Review of willingness-to-pay methodologies

Prepared for
Independent Pricing and Regulatory Tribunal of NSW

Centre for International Economics
Canberra & Sydney

17 August 2001
The Centre for International Economics is a private economic research agency that provides professional, independent and timely analysis of international and domestic events and policies.

The CIE’s professional staff arrange, undertake and publish commissioned economic research and analysis for industry, corporations, governments, international agencies and individuals. Its focus is on international events and policies that affect us all.

The CIE is fully self-supporting and is funded by its commissioned studies, economic consultations provided and sales of publications.

The CIE is based in Canberra and has an office in Sydney.

© Centre for International Economics 2001

This work is copyright. Persons wishing to reproduce this material should contact the Centre for International Economics at one of the following addresses.

CANBERRA
Centre for International Economics
Ian Potter House, Cnr Marcus Clarke Street & Edinburgh Avenue
Canberra ACT
GPO Box 2203
Canberra ACT Australia 2601
Telephone +61 2 6248 6699 Facsimile +61 2 6247 7484
Email cie@intecon.com.au
Website www.intecon.com.au

SYDNEY
Centre for International Economics
Level 8, 50 Margaret Street
Sydney NSW
GPO Box 397
Sydney NSW Australia 1043
Telephone +61 2 9262 6655 Facsimile +61 2 9262 6651
Email ciesyd@intecon.com.au
Website www.intecon.com.au
# Contents

1 Why is willingness to pay information required?  1  
   Information on ‘WTP for quality’ and efficient regulation  2  
   The tension between quality determination and efficient costs  3  
   WTP for quality and departure from ‘postage stamp’ pricing  4  
   The report  6  

2 Quality of service attributes  7  
   An attribute approach to defining quality  7  
   Types of values  8  
   Defining the quality attributes  9  
   Attribute measurement  10  
   Attribute levels  12  

3 Review of stated preference methodologies  13  
   The family of stated preference techniques  13  
   Contingent Valuation  14  
   Choice Modelling  19  
   Conjoint Analysis  21  

4 Best practice guidelines  25  
   Assessing validity and reliability  25  
   Technique selection  27  
   Technique application  28  
   Questionnaire design  29  
   Survey logistics  31  
   Other technical considerations  32  

5 Conclusions and summary  37  

References  42
Boxes, charts and tables

2.1 Examples of quality of service attributes associated with electricity supply. 8
3.1 The family of stated preference techniques. 14
3.2 Embedding effects 18
3.3 Illustrative example of a choice set used in a Choice Modelling questionnaire a 19
3.4 Experimental design 20
3.5 Strengths of Choice Modelling 21
3.6 Illustrative example of a Conjoint Rating question a 22
3.7 Illustrative example of a Conjoint Ranking question. 23
3.8 Example of a Conjoint Paired Comparison question, adapted from an application by Johnston and Desvousges 24
4.1 The equivalence of willingness to pay (WTP) and willingness to accept measures (WTA) 34
Why is willingness to pay information required?

TO PROTECT CONSUMERS FROM the potential abuse of monopoly power most utilities are subject to both price controls and quality of service controls. However, there is tension between these controls. The threat of too little or too much quality accompanies the regulator’s decision on allowable revenue in pricing reviews. Desirably this threat could be reduced by information on the willingness to pay for different levels of service quality and the cost associated with them.

The efficient level of quality of service is found where the cost of supplying additional service quality is equal to the marginal consumer benefits. Typically the utility knows the cost of providing different levels of service. These costs may relate to infrastructure. However, typically neither the regulator nor the utility knows the value consumers’ place on different service levels.

The Independent Pricing and Regulatory Tribunal has commissioned the Centre for International Economics to review the methodologies used to assess willingness to pay for environmental or service attributes of utilities. The review includes:

- an assessment of the strengths and weaknesses of different valuation techniques and appropriate areas of application;
- comment on the robustness of each of the techniques;
- criteria for assessing the reliability of willingness-to-pay studies in the form of a set of guidelines; and
- an assessment of the validity of submissions made to the Tribunal on willingness to pay.

This study is a sequel to work commissioned by the Tribunal from the Allen Consulting Group entitled The Incorporation of Service Quality in the Regulation of Utility Prices (IPART, March 2001). It is a revised version of a paper discussed at a forum organised by the Tribunal and held on 5 July,
1 WHY IS WILLINGNESS TO PAY INFORMATION REQUIRED?

2001 and reflects that discussion and subsequent written comments provided by participants.

Information on ‘WTP for quality’ and efficient regulation

Utilities such as energy distribution networks, transport systems and water and waste water are ‘natural monopolies’ because the physical distribution of power, gas and water involve large fixed costs for capital equipment and relatively insignificant operating costs. The most economical way of distributing power is to have one utility operating in a given geographic area. But this leads to an absence of competition in the market place. In addition, consumer demand for electricity and other ‘essential’ services is inelastic. These factors give rise to the potential for unregulated distributors to reduce service quality and increase price – capturing a greater share of the consumer surplus than would be possible in competitive markets. In New South Wales and elsewhere, a regulatory response to this possibility is provided through the price setting activities of bodies such as the Tribunal.

The efficient regulation of utilities requires hidden information problems to be overcome. The efficient cost levels (including the costs of offering different service quality levels) are typically much better known to the utility than to the regulator, from whom they are hidden. The valuations placed on higher or lower quality by consumers are typically hidden from both. Yet in its efforts to set price or revenue paths for the utilities it regulates, the regulator must allow for the influence of different quality levels on cost and cost recovery mechanisms and on the level of demand.

It is difficult to estimate consumers’ value function for quality. This is because in an environment of only limited competition among service providers there is little, if any, opportunity to observe differentiation between firms in the quality they offer. Consequently, there is little market information on customer preferences and willingness to pay (WTP) for quality. One exception is the ‘green power’ schemes where customers have the choice of paying a premium for electricity that is generated using renewable processes and is perceived by some consumers to be worth more for that reason. Customer demand for ‘green power’ could be used as a guide to WTP for ‘environmentally friendly’ technology.

---

1 Consumers’ surplus is the satisfaction that consumers obtain from a good over and above the price paid. This is the difference between the maximum demand price that they would be willing to pay and the price that they actually pay.
Other factors that complicate the task of estimating consumer preferences are the multi-dimensional nature of service quality and the fact that some attributes of quality are lumpy — that is they cannot be consumed in continuous amounts. Examples include air conditioned transport, underground cabling, fluoridated water and the like. Others, like frequency of transport service, travel time and costs, are continuously variable.

In situations where prices are market determined, quality differentiation sits alongside price as something that firms can control in the interests of influencing market share and profit. The availability of choice on the consumer’s part in such situations does two things. It enables the consumer to make trade offs between price and quality. And the choices so made provide valuable feedback to firms trying to meet the market in terms of the public’s willingness to pay for various aspects of ‘quality’.

But the extent to which preferences are revealed through observed market behaviour is limited, even in unregulated markets. The dynamic nature of tastes and the impact of changing technology drives firms in more or less competitive markets to gather other information on WTP for altered or new products as well as drawing on what the market reveals about consumer preferences. What they do share in common with unregulated markets is a need for reliable information on individuals’ willingness to pay for quality.

Regulated utilities differ from unregulated firms and from each other in the extent to which market-based information on the value placed on quality is useful to them and to the regulator. In the case of urban water and wastewater services, electricity and gas distribution networks (as opposed to energy retail services) consumer choice is often minimal and the opportunity for one business to win market share from others through quality variation (or any other means) is limited. In public transport the situation is different. Alternative transport modes are often available. Gas as an optional and partial household energy choice may also be differ from, say, water and waste water services. As a consequence the best methods for finding out about WTP may differ.

The tension between quality determination and efficient costs

In its role as price regulator for a range of NSW public utilities the Tribunal has seen merit in incentive-based regulation. This approach recognises that if the interests of all stakeholders (consumers, owners, operators and their employees, taxpayers etc) are to be served in a balanced way, the regulatory requirements must embody an element of reward for improved practice. Better cost outcomes have been one critical dimension of im-
proven practice targeted by the Tribunal. But the tension between incentives to extract cost efficiencies and incentives to provide appropriate quality levels is ever present.

The very incentive structure of regulation of price/revenue that places strongest pressure on costs may be the least effective in preserving or enhancing quality of service. So-called ‘strong incentives’, (see Tirole 1994) that allow the utility to retain most or all of the difference between realised costs and a capped revenue in the interests of driving down costs, may provide incentives to skimp on quality. ‘Weak incentives’ that lean more towards cost reimbursement encourage gold plating.

