Appalachian
STATE UNIVERSITY

Department of Economics Working Paper

Number 13-07 | March 2013

From Hopeless to Curious? Thoughts on Hausman's "Dubious to Hopeless" Critique of Contingent Valuation

Timothy C. Haab
Ohio State University

Matthew G. Interis
Mississippi State University

Daniel R. Petrolia
Mississippi State University

John C. Whitehead
Appalachian State University

Department of Economics
Appalachian State University
Boone, NC 28608
Phone: (828) 262-6123
Fax: (828) 262-6105
www.business.appstate.edu/economics

From Hopeless to Curious? Thoughts on Hausman's "Dubious to Hopeless" Critique of Contingent Valuation[1]

Timothy C. Haab
Department of Agricultural, Environmental and Development Economics
Ohio State University
haab.1@osu.edu

Matthew G. Interis
Department of Agricultural Economics
Mississippi State University
interis@agecon.msstate.edu

Daniel R. Petrolia
Department of Agricultural Economics
Mississippi State University
petrolia@agecon.msstate.edu

John C. Whitehead
Department of Economics
Appalachian State University
whiteheadjc@appstate.edu

March 19, 2013

[1] The authors wish to thank Trudy Cameron, Richard Carson, John Loomis, Ted McConnell, and Kerry Smith for reviewing an earlier draft of the manuscript and for their suggestions that greatly improved it.

From Hopeless to Curious? Thoughts on Hausman's "Dubious to Hopeless" Critique of Contingent Valuation

Abstract. Hausman (2012) "selectively" reviews the CVM literature and fails to find progress over the 18 years since Diamond and Hausman (1994) argued that unquantified benefits and costs are preferred to benefits and costs quantified by CVM for policy analysis. In these comments, we provide counter-arguments to the claims made by Hausman. We provide these counterarguments not with the intent to convince the reader that the debate over contingent valuation is settled but rather to urge the community of economists to recognize that the intellectual debate over contingent valuation is still ongoing and that plenty of work remains to be done. We review the literature and argue that (1) hypothetical bias raises important research questions about the incentives guiding survey responses and preference revelation in both real and hypothetical settings that contingent valuation can help answer, (2) the WTP-WTA gap debate is far from settled and the debate raises important research questions about the future design and use of benefit cost analyses of which contingent valuation will undoubtedly be a part, and (3) CVM studies do, in fact, tend to pass a scope test and there is little support for the assertion that an adding up test is the definitive test of CVM validity.

Introduction

The contingent valuation method (CVM) has a tortured history, due to the skepticism among many economists of survey data and the high profile of the role of contingent valuation in litigation following the Exxon Valdez oil spill (Carson 2011). Contingent valuation was developed because revealed preference methods (e.g., travel cost) that had been used to measure non-market values were limited. Benefit-cost analysis, and other types of policy analysis, required empirical estimates of value that were not, to that point, measureable. Contingent valuation surveys contain detailed descriptions of a hypothetical valuation scenario, payment vehicles (e.g., taxes), payment rules (e.g., majority rule) and valuation questions (e.g., referenda). The initial motivation for the method came from three federal government agencies in the United States from the 1940s through the 1970s. The National Park Service was seeking estimates of the economic value of park recreation and the Army Corps of Engineers was seeking estimates of water-based recreation (justifying dam building). Another key force was the need for empirical estimates of the value of risk reduction in environmental and health economics and the related role of the U.S. Environmental Protection Agency in the 1980s.

The major development that launched the current controversy was the recognition that the passive use values described by Krutilla (1967) could be estimated using contingent valuation. Alan Randall et al.'s (1974) article in the first issue of the *Journal of Environmental Economics and Management* introduced more differentiation of contingent valuation surveys from opinion polls with the use of photographs to help describe the valuation scenario.

The mid-1970s through the 1990s was a period of intense methodological development and increased acceptance, with a broadening of applications beyond recreation and health, expansion to developing countries, and the publication of a number of influential overviews.

Two key milestones were the publication of the Mitchell and Carson (1989) book, which first integrated economic theory, survey research methods and social science measurement issues, and the Exxon Valdez oil spill. The Exxon Valdez oil spill first brought contingent valuation to the attention of many economists, government agencies and the courts. The U.S. National Oceanic and Atmospheric Administration assembled a "blue ribbon panel" to assess the method. The panel concluded that the method could provide accurate measures of passive use values under certain conditions and should be used for policy analysis (Arrow et al 1992). The panel's recommendations created a number of testable hypotheses that emerged in the literature of the 1990s and 2000s. In response to the NOAA Blue Ribbon Panel's conditional endorsement of the contingent valuation method, a contingent of primarily Exxon-funded researchers published a compendium of critiques questioning the accuracy and validity of the method (Hausman 1993).

Following this controversy regarding its accuracy, the contingent valuation method was the subject of a symposium in the American Economic Association's *Journal of Economic Perspectives* in 1994. Portney (1994) introduced the important issues. Hanemann (1994) defended the method. Diamond and Hausman (1994) argued that, for policy analysis, unquantified benefits and costs are preferred to benefits and costs quantified by CVM, and that expert opinion is preferred to imperfect contingent valuation estimates of value. While Diamond and Hausman raised a number of important issues, their negative opinion has done little to quell the demand for contingent valuation research. Since 1994, Google Scholar indicates that the number of contingent valuation studies has increased five-fold.

