reforms to enable defensible, credible and robust ex post evaluations to be conducted.
**List of Abbreviations**

- ACCC – Australian Competition and Consumer Commission
- ACG – Allen Consulting Group
- AER – Australian Energy Regulator
- AEMC – Australian Energy Market Commission
- AGE – Applied general equilibrium
- ATE – Average treatment effect
- Capex – Capital expenditure
- CAN – Customer access network
- CBA – Cost benefit analysis
- CES – Constant elasticity of substitution
- CGE – Computable general equilibrium
- CN – Canadian National Railway
- COAG – Council of Australian Governments
- CoPS – Centre of Policy Studies
- CPI – Consumer price index
- CSO – Community service obligation
- DEA – Data envelopment process
- ECTA – European Competitive Telecommunications Committee
- GDP – Gross domestic product
- GBE – Government business enterprises
- GTE – Government-owned trading enterprises
- IC – Industry Commission
- ICT – Information and communication technology
- K-H – Kaldor-Hicks
- LRMC – Long-run marginal cost
- MC – Marginal cost
- MPPK – Marginal product of capital
- MR – Marginal revenue
- MTE – Marginal treatment effect
- NCC – National Competition Council
- NCP – National Competition Policy
- NPV – Net present value
- NRA – National Reform Agenda
- OBPR – Office of Best Practice Regulation
- OECD – Organisation for Economic Co-operation and Development
PC – Productivity Commission
PV – Present value
RAB – Regulatory Asset Base
RIS – Regulation impact assessment
SAM – Social accounting matrix
SCBA – Social cost benefit analysis
SFA – Stochastic frontier analysis
TFP – Total factor productivity
USO – Universal service obligation
WACC – Weighted average cost of capital
WEA – Wheat Export Authority
WMA – Wheat Marketing Act
WTP – Willingness to pay
ΔDWL – Change in deadweight loss
Appendix A: References


ACMA (2005), Consumer Benefits Resulting from Australia’s Telecommunications Sector, prepared for the Australian Communications and Media Authority by ACIL Tasman, November.


AEMC (2009), Review into the Use of Total Factor Productivity for the Determination of Prices and Revenues – Preliminary Findings, 17 December, Sydney.


Allen Consulting Group (2004), Benefits Resulting from Changes in Telecommunications Services, Report to Australian Communications Authority, 20 August.


Cabral, L (2000), Introduction to Industrial Organization, Massachusetts Institute of Technology.


Carr EH (1990), What is History?, Penguin reprint.

Centre for International Economics (1999), Guidelines for NCP Legislation Reviews, report prepared for the National Competition Council.


Dee, P (2005), Quantitative Modelling at the Productivity Commission, Productivity Commission, Melbourne.


Denzin, NK and YS Lincoln (eds.) (2000), *Handbook of Qualitative Research* (2nd ed), Sage Publications.


Diewert, WE (1976), ‘Exact and Superlative Index Numbers’, *Journal of Econometrics*, 4, pp. 115-145.


Hicks, JR (1943), ‘The Four Consumer’s Surpluses’, Review of Economic Studies, 11, pp. 31-41.


Independent Committee of Inquiry (F. Hilmer, Chair) (1993), National Competition Policy, AGPS, Canberra.


NERA (2007), Assessing the Impact of Competition Policy Reforms on Australia’s Infrastructure Performance, A Report for the Australian Competition and Consumer Commission, July,


Patton, MQ (2002), Qualitative Research and Evaluation Methods (3rd ed), Sage, California.

Patton, MQ (2003), Qualitative Evaluation Checklist, Available at: http://www.wmich.edu/evalctr/checklists/qec/ [accessed on 4 May 2010].


Productivity Commission (1999a), Impact of Competition Policy Reforms on Rural and Regional Australia, Report no. 8, AusInfo, Canberra.