The threat of too little or too much ‘quality’ accompanies the typical decision on allowable revenue that the Tribunal must make in its pricing reviews. Where minimum quality standards are set externally by others (as with health guidelines on drinking water quality, EPA acceptance standards on treated sewage discharge or safety standards for rolling stock or Department of Energy technical requirements for cable insulation) consumers’ (un)willingness to pay has been overridden. Consumer sovereignty has been truncated in the wider public interest of public health, clean waterways, safe power lines etc. But the issue of WTP for more than the required minimum is a live issue for the regulator in these circumstances.

Where people are willing to pay more than an amount that would fund the licensed minimum standards, or for some aspect of a service which is not standard-regulated, such information becomes important for regulated pricing decisions. The challenge is to elicit WTP information which is robust enough for the purpose.

**WTP for quality and departure from ‘postage stamp’ pricing**

In its recent discussion paper on ‘Meeting customer requirements under the regulatory framework’ EnergyAustralia has referred to ‘value based user pays’ to distinguish pricing to capture user benefits as distinct from pricing for cost recovery. But the critical question is how much of these user benefits generated by costly increases in quality should the utility capture? And for how long? More fundamental still, what measure of the willingness to pay for various quality attributes should be the basis for this charging and how should it be established?

Ideally, efficient pricing will reconcile the benefits and costs of additional quality. It does so when additional quality demanded is equal to the
incremental costs of providing it. Put another way, this happens when the marginal costs of quality are equal to the marginal benefits.

EnergyAustralia argues for a form of quality based price differentiation which is a mixture of differentiation based on quality driven cost and price discrimination based on benefit received. What is the link between ‘benefit received’ and WTP?

Charging on the lines advocated by EnergyAustralia would see ‘reliability’ emerge as a quality attribute and would see some rural customers with less reliable services paying less per kWh than city consumers with reliable services, undergrounding costing more than overhead services etc. This would be a departure from current arrangements where all (residential) customers who consume the same kWh pay the same price (with the exception of greenpower consumers).

**Cost reflective pricing, differential quality and WTP**

If the avoidable costs of existing services of differing quality differ, there may be a case for price differentiation on the grounds of avoiding internal cross subsidisation. If it is cheaper to provide services to one area which suffers greater interruption than to another, then cost reflective pricing may suggest differential pricing. Similarly if different customers can be identified as having different willingness to pay for additional quality and this WTP exceeds the additional costs this could be grounds for providing it and charging for it. Neither of these departures from existing practice need imply that anything more than cost recovery is involved. However, the Tribunal is being asked to consider allowing revenue recovery over and above the avoidable costs of additional quality.

**Capture of consumer surplus when WTP for additional quality exceeds additional costs**

EnergyAustralia argues, for instance, that mere cost recovery would not provide the incentive for utilities to invest in quality upgrades. They put a case for allowing utilities to capture more than the additional cost for extra quality, at least for more than one regulatory period. The discovery of latent demand and WTP for quality improvements in unregulated markets allows service providers who enjoy some market power to introduce improvements and recover more than the cost of doing so until competition eliminates these margins. It is argued that the prospect of these returns must be sufficient to overcome the risks associated with the investment.
Risks associated with quality improvement in services provided by the regulated utilities in NSW will depend on the degree of choice open to consumers. We can expect this to vary significantly between, say, domestic waste water services and ferry services. Consumer and revenue responses to higher service quality-induced prices can mean that the riskiness of the respective investments is quite different.

Whatever the merits of differential pricing based on quality, the need for reliable, best practice estimates of WTP for any proposed quality change is evident if such pricing practices are to be welfare improving for the community. Regulators and the broader public need to be confident that the methods of estimating WTP that underlie requests for quality related pricing provide robust estimates. Critically, the methods need to offer a meaningful range of such values since a single point estimate will be of limited help.

Unfortunately, as this discussion will reveal, it has not always been usual practice by Australian utilities to disclose key details on methodology and findings in a way that serves these ends.

The report

The objective of this report is to critically review methodologies used to assess the willingness to pay for quality improvements in services provided by utilities. The focus will be on electricity distribution but the applicability of methodologies to other utilities, including gas, water, and public transport are also examined.

Chapter two examines which quality of service attributes are important and outlines the methodologies for defining and selecting attributes. Chapter three reviews the methodologies for estimating consumers’ willingness to pay for an increase in quality levels. Drawing from this review and examples of their application chapter four outlines the best practice guidelines and the conclusions follow this in chapter five.
Quality of service attributes

An attribute approach to defining quality

THE TRADITIONAL ECONOMIC APPROACH to measuring consumer values for goods and services is to use market information to estimate a demand function, which is a schedule of the quantity of product demanded by consumers at different prices. This approach is appropriate for commodities such as milk which are homogenous and can be consumed in continuous amounts.

However, the same approach is inappropriate for valuing quality characteristics of goods and services, as quality is often discrete or ‘lumpy’ in nature and multi-dimensional. For valuing quality, a separate theory of consumer demand has been developed known as multi-attribute theory (Lancaster 1966). This theory assumes that people derive satisfaction or ‘utility’ from consuming a good and that their level of utility is a function of multiple attributes that make up the good, such as price and a range of quality characteristics. In making a choice between two or more alternative goods with similar characteristics, individuals are assumed to evaluate each alternative on the basis of their ‘attribute profile’, and then choose the alternative that maximises utility. In the process of weighing up the different alternatives, individuals are assumed to implicitly trade-off the level of one attribute against another.

Thus, a first step to understanding customer preferences for quality is to define quality in terms of distinct attributes that are meaningful to customers and are measurable, either quantitatively or qualitatively. The stated preference techniques reviewed in this report all use an attribute-based approach to eliciting consumer values. Table 2.1 contains examples of quality attributes that are applicable to electricity distribution.
2.1 Examples of quality of service attributes associated with electricity supply.

<table>
<thead>
<tr>
<th>Attributes</th>
<th>Unit of measurement</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Outcome attributes</strong></td>
<td></td>
</tr>
<tr>
<td>Frequency of outages</td>
<td>Number of outages per year</td>
</tr>
<tr>
<td>Duration of outages</td>
<td>Total minutes per year</td>
</tr>
<tr>
<td>Voltage quality</td>
<td>Number of surges or dips in electricity per year</td>
</tr>
<tr>
<td>Visual amenity</td>
<td>Underground or above ground cables</td>
</tr>
<tr>
<td>Safety</td>
<td>Bare wire or other form of cabling</td>
</tr>
<tr>
<td>Street lighting</td>
<td>Dim or brightly lit streets.</td>
</tr>
<tr>
<td><strong>Process attributes</strong></td>
<td></td>
</tr>
<tr>
<td>Connections</td>
<td>Duration between customer request and connection made (hours)</td>
</tr>
<tr>
<td>Fault service</td>
<td>Duration between fault reported and repair of fault (hours)</td>
</tr>
<tr>
<td>Telephone assistance</td>
<td>Time spent in telephone queue (minutes)</td>
</tr>
</tbody>
</table>

To what extent are the set of electricity supply attributes common to the other utilities? Reliability as a valued attribute is also a feature of gas, water, and public transport. Health, safety and environmental cleanliness likewise. When aggregated upwards to broader characteristics like reliability (measured by the first three attributes in table 2.1) health and safety, they are likely to be relevant across all utilities. Their relative contribution to consumer value will however differ.

**Types of values**

The quality attributes associated with gas, water, transport and electricity utilities are characterised as having both private ‘use values’ and public good values. Private use values refer to those elements of the service that are consumed directly by the individual, such as reliability, personal safety, prompt and polite service, and provision of assistance over the telephone. For electricity distribution, private values may be held for the visual amenity associated with the supply of underground power as opposed to overhead lines.

Consumers may also hold ‘non-use’ values for public good aspects of service provision. For example, customers may be willing to pay for improvements in the overall electricity distribution network, over and above the service delivered to their own household. Other public goods include environmental quality, such as the reduction in greenhouse gases, and public safety. These values stem from altruistic motives and the moral...
satisfaction that some people gain from the knowledge that society or the environment is made better off by improvements in service delivery.

When designing a stated preference survey for estimating customer values for quality, it is necessary to define the scope of the ‘commodity frame’ or the range of attributes to be valued. Should both types of values be included in the analysis, or just private use values? In part, this will depend on the set of attributes that can feasibly be monitored and regulated with financial incentives and on the ultimate purpose of the survey. However, it also depends on the range of concerns that are identified by customers, as omissions of key attributes could bias the value estimates for the included attributes.

Recent willingness to pay studies for NSW utilities have concentrated largely on private benefits. This reflects the preoccupation with producing evidence on willingness to pay grounds for investment in quality improvement.