In the muck of the BP *Deepwater Horizon* oil spill, the *Journal of Economic Perspectives (JEP)* has offered yet another symposium, revisiting the nightmare of the "CVM debate" that followed the Exxon *Valdez* oil spill. Kling, Phaneuf and Zhao (2012) provide a balanced

interpretation of the usefulness of the CVM. Carson (2012) is optimistic about the progress researchers have made in the past twenty years, suggesting that, under the right circumstances, the CVM can be a useful valuation tool. Hausman (2012) "selectively" reviews the CVM literature and fails to find progress over the eighteen years since Diamond and Hausman (1994).

Kling, Phaneuf and Zhao (2012) and Carson (2012) offer some counterarguments to the assertions made by Hausman (2012), and Hausman himself acknowledges a selective reading and interpretation of a vast literature that has developed in the last twenty years.[2] Yet Hausman's claim that "three long-standing problems continue to exist" (p.43) -- hypothetical bias, the divergence between willingness to pay and willingness to accept, and the lack of scope effects -- went largely uncontested. It is to these claims that we respond in this reflection.

We provide counterarguments to these claims not with the intent to convince the reader that the debate over contingent valuation is settled in favor of the method, but rather to ensure that non-specialist economists and others recognize that the intellectual debate over contingent valuation is ongoing and that there remains plenty of research to be done by the intellectually curious, something we would hope even a skeptic would embrace. In short, we fundamentally disagree with, and more importantly provide evidence that contradicts, Hausman's conclusion

[2] A comprehensive bibliography of over 7,500 scholarly papers and studies on contingent valuations is provided by Carson (2011). The vast majority of these papers have been written during the eighteen years since the initial *Journal of Economic Perspectives* symposium on Contingent Valuation in 1994, where Hausman first pronounced upon the method. Hausman himself has published no original contributions to the body of basic research concerning CVM.

that “despite all the positive-sounding talk about how great progress has been made in contingent valuation methods, recent studies by top experts continue to fail basic tests of plausibility (p. 54).” Instead, we find hope in recent studies by respected experts that contingent valuation can indeed provide plausible value estimates as well as information on market extent and public opinion and, as a consequence, should be one (but not the only) tool at the disposal of public policy decision makers. This remains especially true in contexts where important components of the benefits of a proposed public policy lie beyond the reach of evidence from existing markets. Yes, it is harder, sometimes much harder, to measure social benefits of public policies in these contexts, but this does not diminish the imperative that we do so if economic criteria are to be honored in social decision-making.

Three long-standing problems continue to exist?

A. Hypothetical Bias

Hausman examines three main issues which he believes continue to plague stated preference studies. The first issue he discusses is hypothetical bias, where “what people say is different from what they do” (p.44). Most hypothetical bias studies, whether in environmental valuation, or marketing, or public polling, use some form of the stylized testable hypothesis that stated preference responses are identical to responses to analogous questions when money or some other real outcome is at stake. If the hypothesis is rejected, the stated preference study is viewed to be flawed. While we respond more specifically below to Hausman’s specific criticisms, we note that despite issues of dealing with the very hypothetical nature of intended behavior questions, hypothetical questions are regularly used to evaluate real world decisions outside of the realm of environmental valuation. Marketers frequently utilize hypothetical

market situations to assess the demand for new products or the expansion of existing markets beyond traditional demographics, or to assess price changes beyond the realm of market experience. Pollsters regularly rely on hypothetical questions of intended voting behavior to predict election outcomes. Despite the widespread use of hypothetical questions by private parties with real stakes and the commitment of real resources on the outcomes of such surveys, discussions of hypothetical bias draw special ire when placed in the context of benefit cost analysis and the potential for use in public policy decisions.

Hypothetical bias has been examined *ad nauseum* in numerous empirical studies, but two recent papers by Carson and Groves (2007, 2011) have fundamentally changed researchers' thinking about hypothetical bias. Carson and Groves show that without a closer examination of the incentive structure of a particular stated preference survey, we cannot actually predict whether distortions due to hypothetical bias should or should not occur. Further, to be able to make a prediction about hypothetical bias in the first place, the respondent must view his responses as being consequential. That is, the respondent must believe that his responses to the survey will affect something that he cares about. If he does not care about the outcome, then hypothetical bias may or may not occur, but we have no basis in economic theory to make predictions about it. For an inconsequential survey, *any* response to a question (in particular a stated preference question) gives the respondent the same expected utility, precisely because his responses have no effect on anything he cares about. On the other hand, for a consequential survey, we may or may not predict hypothetical bias depending upon the survey's incentive structure.

Hausman cites three studies which find that "intentions to purchase new products" are overstated: Jamieson and Bass (1989), Hsiao et al. (2002), and Morwitz et al. (2007). However, as noted by Carson and Groves (2007, p.188), we should typically expect stated intentions to purchase new products to overstate true intentions. The logic is that by choosing "yes" (I would buy this product), the respondent increases the likelihood of having the product available for possible future purchase, under the reasonable assumption that the survey is being used to determine whether the new product should be brought to market. Utility is derived from an expansion of the individual's future choice set. The Jamieson and Bass (1989) study and the Hsiao et al. (2002) study actually involve products which are already available, but are relatively new to the market.[3] In these cases, it is unclear what exactly the incentives are for the respondent, since responses do not influence whether the product is brought to the market and thereby do not expand the available set of choices. As Morwitz et al. (2007) point out, purchase intention studies concerning existing products are often used by manufacturers and suppliers when they need to make decisions about "whether to increase or reduce production levels, whether to change the size of the sales force, and whether to initiate a price change" (p.347).[4]

---

[3] The Morwitz et al. (2007) study is a meta-analysis, and it is unclear whether their definition of a "new" product could include products already on the market but which are not yet particularly familiar to potential consumers, or only products that are not yet available for purchase.

