Forecasting in Cost-Benefit Studies

Paper prepared by Peter Abelson for Cost-Benefit Forum, 14 October 2015[1]

Economic Society of Australia, NSW Branch

This paper discuses issues in, and methods for, forecasting impacts for CBA.

"If we speak frankly, we have to admit that our basis for knowledge for estimating the yield ten years hence of a railway, a copper mine, a textile factory, the goodwill of a patent medicine, an Atlantic liner, a building in the city of London amounts to little and sometimes to nothing." Keynes. J.M., 1936, *The General Theory of Employment Money and Interest*, Chapter 12 – The State of Long-Term Expectation, MacMillan, London.

1   Introduction

Forecasting the impacts of projects or policies quantitatively over their life, compared to a no-change base case, is a central part of cost-benefit studies, or indeed of any form of evaluation.

In some cases, economists provide the relevant forecasts, especially for market goods and services, and increasingly for willingness to pay for non-market goods. But in numerous cases, forecasting costs and outcomes is carried out by experts in the relevant discipline, be that engineering, traffic modelling, epidemiology, education, social, environmental or whatever. And indeed consulting firms often note that, while they have undertaken the cost-benefit modelling, they disclaim all responsibility for some or all of the critical inputs.

However, there are three strong reasons why economists should be involved actively in the forecasting process. First, at the outset of a CBA, an organised structure of costs and benefits, consistent with a sound cost-benefit methodology, should be established. Economists, or cost-benefit analysts, need to specify these impacts. Economists should be wary of requests, which the writer has sometimes received, to provide a cost-benefit analysis after the data has been collected! Second, economists are responsible for the analysis of risks or uncertainty for the project or policy. The treatment of risk cannot be separated from the provision of forecasts and understanding the distribution of the key forecast inputs to the analysis. This requirement needs to be communicated to, and discussed with, the professionals making the forecast(s). Thirdly, in my view, the authors of cost-benefit studies should have reasonable confidence that the forecast inputs and, where possible, not simply disclaim any responsibility for the data put into the cost-benefit report.

---

[1] I thank Dr. Roselyne Joyeux for helpful comments on an earlier draft. Roselyne is not responsible for any remaining errors.

One further general introductory remark should be made. It should be recognised that it is not always possible to provide quantified forecasts of outcomes of acceptable rigour. Sometimes it may be possible only to provide a qualitative order of magnitude forecast for some impact such as "high", medium" or "low". Depending of the importance of this impact, it may still be possible to make an overall recommendation on a project or policy or it may suggest that more research is desirable before a decision is made.

Turning to the process of forecasting, it is common to distinguish between two main approaches to generating forecasts of outcome: based on natural and experimental data respectively. Natural data are data that occur as a natural part of the economic process.[2] Market data are a prime example of natural data and forecasting outcomes drawing on market data is a common process. Experimental data are data collected from experiments to estimate some outcome(s). Experiments are used most often for non-market goods where there is no relevant natural data to draw on. However, it should be noted that, although the data sources may differ, similar statistical methods may be applied to the different data sets.

The distinction between natural and experimental data is a useful one which we use on this paper. But it should be observed that the distinction is not always a clear one. Natural data sometimes provides situations in which people are effectively randomly assigned to two different groups and these situations are called "natural experiments". On the other hand, governments may initiate programs as part of their strategic platforms, and not formally as experiments, and the observed data on outcomes is, in effect, natural data. However, for the purpose of estimating outcomes, these programs are usually treated as if they are experiments and we maintain this distinction.

As we will see, there are there also other ways to generate forecast outcomes, including by choice modelling, other market research and drawing on expert modelling and views.

In this chapter, we start with a short discussion of estimating production costs. Although forecasting costs is usually less challenging than forecasting outcomes, errors in cost estimates can have critical impacts. Section 3 describes some examples of forecasting problems and some key forecasting issues. Sections 4 and 5 discuss methods of forecasting based on natural data and experimental data respectively. Section 6 discusses variations of these methods. A concluding section briefly highlights the main points.

2   Forecasting Production Costs

In many cases, where the product and technology are well known and any relevant external physical conditions are well understood, forecasting costs of production is straightforward. Thus, forecasting residential construction costs on a flat site with known soil conditions is straightforward and subject to little variance.

However, there are many cases of major cost overruns. The capital cost of implementing national broadband network in Australia is a huge example. Flyvbjerg et al. (2004) report many large cost overruns on mega projects. Cantarelli et al..(2010) report likewise for several large transportation projects.

---

[2] Natural data are sometimes described as observed or observational data. However, experimental data are also observed.