Defining the quality attributes

The attributes selected for a valuation study should satisfy a number of criteria. Firstly, the attributes should be defined in a way that is meaningful to customers. To satisfy this criterion it is essential that focus group discussions be held with customers to find out what issues are relevant to people and how they perceive different aspects of quality. For instance, defining the reliability of electricity supply in terms of ‘minutes of supply per year’ may hold little relevance to household consumers. Households may be more concerned about the number of times it was necessary to reset clocks and other appliances over the past six months.

The second criterion is to ensure that the attributes can be measured objectively and are able to be controlled, at least within a range of confidence, by the service provider. Furthermore, for the attributes to be policy-relevant, the quality dimensions should be defined in a way that allows benchmarks to be established and ongoing monitoring to take place. To satisfy this criterion, the survey design phase should include discussions with the regulator and service providers.

A third criterion is to select attributes that are independent of one another, at least from the respondent’s perspective. This is particularly important if the objective of the valuation study is to estimate the relative ranking of attributes in terms of their importance to consumers, or the contribution of each attribute to customers’ overall utility. An example of interdependence
is the situation where respondents perceive the undergrounding of power cables to be primarily responsible for reducing power 'blackouts'. In this circumstance, the causal relationship between undergrounding cables and outage frequency could induce respondents to place a high value on undergrounding and a low value on improvements to power reliability, even if reliability is the main 'end product' desired by the customer.

**Attribute measurement**

Quality attributes must be measurable in a qualitative or quantitative sense. Attribute levels are required for establishing quality benchmarks against which utilities and regulators can measure performance. Furthermore, attribute levels are required in stated preference surveys for communicating to respondents the level of service provision under alternative options.

The metric for measurement may take a variety of forms, depending on the attribute in question. Several possibilities are:

- a quantitative, continuous scale such as the average waiting time in a telephone queue for customer service;
- a qualitative, discrete and ordinal scale such as a five-point performance rating for customer service (e.g. poor to excellent); and
- a binary 'yes-no' variable, indicating the presence or absence of a service element such as underground power.

All three types of measures can be used to represent attribute levels in stated preference surveys but the appropriate statistical techniques used to analyse the data varies depending on the type of measure used.

**Perceived versus actual quality**

For a variety of reasons, customer perceptions of quality levels may differ from the actual level of service quality provided, as measured using an objective scale. For example, respondents' view of quality performance is often 'skewed' by recent events. Rather than taking a long term average of service provision, people tend to place a greater emphasis on recent events. The divergence between perceptions and actual levels presents two problems. Firstly, value estimates are likely to be biased upward if people are surveyed shortly after a major event in which services have been disrupted, such as a storm event or the contamination of household water. To avoid this problem, surveys should not be conducted during a time of heightened community awareness about incidents that are 'once off' or
infrequent. Unfortunately, some mismatches between perception and objective measurement (e.g., perceived and measured reliability) may be ongoing and this too can contribute to difficulties such as protest responses in surveys.

A second complication arises if a systematic relationship cannot be established between the actual levels of service delivered by utilities and customer perceptions of quality. If there is a large mismatch between objective measures and respondent perceptions, and this divergence cannot be explained in a systematic manner, then it is difficult for the service provider and regulator to devise incentive policies for encouraging a more efficient level of service quality.

Thus, an important first step in understanding consumer preferences is to establish the extent to which perceptions of quality reflect actual performance delivery and, if necessary, to develop a means of calibrating perceptions to objective measures. In submissions to the Tribunal there appears to be confusion about the task of:

- assessing customer perceptions about the existing levels of quality performance; and,
- estimating customer willingness to pay for quality improvements.

These two steps in the evaluation process should be carried out separately. The task of assessing consumer perceptions is necessary for establishing base levels for quality attributes and for calibrating perceptions to actual quality levels. The second task involves the application of specialised valuation techniques.

Quality risk

Some quality attributes that are associated with gas, water, electricity and public transport utilities are not totally within the control of the service provider. In other words, there is a level of uncertainty or risk associated with the supply of some quality attributes. Examples for the electricity industry include:

- power outages, which are frequently caused by street trees falling on overhead lines. The responsibility of maintaining trees may reside with the local council or householder rather than the electricity distributor; and
- voltage quality can be improved within a range of confidence, but it is not technically feasible to eliminate harmonics and spikes altogether.
Examples in public transport include traffic congestion effects on reliability and the quality of bus stop shelters which are maintained by local councils.

Quality risk complicates the valuation task because respondents find it difficult to understand the concept of probability. Instead, it is common practice to present respondents with average levels of quality over a specified period of time and disregard variance. If focus group discussions reveal that variability is an important concern, then it may be necessary to include a measure of quality variance as a separate attribute in scenarios to be valued. In the case of electricity and gas supply, it is possible that businesses and industrial users are more affected by supply variability than household users.

Attribute levels

The application of stated preference techniques requires information about the current levels of each quality attribute and upper bound levels, representing potential quality improvements that are considered to be technically feasible. It is important to ensure that the range of attribute levels selected for a stated preference survey cover the range of possible outcomes that are relevant to the policy question.

Some of the advanced, ‘state of the art’ valuation techniques require respondents to state their preferences for alternatives that are defined in terms of an ‘attribute profile’ (see chapter 3 for details). The profile of an alternative is specified by attribute levels, which are assigned using an experimental design. The process essentially randomises the levels presented to respondents. Because levels are randomised, it is necessary to ensure that combinations of attribute levels are ‘sensible’ and do not appear ‘strange’ to respondents. Unrealistic attribute profiles should be removed from the set of alternatives presented to respondents.

As a general rule the attribute levels should be expressed in absolute terms rather than percentage changes from a base because some respondents have difficulty comprehending percentages. However, there are exceptions. For some attributes that are well understood by respondents, such as price, it may be satisfactory to express changes in percentage terms. The use of percentages should be examined on a case-by-case basis and be tested using focus groups and pilot tests.
3 Review of stated preference methodologies

The family of stated preference techniques

A range of survey techniques have been developed for eliciting consumer preferences and economic values for non-market goods and services, or values for new products that firms may be considering introducing to the market (chart 3.1). All the techniques involve asking respondents to consider one or more hypothetical options and to indicate their preference for these options. However, aside from this general commonality there are significant analytical differences between the techniques.

Contingent Valuation and Discrete Choice Modelling originate from the economics discipline, and have been used widely for valuing a diverse range of non-market goods. Applications include public health (Viscusi et al. 1991), environmental damage assessment (Hanley et al. 1998), outdoor recreation (Adamowicz et al. 1994), and transport modelling (Hensher 1994). These methods are underpinned by utility theory, which allows measures of consumer welfare to be estimated for changes in the supply of non-market goods or services. Because the techniques are based on a theory of choice behaviour, stated preference data can be analysed in the same way as actual choice data observed in the market, such as commuters’ choice of transport mode. Where this is possible it can be a powerful way of checking the validity of stated preferences.

Another survey technique is Conjoint Analysis. It has its origins in the marketing literature and it is mainly focused on gaining an insight to consumer preferences rather than to estimate economic values (Louviere 1988). It requires respondents to make decisions on the relative merits of a series of options by ranking or rating alternatives. The principal difference between Conjoint Analysis and other stated preference techniques is that respondents are not required to make a commitment to selecting a particular option. As such, Conjoint Analysis is one step removed from economic theory and from decisions that are made in a market environment. Consequently, there is debate in the literature about the
extent to which Conjoint Analysis can be used to explain choice behaviour or estimate valid measures of welfare (Roe et al. 1996).

In the sections that follow, each technique is outlined in more detail. The different methodologies are reviewed in terms of their strengths and weaknesses at providing estimates of consumer preferences and values. The review is conducted with a focus on the applicability of these techniques to valuing quality aspects of electricity, gas, water, and public transport utilities.

### 3.1 The family of stated preference techniques.

<table>
<thead>
<tr>
<th>Stated preference methods</th>
</tr>
</thead>
<tbody>
<tr>
<td>Open-ended Contingent Valuation</td>
</tr>
<tr>
<td>Conjoint rating, ranking and paired comparison</td>
</tr>
<tr>
<td>Discrete choice methods</td>
</tr>
<tr>
<td>Referendum Contingent Valuation</td>
</tr>
<tr>
<td>Choice Modelling</td>
</tr>
</tbody>
</table>

### Contingent Valuation

#### Open-ended version

The Contingent Valuation Method has been used widely for estimating non-market values since its first application in the early 1960's by Davis (1963). The Environment Protection Authority’s ENVALUE database [http://www2.epa.nsw.gov.au/envalue/](http://www2.epa.nsw.gov.au/envalue/) provides summary details on valuation studies using a wide range of methods including Contingent Valuation. The technique derives its name from the fact that the value estimates are contingent on a hypothetical scenario that is presented to respondents for valuing. The original form of Contingent Valuation (CV)
constitutes an open ended question, in which respondents are asked to state their willingness to pay (or accept compensation) for a specified environmental improvement (or decrement) or a specified quality change (fluoride addition to water supplies, undergrounding of power lines etc).