[4] This highlights the incongruous argument that stated preference surveys are not good enough for government work when they do appear to be good enough to support the efforts of profit maximizing business firms.

However, it is not clear in this case exactly how responses will affect something the respondent cares about (i.e. product price). Thus we should not be surprised by the findings of these studies (and others cited therein) that stated purchase intentions for existing products have some predictive ability but that it is not particularly strong.[5] It is perhaps surprising that Hausman did not more carefully consider the incentive structure of the surveys used in these cited studies because he points out (p.45) that one possible cause of hypothetical bias is that the survey does not explicitly state how the survey responses will be used.

As a counter-example, consider the role of stated preference methods when they are used to expand the domain of preferences beyond the support of current experience. We can cite at least two examples where stated preference estimates of prospective demand under unexperienced future conditions have been shown to predict with considerable accuracy revealed behavior when the future conditions actually materialize. Grijalva et al. (2002) conduct a predictive validity test of rock climbing trip behavior. Respondents are surveyed about their current revealed preference trip behavior and their stated preference behavior under future access conditions. Following the realization of the hypothetical scenarios, respondents are surveyed

[5] The issue is further muddled when moving from market goods to nonmarket goods as is the case with most environmental applications of CVM surveys. We note with curiosity Hausman's selective choice of a market good case study—cable TV subscriptions—to illustrate some of his criticisms of CVM. It is not clear whether this choice is intended to strengthen the case against CVM (i.e. 'it doesn't even work for market goods') or yet another example of the selective nature of the evidence upon which Hausman bases his critique.

again. With hypothetical closure of rock climbing areas, stated preference rock climbing trips fall. When the areas are actually closed, actual trips differ in the expected direction and by similar magnitudes. Similarly, Whitehead (2005) conducts a predictive validity test of hurricane evacuation behavior. Respondents are surveyed about their revealed preference evacuation behavior after low-intensity storms and about their stated preference behavior after both hypothetical low-intensity and hypothetical high-intensity storms. Two hurricanes followed the survey and respondents were surveyed again to determine their actual behavior. Models using revealed and stated preference evacuation data forecast revealed preference behavior with prediction error of less than 20%.

One of the four key areas of future stated preference research noted by Kling et al. (2012) is the need for validity tests on consequential surveys. They cite Vossler and Evans (2009) and Landry and List (2007), who find no hypothetical bias when responses are consequential, and Vossler and Poe (2011), who conduct four consequential lab experiments and also demonstrate a lack of hypothetical bias. Three other recent studies that support the consequentiality paradigm are Vossler, Doyon and Rondeau (2012) in a field experiment context, and Bulte et al. (2005), Herriges et al. (2010), and Vossler and Watson (forthcoming) in a field survey context. The Carson and Groves paradigm for assessing the predictive ability of stated preference surveys has been extremely influential.[6] However, the concept is relatively new and so indeed more research is needed to examine whether inconsequential surveys suffer from *unexpected* degrees of hypothetical bias.

[6] Carson and Groves (2007) paper has been cited 502 times according to Google Scholar, as of the time of this writing.

In the absence of true consequentiality, two approaches to hypothetical bias mitigation have become popular (Loomis 2011). Some researchers find that the divergence between hypothetical and actual willingness pay is mitigated or eliminated by providing additional instructions to respondents explicitly encouraging them to treat the hypothetical scenario as if an actual monetary transaction were taking place (Cummings and Taylor 1999, List 2001). Still more researchers advocate for various forms of 'scenario adjustment' including controlling for and netting out respondent admitted departures from the conditions described in the offered choice scenarios (Cameron and DeShazo 2013) or taking into account the level of certainty respondents have that they would actually make the same decision if the choice was real (Champ and Bishop 2001, Blumenschein et al. 2007). Adjusting for certainty produces "hypothetical" results that are often close to "real" values and these alternative levels of WTP allow for sensitivity testing when CVM estimates are used in benefit-cost analysis (BCA) which leads us to assert that a range of numbers that likely includes accurate values is certainly better than no number.

B. Willingness to Pay and Willingness to Accept

Researchers have been troubled since the onset of stated-value elicitation surveys by the seemingly simple finding, perhaps obvious to some, that compensation demanded (willingness to accept) regularly exceeds compensation paid (willingness to pay). Based on this common WTA/WTP gap, researchers have split along two lines in their conclusions. Some dismiss contingent valuation based on the naïve premise that the gap represents a violation of basic economic theory and thus invalidates value elicitation surveys. Others remain curious as to the cause of the disparity and wonder whether the gap can be explained either within the neoclassical

framework, or by necessary extensions to this framework that draw upon psychological insights from behavioral economics. In dismissing contingent valuation, Hausman claims that "Basic economic theory suggests that [questions phrased to elicit willingness to pay to avoid a negative outcome and questions phrased to elicit willingness to accept the negative outcome] should give (approximately) the same answer…" Indeed, using income elasticity estimates from over 200 published studies comparing WTP and WTA, McConnell and Horowitz (2003) "conclude that the ratio WTA/WTP is too high to be consistent with neoclassical preferences." However, they remain open to the possibility that neoclassical preferences may not always be an appropriate assumption, noting that their results "should help in developing explanations and alternative models, since income effects are such a prominent part of economic models such as choice under uncertainty." (p. 544)

While Hausman's claim may be true in terms of the "basic economic theory" that prevailed prior to the discovery of the WTP/WTA gap, he ignores an ongoing literature including contributions from those who have been curious about whether there is an explanation that provides both an updated theoretical framework and experimental evidence that the gap can be explained within a slightly more general version of basic economic theory—although the debate remains unsettled.

