Introduction

Daniel McFadden and Kenneth Train

Contingent valuation (CV) is a procedure that attempts to estimate the value to households of public goods. While CV can be used in many contexts, we consider its use for evaluating environmental goods. The method is implemented through a survey of households. Respondents are given a detailed description of a program that will improve the environment, such as protecting wilderness areas from development or repairing coral reefs. Each respondent is asked whether they would vote in favor or against a ballot measure to fund the project at a specified cost to each household. The cost is varied over respondents, and the share of respondents who say that they would vote in favor is tabulated for each cost level. These shares are then used to estimate the mean willingness to pay (WTP) for the program. The method is sometimes revised to ask each respondent to make choices among several different programs at different costs, instead of just one. For convenience, we use the term CV as encompassing the traditional referendum method as well as these variations.

This book is born of our concern about the reliability of CV. We have collected a series of papers, two previously published and nine new in this book, that tell a unified story about CV. We describe each of the studies briefly below, as a way of introducing them to the reader. Bringing the papers together in one volume allows a picture of CV to emerge that could not, we think, be obtained from any one paper alone. Our narrative is intended to bring out the connection among the papers.

INADEQUATE RESPONSE TO COST

CV studies ask respondents whether they are WTP a specified dollar amount for a program or improvement that has been described to them. Different dollar amounts – called cost prompts, or bids – are asked of different respondents to obtain the variation in cost that is needed to estimate mean WTP. The question arises: how sensitive are CV estimates to the researchers' choice of cost prompts? The answer seems to be: tremendously.

Burrows, Dixon, and Chan (Chapter 1) examined this issue for a prominent CV study conducted by NOAA[1] on WTP for the preservation of marine species, using the original study's data. The survey included several designs with cost prompts that were twice and half, respectively, the costs used in the main survey design. In these variants, one sample of respondents was presented with prompts that ranged from $5 to $50, and another sample of respondents was given prompts ranging from $20 to $200. The original study's report only presented results for the main design; it did not report how the estimates differed under the alternative sets of prompts. Burrows, Dixon, and Chan performed the relevant calculations and found that the estimated WTP was three times greater with the higher-cost prompts than with the lower prompts. That is, raising the cost prompts by a factor of four raised the estimated WTP for the program by a factor of three. This result is consistent with the view that respondents take the cost prompts as a suggestion of the amount that is reasonable to pay and adjust their concepts of their own WTP in relation to these prompts. As a result, CV is not actually estimating a true WTP, but rather is creating an estimated WTP through the researcher's choice of the cost prompts.

Parsons and Myers (Chapter 2) examine the issue of cost prompts from a different perspective. They review CV studies and find that the estimated WTP depends greatly on the highest cost prompt. They find that the share of "yes" votes – that is, the share of respondents who say they are WTP the specified cost prompt – stays relatively high no matter how large the cost prompt is. They call this the "fat tails" phenomenon. To investigate how far the fat tail extends, Parsons and Myers conducted a study about protecting the red knot, a migratory shorebird whose population has declined in recent years. In this study, they kept raising the highest cost prompt, asking new samples of respondents ever-higher prompts, and found that the yes share never approached zero. They raised the prompt as high as $10,000, and still 23% of the CV respondents said that, yes, they would be willing to pay $10,000. The estimated mean WTP ranged from $102 to $2,254, depending on the highest cost prompt that they used. Their study suggests that (essentially) any estimated WTP can be obtained through specification of the highest cost prompt.

[1] National Oceanic and Atmospheric Administration.

INADEQUATE RESPONSE TO THE NUMBER OF PAYMENTS

CV studies can specify the cost prompt as a one-time payment, annual payments over a period of time, or other payment schedules. The question arises: is the estimated present value willingness to pay (PV WTP), which is the relevant measure for resource allocation decisions, sensitive to the payment schedule that is specified in the CV study? The answer is: yes.

Myers, Parsons, and Train (Chapter 3) review studies that have examined this issue. All of the past studies find that results differ greatly depending on how the payment schedule is specified, with the estimated PV WTP being far greater when the researcher specifies a series of periodic payments rather than a one-time, lump-sum payment. The implicit discount rate that reconciles the CV responses under different payment schedules has been found in all studies to be implausibly high. In addition to their literature review, Myers, Parsons, and Train implemented a CV study to compare one-time and annual payments. They found that the estimated PV WTP is 32 times larger when the cost prompt is specified as annual payments than when the cost prompt is specified as a one-time payment.