Statistical analysis of the data generated by this technique is relatively straightforward. The simplest approach is to calculate mean willingness to pay (WTP) or willingness to accept compensation (WTA) for the sample of respondents. Alternatively, an attempt can be made to ‘explain’ the variation in values across individuals by regressing the ‘bids’ against a range of independent explanatory variables such as socio-economic characteristics and respondent attitudes. Ordinary Least Squares regression is usually employed for this task.

A study by Dwyer Leslie (1991) is an example of the open-ended CV approach. The open-ended CV questionnaire was supplemented by a referendum CV questionnaire, as discussed in this report. In the Dwyer–Leslie study, the technique was used to estimate consumers’ WTP for further treatment of Sydney’s drinking water which was to be facilitated by the construction of the Prospect water treatment facility and two smaller plants. The question was couched in terms of marginal willingness to pay (over and above current bills) ‘for the guarantee of maintaining high quality drinking water into the future’ and respondents were asked to select from a range of values. The survey gave results with an average willingness to pay of an additional $66 per annum at the household level and more than $1300 for industry. These contrast with $49 ($804 for industry) as the average result when the question was put to respondents on a ‘take it or leave it’ basis.

The open-ended CV method is now rarely used because it has been found to be vulnerable to a range of biases. It faces special problems where the good in question is not purchased directly by the public. The principal problem in the context of public or environmental goods which are not privately purchased is that respondents find open-ended questions too difficult to answer because they are not accustomed to paying for non-market goods and services (Haneman 1994). Respondents may have a preference for one alternative over another but do not know their maximum willingness to pay for a good. Early applications of the technique sought to overcome this problem by asking respondents to select their bid from a list of suggested payment options. While this approach reduces the cognitive burden on respondents, it tends to induce starting point bias, meaning that respondents anchor their responses to the midpoint of the set of values listed on the payment card.
Another bias associated with CV is payment vehicle bias. This describes the situation whereby a respondent has a particular dislike, for the payment mode, such as taxes, electricity bills or entry fees. Sandry (1996) concluded that a considerable degree of vehicle bias is introduced if unsuitable payment instruments are used.

Referendum version

Owing to the problems of eliciting values using an open-ended question, most CV studies are now undertaken using the referendum CV method. This technique involves asking respondents to make a discrete choice between two alternatives: Pay nothing (extra) and maintain the status quo level of quality, or pay a specified 'bid amount' in return for an improved level of quality. For example, a typical question might be:

‘Would you be willing to pay $A to secure a quality improvement of X units? (yes or no)’

The preference data generated using this method is encoded in binary form, as respondents are only given the option of answering ‘yes’ or ‘no’. To translate this data into respondent values for a specified change in quality, a theoretical construct known as ‘random utility theory’ is assumed to underpin respondent choices (McFadden 1974). Under this construct, individuals are assumed to accept or reject the offer amount on the basis of maximising their utility (or satisfaction) subject to a budget constraint. Thus, the probability of a respondent answering ‘yes’ is given by the probability of $U_1$ being greater than $U_0$, where $U_1$ is the utility associated with the ‘change’ option and $U_0$ is the utility associated with the status quo option.

The two principal factors influencing an individual’s utility are the quality level ‘X’ and the bid amount ‘A’. In a CV application the levels of these factors are varied systematically across respondents. The resultant choice data are used to estimate a binary logit model of respondent choice, where the model parameters serve to define the utility function

$\begin{align*}
\text{Pr} \{ \text{yes} \} &= \frac{e^{U_1}}{e^{U_1} + e^{U_0}} \\
\text{Pr} \{ \text{no} \} &= 1 - \frac{e^{U_1}}{e^{U_1} + e^{U_0}} 
\end{align*}$

Having estimated the model, it is a straightforward task to calculate respondent values for a specific quality improvement. (A straightforward explanation of this application is given in Louviere et al. 2000, p. 62.)

---

2 Data from the referendum CV technique constitutes the choice indicator, which is a binary variable (0 or 1), and values for the bid amount and the level of quality improvement. Parameters of the binary logit model are estimated using a Maximum Likelihood statistical procedure, which is used to find parameters that maximise the model’s prediction of the observed sample of choices.
As mentioned, the Dwyer Leslie study on Sydney water quality used the referendum approach (as a supplement to open-ended CV) for obtaining estimates of consumer WTP for water quality improvements. The survey administration included backgrounding of respondents on the implications of allowing the status quo to continue. People were asked about their knowledge of their current bill. The survey also tested attitudes to the supplier (then the Sydney Water Board) and so was in a position to test for a protest response when people were confronted with a single nominated percentage increase in the quarterly bill as the price to pay for guaranteed quality. Different subsamples of consumers were confronted with different percentage increases and were asked whether they would pay the nominated increase. This application of the referendum CV resulted in substantially (and statistically significantly) lower estimates of mean willingness to pay than the open ended approach of allowing respondents to choose from a ‘menu’ of values. Further examples of the referendum CV method are cited in a report commissioned by the Office of the South Australian Independent Industry Regulator (SAIIR, June 2000). For instance, the report outlines the results of a Canadian study (Tollefson, 1994) where customers from Canadian electricity utilities were asked to declare their willingness to pay (extra) for ‘assured systems’ as an improvement on a given number of interruptions per month.

While the referendum CV technique is considered to be ‘state of the art’ for this type of application, it does have some limitations for estimating values. Firstly, only one quality attribute or scenario can be presented to a sample of respondents for valuation. This may not be a limitation if the objective of the study is to estimate values for a one-dimensional change in quality, for instance an improvement in the reliability of electricity supply. However, it is an inefficient method of value estimation if multiple attributes are involved and we are interested in the values attached to each and trade-offs between them, rather than ‘quality improvement’ in aggregate. Furthermore, where the technique focuses the respondent’s attention on a single attribute, the values estimated for that attribute may be over-estimates if there are other important attributes that are not presented to the respondent for valuation (see box 3.2 for a discussion on embedding effects). The challenge for the analyst is to select a valuation technique and ‘framing’ strategy that is appropriate for the policy under investigation.

A second criticism frequently levelled at CV is its hypothetical nature. Critics of the technique hold the view that it is a poor method for estimating consumer values because respondents are unlikely to provide an accurate response when presented with a hypothetical scenario. Hypothetical bias is potentially a problem for valuing attribute changes that are unfamiliar to respondents, or for changes that do not have a
natural market mechanism for bringing about the change. It is less likely to be a problem for valuing services provided by the electricity, gas and water utilities because consumers are already accustomed to paying for a base level of service.

An interesting recent development in trying to deal with problems created by unfamiliarity in discrete choice (‘yes/no’) CV valuation studies, is to take explicit account of respondents’ uncertainty in making the responses they do make. Not unlike the preference ratings discussed below, this approach follows up a referendum type question by asking respondents how certain (on a confidence rating scale from say 0 to 100) they were in accepting or rejecting the bid. The resulting information is then used to adjust estimated WTP from the sample. One application of this suggests that failure to take account of this ‘preference uncertainty’ can lead to serious upward bias (by as much as a factor of six times) in apparent willingness to pay (see Li and Mattsson 1995).

A third potential weakness of CV is that it may induce some respondents to behave strategically, particularly when public goods are involved, such as the environment or streetscape aesthetics. For these goods, there is scope for individuals to ‘free ride’ and reduce their bid in the belief that someone else will pay. Alternatively, respondents who are pro-environment could over-bid or ‘yea-say’ in the hope of influencing policy outcomes if they believe that they will not have to pay the stated amount. Overall, the empirical evidence suggests only weak forms of strategic bias occur (Morrison et al. 1996). The bias is reduced if careful attention is paid to making the valuation scenario realistic.

### 3.2 Embedding effects

An embedding effect is said to occur when the estimated mean willingness to pay for an attribute is lower when it is valued as part of a more inclusive set of attributes, rather than on its own (Bennett et al. 1998). At the extreme, the summation of values for individual attributes, when evaluated separately is sometimes observed to exceed an individual’s total income. The embedding effect is not a bias. Rather, the value of all goods (market and non-market) are dependent on the context in which they are framed. Thus, the wider the array of substitute goods available to a consumer, the lower the value placed on any individual good.

