Ecological Economics 131 (2017) 475–484

Contents lists available at ScienceDirect

Ecological Economics

journal homepage: www.elsevier.com/locate/ecolecon

Using the Delphi method to value protection of the Amazon rainforest

Jon Strand [a,*], Richard T. Carson [b], Stale Navrud [c], Ariel Ortiz-Bobea [d], Jeffrey R. Vincent [e]

[a] Inter-American Development Bank, 1300 New York Avenue, N.W., Washington, D.C. 20577, USA
[b] Department of Economics, University of California, San Diego, La Jolla, CA 92093, USA
[c] Department of Economics and Business, Norwegian University of Life Sciences, 1430 Ås, Norway
[d] Charles H. Dyson School of Applied Economics and Management, Cornell University, Ithaca, NY 14850, USA
[e] Nicholas School of the Environment, Duke University, Durham, NC 27708, USA

ARTICLE INFO

Article history:
Received 22 October 2015
Received in revised form 15 September 2016
Accepted 21 September 2016
Available online 12 October 2016

Keywords:
Amazon rainforest
Delphi method
Stated preference
Contingent valuation
Nonmarket valuation methods

ABSTRACT

Valuing global environmental public goods can serve to mobilize international resources for their protection. While stated-preference valuation methods have been applied extensively to public goods valuation in individual countries, applications to global public goods with surveys in multiple countries are scarce due to complex and costly implementation. Benefit transfer is effectively infeasible when there are few existing studies valuing similar goods. The Delphi method, which relies on expert opinion, offers a third alternative. We explore this method for estimating the value of protecting the Amazon rainforest, by asking more than 200 environmental valuation experts from 37 countries on four continents to predict the outcome of a contingent valuation survey to elicit willingness-to-pay (WTP) for Amazon forest protection by their own countries' populations. The average annual per-household values of avoiding a 30% forest loss in the Amazon by 2050, assessed by experts, vary from a few dollars in low-income Asian countries, to a high near $100 in Canada, Germany and Norway. The elasticity with respect to average (PPP-adjusted) per-household incomes is close to unity. Results from the Delphi study match remarkably well those from a recent population stated-preference survey in Canada and the United States, using a similar valuation scenario.

1. Introduction

The Amazon rainforest is the world's largest, and is widely recognized as a crucial natural resource for all of mankind. Most of its area (about 60%) is located in Brazil, with the remainder spread across eight other South American countries (collectively, the Amazon region). The average annual area deforested in the Amazon region during 2005–10 was nearly five times as large as the annual area deforested in Indonesia, which had the highest deforestation rate of any tropical country outside of the Amazon (Table 3 in FAO, 2010). Deforestation remains a serious concern in the Amazon region, despite an 80% reduction in the annual area deforested in the Brazilian portion between 1995 and 2005 and 2014 (Nepstad et al., 2014).

Surprisingly little is known about the global value of the Amazon rainforest, most reasonably measured by the willingness-to-pay (WTP) to protect the remaining forest, in spite of the conceptual framework for examining this issue laid out in Carson (1998). Two early contingent valuation (CV) studies, however, do shed some light on the issue. Kramer and Mercer (1997) conducted a random population survey of the U.S. population in 1995, to elicit WTP for protecting 5% of *global* rainforests, not specifically the Amazon rainforest. They found that the average U.S. household was willing to make a *one-time* payment of $21–$31 (1995 dollars) for this purpose. Horton et al. (2003) surveyed outdoor recreationists at a small number of recreation sites in Italy and the U.K. in 1999. They found evidence of much higher WTP, around $45 per household *per year* for a program to protect 5% of the Amazonian rainforest, and $60 for a 20%-protection program. Apart from differences in the populations sampled, these differences could reflect preference differences between Europe and the U.S.; a relatively higher value placed on protecting the Amazon than other tropical rainforests; or increasing public attention to, and support for, rainforest protection over the period in question.

A more reliable comparison is achieved from a recent national population WTP survey, Siikamäki et al. (2015), in Canada and the U.S. using essentially the same survey instrument as in this Delphi study. As discussed at more length in the final section of this paper, that survey gives higher per-household valuations than the two earlier surveys noted above; and values reasonably close to estimates obtained from these countries in the Delphi study presented here.

Conceivably, one could estimate global WTP to protect the Amazon rainforest by conducting a CV survey of a globally representative population sample or by sampling the populations of a large number of countries that collectively account for much of the world's population.