One illustrative strand of this literature (but by no means the only set of contributions) begins with Hanemann's (1991) *American Economic Review* piece in which he lays out a neoclassical explanation for the WTP-WTA gap. Hanemann notes that the difference between WTP and WTA, or more precisely the difference between Hicksian compensating and equivalent variations, hinges on the ratio of the income elasticity to the Allen-Uzawa elasticity of

substitution. Hanemann goes on to argue that it is likely that the elasticity of substitution for unique goods, such as those often valued in contingent valuation surveys, is likely to be particularly small, thus increasing the expected difference between neoclassical compensating and equivalent variation. Shogren et al. (1994) use a series of market and nonmarket experiments to test Hanemann's neoclassical substitution hypothesis and conclude: "Our experimental results support [Hanemann's] argument that the degree of substitutability between goods may drive the difference between WTA and WTP measures of value" (p. 266). Revisiting the experiments of Shogren et al. (among others), Morrison (1998) finds that "even when allowing respondents to learn through repeated trials, controlling as much as possible for Hanemann's substitutability argument…the result of WTA exceeding WTP not only remains, but remains strongly significant" (p. 193). Shogren and Hayes (1997) counter "that the WTP-WTA disparity depends on the auction institution, not on a deviation from neoclassical rationality as suggested by the endowment effect" (p. 243).[7] Hausman's selective oversight[8] of such work

[7] Dependence upon the auction institution again highlights the importance of understanding incentives.

[8] We debated Hausman's strategy in choosing not to include any reference to well-known strands of the literature that are contrary to his conclusions. In the end, we decided that it is not our place to infer intent and we have chosen the most charitable interpretation: mere oversight. However, we do find it objectionable that well-known and respected works such as those of Hanemann (973 Google Scholar citations) and Shogren et al. (565 Google Scholar citations) are so readily and summarily dismissed under the terse label of "Various efforts."

allows him to dismiss contingent valuation on the grounds of violations of basic economic theory noting that:

> Various efforts have been made to extend the neoclassical framework in a way that rationalizes the gap. Proponents of contingent valuation have attempted rationalizations of these differences, but have not overcome the findings of Diamond and Hausman (1994) or the results of Milgrom (1993). Both papers demonstrate that the attempts to rationalize the well-recognized and persistent disparity between willingness to pay and willingness to accept fail as a matter of economic theory and observed empirical outcomes (p 47-48).

We find it troubling that Hausman fails to acknowledge the deep and continuing debate in the literature that is in large part responsive to his 1994 work with Diamond, yet has chosen to change his own opinion of contingent valuation from "dubious" to "hopeless." In a recent exploration of the impacts of reference dependence (i.e. endowment effects), Knetsch (2010) notes in a particularly relevant footnote: "One somewhat unfortunate consequence of the many reports of large disparities between WTA and WTP valuations in contingent valuation surveys, is that many, especially environmental economists, have taken these disparity findings to be the result of a problem of the inadequacy of contingent valuation methods to accurately measure people's values rather than taking it more seriously as one of people valuing losses more than gains" (p. 180). Knetsch goes on to argue that the appropriate measure of value (WTP or WTA) will depend on the initial assignment of property rights, perceived or otherwise. This creates potential difficulties in the interpretation of benefit-cost analysis (see Sugden 2005) for a proposal for dealing with preference anomalies such as the endowment effect within the context

of benefit cost analysis). However, the prevalence of a WTP-WTA gap, or the existence of endowment effects, is not a fatal flaw in the contingent valuation method in and of itself.

Based on our own reading of the literature, we are not ready to close the WTP-WTA debate. And apparently, we are not alone:

> A subtle controversy exists in the literature. At issue is the existence and interpretation of a possible gap between willingness to pay ("WTP") and willingness to accept ("WTA")…. In spite of the enthusiastic interpretations of the WTP-WTA gap as a fundamental feature of human preferences…in fact there is no consensus about whether the literature, considered in its entirety, supports such interpretations. (Plott and Zeiler, 2005, p. 531).

C. Scope Test

Hausman argues that contingent valuation studies tend to fail the "scope test," and those that pass the test fail to pass it "adequately." At the same time, he acknowledges that "We do not know how large scope effects should be" (p. 48). We share Hausman's ignorance about what would constitute scope effect "adequacy." The simplest model of willingness to pay, a difference in expenditure functions with changes in quality or quantity, can be used to show that willingness to pay is nondecreasing in quality or quantity (Whitehead, Haab and Huang 1995). The size of scope effects may be limited by diminishing marginal utility or substitution among nonmarket goods or between nonmarket and market goods. For example, Rollins and Lyke (1998) find that people do distinguish between existence goods of different scope, but due to diminishing marginal valuations for larger scope goods, estimated differences between the values of larger

scope goods may be negligible unless sample sizes are adjusted accordingly. Additionally, Amiran and Hagen (2010) develop a formal model of scope with bounded utility functions and show that relatively small scope effects are not inconsistent with economic theory. The size of the scope effect in this model depends explicitly on the substitutability between market and nonmarket goods.