INADEQUATE RESPONSE TO SCOPE

One of the most important issues in CV is whether CV estimates reflect the scope of the environmental good that is described to respondents. An early influential study (Boyle et al., 1994) found, for example, that CV estimates of WTP to protect birds were essentially the same whether respondents were told that 2,000 birds would be saved or 200,000 birds. Controversy about this issue led NOAA to convene an expert panel to provide guidelines for CV studies. The panel stated (Arrow et al., 1993, p. 38) that a CV study would be deemed unreliable if it exhibited "[i]nadequate responsiveness to the scope of the environmental insult." The panel stated that the burden of proof for demonstrating adequate response must rest with the researchers who conducted the CV study.

As discussed by Desvousges, Mathews, and Train (Chapter 4), out of the hundreds of CV studies that have been conducted, only one has tested for adequate response to scope – despite the expert panel's requirement. This one study concluded that its CV responses evidenced inadequate response. To extend this line of inquiry, Desvousges, Mathews, and Train implemented an "adding-up" test on a prominent and well-funded CV survey. The test examines whether the estimated WTP for each component of a multi-part program, when evaluated incrementally, sum up to the

estimated WTP for the whole program – as required by the definition of WTP. They found that the test fails: the sum of the parts is estimated to be valued three times more than the whole. This finding suggests that respondents' answers to CV questions reflect their expression of interest in the *concept* of an improvement, rather than the scope of the actual improvement that is described to them.

A scope test has often been applied to determine whether there is *any* response to scope, as opposed to the response being adequate in magnitude. Burrows, Newman, Genser, and Plewes (Chapter 5) review the CV studies that have conducted external scope tests[2] and find that more studies fail the test than pass it. That is: more often than not, CV studies don't find *any* response to scope, much less an adequate response. The authors show that previous reviews that have found otherwise (i.e., that passing a scope test is more common than failing) have ignored many studies that failed, have inappropriately included internal tests, and have interpreted results as representing a pass when there is insufficient or contradictory evidence for this inference. Interestingly, the incidence of scope failures has risen over time as the quality of studies has presumably improved, which suggests that the failures cannot be attributed in general to faulty design of the studies but seem instead to be intrinsic to the CV procedure.

DIFFICULTY ANSWERING CV QUESTIONS

Why do CV studies evidence inadequate response to cost, the frequency of payments, and the scope of the program? McFadden (Chapter 6) reviews the history of stated preference (SP) elicitation in general and examines studies that have used these methods in various fields. He identifies the features of a study, and of the good being evaluated, that affect the reliability of the method. He concludes that CV studies of environmental goods possess the very features that make SP elicitation least reliable. The main problem is that respondents are unfamiliar with making choices about environmental goods. The respondent, struggling to provide meaningful answers to CV questions, is susceptible to suggestion by the survey instrument (especially the cost prompt) and to substituting general political concerns for the specific, but unanswerable, personal valuation question.

The difficulty that respondents have in answering CV questions about

[2] An external test uses a split-sample design, where one sample is asked about a program with a specified scope and another sample is asked about a program with a greater (or smaller) scope.

environmental goods seems to be evidenced neurologically. Khaw et al. (2015) measured brain activity of respondents in choice exercises for four classes of goods: snack food, market goods, daily activities, and environmental proposals. For the first three classes, activity was evidenced in the traditional valuational area of the brain, as expected. However, for the environmental proposals, activity was not evidenced in this valuational area. Instead neurological activity appeared in a region of the brain that is associated with cognitive control and shifting decision strategies. Neural measurement is fairly new in economics, and further research is required before conclusions can be drawn. But at face value, the results are consistent with McFadden's assessment that respondents do not know how to approach the CV questions about environmental goods and are struggling for ways to approach the task.

What makes the task so difficult? At least part of the problem is thinking about a budget constraint in the context of environmental goods. A respondent can think that paying $100 to clean up a polluted lake sounds reasonable but then might start to wonder about the thousands of other lakes that need cleaning up, and realize that paying $100 for each of them is impossible. The respondent might then remember all the birds and other species that need protection, and people dying of curable diseases who could be helped with some money for medicine. The respondent faces a quandary about allocation among public goods that the CV survey ignores by asking about only one public good.[3]

Kemp, Leamer, Burrows, and Dixon (Chapter 7) examine the issue of respondents' budget awareness by asking WTP in several ways, including the traditional CV single-focus referendum and by walking the respondent explicitly through a budget allocation task for components of a much larger environmental protection program. The authors found that the estimated mean WTP for a specified project is about $120 when asked in the traditional way but only $2 to $3 when the respondent budgets components of the composite good. And several findings of their study point to pervasive respondent difficulties in thinking about the costs of environmental goods in relation to one another and to other public goods.