Forecasting problems may arise with new products, uncertain technology and unknown physical conditions. Also project scale often changes. For one or a combination of all these reasons, high cost overruns are common in procurements of new military products. Uncertain technology can affect the costs of producing renewable energy facilities or indeed the cost of any research program. Uncertain physical conditions can also create wide errors in cost estimates, especially in the case of underground conditions that may affect the costs of mining operations, undergrounding power lines or building dam walls. As Boardman et al. (2014) remark: "Forecasts of the real resources required for large and complex infrastructure projects are often too low because of the need for redesign as information about the site and actual performance of the capital equipment employed".

Uncertainty over prices and labour costs may also affect production costs. Generally, production costs rise broadly with consumer prices and so estimates of costs in constant prices can ignore possible changes in production costs. However occasionally changes in relative prices are significant especially for the costs of imported materials which reflect changes in exchange rates.

Forecasts of costs (as well as of benefits) for Cost-benefit studies must also be on guard against optimism bias, sometimes referred to as appraisal optimism or cognitive bias. This bias may occur through excessive confidence in the technology of production or when project proponents have an unrealistic view that they will be able to control production costs against any cost blowouts.

The main check against both technical error and optimism bias is independent expert reviews. However, it is not always easy to find genuinely independent experts. In the defence field, for example, many relevant experts may derive considerable income as advisers to the military establishment and be reluctant to put this income at risk with a genuinely independent review of the military's favourite projects. And government agencies may also be inclined to contract advisers known to be sympathetic to the project.

Importantly, it may be noted that there is often a skewed distribution of cost estimates. This arises because the lowest cost outcomes are rarely more than 20 or 30 per cent below the mean, whereas the highest cost outcomes may be 50 or even 100 per cent higher than the mean.

This leads some people, mainly non-economists, to suggest guarding against errors in cost estimates by working with P90 cost forecasts, i.e. costs that will be at this level or lower 90 per cent of the time. However, as discussed elsewhere, the preferred approach in CBA is to work with mean estimates and deal with possible cost variations through an explicit risk methodology.

Finally, we should also note that in some situations costs may be overestimated. As Harrington et al. (2000) report, costs to respond to environmental regulations are often over-estimated for two reasons. First, to minimise regulations, firms have an incentive to over-estimate their cost of compliance to the regulator. Second, firms may adapt to regulations by changing their methods of production in ways that are not anticipated by regulators.

3   Examples and Issues in Forecasting

Table 1 provides some examples of forecasting problems for capital projects, recurrent programs and for regulations. In each case, it describes some intended outcomes and some issues arising.

Table 1   Examples of issues in forecasting impacts

Area / drivers	Outcomes	Comments
Investment projects		
Investment in infrastructure e.g. in water, power or public transport	Increased consumption of water, power or public transport over time with lower prices	Estimating income and price elasticities from past experience
Investment in road infrastructure in cities	Forecasting traffic use over the network including mode shifts	Complex modelling of network traffic and land use changes
Investment on national broadband	Greater capacity and speeds for internet communications	Unknown capacities of alternative future technologies
Investment dam construction for flood mitigation	Reductions in flood damages and changes inland uses	Forecasting impacts of some very low probability events
Recurrent programs		
Early childhood and family assistance programs	Changes in behaviour, educational and social outcomes	Separating impacts of other factors on outcomes and estimating length of outcome
Class sizes	Changes in educational outcomes	Separating impacts of other factors on educational outcomes
Training and other programs for unemployed workers	Length of time in unemployment, income outcomes	Separating impacts of other factors on outcomes and estimating length of outcome
Plain packaging of tobacco products	Reduced tobacco consumption and improved health	Hard to estimate impact on tobacco consumption and length of impact
Regulations		
Compulsory fencing of backyard swimming pools	Reduced swimming pool fatalities and other accidents	Hard to identify cause and effect of pool accidents
Regulations over hotel / pub operating hours	Reduced alcohol consumption and social violence	May relocate alcohol consumption and social violence
Regulations for food safety	Improved health outcomes	Hard to estimate costs to businesses which change production methods in response
Climate change policies, including subsidies for investment in renewal energy sources	Impacts on global climate and vast array of environmental and economic outcomes	Relies on expert modelling of climate change, hard to forecast impacts of alternative technologies and impacts on biodiversity and production functions

These examples indicate several practical problems for forecasting. These include:

- Uncertain future technologies, for example in communications technology or renewable energy

- The need in some cases to forecast the likelihood (probability) of very low occurrence events (1 in 1000 or 2000 year events) such as extreme floods and concurrent levels of fatalities.

- The need to identify the program effect when there are many possible causes of outcomes.

- The need to forecast both changes in behaviour and the effects of these changes in behaviour.

- The increasing need to forecast non-market outcomes.

- The complexity of some modelling, such as traffic modelling across a network of close substitutes where there are typically four steps in the forecasting model: trip generation, trip distribution, mode choice and route choice (and even this may not allow fully for trip - land use interaction).