Hausman considers the "Diamond-Hausman adding up test," developed by Diamond (1996) to be the definitive scope test. To illustrate the adding up test and other scope tests, suppose there are three samples of CVM respondents:

	Sample		
WTP Question	1	2	3
First	A	B | A	A + B
Second	A + B		

Sample 1 is presented a choice or WTP question for A and then a second question for A + B. Sample 2 is presented a choice or WTP question for B, given that A is already provided as part of the consumption bundle and payment has already been extracted (labeled as B | A). Sample 3 is presented with a single choice or WTP question for the joint provision of A + B. Success in an adding up test occurs when the sum of WTP from the first question posed to Sample 1 and Sample 2 is equal to the WTP of Sample 3. The so-called external (i.e., split-sample) scope test is a comparison of values across the first question for Samples 1 and Sample 3. The so-called internal scope test is a comparison of the first and second WTP values for Sample 1.

Hausman highlights the evidence provided by Desvousges, Mathews and Train (2012) who review 109 studies that report a scope test and find that 36% pass the test, 15% fail the test and the rest have mixed results. However, Desvousges, Mathews and Train (2012) find only a few studies that provide enough information to conduct an adding up test. The axiom of revealed

preference (on the part of CVM researchers) suggests that internal and external scope tests are preferred to the adding up test. One reason for this is that the adding up test asks respondents in Sample 2 to assume that B has been provided by the policy process and payment has already been extracted. This more-complicated scenario substantially increases the cognitive difficulty of the valuation task (Hanemann, 1994).[9]

Given the absence of any explicit adding up test for most of the 109 scope-effect papers in the literature, and the fact that the adding up test is not the only valid test for adequate sensitivity to scope, it is necessary to examine more closely the types of scope tests that are routinely conducted. When considering whether an individual study passes a scope test adequately, any meta-analysis should consider characteristics of the study that might drive scope-testing results. For example, the type of valuation question might play a role. Carson and Groves (2007) argue that a consequential referendum is more likely to produce scope effects. Smith and Osborne (1996) conduct a meta-analysis of studies focused on the value of changes in visibility at U.S. parks and find clear responsiveness of CVM estimates to scope. They warn, however, that scope predictions are sensitive to theoretical assumptions imposed during model estimation. Richardson and Loomis (2009) find that the size of the population change is statistically and economically significant in a meta-analysis of 67 willingness to pay estimates of threatened and endangered species allocation. In two specifications the scope elasticity of willingness to pay is not statistically different from one. Ojea and Loureiro (2011) undertake a

[9] Also, the adding up test requires three subsamples, increasing the cost of a study substantially It is apparently not clear to many researchers whether the benefit of conducting this test, relative to the more straightforward external scope test, is worth the cost.

meta-analysis of 355 biodiversity contingent valuation studies with a focus on the issue of scope-test satisfaction. They find that scope effects are more likely to be found when scope is measured absolutely instead of relatively.[10]

Other Issues

A. The Use of Experts

In response to questions about how to handle nonuse values in decision-making without recourse to stated preference surveys, Hausman advocates the use of experts: "[P]ublic policy will do better if expert opinion is used to evaluate specific projects, including non-use value…" (p.44). This argument is also made in the original Diamond and Hausman (1994) paper.

The first question this approach raises is: what kinds of experts these should be? It is hard to imagine that non-economists will be better equipped than economists to assess passive use value, or even to be familiar with the concept that it is appropriate to consider passive use values in the first place. For ecological services, for example, are we to substitute ecologist's subjective opinions about "intrinsic value" for the instrumental values required for benefit-cost analysis? How are these intrinsic values to be monetized? The second question is: why would experts be better than non-experts at determining the monetized nonuse values accruing to society? After all, decision-makers are usually interested in the preferences of the general

[10] Amiran and Hagan's (2010) scope elasticity might be a useful way of summarizing the adequacy of scope effects in these studies.

public, so it is difficult to see how experts, as a highly selected sample from the general population, should be better informed about social values than non-experts, that is, a representative sample from the general public stating their own preferences directly. If we are discussing whether a given environmental project does or does not deliver a specific increment of physical ecological function, then generally an expert is better informed than the average person. But value is distinct from quantity. We might expect that the average person knows his own preferences better than the expert, conditional on full information.

The more practical question, and really the most fundamental one, is how can these alleged experts determine nonuse values without having first conducted at least an implicit stated preference study as the basis for their expertise in judging societal nonuse values? It is well known that the distinctive characteristic ofpassive use value is that there is no observable related behavior from which to judge people's willingness to make tradeoffs for the benefit in question. The role of the stated preference survey, explicitly, is to create such a situation in which we can observe this willingness. Hausman's argument is that because stated preference surveys cannot reliably elicit information on preferences, then an expert panel is preferred. The problem is that no expert can claim to know the preferences of a particular population without having first made some informal observations about people's willingness to make tradeoffs.[11] Thus, the actual

[11] Although in a different context, Smith's (2010) comment as to the prospect of relying on expert panels is worth noting: "This suggests that all we need to do is convene an expert panel that will somehow come to a consensus. I have lost count of how many conferences and conference sessions have been charged with this task and have failed to arrive at a consensus value."

choice of passive use value estimates we are confronted with is not, as Hausman would have us believe, between those of a wholly unreliable stated preference survey and those of a reliable expert measure of passive use value. Instead, it is between the values advocated by a statistically uninformed, highly self-selected and potentially biased "expert" panel, and those implied by a stated-preference survey that is possibly flawed, but nevertheless broadly and systematically informed (in the sense of actually being based on direct input from a large representative sample drawn from the population of interest).