If, given the limitations of Contingent Valuation, utilities nevertheless consider it useful to include in their appraisal of customer WTP, it is recommended that attention be paid to such issues as the choice of the number of value bid points to include in the survey and the merits of giving respondents a follow up bid when they accept or reject the first one. These issues of optimal design are dealt with in detail in Albernini (1995).
Choice Modelling

Choice Modelling is another form of discrete choice analysis and, consistent with the referendum CV, random utility theory is assumed to underpin respondent choices. However, compared with referendum CV, the Choice Modelling (CM) technique produces a much richer set of preference data because it involves asking respondents to choose between more than two alternatives, with each alternative being described by multiple attributes. In a CM application, respondents are presented with a series of questions (or choice sets). Each question requires respondents to select one option from three or more alternatives. Chart 3.3 contains an illustrative example of a choice set that could be used for valuing quality attributes associated with electricity distribution.

3.3 Illustrative example of a choice set used in a Choice Modelling questionnaire

<table>
<thead>
<tr>
<th>Frequency of outages</th>
<th>Total duration off supply</th>
<th>Telephone assistance</th>
<th>New connections service</th>
<th>Underground or above ground power</th>
<th>Increase in fixed charges on quarterly power bill</th>
</tr>
</thead>
<tbody>
<tr>
<td>10 per year</td>
<td>2 per year</td>
<td>15 per year</td>
<td>20 min per year</td>
<td>10 min per year</td>
<td>5 min per year</td>
</tr>
<tr>
<td>20 min per year</td>
<td>10 min per year</td>
<td>10 min per year</td>
<td>&lt; 30 sec in cue</td>
<td>&lt; 5 min in cue</td>
<td>no wait</td>
</tr>
<tr>
<td>&lt; 30 sec in cue</td>
<td>Within 24 hours</td>
<td>Within 1 hour</td>
<td>Aboveground</td>
<td>Underground</td>
<td>Within 3 hours</td>
</tr>
<tr>
<td>Within 24 hours</td>
<td>Aboveground</td>
<td>Underground</td>
<td>No increase</td>
<td>Plus $75</td>
<td>Aboveground</td>
</tr>
<tr>
<td>No increase</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

I would choose..... (please tick one box)

Typically, five to eight choice sets are included in a questionnaire. That is respondents are asked to make between five and eight separate choices. Usually three to five alternatives are specified in each choice set. The alternatives are differentiated from one another by varying systematically the attribute levels, which are assigned using an experimental design (box 3.4). The experimental design is used to ensure that the range of options presented to respondents in the choice sets is adequate. As a general rule, the maximum number of attributes that can be included in a
CM experiment is limited to six or seven. Respondents have difficulty processing more than this number.

3.4 Experimental design

To determine how variations in each attribute alter the choices made by respondents a large number of systematic changes in the level of different attributes must be examined. Specifically the set of changes should include all possible combinations of attribute levels - known as a ‘full factorial’. In many cases this is not possible due to the large number of possible combinations. In such cases a ‘fractional factorial’ is used. The structured way in which the attribute levels are transformed into choice sets is called the ‘experimental design’ (Bennett 1999).

Each choice set includes a ‘status quo’ option that describes the current level of service provision. It serves as a base against which respondents’ willingness to make trade-offs in securing change can be measured. The other options are deviations from the status quo. Data from a CM application are used to estimate a probabilistic model of product choice. This model allows the analyst to estimate the extent to which individuals are prepared to trade-off one attribute against another. Provided one of the attributes is measured in dollar terms (eg. the price of electricity), it is possible to estimate the amount of money people are prepared to pay for improving a non-monetary attribute by one unit. This value is known as a ‘part worth’, and it indicates the implicit price of an attribute.

In addition to estimating attribute implicit prices, the CM technique enables a measure of consumer welfare (compensating surplus) to be estimated for various packages of service improvements. Valuation is not restricted to the set of alternatives presented in the questionnaire. Rather, the values associated with a whole range of improvements can be calculated once a model of respondent choice has been estimated. The CM application need only employ a range of attribute levels sufficient to cover the range of quality improvements under investigation.

CM has a number of potential strengths over other stated preference techniques for eliciting consumer willingness to pay for service quality (box 3 The CM technique generates panel data, which constitutes a choice indicator variable for each alternative (0 or 1), together with attribute levels corresponding to each alternative. The data are used to estimate a multinomial logit model, which is a probabilistic model of product choice (other functional forms include the nested logit, HEV, GEV models — see Adamowicz et al. 1998). The estimated parameters of the choice model define the utility functions for each alternative.)
3.5. It enables the researcher to define better the range of attributes under investigation and to communicate the frame of reference to respondents in a meaningful way. The technique could also reduce the incentive for respondents to behave strategically or ‘yea-say’, as the questionnaire forces respondents to consider multiple trade-offs between attributes.

A drawback of the technique is that, relative to CV, it places a greater cognitive demand on respondents. This has at least two implications. Firstly, the questionnaire cannot be administered by telephone. Instead, questionnaires must be mailed out or administered face-to-face (see chapter 4) substantially increasing the cost of the survey. Secondly, the added complexity means that CM surveys tend to yield a lower response rate than CV surveys. The combination of added complexity and cost constraint can mean that some compromise on sample size may be required.

### 3.5 Strengths of Choice Modelling

- Forces respondents to consider trade-offs between attributes;
- Makes the frame of reference explicit to respondents via the inclusion of an array of attributes and product alternatives;
- Enables implicit prices to be estimated for attributes;
- Enables welfare impacts to be estimated for multiple scenarios;
- Can be used to estimate the level of customer demand for alternative ‘service products’ in non-monetary terms; and
- Potentially reduces the incentive for respondents to behave strategically.

A number of Australian utilities have reported the use of what they describe as CM investigations of WTP for quality change. Unfortunately, full details of the modelling approach are rarely disclosed by the consultants undertaking the customer surveys so critical analysis is constrained. For example, EnergyAustralia’s discussion paper (1999) reports that those respondents who already were supplied by underground infrastructure had a stronger preference for it, however there is no discussion on the methodology used to determine this.

### Conjoint Analysis

#### Conjoint Rating

In a conjoint rating questionnaire, respondents are asked to evaluate a series of alternatives, one at a time, using a numerical ratings scale. As with
CM, each ‘product’ is defined in terms of a set of attributes whose levels are varied across questions according to an experimental design. Unlike CM, respondents are not required to make comparisons between alternatives or to select a preferred product. Instead, the technique forces respondents to examine each alternative separately and give it a preference rating (chart 3.6).

3.6 Illustrative example of a Conjoint Rating question

<table>
<thead>
<tr>
<th>Frequency of outages</th>
<th>10 per year</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total duration off supply</td>
<td>20 min per year</td>
</tr>
<tr>
<td>Telephone assistance</td>
<td>&lt; 30 sec in cue</td>
</tr>
<tr>
<td>New connections service</td>
<td>Within 24 hours</td>
</tr>
<tr>
<td>Underground or above ground power</td>
<td>Underground</td>
</tr>
<tr>
<td>Increase in fixed charges on quarterly power bill</td>
<td>plus $25</td>
</tr>
</tbody>
</table>

Please circle one of the numbers below to show your strength of preference for the following alternatives

Weakly preferred 1 2 3 4 5 Strongly preferred

The attributes and levels in this chart are for illustrative purposes only.

The data are usually analysed by regressing the rating scores against the attributes and using Ordinary Least Squares (OLS), the most commonly used technique, to estimate regression parameters. The resultant parameters are used to calculate implicit prices for each quality attribute. While this practice is widespread in the literature, the resultant value estimates are generally biased because OLS is an inappropriate estimator. Ratings data are ordinal (only the ordering matters: the difference in ratings does not measure the strength of preference for one alternative over another) and discrete (as opposed to continuous variables) and these characteristics of the data violate the assumptions underlying OLS. Furthermore, because respondents are not required to choose a particular alternative, but to simply rate each one on a preference scale, the model cannot be used to predict choice behaviour or level of demand for a particular alternative (Adamowicz et al. 1998).
Conjoint Ranking

Conjoint ranking is not dissimilar to the rating method described above. The primary difference is that respondents are presented with three or more alternatives in one question and asked to rank the alternatives from most to least preferred (chart 3.7). A series of these ranking exercises is administered to the respondent. Conjoint ranking is no longer widely used because of theoretical difficulties in analysing the data (Louviere and Timmermans 1990). Furthermore, the technique shares the weakness of conjoint rating in that respondents are not required to commit to selecting one alternative.