- And the interactions between ecological resources, such as groundwater quality, and production functions for say crop yields while controlling for the effects of unobserved characteristics of climate and topography.

Issues in Forecasting Outcomes

There are several problems in forecasting outcomes. The first one, noted in the introduction, is to identify the outcomes to be predicted. This is particularly important in considering third party and indirect market effects. To determine the likely possibility of indirect market effects, we typically need to draw on economic theory and identify where prices are not equal to marginal costs, and thus markets where a producer surplus or deficit may occur. We would then need to draw on a partial equilibrium model of the related market or, in some cases, a general equilibrium model of multiple markets.

However, in most cases, the key problem is to identify causes and effects in the primary market or activity. A general problem with using natural data, as well as with some experimental data, is the lack of randomness of the data. This puts a lot of weight on econometric methods (the application of statistics to economic data) to sort out cause and effect. A commonly cited case is estimating the impact of class size on child performance. The estimation problem arises because more education-motivated parents may send their children to smaller classes. We discuss below various econometric methods to identify relationships between the dependent and independent variables.

Another major general issue is that forecasts are generally based on an analysis of past events and behaviour. But the object of some policies is to change behaviour, which is in turn expected to change outcomes. Often there is a long chain of events between cause and effects.

There are related issues of timing. How long do the effects last? Responses to some changes, such as price changes, tend to increase over time, as firms and customers have more time to respond to changes in prices. But responses to other changes, such as advertising, may decline over time.

There are often issues of extrapolation to other population groups with possibly different social or demographic attributes. These issues arise in many sectors but are especially important as CBA expands into social fields.

As also noted above, ideally we would like to have the distributions of forecasts as well as means of key variables. But in any case, where possible we would like to have estimates of standard errors to report prediction intervals.

In general, more evidence is preferred to less. Evidence from more research studies is more reliable than findings from fewer studies and safer to extrapolate from. We also need to be aware of publications bias (the tendency of journals to publish positive results rather than negative one).

And finally, again we need to mention appraisal optimism bias. As far as possible, the experts providing the forecasts should be independent of the agency promoting the project.

4   Forecasting with Natural Data

Much empirical work in the social and natural sciences is based on natural data. Economists often draw on natural (unplanned) differences or changes in situations to estimate behavioural responses including price elasticities. These methods are generally guided by economic reasoning or theory. The core issue in each case is to identify casual relationships between the dependent and independent variables.

Natural data may be time series, cross-sectional or a combination of time series and cross-sectional data. The data may be obtained via regular or specific one-off surveys, but they are distinguished from data obtained from experiments.

Time series analysis typically draws on aggregate or average data over time. The use of averages may limit the inferences that can be drawn about individual behaviour.

Cross-sectional data includes data across jurisdictions, various community groups or individuals. Cross-sectional studies based on individual data greatly increase the number of observations. However, in this case it is important to sort out what characteristics of individuals may affect the outcomes.

Combined time series and cross-sectional data may draw on aggregate comparisons as between countries or between states within a country or on longitudinal data on individuals typically collected as panel data. In cross-country or cross-state studies the analyst explores the effects of national or state differences in benefits and institutional arrangements on labour supply or other behaviour. However, differences in other economic or social factors may also influence behaviour and be difficult to model precisely.

With longitudinal panel data, the aim is to infer how changes in individual circumstances influence individual decisions. Given the need for extensive longitudinal data sets, there have been few such studies

although their use has increased greatly over the last 10 or so years. The Melbourne University Hilda survey is the well-established and outstanding Australian longitudinal survey. As with other approaches, modelling all the critical factors and their inter-relationships is not easy.

We discuss below the basic statistical method for analysing natural data (regression analysis) and three related and quite often used statistical methods.

Regression analysis

Regression analysis is the basic and powerful statistical method used to infer causal relationships between variables and their causes, for example the effects of incomes and prices on consumption, of education on incomes or of employment programs on the level of unemployment.

To estimate these relationships, the analyst assembles data on the variables of interest and employs regression to estimate the quantitative effect of the causal variables (independent variables) upon the variable (the dependent variable) that they are expected to influence. The analyst also typically assesses the "statistical significance" of the estimated relationships, that is, the degree of confidence that the independent variables have an effect on the dependent variable.

However, as will be noted, there are several issues to be aware of. Importantly a correlation between variables does not necessarily imply causation.

Thus, regression models involve the following variables:

- The dependent variable, Y
- Independent variables, X
- Estimated parameters, denoted as β, which may represent a vector.

A regression model relates Y to a function of X and β. The key task of the regression analysis is to estimate accurate values of β. In a linear model, this is expressed as

$$Y = \alpha + \beta X + \varepsilon \tag{1}$$

where α is a constant, X may be one or more independent variables and ε is the error term. The ε reflects other factors that may explain the dependent variable (including omitted variables, unpredictable nature of human behavior, errors of measurement).