If Hausman is instead arguing only that the possible *existence* of passive use values should be acknowledged by these experts but that they should go no further in quantifying the magnitudes of these values, then his advice is even less useful. Mere acknowledgement of passive use values cannot be incorporated into the decision process in any systematic and meaningful way. Given the mandate in the U.S. for formal benefit-cost analyses for significant policies or regulations (Smith 1984), failure to monetize the whole category of passive use benefits creates the risk that the value of these benefits will be defaulted to zero in arriving at the bottom line of a net benefits calculation. When expert opinions about values are substituted in lieu of stated-preference evidence, it is inevitable that the process of high-stakes policy decision-making will produce challenges to the basis for their expertise about societal values. Evidence-based policy-making requires just that: evidence. The comparison of all the costs of an action to all the benefits of an action is the decision criterion that separates economics from other disciplines. Giving up on the proper measurement and incorporation of some (potentially sizeable) benefits means giving up on proper BCA. It has been acknowledged repeatedly that BCA should not necessarily be the only criterion used in decision-making, but it is the economic one. And although it is always necessary to determine how much weight BCA actually ends up

having in any decision process, Hausman has no greater authority than anyone else in determining whether passive use values should or should not be formally incorporated into the decision process. *B. Altruism and BCA*

Hausman also argues that values deriving from altruism should not be incorporated into BCA, and that it is difficult to identify and exclude altruistic values from stated preference studies. Bergstrom (1982) indeed showed that the rule for the Pareto efficient level of a public good (the Samuelson rule) is the same in the presence or absence of altruism. This implies that we require purely "selfish" values - excluding altruistic values - for determining optimal outcomes. However, one of the assumptions made by Bergstrom was that the considered change be small and that the economy be already near the Pareto efficient level of the public good. Flores (2002) showed that for discrete (large) changes in the public good, such as are often the subject of CV studies, the selfish benefit-cost test is sufficient but not necessary for making an efficient decision. That is, it is possible to fail to accept a good proposal (from a BCA point of view) using only selfish values. Bergstrom (2006) agrees with this conclusion.

C. Unstable and Inconsistent Preferences

A recurring theme throughout Hausman's paper is the assertion of unstable preferences as an indictment of CVM methods. For example:

> "…the primary argument that is relevant for thinking about contingent valuation methods as a whole is that the answers from such studies are unstable and inconsistent, invented for the moment of the survey, and cannot be treated as preferences in the sense that economist understand that term." (p. 53)

However, although the papers in the *JEP* symposium focus mostly on the validity of the CVM (e.g. hypothetical bias and scope), the assertion that CVM responses are "unstable and inconsistent" is a statement about their reliability, not their validity. Yet Hausman does not properly support his assertion about the unreliability of CVM with references to the literature.

Validity is the extent to which a valuation method generates a measure that is unbiased, that is, an estimate centered around the true value, if it were known. Validity is difficult to demonstrate when valuing nonmarket goods and services because, by their nature, their "true" value is unknown. A valid method for estimating these values is thus one that attempts to provide an unbiased estimate around an unknown and unobservable quantity.

Reliability is the extent to which a valuation method consistently generates the same measure. Reliability tests focus on the within and across study *variation* in estimates rather than the ability of studies to produce *unbiased* estimates of value. The lower the variability in estimates, the more consistent and the less influenced by researcher decisions are the estimates. High variability allows seemingly innocuous decisions by practitioners to significantly influence results in one direction or another, so unreliable results cannot be defended against the criticism that the researcher may be determining the result.

Most CVM research finds that willingness to pay values have some degree of stability over time. Test-retest, or temporal, reliability involves conducting more than one CVM survey with time between surveys with the same or a different sample of respondents. If the magnitude of willingness to pay is consistent across time then willingness to pay is considered temporally reliable (e.g., Carson et al., 1997). However, a difference in willingness to pay over time does not necessarily indicate unreliable results. If willingness to pay changes over time in response to

changing factors that would be expected to affect willingness to pay, such as bubbles in housing markets or a significant recession, for example, then the researcher may still conclude the results are reliable (e.g., Whitehead and Hoban, 1999). Overall, it is important that CVM studies demonstrate some degree of both validity and reliability.

We are also not convinced that unstable or constructed preferences are unique to CVM. The formation of preferences has to begin somewhere. The formation of preferences on the spot in a survey does not necessarily invalidate them for those who have no prior preferences or experience. When confronted with unfamiliar but interesting products in real markets, people also sometimes make impulsive, spontaneous, or ill-advised purchases that they may subsequently regret. Perhaps the preferences elicited by CVM are in their "infancy" stage, but this can be true for market goods and services as well (new products are introduced all the time). It is also probably safe to say that (almost) everyone has well-formed preferences for money, so one can at least make a “no-purchase” decision based on well-defined preferences for "all other goods" relative to the proposed public good with unfamiliar qualities (expressed by a vote against the provision of that good in a consequential referendum).