* Corresponding author.
E-mail address: jstrand1344@gmail.com (J. Strand).

http://dx.doi.org/10.1016/j.ecolecon.2016.09.028
0921-8009/

However, such a survey would be very costly. Benefit transfer (BT), to extrapolate from existing studies (Brookshire and Neill, 1992; Ready and Navrud, 2006; Wilson and Hoehn, 2006, can be an alternative to original valuation work; but is often unreliable when the base of studies upon which to draw is small, the correspondence between the good of interest and those valued in earlier studies is poor, the time separating the present and the early studies is long, and there are substantive differences between the population of interest and the populations sampled in earlier studies; all these are relevant issues here. We consider a third option: a Delphi study that relies on expert judgment to estimate WTP to protect the Amazon. We asked 216 environmental valuation experts from 37 countries outside South America to predict the outcome, expressed in mean and median WTP per household per year, of a hypothetical CV study on Amazon protection administered to their respective national populations. Experts participating in the study came from Europe (49 experts from 21 countries), the U.S. and Canada (82 experts), Australia and New Zealand (16 experts), and Asia (69 experts from 12 countries). The combined populations of these 37 countries account for more than 60% of the world's population, and about 70% of the global population outside of Latin America. Our experts also comprise a reasonably large subsample of researchers actively conducting nonmarket valuation studies.

The Delphi method was developed by the RAND Corporation during the 1950s and 60s, with key contributions by Dalkey (1967, 1969) and Dalkey and Helmer (1963). It has a long background and tradition as a management decision tool (Linstone and Turoff, 1975). The key elements are: (a) anonymous responses by experts to multiple rounds of formal questionnaires; (b) an exercise incorporating iterative, controlled feedback with respect to the information provided at each round; and (c) statistical summary of the group's responses. The approach was designed to minimize the influence of dominant individuals, group pressure, and irrelevant communication and to reduce (statistical) noise. By the early 1970s, hundreds of studies had appeared from around the world. After the mid-1970s, methodological development stalled, as the method was criticized as unscientific and its results speculative (Sackman, 1975). Rebuttals of the critique in the 1990s (e.g., Ziglio, 1996; Landeta, 2006) led to various new applications including Holtsapple and Joshi (2002) and Scholl et al. (2004) for knowledge management; Evans (1997) for pharma-economics; and Okoli and Pawlowski (2004) for e-commerce diffusion in Africa. The method is especially useful when it can produce information not readily obtainable in other ways.

There are few applications of the Delphi method to environmental valuation. Hufschmidt et al. (1983) seem to be the first to mention the method as suitable for such valuation, albeit with no reference to particular applications. Some examples of expert-based work on environmental or related topics still exist. Weitzman (2001) asked more than 2000 Ph.D.-level economists to state the appropriate discount rate for future climate-related damages. León et al. (2003) considered whether environmental values elicited through expert opinion could be used as the basis for benefit transfer. Their study dealt with outdoor recreation at national parks in Spain, and it compared experts' predictions to the results of actual CV studies at the sites. Interestingly, they found a high degree of consistency between experts' average valuations, and the outcome of a subsequent CV study, for valuing these parks among visitors; despite the fact that individual experts' valuations varied substantially. Roman et al. (2012) conducted an intensive study that involved only three experts concerning their assessments of the appropriate value of statistical life to use in the United States for valuing health damages due to air pollution. Two papers (Curtis, 2004; Scolozi et al., 2012) have used the approach for valuing biodiversity. Most similar to our study is a 1998 Delphi study applied to a cultural resource, the Fez Medina in Morocco (Carson et al., 2013). In both that study and ours, experts were asked to predict the outcome of a hypothetical CV survey of national populations of countries outside the one where the resource is found.

This study has two main objectives. The first is to learn about the application of the Delphi method to a global public good. Applying the method to a global public good requires a large group of experts drawn from multiple countries, but levels of expertise in environmental valuation vary across countries. Environmental valuation expertise is concentrated in higher income countries, but lower income countries account for more of the world's population. This leads us to examine the effects of different levels of expertise on experts' WTP predictions. We ask the question: what constitutes an "expert?" investigating this using various indicators of expertise. Note that Delphi studies for other global public goods would face similar issues of heterogeneity of expertise across countries.

The second objective is to use experts' WTP predictions for a preliminary assessment of global WTP to protect the Amazon rainforest and how it varies around the world. We also assess the relationship between the WTP predictions and per-capita national income. This relationship could be useful for BT purposes to predict WTP in countries not included in the study.

2. Design and Implementation of the Delphi Study

2.1. Overview

We implemented the Delphi study by email, in two rounds. In Round 1, we sent each expert a cover letter, which described the purpose and organization of the study, and a study booklet. The booklet provided background information on the Amazon rainforest, described the hypothetical CV study, and asked the expert questions related to the study (in particular, their WTP predictions) and the expert's experience with environmental valuation studies. The booklets were in English in all regions, and are available upon request. An effort was exerted to make them as similar as possible across the regions. Some changes to the order of information and the words used to describe it were made in the Asian booklet, on the basis of cognitive interviews with a small number of Asian economists. English language skills were weaker on average among the Asian experts than among the experts from the other regions.