When there is more than one independent variable, as is usually the case, the statistical exercise is known as multiple regression analysis. This allows additional factors to enter the analysis so that partial effects can be estimated. Of course, the functional relationship (*f*) may be linear or non-linear. If no knowledge on the relationship is available, a flexible form for *f* is chosen.

In addition to wishing to know the value (β) of the key parameters, we also want to know whether the real value is likely to be significantly different from zero. The "conventional" test of significance is that the

value of β should be over twice the standard error. This indicates that there is at least a 95% chance that β is different from zero.

To cite two examples of regression, consider first estimates of the impacts of income and price on demand for a good. These are the key components of forecasts of consumption for market goods. Note that in this case, economists commonly estimate the **percentage change in the dependent variable due to a one percentage change in the independent variable** (this measure is described as an elasticity).[3] Accordingly, the analyst estimates a log equation of the following kind:

$$\mathrm{Ln}C = \alpha + \beta_1 \ln Y + \beta_2 \ln P + \beta_3 \ln Pop + \varepsilon \tag{2}$$

where C is consumption, Y and P represent income and price respectively and Pop equals population. Here, the estimated β_1 represent elasticities.

To take a second example, suppose that we want to understand the effect of education on earnings (Y). The independent variables could include years of education (E), years in the workforce (W) and a variable reflecting level of parental educational qualifications (P). Holding E constant, earnings are likely to rise with W and P. The hypothesised equation in this case could be

$$\mathrm{Ln}Y = \alpha + \beta_1 E + \beta_2 W + \beta_3 P + \varepsilon \tag{3}$$

where α = a constant amount that someone earns with zero education, no experience in the workforce and zero parental educational qualification, the β_i's are the percentage increases in earnings associated with one extra year of education or in the workforce or an extra level of parental qualifications. In this case, ε reflects other factors that influence earnings.

There are several other critical requirements for an unbiased efficient estimation of a regression equation. Chief among these are:

- The key relationship reflects cause and effect
- There is no significant omitted variable (omitted relevant variable bias)
- Multicollinearity between independent variables.

Technically a regression analysis just demonstrates a statistical relationship. A correlation does not itself imply causation. Drawing on the generic equation (1) above, it is always possible that the variable Y is really the independent variable and X is the dependent one. For example, does GDP drive unemployment or does unemployment drive GDP? Some additional theory is generally required to assert a causal relationship.

---

[3] For example, price elasticity shows the relationship between the quantity of a good demanded or supplied and its price as follows: $\Delta Q / Q = \eta (\Delta P / P)$, where Q and P stand for quantity and price respectively, Δ stands for change and η for elasticity. Suppose that $\eta = 1.0$, the percentage change in quantity demanded or supplied is the same as the percentage change in price (this is known as unitary elasticity).

Omitted variables may cause the results to be biased. An omitted variable may be the real driver of the dependent variable and be correlated with one or more expelanatory variables. Thus, omitting a key variable may result in a variable having undue weight in the regression. For example, earnings are affected by a variety of factors in addition to years of schooling. If these factors are not included in the analysis, the role of education in earnings is likely to be over-estimated.

Multicollinearity occurs when two or more independent variables are closely correlated, creating a situation in which their effects are difficult to separate. Multicollinearity between variables increases the standard error of the estimates and thus reduces the degree of confidence in these variables.

In most instances these conditions are necessary conditions for the least-squares estimators to possess desirable properties; in particular, these assumptions imply that the parameter estimates will be unbiased, consistent, and efficient in the class of linear unbiased estimators. It is important to note that actual data rarely satisfies the assumptions. That is, the method is used even though the assumptions are not true.

Two final points should be made. "Statistical significance", as defined above, means that, by conventional tests, we can reject the hypothesis that the true value of some variable is zero. But this does not mean that the regression will produce a perfectly accurate estimate of any parameter. What we can say is that parameter estimates with proportionally low standard errors are closer to the mark than others.

Ultimately, statistics does not say how much weight the results of a regression study ought to be given, or whether it is reasonable to use a particular parameter estimate for some prediction purpose. These assessments are inevitably entrusted to users of the regressions.

Finally, for the purpose of prediction, it is critical that the sample in the regression equation is representative of the relevant population for which forecasts are to be made. This is of course critical if we are to adopt the regression results for other populations.

Natural experimental studies

Natural experimental studies draw on situations in which individuals happen as a result of external or natural circumstances to be randomly assigned to different groups.[4]. This reflects the quality of random assignment in the natural data. There are three main kinds of natural experimental methods.

Instrumental variables analysis is an important form of natural experimental study method. The general problem arises because an omitted variable may affect both the independent variable and the dependent variable. Thus we need to find an (instrumental) variable that influences the independent variable without any necessary effect on the dependent variable.