3.7 Illustrative example of a Conjoint Ranking question.

<table>
<thead>
<tr>
<th>Frequency of outages</th>
<th>Current service</th>
<th>Product A</th>
<th>Product B</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total duration off supply</td>
<td>10 per year</td>
<td>2 per year</td>
<td>15 per year</td>
</tr>
<tr>
<td>Telephone assistance</td>
<td>20 min per year</td>
<td>10 min per year</td>
<td>5 min per year</td>
</tr>
<tr>
<td>New connections service</td>
<td>&lt; 30 sec in cue</td>
<td>&lt; 5 min in cue</td>
<td>no wait</td>
</tr>
<tr>
<td>Underground or above</td>
<td>Within 24 hours</td>
<td>Within 1 hour</td>
<td>Within 3 hours</td>
</tr>
<tr>
<td>ground power</td>
<td>Aboveground</td>
<td>Underground</td>
<td>Aboveground</td>
</tr>
<tr>
<td>Increase in fixed charges</td>
<td>No increase</td>
<td>Plus $75</td>
<td>Plus $15</td>
</tr>
<tr>
<td>on quarterly power bill</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Please rank the three alternatives from most to least preferred by placing the numbers 1, 2, and 3 in the boxes

Paired Comparison

In a paired comparison, respondents are presented with two alternatives at a time and asked to rate their preference for the alternatives on a five or ten-point scale. A series of these questions is administered to each respondent. In most applications it is standard practice to vary both the levels and the types of attributes over the series of questions, such that respondents are only required to consider two or three attributes at a time.

A study by Johnston and Desvouges (1997) used a paired comparison technique to estimate public preferences and values for various electricity
generation scenarios (chart 3.8). The impacts of alternative scenarios were described in terms of health, environmental, and employment attributes. The application required respondents to make tradeoffs between these impacts and the cost of different forms of power generation, which was specified in terms of changes in the price that respondents would have to pay for electricity.

This is the most frequently used version of conjoint analysis, and commercial software packages are available for producing an experimental design and analysing the data. For example, Adaptive Conjoint Analysis (ACA) is a computer algorithm that updates the experimental design (the combination of attribute levels) as respondents work their way through the questionnaire, which is displayed on a computer video monitor.

As with the other conjoint techniques, the paired comparison method is not underpinned by a full economic model of consumer choice and the statistical analysis of rating data using OLS is inappropriate. However, some recent advances have been made in overcoming these difficulties. In the studies by Roe et al. (1996) and Johnston and Desvousges (1997) a new method is devised for analysing ratings data that is superior to earlier applications.

3.8 Example of a Conjoint Paired Comparison question, adapted from an application by Johnston and Desvousges (1997)

According to SAIR (2000) this technique of adaptive conjoint analysis has been applied on behalf of Victorian utility Powercorp by the KBA Consulting Group Ltd. (1999).
4

Best practice guidelines

Assessing validity and reliability

A concern that is frequently voiced as a criticism of stated preference methods is that respondents to the various types of questionnaires answer in ways that diverge from their ‘true’ preferences. Two central questions are:

- what evidence is there to indicate that stated preference techniques are a valid means of estimating consumer values; and
- what procedures should be followed to obtain valid and reliable estimates from stated preference surveys?

With respect to the first question, there is an increasing body of evidence to indicate that a well executed stated preference survey can accurately reflect consumer preferences and produce value estimates that are statistically equivalent to market-based estimates. The most convincing evidence has come from tests of ‘convergent validity’ where the results from choice experiments (within a CM framework) have been compared with actual consumer choices of market products (eg. Blamey et al. 1999) or with observed visits to recreation sites (eg. Adamowicz et al. 1994).

In the study by Blamey et al. (1999), the CM technique was used to model the demand for different brands of toilet paper. Two data sets were involved in the test. The first was supermarket sales records of various brands of toilet paper. The second was a stated preference data set generated from a CM questionnaire that asked shoppers to nominate their preferences for different brands of paper, described by a specified set of attributes. It was found that the choice model predicted accurately the market share of each brand of paper, implying that the CM technique is capable of validly eliciting consumer preferences for a market good.

Encouraging results were also obtained by Adamowicz et al. (1994). In this study the CM technique was used to elicit moose hunters’ preferences for different hunting sites, and to estimate values for various quality attributes associated with a hunting trip. The questionnaire was designed so as to
allow the stated preference data to be compared directly to actual visitation data. The study found a high degree of correspondence between the value estimates obtained from the choice model based on stated preferences and the model based on actual visits. The results support the use of CM as a technique for estimating non-market values - at least those associated with outdoor recreation.

Given that there is a body of evidence to suggest that stated preference surveys are capable of eliciting consumer preferences and values under some circumstances, what can be gleaned from the literature about best practice guidelines for applying stated preference techniques?

In 1992 an expert panel was assembled in the US by the National Oceanic and Atmospheric Administration (NOAA) to develop a set of guidelines for applying the CV Method (Arrow et al. 1992). This study was commissioned in response to a growing demand for non-market valuation assessments by prominent court cases, such as the Exxon Valdez oil spill litigation case that sought to establish the cost of environmental damage caused by the spill.

Briefly, the NOAA panel concluded that referendum CV studies could provide valid and reliable results and gave several specific and stringent recommendations on how stated preference surveys should be designed to ensure reliability and validity. The principal NOAA recommendations are:

- Respondents need to understand and believe the context in which ‘goods’ are being presented for valuation.
- The payment vehicle (or money attribute) must be meaningful to respondents.
- Respondents should be reminded of budget constraints and of available substitute ‘goods’.
- The questionnaire should always be pre-tested.
- Follow-up questions and checks on respondents’ understanding and acceptance of the valuation task should be included in the questionnaire.
- The referendum CV should be used rather than an open-ended CV technique, and the questionnaire should be administered by an in person interview.
- A split-sample test should be conducted to examine whether respondent values are sensitive to the scope (or amount) of good under investigation.
Most of these recommendations are still relevant today, although significant advances have been made with refining stated preference techniques since the guidelines were written (US EPA 2000). Perhaps the most important development has been the emergence and widespread application of CM to a variety of issues (e.g., public health, transport, environment and recreation). The remainder of this chapter provides an overview of the main technical and logistical factors that need to be considered when undertaking a stated preference survey.

**Technique selection**

Different techniques will be suitable for different situations. In most cases the lack of market information, or revealed preferences, will necessitate a stated preference technique. However, in some cases market data will be available. For example in public transport alternative modes of transport are often available, providing market information. The existence of market data gives rise to the potential for revealed preferences to be combined with stated preference data from surveys. Adamowicz et al. (1994) present an application of this approach for recreational site choice. Hensher (1994) also used both revealed and stated preference data to produce a matrix of direct and cross-share elasticities for commuting travel. If there is scope for combining revealed and stated preferences then this should be used.

In general, it is no longer considered best practice to use open-ended CV questions to estimate consumer values. This is certainly the case for goods and services that people are unaccustomed to paying for or are unfamiliar with. In one proposal put to the Tribunal it was proposed that willingness to pay estimates for improvements to electricity distribution be elicited using open-ended questions. This practice is not recommended because of the problems outlined in chapter three.

The referendum version of CV is a useful technique for valuing single-outcome changes such as a proposal to underground power or a proposal to introduce a new piece of technology, such as an advanced ticketing system for all modes of public transport. However, the technique is inappropriate if multiple quality attributes are involved, or if substitute ‘products’ are available to consumers. Under these circumstances, CM or conjoint analysis are better equipped to provide reliable estimates of consumer preferences. If value estimation is the key objective, CM is recommended as the best technique as it is based on sound economic theory.
Substitution is a particularly important factor to consider when valuing transport service quality because people have the choice of a variety of transport modes. Numerous studies have found that value estimates for specific goods are biased upward if substitute alternatives are not presented explicitly to respondents within the valuation framework (Boxall et al. 1996). The technique best suited to dealing with multiple substitutes is CM.

This review has not typically found satisfactory documentation of methodology — either used or proposed. For example, in one recent submission to the Tribunal it was proposed that a study be undertaken using a form of CM that used the design principles of CM but not the analysis techniques. This lack of attention to detail is not satisfactory because it is critical that the details of the proposed methodology, provided in proposals of this kind, are well documented if a regulator is to have confidence that the survey procedures are appropriate for estimating customer values.

**Technique application**

The application of stated preference techniques involves a series of steps. A report by Adamowicz et al. (1998) identifies seven steps that are involved in applying the CM technique. These are:

1. **Characterisation of the decision problem**

   This involves the identification of the problem at hand, for example what are customers willing to pay for an improvement in the quality of public transport?

2. **Attribute level selection**

   This step involves consulting with customers, the utility, and the regulator to define a number of quality attributes for valuation, and establishing a range of levels for each attribute. The levels should be appropriate for addressing the decision problem at hand.