The CV study was described as a survey of a representative sample of the population of the expert's home country. The CV scenario was described as a plan to protect the Amazon rainforest from further deforestation. Two variations on the scenario (i.e., two protection plans) were presented, differing in the extent of protection. The experts were asked to predict mean and median WTP for both plans: i.e., to predict the outcome of the CV study if it were actually implemented in their respective countries. We emphasized in the cover letter and booklet that all experts were asked to assess WTP for a representative sample of their countries' populations, not their own personal WTP.

In Round 2, mean and median predictions across the respective experts in a given country or region were reported back to those experts, who were then given an opportunity to adjust their predictions. National values were reported to the experts in the U.S., Canada, Australia, and New Zealand. In Europe and Asia, values were reported for country groups, shown in Table 1, with the low- and lower-middle-income groups combined for Asia due to the small number of experts from low-income Asian countries. For increased clarity in the Asian survey, we also provided the distribution of responses by broad WTP ranges in addition to mean and median values. We expected the national or regional summary information provided to the experts to draw their Round 2 responses toward the Round 1 summary statistics. An objective of a Delphi study, often implicit, is to achieve an outcome close to a group consensus if one appears to exists, while at the same time not unduly influencing participants to change their predictions if there are strongly held differences in beliefs.

2.2. The Delphi CV Scenario

The two protection plans differed in terms of the area of forest that would remain as of 2050. Under the more extensive Plan A, no additional forest would be lost by 2050. Under the less extensive Plan B, 12% of the current forest area would be lost by that time. Both plans were compared to a costless (in terms of incremental *protection* costs) "business as usual" (BAU) alternative, under which 30% of the current area would be lost by 2050. The BAU alternative and Plan B were based on deforestation scenarios presented in Soares-Filho et al. (2006). The study booklet included maps showing the 2050 area of forest and the area lost during 2012–2050 for both protection plans and the BAU alternative.

Biodiversity loss was presented as a major consequence of deforestation. The information provided to experts read:

> Along with this forest loss there are likely to be losses of species, some of which are found only in the Amazon. If nothing is done to slow the rate of deforestation in the Amazon, scientists estimate that 105 mammal species, out of 442 currently known to be found there, will under the Business as Usual alternative face a high risk of extinction by 2050. Eighty three (83) of these endangered species are found **only** in the Amazon. … A similar fraction (about 20%) of other animal species, such as birds and amphibians, will also in the same way be threatened. Under **Plan A,** which preserves all (100 %) of the current Amazon rainforest by 2050, none of these species would be lost by 2050. Under **Plan B,** which preserves 88% of the current Amazon rainforest by 2050, 41 of these species would face a high likelihood of extinction by 2050.

Table 1
Distribution of experts by region and country.

Region (no. experts)	*Country (no. experts)*	*GDP* per capita *(PPP), US$/year*
North America (82)	U.S. (71)	50,900
	Canada (11)	41,500
Europe: Nordic countries (11)	Denmark (2)	41,500
	Finland (3)	38,000
	Norway (3)	62,900
	Sweden (3)	41,800
Europe: Northern and Central (18)	Austria (2)	43,100
	Belgium (1)	39,500
	Germany (3)	42,000
	Ireland (2)	42,900
	Netherlands (3)	42,500
	Switzerland (2)	51,300
	U.K. (5)	34,700
Europe: Southern (12)	France (3)	36,100
	Greece (2)	25,200
	Italy (4)	33,700
	Portugal (1)	25,100
	Spain (2)	31,200
Europe: Eastern (8)	Croatia (2)	19,900
	Czech Republic (2)	26,700
	Hungary (1)	22,100
	Poland (2)	22,200
	Romania (1)	17,200
Oceania (16)	Australia (9)	42,400
	New Zealand (7)	32,200
Asia: upper-middle-income (24)	China (6)	10,800
	Malaysia (11)	21,900
	Thailand (7)	13,700
Asia: lower-middle-income (39)	India (9)	5100
	Indonesia (6)	8900
	Pakistan (3)	4400
	Philippines (11)	6000
	Sri Lanka (2)	8900
	Vietnam (8)	4900
Asia: low-income (6)	Bangladesh (1)	2400
	Cambodia (1)	2800
	Nepal (4)	2100

This text is from the North American (U.S. and Canada) and Australia/New Zealand survey instruments, and was nearly identical to that in the European instrument (Navrud and Strand, 2013, 2016). The phrasing in the Asian instrument differed somewhat, albeit not the content.