An example of is the impact of class size on child performance. The problem arises here because more education-motivated parents may send their children to smaller classes and also influence educational outcomes. Thus the analyst needs to find a variable which influences participation in smaller classes but

---

[4] These situations are sometimes described as natural experiments (as here) and sometimes as quasi-experimental methods because they are like experiments.

which itself will not affect the educational outcome. In the United States, Hoxby (2000) found that birth date (which determine school entry) vary randomly over time so that class sizes also vary randomly from year to year. She used these random annual differences in kindergarten class sizes within the same schools to test whether differences in class size affected educational outcomes.

Another natural experimental method is *difference-in-differences analysis.* Suppose there are two jurisdictions *A* and *B* and that one jurisdiction (*A*) applies a policy change (say a higher unemployment benefit) and that the other (*B*) does not apply this policy change. The difference-in-difference method examines how, if at all, the *difference* in the unemployment rates has changed as a result of the policy change. This recognises that other factors may cause systemically different levels of unemployment in the two jurisdictions. But this does not eliminate the possibility that changes concurrent with the policy change could have affected the differences. Thus this possibility needs to be examined.

A much quoted example is the study by Card and Krueger (1994) of the effect of changes in the minimum wage employment.   Card and Krueger compared employment in the fast food sector in New Jersey and in Pennsylvania before and after the minimum wage in New Jersey rose from $4.25 to $5.05 per hour in April 1992. Had the authors simply considered any change in employment only in New Jersey, before and after the treatment, they could have failed to control for omitted variables such as changes in the economic conditions of the region. By including Pennsylvania in a difference-in-differences model, variables common to New Jersey and Pennsylvania are implicitly controlled for, even when these variables are unobserved. Assuming that New Jersey and Pennsylvania experience broadly similar economic conditions over the observed period (February to November 1992) the change in employment in Pennsylvania can be interpreted as the change New Jersey would have experienced, had they not increased the minimum wage. The evidence suggested that the increased minimum wage did not induce an increase in unemployment in New Jersey compared with Pennsylvania, as standard economic theory could suggest.

Regression-discontinuity (RD) analysis is a third form of natural experiment study. This method estimates the impact of programs by comparing outcomes for program participants who just meet the threshold criterion for participation with outcomes for non-participants who just fail to meet the threshold criterion but who can be assumed in other attributes to be similar to the program participants. Because both sets of individuals or households are close to the threshold, it is assumed that they are comparable individuals or households in most important respects. In effect the treatment assignment is random. Thus if outcomes differ for the two groups the difference may be ascribed to the new program (or changes in the program).

Regression discontinuity analysis has been widely used in evaluations of educational and social interventions. Education programs are frequently provided to schools or students who score below a cut-off on some scale (student performance, poverty), and school and program funding decisions are often based on allocation formulas containing discontinuities. The design has also proven useful in evaluating the socioeconomic impacts of a diverse set of government programs and laws. Van der Klaauw (2008) cites numerous examples. In all these applications, the treatment variable or the probability of receiving treatment changes discontinuously as a function of one or more underlying variables, which is the defining characteristic of RD data designs.

5   Forecasting with Experimental Data

Experimental studies are studies based on data gathered from experiments. Boardman et al (2014) identify five commonly used program evaluation designs using experimental data. Adopting our own ordering, these are:

1. Random design experiments with baseline data
2. Random design experiments without baseline data
3. Non-random experiments with baseline data
4. Non-random experiments without baseline data
5. Before and after studies.

The first two experimental designs are sometimes called true experiments. A true experiment includes, a treatment group and a control group and random assignment of study participants between the two groups, preferably with pre-and post-test data. The third and fourth, non-random, approaches are called quasi-experiments because they lack the characteristic of random assignment of study participants.[5] This may create selection bias. This problem is avoided in experimental studies where a sample of individuals is randomly assigned to a treatment and to a control group.

Random design experiments with collection of baseline data. In these experiments individuals are allocated randomly to a treatment or control (non-treatment) group. If the sample size is large enough, random assignment ensures that characteristics of individuals in each group are similar. This ensures *internal validity (*there is no selection bias). Because, on average, the individuals in the two groups may be viewed as similar, any differences in outcomes can be attributed reliably to the treatment. Also in this set of experiments, data are collected on outcomes before and after the treatment starts. The evaluation then assesses the **net** changes in selected outcomes. This "true" experiment is widely used in the health domain with testing of drugs where individuals are unaware whether they are in or not in the drug program.

Another example comes from unemployment benefits. Hotz *et al*. (2002) analysed the effect of a policy experiment in California in the 1990s where one-third of families randomly selected received about 15 per cent higher non-work (unemployment) benefits than other families. They found that there was about a 10 per cent increase in employment in families receiving the lower level of benefits, indicating an elasticity of employment relative to benefits of about –0.67. Zabel *et al*. (2010) reported on a Canadian randomised trial in which the treatment group of unemployed workers were given a subsidy that roughly doubled their pre-tax salary for three years if they found long-term full employment within 12 months. Four years later, the employment rate for the treatment group was approximately 25 per cent higher than for the non-treatment group.