Conclusions

Hausman (2012) makes an effort to debunk the CVM once and for all, thereby removing it from the research agenda and from use in decision-making about public policy. We argue instead that (1) the existence (or nonexistence) of hypothetical bias continues to raise important research questions about the incentives guiding survey responses and preference revelation in real as well as hypothetical settings, and contingent valuation can help answer these questions, (2) the WTP-WTA gap debate is far from settled and the debate raises important research

questions about the future design and use of benefit cost analyses in which contingent valuation will undoubtedly play a part, and (3) CVM studies do, in fact, tend to pass a scope test and there is little support for the argument that the adding up test is the definitive test of CVM validity. It is our conclusion that Hausman's "selective review" of the literature is demonstrably biased in favor of studies that are negative towards the CVM. As a result, Hausman's conclusion is preordained. We are in complete agreement with Carson (2012) who concludes "the time has come to move beyond endless debates that seek to discredit contingent valuation and to focus instead on making it better."

References

Amiran, Edoh Y., and Daniel A. Hagen, The scope trials: Variation in sensitivity to scope and WTP with directionally bounded utility functions, Journal of Environmental Economics and Management 59:293–301, 2010.

Bergstrom, Theodore C. When is a man's life worth more than his human capital? In: The Value of Life and Safety: Proceedings from a Conference Held by the Geneva Association, M.W. Jones-Lee (ed.). North Holland Press, Amsterdam: 3-25, 1982.

Bergstrom, Theodore C. Benefit-cost in a benevolent society. American Economic Review 96(1): 339-51, 2006.

Blumenschein, Karen, Glenn C. Blomquist, Magnus Johannesson, Nancy Horn, and Patricia Freeman, Eliciting willingness to pay without bias: Evidence from a field experiment. The Economic Journal, 118: 114–137, 2008.

Bulte, Erwin, Shelby Gerking, Jonh A. List, and Aart de Zeeuw, The effect of varying the causes of environmental problems on stated WTP values: evidence from a field study, Journal of Environmental Economics & Management 49: 330-42, 2005.

Cameron, Trudy Ann and J.R. DeShazo, Demand for health risk reductions, Journal of Environmental Economics and Management, 65: 87-109, 2013.

Carson, Richard T. Contingent Valuation: A Comprehensive Bibliography and History. Edward Elgar, Cheltenham, UK, 2011.

Carson, Richard T., Contingent valuation: A practical alternative when prices aren't available, Journal of Economic Perspectives 26(4):27-42, 2012.

Carson, Richard T., Nicholas E. Flores, Norman F. Meade, Contingent valuation: controversies and evidence, Environmental and Resource Economics, 19(2):173-210, 2001.

Carson, Richard T., W. Michael Hanemann, Raymond J. Kopp, Jon A. Krosnick, Robert C. Mitchell, Stanley Presser, Paul A. Ruud, and V. Kerry Smith, Temporal reliability of estimates from contingent valuation, Land Economics 73(2):151-163, 1997.

Carson, Richard T., and Theodore Groves, Incentive and informational properties of preference questions, Environmental and Resource Economics 37:181-210, 2007.

Carson, Richard T. and Theodore Groves, Incentive and information properties of preference questions: Commentary and extensions, in Jeff Bennett, ed., International Handbook of Non-Market Environmental Valuation (Northampton, MA: Edward Elgar, 2011).

Champ, Patricia A. and Richard C. Bishop, Donation payment mechanisms and contingent valuation: An empirical study of hypothetical bias, Environmental and Resource Economics 19(4):383-402, 2001.

Cummings, Ronald G. and Laura O. Taylor, Unbiased value estimates for environmental goods: A cheap talk design for the contingent valuation method, American Economic Review 89(3):649-665, 1999.

Desvousges, William, Kristy Mathews and Kenneth Train, Adequate responsiveness to scope in contingent valuation, Ecological Economics 84:121-128, 2012.

Diamond, Peter, Testing the internal consistency of contingent valuation surveys, Journal of Environmental Economics and Management, 30(3):337-347, 1996.

Diamond, Peter A., and Jerry A. Hausman, Contingent valuation: Is some number better than none? Journal of Economic Perspectives, 8(4):45-64, 1994.

Flores, Nicholas E., Non-paternalistic altruism and welfare economics, Journal of Public Economics 83:293-305, 2002.

Grijalva, Therese C., Robert P. Berrens, Alok K. Bohara and W. Douglass Shaw, Testing the validity of contingent behavior trip responses, American Journal of Agricultural Economics. 84(2):401-414, 2002.

Hanemann, W. Michael, Willingness to pay and willingness to accept: How much can they differ? American Economic Review 81(3):635-647, 1991.

Hanemann, W. Michael, Willingness to pay and willingness to accept: How much can they differ? Reply, American Economic Review 93(1):464-464, 2003.

Hanemann, W. Michael, Valuing the environment through contingent valuation, Journal of Economic Perspectives 8(4):19-43, 1994.

Hausman, J.A., Contingent Valuation: A Critical Assessment, North Holland, 1993.

Hausman, Jerry, Contingent valuation: From dubious to hopeless. Journal of Economic Perspectives 26(4):43-56, 2012.

Herriges, Joseph, Catherine L. Kling, Chih-Chen Liu, and Justin Tobias, What are the consequences of consequentiality?, Journal of Environmental Economics & Management 59: 67-81, 2010.