The experts were asked to base their WTP predictions on only the impacts of protection on biodiversity and other non-carbon values of rainforest protection that they believed were important to their national populations. The exclusion of carbon-related values was justified by a stated assumption that a global political process would identify carbon values separately. But, whether survey respondents are willing to believe this turns out to be a major issue in the implementation and analysis the actual survey that we were aware of from the start and which a number of experts raised in comments on the draft survey instrument.

Experts were told to assume that payments would be collected as a mandatory national tax on all households in their country if a protection plan were enacted. The payments would be submitted to an international Amazon Rainforest Fund, which would be controlled by an international governing body that would use the funds exclusively for the selected plan. Payments would be collected only if a majority of the national population supported the plan, and they would continue only as long as the Amazon countries complied with the plan's terms. This scenario was designed to convey to experts that the CV study would be viewed as consequential by respondents.

Each expert was asked to provide four numbers in each round of the study: their predictions of mean and median annual WTP by households in their country for implementing each of Plans A and B.

2.3. Experts' Characteristics and Attitudes

The Delphi instrument contained several other questions with various aims: inform about expert characteristics and other variables that could be used to explain variation in the WTP predictions; investigate whether the experts' predictions were distorted due to misunderstandings or lack of knowledge; and gather information that could be useful in designing actual SP studies in some of the countries. Among issues questioned were:

- Expert's highest educational degree
- Experts' background from environmental valuation research, in terms of numbers of studies conducted, papers published and reviewed, work time allocated to such research, knowledge of the literature, and methods employed (stated preference; revealed preference; benefit transfer)
- Familiarity with specific journals covering environmental economics
- Perceived difficulty in answering the questions; whether the overall exercise was viewed as meaningful; and whether experts viewed their own answers as reliable
- Gender

Experts also completed a final debriefing section where they were, among other things, asked about possible reformulations of the survey; anticipated problems in implementing the survey; and the possible effects of presenting protection plans in terms of forest area gained instead of the area of loss prevented.

2.4. Selection of Experts and Survey Implementation

All participants in the study were environmental economists with experience from applying nonmarket methods to environmental and natural resource valuation studies in their respective countries. We invited 308 individuals to participate in the study, identifying them from membership lists of regional environmental and resource economists' associations, and our own contact lists. In Asia, we were assisted by the directors of two regional environmental economics capacitybuilding networks, the Economy and Environment Program for Southeast Asia (EEPSEA), and the South Asian Network for Development and Environmental Economics (SANDEE). We aimed at identifying individuals who could be viewed as experts within their own countries. We preferred researchers with experience applying stated-preference (SP) methods but did not require it; we also preferred researchers who had some experience from natural resource valuation (e.g., forests, biodiversity, ecosystem services). We did not require the researchers to be experts on the Amazon rainforest as the number of experts on both nonmarket valuation and the Amazon would be small, and zero in many countries. Instead, we provided essential information on the Amazon in the study booklet which was similar to what respondents would see.

The European study was carried out during the spring and summer of 2012, and in the other regions during May–September 2013. Incentive schemes, differing by region, were used to encourage experts to complete both rounds of the survey: European experts received a small cash award; experts from North America and Oceania a gift card on Amazon.com; and Asian experts an author-signed copy of a book (*Economics: A Very Short Introduction*, by Partha Dasgupta).

3. Results

3.1. Response Rate and Expert Characteristics

More than two-thirds (217; 70%) of the invited researchers completed the study. The response rate was the highest in Europe (83%) and the lowest in Oceania (53%). We did not aim at a 100% response rate, as we wanted to exclude individuals who felt they lacked sufficient expertise or information to reliably predict WTP in their own countries. On the other hand, an extremely low response rate could be taken as evidence that our process for identifying national experts set the bar too low, or that our study design was viewed as flawed by those invited. The final number of participants was 216 (we dropped one Asian researcher who did not respond to our request to clarify some of her responses). Table 1 provides information on the distribution of experts across countries, and GDP per capita levels in the year of the study. The appendix lists the participating experts by country. The sample sizes for some of the results presented later in this paper are a bit <216 because some experts did not answer some of the questions.

Table 2 provides information on some important characteristics of experts, by region, that may have systematically influenced valuation answers. Women comprised 39 of 147 experts (26.5%) from other regions than Asia, and 29 of 69 Asian experts (42.9%). All experts except 13 had PhDs; most of those without a PhD were from Asia. Asian experts are relatively less experienced than experts from the other regions also in two other respects: numbers of SP-related papers written and refereed during the last 5 years; and numbers of empirical environmental valuation studies of all types conducted over their careers. The discrepancy is particularly striking for the latter measure, which reflects mainly Asian experts being young on average relative to those from other regions.