It is widely agreed that randomised experiments represent the 'gold standard' method of isolating and identifying cause and effect. However, there are several drawbacks. These include the need for a large sample size, problems associated with drop-outs that may not be random, making inferences from behaviour in experiments to longer-term behaviour, the high costs and length of time of the experiments and the ethics of providing different levels of benefits or services to different groups in the community even if this is done

---

[5] As noted in footnote 3, some natural experiments are also sometimes called "quasi experiments".

randomly. Finally, the experiment does not guarantee full *external validity*—that the results will apply to any other groups in communities with different socio-economic characteristics. The effectiveness may vary with alternative groups.

Random design experiments without baseline data. This method is similar to the above approach except that there is no collection of pre-treatment (baseline) data. Similar observations therefore apply about such matters as ethics and external validity. However, in this case there is no baseline data on possible differences in attributes or behaviour between the treatment and control group. If the groups are relatively small, with say one or two hundred people, or if the randomisation process is incomplete, there may be some significant differences between the treatment and control groups. But these differences will not be known.

Non-random experiments with baseline data. In this case pre- and post-treatment data are collected. Therefore, where post-treatment outcomes vary, this finding can be adjusted for any pre-treatment differences. This assists with internal validity.

But where the treatment and control groups are not randomly selected (the control group is sometimes called a quasi-control group), there is a risk of sample selection bias and it may be difficult to determine the real causes of differences between the groups. Motivations for behaviour change may be greater in one group than in another.

Suppose a program dealing with unemployed persons leads to various outcomes. This involves a non-random assignment with baseline evidence. Can we assume that extending the program to other unemployed persons would have similar outcomes? Not, if the initial group is selected rather than a random experiment. In the absence of a random comparison group, extrapolating the results to other groups would need to be done cautiously.

Non-random experiments and no baseline data. This approach suffers from both a potential sample selection bias and a lack of baseline data. The combination of non-random assignments to treatment and control groups and an absence of baseline data mean that internal validity is hard to obtain. Some results will be the consequence of sample selection bias.

Before and after studies of the impacts of programs with the same group of persons. These studies have the advantages of being relatively easy and inexpensive to conduct and are therefore a common way to estimate program effects. In any case there may be no convenient comparison group. However, there is a major problem in attributing cause and effect. There is no information on what would have happened without the treatment. Any changes in behaviour and outcomes may be affected by other factors than the program.

In general, randomised experiments are preferred to non-randomised experiments and experiments with baseline data are preferred to those without baseline data. However, the optimum experiment is not always possible for various reasons, including cost, time and ethics.

Researchers conducting experiments may attempt to address non-random selection of treatment and control groups in quasi-experimental studies in various ways (e.g., by matching treatment groups to like control groups or by controlling for these differences in analyses). Drawing on existing programs in place

to specific population groups may allow extrapolations to other population groups. They do provide a basis for estimating the resources required for any program and some indication of the likely impacts of programs.

However, there does need to be a comparison or control group. This could be the baseline pre-program outcome for the group now receiving the program services. Programs may be broadly similar but differ in intensity, types of inputs or populations served. Confidence in extrapolating would depend on the number of differences and their likely impacts. If extrapolating from one or two cases, judgement as well as statistical analysis may be needed.

7   Other Forecasting Methods

Other forecasting methods include meta-analysis or generic studies, expert simulation modelling, market research including choice modelling, and simply reliance on expert advice.

Meta-analysis and generic income and price elasticities

Meta-analysis draws on a pool of studies to obtain estimates on mean impacts and variations over the pool. Of course, many of the studies in the pool will draw on either natural or experimental data. To conduct a meta-analysis, comparable quality studies have to be selected and, critically, a standardised measure of effect determined so that the findings can be compared.

Simple meta studies find an average effect size and variance. More detailed studies use multivariate regression analysis to estimate an average size effect and variance controlling for quality of study, variations in the study populations, and other details in study implementation. See for example Washington State Institute for Public Policy (www.wsipp.wa.gov)

Meta-analysis reduces the bias that can result from reliance on a single study. However, the initial studies must be of good repute. Where possible, the circumstances of the studies should be similar to those of the project or program under consideration. Any differences between the socio-economic conditions of the research studies and the project case need to be considered and possibly adjusted for.

Deriving and using generic elasticities from multiple studies is a related form of meta-analysis. There are many publications of price elasticities of demand for goods including, for example, water (Dalhuisen et al, 2003), electricity (Espey and Espey, 2004), public transport (Holmgren, 2007), and petrol (Brons et al, 2008). There are also many studies of the impacts of taxes on corporate investment and on labour supply.