Hsiao, Cheng, Baohong Sun, and Vicki G. Morwitz, The role of stated intentions in new product purchase forecasting, Advances in Econometrics 16:11-28, 2002.

Jamieson, Linda F., and Frank M. Bass, Adjusting stated intention measures to predict trial purchase of new products: A comparison of models and methods, Journal of Marketing Research 26(3): 336-45, 1989.

Kling, Catherine L., Danief J. Phaneuf, and Jinhao Zhao, From Exxon to BP: Has some number become better than no number? Journal of Economic Perspectives 26(4):3-26, 2012.

Knetsch, Jack, Values of gains and losses: Reference states and choice of measure, Environmental and Resource Economics, 46(2):179-188, 2010.

Krutilla, John V., Conservation Reconsidered, American Economic Review 57(4):777-786, 1967.

Landry, Craig, and John List, Using ex ante approaches to obtain credible signals for value in contingent markets: Evidence from the field, American Journal of Agricultural Economics 89:420-429, 2007.

List, John A., Do explicit warnings eliminate the hypothetical bias in elicitation procedures? Evidence from field auction experiments, American Economic Review, 91(5):1498–1507, 2001.

Loomis, John B., What's to know about hypothetical bias in stated preference valuation studies, Journal of Economic Surveys, 25:363–370, 2011.

McConnell, Kenneth E., and John K. Horowitz, Willingness to accept, willingness to pay and the income effect. Journal of Economic Behavior and Organization 51: 537-545, 2003.

Morrison, Gwendolyn C., Understanding the disparity between WTP and WTA: Endowment effect, substitutability, or imprecise preferences? Economics Letters 59(2):189-194, 1998.

Morwitz, Vicki G., Joel H. Steckel, and Aloka Gupta, When do purchase decisions predict sales? International Journal of Forecasting 23(3):347-64, 2007.

Ojea, Elena, Mario L. Loureiro, Identifying the scope effect on a meta-analysis of biodiversity valuation studies. Resource and Energy Economics 33:706-724, 2011.

Plott, Charles R., and Kathryn Zeiler, The willingness to pay–willingness to accept gap, the endowment effect, subject misconceptions, and experimental procedures for eliciting valuations, American Economic Review 95(3): 530-545, 2005.

Portney, Paul R., The contingent valuation debate: Why economists should care, Journal of Economic Perspectives 8(4):3-17, 1994.

Randall, Alan, Berry Ives, Clyde Eastman, Bidding games for valuation of aesthetic environmental improvements, Journal of Environmental Economics and Management 1(2:132-149, 1974.

Richardson, Leslie, and John Loomis, The total economic value of threatened, endangered and rare species: An updated meta-analysis, Ecological Economics 68:1535-1548, 2009.

Rollins, Kimberly and Audrey Lyke, The case for diminishing marginal existence values, Journal of Environmental Economics and Management 36(3): 324-44, 1998.

Scopus search. SciVerse Scopus. Accessed Nov. 25, 2012. <www.scopus.com>

Shogren, Jason F. and Dermot J. Hayes, Resolving differences in willingness to pay and willingness to accept: Reply, American Economic Review 87(1):241-244, 1997.

Shogren, Jason F., Seung Y. Shin, Dermot J. Hayes and James B. Kliebenstein, Resolving differences in willingness to pay and willingness to accept, American Economic Review 84(1):255-270, 1994.

Smith, V. Kerry, Environmental Policy under Reagan's Executive Order: The Role of Benefit-Cost Analysis, University of North Carolina Press: Chapel Hill, 1984.

Smith, V. Kerry, Reflections – legacies, incentives, and advice, Review of Environmental Economics and Policy 4(2): 309-24, 2010.

Smith, V. Kerry and Laura L. Osborne, Do contingent valuation estimates pass a "scope" test? A meta-analysis, Journal of Environmental Economics and Management 31(3): 287-301, 1996.

Sugden, Robert, Coping with preference anomalies in cost–benefit analysis: A market-simulation approach, Environmental and Resource Economics 32(1):129-160, 2005.

Vossler, Christian, and Mary Evans, Bridging the gap between the field and the lab: Environmental goods, policy maker input, and consequentiality, Journal of Environmental Economics and Management 58:338-345, 2009.

Vossler, Christian, and Gregory Poe, Consequentiality and contingent values: An emerging paradigm, in The International Handbook on Non-Market Environmental Valuation, ed. Jeff Bennett. Northhampton, MA: Edward Elgar, 2011.

Vossler, Christian A., Maurice Doyon and Daniel Rondeau, Truth in consequentiality: Theory and field evidence on discrete choice experiments, American Economics Journal: Microeconomics 4(4):145-171, 2012.

Vossler, Christian A., and Sharon B. Watson, Understanding the consequences of consequentiality: Testing the validity of stated preferences in the field, Journal of Economic Behavior and Organization, forthcoming.

Whitehead, John C., Environmental risk and averting behavior: Predictive validity of jointly estimated revealed and stated behavior data, Environmental and Resource Economics. 32:301-316, 2005.

Whitehead, John C., Timothy C. Haab and Ju-Chin Huang, Part-whole bias in contingent valuation: Will scope effects be detected with inexpensive survey methods? Southern Economic Journal, 65(1):160-168, 1998.

Whitehead, John C., and Thomas J. Hoban, Testing for temporal reliability in contingent valuation with time for changes in factors affecting demand, Land Economics 75(3), 453-465, 1999.