Differences between the other regions are smaller, except that the average number of lifetime SP valuation studies carried out is lower for the North American experts than for experts from Europe and Oceania. This likely reflects the lower average age of North American experts relative to experts from Europe and Oceania. Mean perceptions of the accuracy of the Delphi estimates varied little across the regions. More experts from Oceania than others however considered an actual CV study on Amazon protection to be feasible in their countries.

We also found regional differences in journal reading patterns. We asked experts whether they regularly read eight environmental economics journals, and gave them the option to list other journals they regularly read. We created a set of dummy variables for the various journals and performed a principal components (PC) analysis to identify systematic reading patterns. We focused on the first four principal components, which accounted for two-thirds of the variation embodied in the journal dummies. The loadings indicate that PC1 (27% of the variation) represents a reading pattern that emphasizes the four well-established journals (*AJAE*, *ERE*, *JEEM*, *Land Economics*), while PC2 (15%) represents a pattern that emphasizes two journals focusing more on developing-country topics (*EDE*, *Ecological Economics*). PC3 (14%) represents a pattern that emphasizes the two journals with the greatest effort to target a policy audience (*EDE*, *REEP*). Finally, PC4 (12%) represents a pattern that emphasizes miscellaneous other journals. From Table 2 reading habits differ systematically by region; in particular, North American experts read PC1 journals more, and PC2 journals less than the average; and Asian experts read PC4 journals less than the average.

3.2. Predictions of National WTP for Amazon Forest Protection: Regional Summary Results

Summary information on regional WTP predictions from the study are presented in Fig. 1. Panel a of Fig. 1 suggests that the experts believed that respondents would see a clear difference between Plans A and B, as the Round 2 predictions of mean WTP for Plan A exceed those for Plan B, generally by a margin of about 25–35%. Round 1 and median predictions showed similar patterns.

Panel b shows that the experts in all regions believed that the mean WTP for Plan A in Round 2 would be substantially larger than median WTP, which is the typical result seen in empirical studies. Round 1 and Plan B predictions showed the same patterns.

Panel c shows that the predictions are lower on average in Round 2 than in Round 1. This holds for all regions and for both mean and median WTP, with this shift being more pronounced for mean WTP. As

Table 2
Mean characteristics of experts, by region. CV = contingent valuation, CE = choice experiment, SP = stated preference, RP = revealed preference.

Characteristic	*North America*	*Europe*	*Oceania*	*Asia*
% Female	23	35	25	43
% with PhD	99	94	100	87
No. CV surveys (lifetime)	8.8	10.8	5.5	4.1
No. CE surveys (lifetime)	4.4	6.1	14.0	1.7
No. SP surveys on biodiversity and ecosystem services (lifetime)	3.8	5.7	8.4	2.6
No. RP valuation studies (lifetime)	7.8	4.9	3.2	2.6
No. benefit transfer exercises	3.9	3.9	2.3	0.9
No. SP papers published in national and international journals (past 5 years)	4.9	6.7	7.3	1.7
No. SP papers reviewed for national and international journals (past 5 years)	10.4	12.0	15.0	3.1
Journals read regularly: 1st principal component	0.36	0.17	0.44	−0.64
Journals read regularly: 2nd principal component	−0.58	0.34	0.68	0.30
Journals read regularly: 3rd principal component	−0.06	−0.35	−0.35	0.40
Journals read regularly: 4th principal component	0.05	−0.06	1.15	−0.27
Proximity of Delphi estimate to estimate from an actual CV study on this topic (1–10, with 10 being very close)	6.1	6.3	5.6	6.5
Difficulty of successfully implementing a CV study on this topic (1–10, with 10 being very difficult)	5.9	5.6	4.1	6.4