As Boardman et al (2014) point out, there are also many data bases available such as ECONLIT, Google Scholar, JSTOR and Proquest Digital Publications, as well as subject specific data bases such as ERIC for education and PubMed for health.

However, many factors affect income and price elasticities and care must again be taken in transferring any generic or average elasticity to the particular case study. In demand functions, critical factors include differences in demographic factors, income levels and the availability of substitute goods. Also demand and supply elasticities may increase over time, as customers and firms have more time to respond to

changes in prices. All such factors should be considered in adopting a generic price elasticity in a CBA for a specific project or program.

Expert simulation models

Many forecasts are the output of complex simulation models which are in turn usually based on models of causes and effects often based on evidence collected and analysed over many years. Examples include computable general equilibrium (CGE) models, transport models, climate change and flood modelling. We briefly describe three of these.

CGE models are complex models of an economy (this may be national, regional or international). They model household demands for goods and services through equations typically modelling a large number of industries, often over a hundred. Household demands are a function of household types and income and price elasticities. The industry impacts are based on actual data usually from extensive input-output tables which show the inputs required to provide output in each industry. However, in CGE models, the outcomes are constrained by domestic resource availability and assumptions about capital movements and net immigration. Most CGE models assume competitive markets and market clearing. But they can allow for imperfect competition and pricing and non-market clearing,

CGE models can show how an economy as a whole may react to changes in policy, technology or other exogenous factors. For example, CGE models are routinely used to forecast the impacts of monetary and fiscal policy, trade policy, climate change and changes in international prices on an economy. They can show employment and income effects.

Perhaps three caveats should be noted. First CGE models are dependent on the quality and timing of the data available. The input-output data are sometimes a few years old and are not always available at sub-national levels. Secondly, the models depend upon the assumptions about household and business behaviour and the operations of markets. This means that different CGE analysts can produce quite different forecasts using similar inputs. Thirdly, they are models of the market economy, not of household welfare.

Urban transport models are another example of a highly complex model. Typically, this involves four stages: a trip generation model of the number of trips per household by type of trip in some hundreds of urban zones, a trip distribution model which predicts the combinations of origins and destinations, a modal split model and a route choice model that allows for tolls where appropriate. The model is typically run for a week-day morning peak for some future year, some 10 years or more ahead. This implies modelling a future unknown network in the base case, future service levels and factoring up the peak hour estimates to annual figures. While based on current observed survey data, behaviours may change. It is not surprising that traffic forecasts are subject to a high degree of variance.

Modelling floods is another complex area because we are often dealing with very low probabilities. There has been a lot of discussion about the need to raise the Warragamba dam in Sydney to avert possible very large flood damages. But the flood damages rise exponentially with high, but very low probability, flood events. We are talking here about the impacts of 1 in 500 or 1 in 1000-year events, or even 1 in 5000 year events. Half the expected benefits from raising the Warragamba dam wall arise from lower damages in these extreme events. But the highest flood experienced in the 200 years of European settlement is

estimated to be a 1 in 200-year flood. So the modelling of more extreme but important events in this case is based on mathematical assumptions rather than observed data.

Market research

Market research, in the form of focus groups or more extensive quantitative surveys, is of course the widely used method for forecasting consumer demand in the private sector. It can also be used to forecast how people will respond to changes in public programs or policies. For example, before requiring packaging of tobacco products to depict various forms of serious diseases in 2005, the Australian Government commissioned extensive focus group research to determine whether tobacco smokers would change their smoking behaviour as a result of such pictures.

Choice modelling is another major forecasting tool today. As we have seen, when major changes in the transport network are planned, complex modelling of trip choices may be needed. We cannot assume that trip makers will automatically choose the least cost mode for various reasons. One is that individuals have differing values of travel time. Another is that trip makers have multiple reasons for making trips and diverse preferences over trip attributes. Some people may be willing to stand in public transport. Others will travel only if they can obtain a seat. Choice modelling provides a method for understanding and forecasting the decisions that individuals are likely to make when facing various trade-offs in their travel choices.

Expert advice

Another traditional and contemporary approach to forecasting impacts is simply to seek expert advice. This is essentially an umbrella description which encompasses a range of methods from the application of highly professional and exhaustive research in a similar field to the proposed program or in a closely related one to the professional, but essentially speculative, opinions of experts on a largely or wholly un-researched area.

8   Conclusions

Forecasting outcomes is critical to CBA. As a minimum, economists should be involved in setting up the structure of the CBA and the outcomes to be estimated. In some cases, employing econometric methods, economists will make the relevant forecasts. In many other cases, experts from other fields will make the forecasts. But the economist, or CBA analyst, should be aware of the various forecasting methods and able either to provide a review of the forecasts or to commission a review. In any case, the CBA analyst will need to understand the distributions involved in the major variables in order to conduct or manage an analysis of the risks involved as part of the CBA.