3. **Experimental design development**

   Once attributes and levels have been determined, experimental design procedures are used to construct the alternatives that are to be presented to respondents.
4. Questionnaire design and delivery

The choice sets are only one component of a CM questionnaire. Other important design considerations include style of presentation, the type and quantity of background information for the respondent, and supplementary questions to collect information about respondent attitudes and demographics. (See below for further details on questionnaire design and delivery.)

5. Survey logistics

This step includes the establishment of a sampling frame (e.g., a utility’s customer accounts), the selection of a sample size, and the overall management of the survey process.

6. Statistical analysis and model estimation

The statistical methods used to analyse data from stated preference surveys depend on the type of valuation technique. For referendum CV, the binary logit model is appropriate. For CM, the multinomial logit or nested logit models are frequently used. The parameters of these models are estimated by maximum likelihood. In addition to model estimation, it is good practice to provide descriptive statistics.

7. Policy application

At this step, the estimated models are used to generate implicit prices for the attributes, welfare measures for different scenarios, and/or predictions of consumer behaviour induced by a change in quality and price (e.g., commuter choice of transport mode).

Questionnaire design

Focus groups and pilot testing

It is imperative that the questionnaire be designed in consultation with focus groups comprising a cross section of people who are representative of the population that will be sampled. Market research firms specialise in recruiting participants for these groups. The role of the focus group is to assist with:

- defining the attributes;
BEST PRACTICE GUIDELINES

- checking communication aspects of the questionnaire;
- checking that the scenarios are plausible and understood; and
- ensuring that the payment vehicle (i.e., the money attribute) is appropriate.

Once the questionnaire has been developed, it should be pilot tested using a randomly selected sample from the population of interest. At this stage in the process, the questionnaire should only require minor changes. The role of the pilot test is ‘pick up’ any problems that were not detected during the focus group discussions.

Attitudinal and demographic information

In addition to the stated preference questions for estimating values, the questionnaire should contain questions to collect information about respondent attitudes and demographics. The attitudinal information is useful for two reasons. Firstly, it can be used to check the validity of valuation results by cross tabulating respondent attitudes against the value estimates. For example, respondents who are identified as being risk averse should have an above-average willingness to pay for safety improvements. If not, there could be a problem with the survey instrument. Secondly, information on respondent attitudes can be incorporated as explanatory variables into the stated preference model.

Demographic and socio-economic information is required as an input into the modelling phase and is also useful for checking how well the sample represents the population of interest. At a minimum, data should be collected on age, income, sex, educational status, and occupation.

Follow-up questions

The questionnaire should contain follow-up questions after the stated preference exercise for picking up any response aberrations such as:

- payment vehicle protests;
- protests that constitute free riding behaviour; and
- lexicographic preferences (e.g., options that include an improvement in safety are always chosen irrespective of the cost).

The follow-up questions should identify whether the respondent had any specific problems with the questionnaire, including concerns about bias and plausibility and the respondent’s level of understanding about the task. In cases where it is clear that a respondent has a non-zero willingness
to pay but has given a protest response (that is, consistently chooses the status quo option), it is usual practice for these respondents to be removed from the estimating sample.

Protest behaviour is not expected to be a major problem in the application of stated preference techniques to value quality aspects of utilities such as reliability and customer service. This is because these service improvements are essentially market goods that have private use value. Furthermore, respondents are accustomed to paying for different service ‘packages’, such as the alternatives offered by the various mobile telephone companies. Protests and free-riding behaviour are more likely to be a problem for valuing services that are perceived to be ‘public goods’ such as the purity of household water or safety issues related to public transport. For these goods, there is greater potential for free riding because respondents could perceive that these standards should be maintained with funds collected through the tax system.

Minimum safety standards are normally mandated. Regulatory requirements that raise these standards will generate cost pass through issues that are distinct from the trade offs between price and other quality attributes being discussed in this paper.

Survey logistics

Questionnaire delivery

The multi-attribute techniques of CM and Conjoint Analysis are too complex to be administered by telephone. Instead, the questionnaire must be delivered by:

- mail-out/mail-back;
- personal drop-off with a later personal pick-up;
- personal interview; or
- centralised administration of the questionnaire, where respondents meet at a central location and complete the survey on computer terminals or using pen and paper.

Each delivery mode involves both benefits and costs (Bennett 1999). Personal interviews are relatively expensive but do generate higher response rates. Postal delivery is relatively low cost but can be prone to low response rates and consequential sampling bias. Using the post does however allow respondents the time to contemplate their answers more
completely and removes the prospect of ‘interviewer bias’. The drop-off and pick-up method is almost a hybrid of the mail out and personal interview approaches. It yields a higher response rate than mail-out but comes at a higher cost. The centralised administration of questionnaires can induce sample selection bias because it requires survey participants to make an effort to attend a meeting.

Sample size

In CM applications it is common practice to divide the respondent sample into ‘blocks’, with each block being administered a subset of the total number of choice sets. For example, five blocks would be developed for an application that involves 25 choice sets, such that respondents belonging to a given block are all administered the same set of five choice sets. For reliable statistical estimation, a minimum of ten respondents is required per choice set (Adamowicz et al. 1998). Thus, 250 respondents are required for an application involving 25 choice sets. Response rates in Australia for mail-out CM surveys typically range between 20 to 40 per cent. Assuming an average response rate of 30 per cent, a total sample size of 830 households is required to achieve the minimum number of 250 respondents.

Estimated differences in WTP usually refer to group averages of some sort — as particular market segments, constraints in one geographical area versus another etc. To address this disparity of consumer preferences the sample size must also be sufficient to allow for reliable statistical estimation for each market segment (rural households, urban household and the different socio-economic backgrounds). Each willingness to pay study should include a detailed breakdown of how the total sample was allocated to each market segment.

Other technical considerations

Willingness to pay versus willingness to accept

There are two alternative approaches to specifying the payment scenario or ‘money’ attribute in stated preference surveys. One alternative is to estimate the amount respondents are willing to pay (WTP) for an improvement in quality from $Q^1$ to $Q^0$. The alternative approach is to estimate how much compensation respondents are willing to accept (WTA) for a deterioration in service levels from $Q^1$ to $Q^0$ (or to remain at an inferior standard). Standard theory predicts that the two approaches should produce the same value estimate, as illustrated in chart 4.1. That is WTP
should be equivalent to WTA. However, there is a large body of empirical
evidence to show that WTA compensation frequently exceeds WTP for an
equivalent change in quality (Shogren et al. 1994).

Whilst this finding has troubled economists for some time, there is now an
emerging consensus in the literature that divergence between the two
measures can be explained by income and substitution effects (Shogren et
al. 1994). Properly defined, WTP measures recognise that customers’
incomes have a constraining influence. Income effects become potentially
important where the cost of an improvement is large relative to an
individual’s total income. In contrast, the WTA measure, which tries to
estimate the compensation required to achieve a given level of satisfaction
in return for foregoing beneficial changes, is free from direct income
constraints. Income effects should not play a large role in the valuation of
services provided by utilities because the changes involved are likely to
impose a relatively small impost on households when compared with total
household budgets.

With respect to substitution effects, it has been shown that the divergence
between WTP and WTA is greatest for goods that have few or zero
substitutes, such as personal health. That is, when there are few substitutes
it is possible that money does not perfectly compensate for losses.
Depending on the set of quality attributes in question, substitution effects
could be important in the context of valuing services provided by public
utilities. For example, the WTA compensation measure is likely to exceed
the WTP measure for changes in water quality attributes that affect human
health. In these circumstances, it is recommended that a WTP measure is
used as this is likely to produce a conservative, lower-bound value
estimate.

**Embedding effect**

The embedding effect, discussed in chapter three, exists if individual
attributes are evaluated separately and the sum of these values exceeds the
value given for the whole package of attributes. Bateman et al. (1997) found
evidence of this effect for private consumption goods bought and sold
using incentive compatible mechanisms, suggesting that this phenomenon
may not be attributable to the stated preference method, or application of it.
Instead it may be a symptom of some fundamental property of individuals’
preferences which conventional consumer theory does not allow for.
However, willingness to pay studies should report on the embedding effect
and provide a full explanation as to the context in which the quality
improvements were embedded or ‘framed’.
4.1 The equivalence of willingness to pay (WTP) and willingness to accept measures (WTA)

Design complexity issues

CM has to deal with potentially complex designs of choice sets when the number of attributes is relatively large. Block design is one method for overcoming this problem. Block design ensures that while there may be many attributes in the total choice design, only some attributes vary in each choice presented to respondents. Some of the attributes are the same between each choice, allowing respondents to focus on the ones that are different. By varying the selection of attributes that are held fixed between respondents it is possible to expose the target population to a wide range of choices without over complicating the choices put to any one individual.