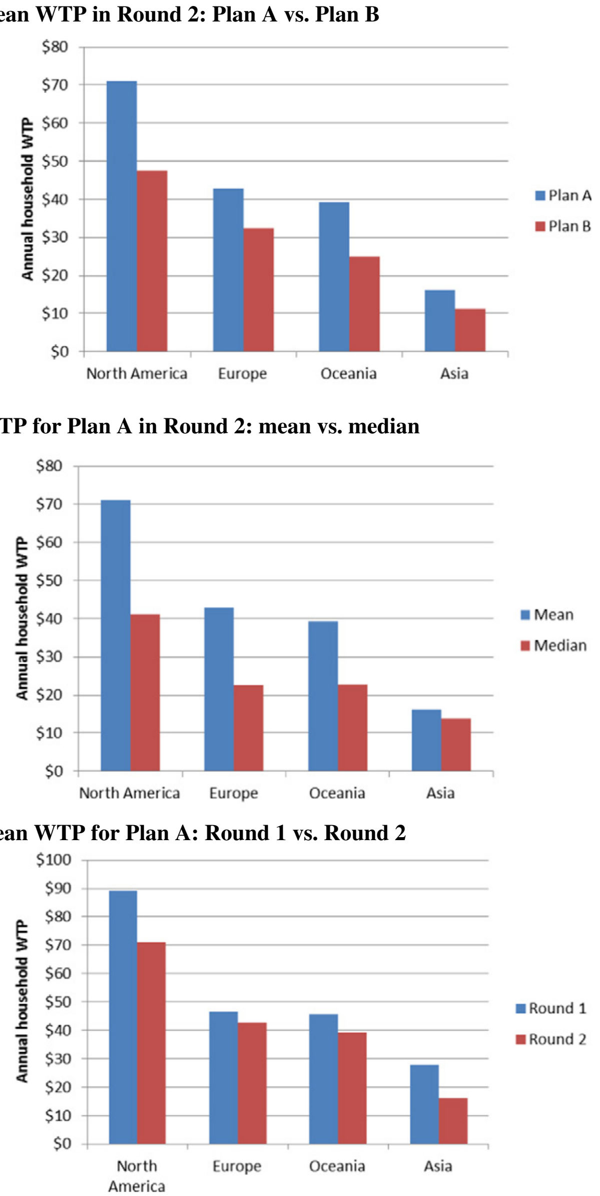

Fig. 1. Mean WTP predictions by country group (US$ per household per year). a. Mean WTP in Round 2: Plan A vs. Plan B, b. WTP for Plan A in Round 2: mean vs. median, c. Mean WTP for Plan A: Round 1 vs. Round 2.

detailed in the next section, this shift occurs because many experts changing their predictions from Round 1 to Round 2, with more experts changing, and with on average greater absolute changes, down than up.

The four regions are shown in Fig. 1 in order of average GDP per capita, so that the pattern suggests a positive relationship between WTP and income. We investigate this relationship more explicitly in Section 3.4 below.

3.3. Changes in Predictions from Round 1 to Round 2

Roughly half of the experts in each region changed their WTP predictions for Plan A from Round 1 to Round 2: 6 of 16 in Australia/New Zealand, 32 of 69 in Asia, 24 of 48 in Europe, and 42 of 81 in Canada/U.S. Among those who made changes, predictions were reduced by five-sixths of European experts; by about two-thirds of experts from Australia/New Zealand and Asia; and by about half of Canada/U.S. experts. Overall, more experts reduced their predictions (65) than raised them (40). Patterns were similar for Plan B (61 reduced, 35 raised).

Table 3 provides an overview of changes in mean WTP by the main regions for both plans. Focusing on Plan A figures, the table shows that, across all experts (those who did and did not revise), the average downward revision was US$12.70. The average revision among those 65 respondents who revised their answers downward was much greater, −$49.50. Among those 40 who revised upward, the average revision was more moderate, +$13.00. Average (downward or upward) revisions were the largest among North American experts (who also had the highest average WTP in Round 1); and the smallest among European experts. The structure of Plan B revisions was similar to that for Plan A, except that revisions were smaller for Plan B (in line with average WTP figures expressed in Round 1 being lower).

A skeptic might argue that Round 2 answers will be biased downward (upward) for those with high (low) round 1 figures, almost by construction via the information provided about average expert responses in round 1. An alternative view, more favorable to our main view, on the constructive nature of the Delphi method, as one for group decision making which is more robust than that by any individual, is that Round 2 provides useful information to individual experts, who were initially less sure of their answer, resulting in adjustments to their Round 1 estimates. Indeed, this motivation behind the Delphi method implies that it will involve some degree of learning by, initially, less than fully informed experts. This is also our own main view. In this particular case our two-rounds procedure implies a measure of conservatism, as average WTP estimates based on Round 2 values are lower than those based on Round 1 values. Note also that the Round 2 values exhibit less uncertainty than Round 1 values: the overall coefficients of variation drop from Round 1 to Round 2, from 1.444 to 1.097 for Plan A, and from 1.482 to 1.111 for Plan B. Interestingly, the median values are the same in Rounds 1 and 2.