Thus the economist needs to be familiar with the main approaches to forecasting described in this paper, methods of analysis of natural and experimental data, as well as various other methods including meta analysis or generic studies, expert simulation modelling, market research including choice modelling, and simple reliance on expert advice.

A general problem with using natural data is their lack of randomness. This puts a lot of weight on the econometric analysis to sort out cause and effect typically with some form of regression analysis. This problem is mitigated when the natural data effectively assigns individuals randomly to a treatment and to a control group. These are sometimes described as quasi-experimental situations. Techniques of analysis include instrumental variables, difference in differences regression analysis and regression-discontinuity analysis.

Experiments with random assignments of subjects to treatment and control (non-treatment) groups along with baseline data are regarded as the gold standard for predicting outcomes. However, these experiments are often not feasible due to the large sample size required to ensure robust extrapolation to other groups in society, the costs and length of time of the experiments and the ethics of providing different treatments to different groups in the community.

Thus researchers often adopt experiments with non-random treatment groups. These often provide useful insights into likely outcomes, but generally provide less reliable results or results that cannot be reliably applied to other groups of individuals.

Of course, these methods are not mutually exclusive. Impacts may be forecast using more than one method. And, as noted, other methods include various forms of meta-analysis, including generic elasticities, expert modelling and various kinds of market research. Special risks occur in extrapolating mean results from meta studies when there are large variances or in extrapolating from any single study with a small sample size.

Two final observations. First, it is not always necessary to forecast outcomes for all impacts. While quantification of impacts is desirable, where impacts are reasonably regarded as minor, they can be treated as unquantified. The CBA analyst may then report on whether the unquantified items are of such magnitude that they are likely to change significantly the estimated net present value or benefit-cost ratio.

Secondly the analyst needs to be aware that appraisal optimism is a well-documented phenomenon whereby some stakeholders may consciously, or unconsciously, tend to underestimate costs or overestimate benefits. Thus, independent peer review of forecasts in major projects undertaken **before** political decisions are made is highly desirable.

References

Angrist, J.D. and Pischke, J-S., 2015, *Mastering Metrics, The Path to Cause to Effect*, Princetown University Press, Princetown, New Jersey.

Boardman, A., Greenberg, D., Vining, A. and D. Weiner, 2014, *Cost-Benefit Analysis, Concepts and Practice*, Chapters 11 to 13, Pearson Education, Essex, UK.

Brons, M., Nijkamp. P., Pels, E., and P. Rietvald, 2008, "A meta-analysis of the price elasticity of gasoline demand: an SUR approach", *Energy Economics*, 30(5), 2105-22.

Card, D. and A. Krueger, 1994, "Minimum Wages and Employment: A Case Study of the Fast-Food Industry in New Jersey and Pennsylvania", *American Economic Review*, 84(4), 772-792.

Canterelli, C., Flvbjerg, B., Molin, J. and B. van Wee, 2010, *Cost Overrruns in Large-Scale Transportation Projects: Explanations and their Theoretical Embeddedness*, EJTIR, 10, (1), 5-18.

Dalhuisen. J., Florax, R., de Groot, H. and P. Nijkamp, 2003, "Price and income elasticities or residential water demand: a meta-analysis", *Land Economics*, 79(2), 292-308.

Espey, A. and M. Espey, 2004, "Turning on the lights: a meta-analysis of residential electricity demand elasticities", *Land Economics*, 79(2), 292-308.

Flvbjerg, B., Bruzelius, N., and W. Rotengatter, 2003, *Megaprojects and Risk,* Cambridge University Press, Cambridge, UK.

Harrington, W., Morgenstern, R.D. and P. Nelson, 2000, "On the accuracy of regulatory cost estimates", *Journal of Policy Analysis and Management*, 23 (3), 297-322.

Holmgren, J., 2007, "Meta-analysis of public transport demand", Transportation Research: Part A; Policy and Practice, 41(10, 1021-35.

Hotz.V., Mullin, C. and J. Scholz, 2002, "Welfare, employment and income: evidence on the effects of benefit reductions from California. *American Economic Review,* Vol. 92 (2), Papers and Proceedings of 114th Annual Meeting of American Economic Association, 380-384.

Hoxby, C.M., 2000, "The effects of classroom size on student achievement: new evidence from population variation", *Quarterly Journal of Economics*, 115(4)), 1239-85.

Van der Klaauw, W., 2008, "Regression discontinuity analysis, a survey of recent developments in economics", *Labour*, 22(2), 219-45.

Zabel, J., Schwartz, S. and S. Donald, 2010, "The impact of the self-sufficiency project on the employment behaviour of former welfare recipients", *Canadian Journal of Economics*, 43(3), 882-918.