But does a budget constraint even come into play when people answer CV questions? The fundamental assumption of CV is that respondents, in giving their response to the cost prompt, are trading off the costs of the program with the benefits. However, as explained above, respondents seem

[3] If the respondent gets so far as to think of the "ordering" problem for public good allocations, then the respondent will also realize that the socially optimal order is not reflected in what projects CV researchers happen to do surveys about.

to have a hard time answering the CV question about WTP for environmental goods. The problem of how to think about the budget constraint in this context, and respondents' sense that the survey is an opportunity to send messages (about, say, culpability or politics) can lead the respondent to answer in ways that do not represent a trade-off of the benefits of the specified program with the cost prompt that they are offered. To examine this issue, Leamer and Lustig (Chapter 8) estimated a latent class model in which each class represents a decision-making process that the respondent might use. The traditional compensatory utility model with a trade-off between costs and benefits is represented by one class, and other decision heuristics are represented by other classes. The shares of respondents using each decision process were estimated as parameters. The authors found that fewer than 25% of respondents seem to be trading off benefits and costs; the other 75% are using decision rules that do not incorporate trade-offs and for which there is no WTP.

THE SEARCH FOR APPROPRIATE CORRECTIONS

It has been suggested that CV samples can be restricted, through the use of follow-up questions, to the "core" of respondents who seem to be answering the CV question appropriately. Past studies have considered eliminating respondents who say they are unsure of their answer, or say that they think the survey is inconsequential, or say that they considered the impact of the program on jobs or other non-environmental outcomes. Each of these studies has generally looked at one issue only, determining the effect of eliminating respondents who do not adhere correctly with respect to that one issue. Myers, MacNair, Tomasi, and Schneider (Chapter 9) apply the procedure to all the issues in combination. They use follow-up questions to address the various issues that past articles have examined only one by one. They find that, out of a sample of 1,224, only two respondents are not eliminated. That is, the "core" group of respondents that seem to be answering the CV question appropriately consists of only two people. And both of these people voted against the specified program.

Similarly, it has been suggested that CV estimates can perhaps be adjusted to account for hypothetical bias, that is, for the bias that arises because the data are for hypothetical programs and payments rather than real ones. The idea behind this suggestion is that, for some kinds of goods, estimates of value can be obtained in both hypothetical and actual settings, and the ratio of these estimates (called the "bias ratio") can perhaps be used to adjust CV estimates for their inherent hypothetical nature. Foster

and Burrows (Chapter 10) examine this possibility, using 432 comparisons between paired estimates in hypothetical and real settings drawn from previous studies. They find that the bias ratios vary greatly, with no ratio being "typical" or common. Using regression analysis, they find that only a small portion of the variation can be explained by attributes of the study or product. The bias ratios in past studies vary so greatly and with so little explainable pattern that they provide no reliable guidance for adjusting CV estimates.

LEGAL ISSUES

Given these issues about the reliability of CV to estimate WTP, how have CV results actually been used in decision-making processes, especially litigation? Lawyers Israel, Martin, Fayne, and Daniel (Chapter 11) review the history of CV in litigation and find that CV results have not been relied upon by any court and have been explicitly rejected in a few cases. The authors describe the legal requirements for reliability of damage estimates in court cases. They conclude that it is doubtful that CV estimates of environmental damages can, or could, meet these requirements. On the regulatory front, the authors point out that natural resource damage (NRD) regulations strongly disfavor CV, allowing it as a last resort to be used only when other methods are not possible. The authors show that the other, more favored methods can practically always be applied. They describe the *Deepwater Horizon* oil spill as a case in point, where the trustees were able to (and did) use non-CV methods of valuation despite the extent and variety of resources affected by the spill.