We conducted simple *t*-tests of mean differences in selected characteristics of the experts who reduced their WTP predictions for Plan A from Round 1 to Round 2 (Group 0), and those who increased their predictions (Group 1). Table 4 shows these results. As expected, the mean value of the Round 1 WTP estimates was much higher for Group 0 ($116.47) than for Group 1 ($20.16): experts with high Round 1 WTP estimates tended to reduce their estimates, while experts with low Round 1 WTP estimates tended to increase them. Also as expected, Round 1 estimates of experts who reduced their estimates were above-average for their country (Canada, U.S., Australia, New Zealand) or region (Europe, low-income/lower-middle-income Asia, upper-middle-income Asia) (Group 0 mean difference from average = +$61.43); while it was opposite for experts who increased their estimates (Group 1 mean difference from average = −$44.38). This suggests that the Delphi method "worked" in terms of reducing the variance of the country/regional means.

Table 4 also shows differences for the expert characteristics given in Table 2. Among the other variables tested, three had means that differed significantly ($p < 0.05$, one-sided test) between Groups 0 and 1: experts who had published more SP papers were more likely to lower their predictions, as were those who believed more strongly that the Delphi

Table 3
Average changes in predictions of mean WTP from Round 1 to Round 2. All values are in US$ per household per year. Numbers of experts in parentheses.

Plan	*Region*	*All experts*	*Downward revisions only*	*Upward revisions only*
Plan A	All	−12.7 (214)	−49.5 (65)	13.0 (39)
	North America	−18.0 (81)	−94.1 (20)	19.3 (22)
	Europe	−7.1 (48)	−19.7 (20)	12.9 (4)
	Oceania	−5.3 (16)	−23.6 (4)	4.4 (2)
	Asia	−11.9 (69)	−40.2 (21)	1.9 (11)
Plan B	All	−7.5 (215)	−31.7 (60)	8.4 (35)
	North America	−9.4 (81)	−62.4 (16)	10.8 (22)
	Europe	−4.0 (49)	−15.6 (15)	7.9 (5)
	Oceania	−3.3 (16)	−14.3 (4)	4.7 (1)
	Asia	−8.6 (69)	−24.3 (25)	1.8 (7)

Table 4
Tests of mean differences between experts who decreased their predictions for Plan A between Rounds 1 and 2 (Group 0) and experts who increased their predictions (Group 1). Table shows means for each group and *P*-values for one-sided *t*-tests (null hypothesis: Group 0 mean equals Group 1; alternative hypothesis: Group 0 mean is greater than or less than Group 1 mean, depending on which mean is larger), with unequal variances between groups. Sample size varies slightly from variable to variable, but in most cases is 65 for Group 0 and 39 for Group 1.

Variable	*Group 0 mean*	*Group 1 mean*	*P-value*
Round 1 prediction	$116.47	$20.16	0.000
Discrepancy between expert's Round 1 prediction and mean for his/her country/region	$61.43	$44.38	0.000
Ratio of expert's Round 1 prediction to mean for his/her country/region	2.17	0.27	0.000
Gender: female	0.354	0.282	0.224
Highest degree: non-PhD	0.092	0.026	0.068
No. CV surveys (lifetime)	7.22	7.18	0.490
No. CE surveys (lifetime)	3.77	3.63	0.448
No. SP surveys on biodiversity and ecosystem services (lifetime)	3.22	3.56	0.363
No. RP valuation studies (lifetime)	3.79	6.04	0.038
No. benefit transfer exercises	2.58	2.20	0.329
No. SP papers published in national and international journals (past 5 years)	4.65	2.72	0.012
No. SP papers reviewed for national and international journals (past 5 years)	9.48	8.66	0.379
Journals read regularly: 1st principal component	−0.079	−0.272	0.261
Journals read regularly: 2nd principal component	−0.118	0.112	0.166
Journals read regularly: 3rd principal component	−0.104	−0.215	0.291
Journals read regularly: 4th principal component	0.072	0.148	0.352
Proximity of Delphi estimate to estimate from an actual CV study on this topic	6.48	5.78	0.043
Difficulty of successfully implementing a CV study on this topic	5.47	5.85	0.230

estimates were close to actual WTP in their countries. Experts who had conducted more RP studies were by contrast more likely to raise their predictions. Experts whose highest degree was not a PhD were more likely to lower their predictions ($p = 0.068$).

3.4. Effect of National Income on WTP: No Control for Expert Characteristics

If properly identified, the relationship between WTP and national income may provide one way to use results from the Delphi study for valuation purposes through a benefit transfer approach, to countries where we interviewed valuation experts. There are several ways to characterize this relationship. We start with simpler ones before moving on to more complicated ones.

Fig. 2 provides a visualization of the relationship between PPP-adjusted GDP per capita and the Round 2 expert assessments of regional mean WTP per household per year for Plan A. We find a reasonably regular increasing relationship between mean WTP and average income, which is not entirely regular due to relatively high variability of WTP estimates across individual experts.