Allowance for consumer heterogeneity

The design of policies to promote or maintain quality standards among service providers may require an understanding of how different sectors of the community value quality attributes, rather than just an estimate of the aggregate value of quality improvements. For instance, it may be necessary to identify how values vary across businesses and households, or rural and urban households, or among households with different socio-economic characteristics.

There are a variety of approaches available for identifying whether value estimates are statistically different across different strata in a population. In
a CM or conjoint application, the simplest approach is to define segments within the sample using distinguishing individual characteristics such as income, occupation, or whether the respondent is a business or household. The individual-specific variables are then interacted with various attributes of the choices to produce a model that is specific for a given segment of the sample. Alternatively, more sophisticated methods of accounting for respondent heterogeneity can be used (Adamowicz et al. 1998). These include:

- performing a cluster analysis on the sample of data using individual characteristics and then developing choice models for each cluster;
- estimating a random parameters logit model that accounts for heterogeneity by allowing model parameters to vary randomly over individuals; and
- developing a latent class model, which is used to jointly predict the probability of a respondent belonging to a population segment (based on socio-economic characteristics) and the probability of an individual choosing a particular alternative.

None of the recent submissions put to the Tribunal discuss these issues in any depth. However, to make judgements about the appropriateness of introducing ‘WTP for quality’ based price differentiation, regulators need some evidence that any purported differences in WTP between customer segments are not the result of chance. Regulators also need some guidance on the spread of valuations placed on incremental quality within any one customer group to which a single price applies.

**Diminishing marginal values for quality**

Willingness to pay studies must differentiate between total value and marginal value. Submissions to the Tribunal have failed to do this. The focus of WTP studies should be on marginal valuation — the value attached to successive increments of quality — as it is this information that is important for the Tribunal in its pricing review decisions.

It is reasonable to expect that most respondents will have a diminishing marginal value for quality improvements. That is, the first unit of quality improvement will be valued more highly than the second unit, and so on. However, it may not be possible to detect diminishing values if only small ranges of improvement scenarios are presented to respondents for valuation.
The analysis of data from a stated preference application should include an assessment of whether or not the total value function is linear. In a CM application, this can be undertaken by testing whether attributes enter the utility function in a linear or quadratic form. This is an important step in the analysis because incorrect assumptions about functional form could lead to large errors in the estimation of values.
Conclusions and summary

RELIABLE INFORMATION ON CONSUMERS’ willingness to pay for additional levels of quality provided by utilities has the potential to assist regulators moving towards the optimum level of service quality. This is found where the cost of supplying a higher level of service is equal to the additional benefit that the consumer receives from this improved service level. But regulators are still justifiably cautious in translating evidence on willingness to pay for enhanced quality into increases in regulated utility prices.

The information problem

To form a view on the efficient level of quality of service the regulators need two key pieces of information: the cost of supplying additional service quality and the additional benefits consumers receive from the increase in quality. Cost information is typically known by the utility. But due to the nature of the utility sector, characterised by limited competition and little if any observable differentiation in service quality, there is generally little market (as opposed to stated preference) information on the value people place on quality.

Uncovering reliable WTP information

Due to the lack of market data, or revealed preferences, for the level of service quality, it is necessary to use stated preference techniques to estimate the willingness to pay for quality improvements. Estimates of the willingness to pay for enhanced quality, derived from these techniques, are increasingly being used to back up urgings for quality directed expenditure with cost pass through to consumers plus incentive related margins being advocated by some utilities. The Tribunal has to have confidence in these WTP estimates, and therefore the methodologies used to derive them, before they can be considered in pricing policy.

The consequences of incorporating incorrect estimates in pricing review decisions have the potential to be significant. If the estimates overstate
consumers’ willingness to pay and result in a price increase for the utility service then consumer welfare will decrease. Consumers will be paying more for the service than the value they place upon it, and with limited recourse to alternatives in some cases.

Although there is evidence that stated preference techniques have the potential to accurately reflect consumer preferences this potential may not be realised if the methodology employed is flawed. Even the best available techniques can struggle to provide willingness to pay estimates within narrow confidence intervals.

**Quality of service attributes**

The first step in attempting to uncover consumers’ willingness to pay is to define the quality of service. This allows different levels of service to be identified. Most willingness to pay studies define quality as a set of attributes related to the service.

The attributes selected should satisfy a number of criteria:
- the attributes must be meaningful to customers;
- the attributes must be able to be measured objectively and be controlled within a range of confidence by the service provider; and
- the attributes must be independent of each other, at least from customers’ perspective.

It is important to establish the extent to which perceptions of quality reflect actual performance delivery. This is a separate task in the valuation process from actually estimating customer willingness to pay for quality improvements. The task of assessing consumer perceptions is necessary for establishing base levels of quality attributes and for calibrating perceptions to actual quality levels.

**Technique selection**

If market data is available we recommend that it be combined with data derived from stated preference techniques. It is unlikely that market data will be available for estimating consumers’ willingness to pay for improved service quality of electricity supply. However, data from the ‘greenpower scheme’ may be able to be used to estimate the environmental attribute of electricity supply.
Transport studies would seem to offer this alternative when choice of transport mode is relevant. However, recent State Transit Authority submissions on WTP do not use this combined approach, and the Authority suggests that it is not appropriate in the context of service quality WTP.

Our review of the methodologies used in the different stated preference techniques concluded that Choice Modelling is the most suitable method for estimating consumers’ willingness to pay for quality improvements in multiple dimensions. This is because it has less inherent biases than the other techniques enabling the implicit price to be estimated for attributes and enabling welfare impacts to be estimated for multiple scenarios.

Criteria for assessing the reliability of WTP studies

The large number of stated preference technique applications has allowed us to develop a set of criteria for assessing the reliability of willingness to pay studies. The criteria are in the form of the guidelines presented in the previous chapter. The key features are:

Survey design and logistics

- consultation with a suitable focus group, preceding survey design followed by a pilot survey to test design adequacy;
- questions to collect attitudinal and demographic information; and
- use of follow up questions on the questionnaire.

Survey logistics

- For other than simple CV techniques or preliminary attitudinal testing where telephone surveys are appropriate, questionnaires to be administered by one of the following methods;
  - mail out/mail back;
  - personal drop off and pick up;
  - personal interview; or
  - centrally administered.
- The sample size must allow reliable statistical estimation for all choice sets and customer segments.
Reporting

- detailed evidence that any purported differences in WTP between customer segments are not the result of chance;
- discussion on any evidence of the embedding effect and a detailed explanation of the frame of reference in which quality improvements were presented to respondents;
- distinction between marginal value and total value; and
- full reporting of the methodology and accompanying to derive the estimates.

The validity of submissions made to the Tribunal

On the available evidence, the submissions made to the Tribunal on the willingness to pay for additional service quality fail to meet the above criteria. Without exception, a full explanation of the methodology used to produce the estimates is absent. A full report on methodology is required to enable the Tribunal to assess the estimates. If the willingness to pay estimates cannot be assessed the Tribunal will not be able to decide if the estimates are valid.

Robust willingness to pay estimates have the potential to assist the Tribunal to determine the efficient level of service quality. Adherence to the best practice guidelines, particularly the full reporting of methodologies will help the validity of willingness to pay estimates for improvements in services supplied by utilities.

This review has emphasised the strengths and limitations of a number of techniques that have been in common use in quantifying the benefits of quality. Considerable gains in rigour have been achieved over former practices using open-ended contingent valuation techniques. However, even the best practice techniques of carefully constructed choice modelling leave room for considerable uncertainty around estimates of ‘marginal utility’ or ‘willingness to pay’ for additional quality. And even where there is compelling evidence that there is on average a positive willingness to pay, the steps from these to a judgement on the appropriateness of a regulated price increase to fund improved quality is not easy. For one thing the regulator is typically setting a charge in a single price that affects a customer population with a wide range of valuations placed on additional quality. Consequently, it is important that wherever possible, the spread of valuations among the underlying population is reported.
At a minimum, therefore, submissions on WTP for enhanced quality based on choice modelling should include demographic characteristics in choice equation estimation to allow for the impact of different user characteristics on WTP.

Finally, framing problems may arise when individual utilities independently seek WTP responses to their own proposed quality improvements without reference to similar choices being posed by other utilities to the same broad population. If the prospective choice facing consumers is not just one relating different electricity prices to different service standards but these options and equivalent choices for public transport then a higher level choice problem arises.

There may be merit in conducting a broader ranging choice modelling experiment to test the trade offs the public is willing to make across utilities' service attributes before undertaking individual studies.
References


Allen Consulting Group. 2001, The Incorporation of Service Quality in the Regulation of Utility Prices, report to IPART.


Hanemann, W. M. 1994, Contingent Valuation and Economics, Department of Agricultural and Resource Economics, University of California, Berkley.