A CONCLUDING THOUGHT

One additional issue needs to be discussed, because it seems to get to the heart of the CV debate. There seems to be a view that supporting CV is pro-environmental and criticizing CV is anti-environmental. This is a deeply dangerous view. Importantly, results-driven science has an uncanny tendency to circumvent the instigators' intentions. CV can indeed be used to claim large damages against responsible parties (RPs), which seems, in itself, to be a pro-environmental outcome. But CV is used for restoration programs as well as environmental injury, and it gives large benefits for restoration programs. This side of CV provides an incredible boon to RPs, by allowing them to pay off their debts to society at pennies on the dollar. RPs are legally allowed, and in fact expected, to implement restoration

projects to compensate for the environmental damage that they inflicted. CV studies estimate large benefits for environmental projects that cost very little.

Consider, for example, NOAA's study of reef protection (Bishop et al., 2011). This CV study estimated that a program to repair five acres of reefs a year provides a social benefit of $7.3 billion per year. The cost of repairing five acres of reefs has been estimated to be $13.2 million dollars or less (Edwards and Gomez, 2007) – giving a benefit–cost ratio of 553. If CV is actually considered to be reliable, then an RP can rightfully claim $7.3 billion in compensatory restoration by spending $13.2 million on reef repair.

Let's put this into the context of the *Deepwater Horizon* oil spill. A CV study of the spill (Bishop et al., 2016) estimated that households' WTP to avoid a future spill was $15.3 billion or $17.2 billion, depending on how the harm from the spill was described to respondents.[4] Assuming the restoration benefits derived in the study by Bishop et al. (2011), a responsible party could repair five acres of reef per year for three years, creating compensatory benefits of $21.9 billion. So, if CV estimates are believed to be reliable, a responsible party would be able to more than fully compensate the public for the entirety of the *Deepwater Horizon* natural resource damages by paying just $13.2 million per year for three years – less than $40 million in total.

Anti-environmental outcomes like this are the inevitable consequence of CV's inadequate response to scope. The ramifications are wider than the issue of assessing compensation by responsible parties. In benefit–cost analysis, CV tilts the calculations against large environmental improvements. Small measures with relatively little environmental impact (e.g., repairing 15 acres of reef) obtain higher benefit–cost ratios than larger projects with substantial impact (preventing another Gulf spill) because, by CV, the former have about the same benefit as the latter but cost far less. Recognizing CV's unreliability – especially the form it takes – is not just scientifically responsible: it is ecologically responsible.

[4] As we stated above, the trustees did not use CV results in their valuation; see the discussion in Israel, Martin, Fayne, and Daniel in Chapter 11 this volume. Nevertheless, the trustees funded a CV study that obtained these numbers, which were then not used in the NRD valuation.

REFERENCES

Arrow, K., R. Solow, P. Portney, E. Leamer, R. Radner, and H. Schuman (1993), *Report of the NOAA Panel on Contingent Valuation*, accessed November 25, 2016 at http://www.economia.unimib.it/DATA/moduli/7_6067/materiale/noaa%20report.pdf.

Bishop, R. and the Total Value Team (2016), "Technical Memo TM-11: Aggregate estimate of total loss value" (Revised Draft), dated January 25, 2016, to Katherine Pease, NOAA, accessed November 25, 2016 at https://www.fws.gov/doiddata/dwh-ar-documents/980/DWH-AR0302133.pdf.

Bishop, R., D. Chapman, B. Kanninen, J. Krosnick, B. Leeworthy, and N. Meade (2011), *Total Economics Value for Protecting and Restoring Hawaiian Coral Reef Ecosystems (Final Report)*, accessed November 28, 2016 at http://docs.lib.noaa.gov/noaa_documents/NOS/CRCP/TM_CRCP/TM_CRCP_16.pdf.

Boyle, K., W. Desvousges, F. Johnson, R. Dunford, and S. Hudson (1994), "An investigation of part-whole biases in contingent valuation studies," *Journal of Environmental Economics and Management*, **27**(1), 64–83.

Edwards, A. and E. Gomez (2007), *Reef Restoration Concepts and Guidelines: Making Sensible Management Choices in the Face of Uncertainty*, St Lucia, Australia: Coral Reef Targeted Research & Capacity Building for Management Programme.

Khaw, M., D. Grab, M. Livermore, C. Vossler, and P. Glimcher (2015), "The measurement of subjective value and its relation to contingent valuation and environmental public goods," *PLOS ONE*, **10**(7), e0132842, DOI:10.1371/journal.pone.0132842.