Table 5 presents results from estimating full-sample models where we regress the experts' predictions of mean and median WTP on GDP per capita for their countries. The models were estimated separately for mean and median WTP, but the standard errors were estimated by pooling the residual variation across the mean and median models ("seemingly unrelated estimation", which is distinct from seemingly unrelated regression). The sample for these models included only WTP predictions from Round 2, which are assumed to be more accurate than predictions from Round 1. It included WTP predictions for both Plan A and Plan B, with a dummy variable used to control for the larger area protected under Plan A. The dependent variables in the models were ln(1 + WTP), not ln(WTP), as some experts set WTP = 0 (in most cases, for median WTP). PPP-adjusted GDP per capita was used throughout as the income measure; other measures, including regular GDP and GNI per capita and PPP-adjusted GNI per capita, yielded similar

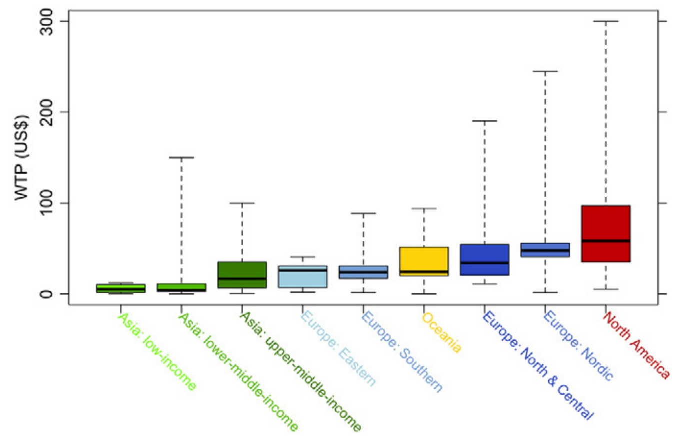

Fig. 2. Distributions of WTP predictions for Plan A (Round 2), with regions ranked by increasing GDP per capita (PPP). The solid band inside the box represents the median. The box extremities represent the first and third quartiles. Whiskers extend to data extremes.

results. The models also include regional dummy variables to control for regional differences in the timing of the surveys, the wording of the questionnaires, and the distances of the regions from the Amazon. Europe was the reference region.

The GDP estimates within a given country are the same across experts from that country, which potentially causes the standard errors of regression coefficients to be biased downward exaggerating the precision of the estimates (Moulton, 1986, 1990). We clustered the standard errors by country to address this problem, all though we note that it would be desirable to have valuation experts from more countries to improve the statistical properties of this standard error adjustment procedure (Angrist and Pischke, 2009).

We used regression weights to account for the high variation in the number of experts by country, from only one from Bangladesh, Belgium, Cambodia, Hungary, Portugal, and Romania to a high of 70 from the U.S. Countries are the implicit units of observation when estimating the effect of GDP per capita on WTP, but countries with more experts have greater influence on the estimation results than countries with fewer experts. Treating countries equally requires estimating weighted

Table 5
Effect of income on mean and median WTP, with regional differences controlled by dummies. Dependent variables: ln(1 + mean WTP) and ln(1 + median WTP). Observations: individual respondents in the models in the first two columns; countries in the models in the last two columns.

	Individual predictions		*Country means*	
Variables	*Mean*	*Median*	*Mean*	*Median*
ln(GDP per capita, PPP)	0.707***	0.735***	0.905***	0.932***
	(0.210)	(0.154)	(0.195)	(0.157)
Dummy: plan A	0.285***	0.293***	0.301***	0.317***
	(0.0347)	(0.0367)	(0.0299)	(0.0315)
Dummy: Asia	0.0304	0.369	0.608	0.894**
	(0.405)	(0.322)	(0.403)	(0.368)
Dummy: North America	0.575**	0.757**	0.535**	0.661***
	(0.284)	(0.314)	(0.216)	(0.253)
Dummy: Oceania	−0.234	−0.298	−0.137	−0.141
	(0.243)	(0.476)	(0.169)	(0.251)
Country FEs?	No	No	No	No
Weighted regression?	Yes	Yes	No	No
Income elasticity	0.729	0.775	0.940	0.996
P-value for H_0: elasticity = 1	0.211	0.167	0.767	0.982
R^2	0.359	0.283	0.612	0.611
F statistic	47.4	33.3	21.2	21.1
Observations	429	428	73	73

*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. Standard errors are in parentheses and were estimated jointly for the pair of equations (mean, median), with clustering by country.
Weights: inverse to number of respondents from a given country.

models, with the inverse of the number of experts from each country used as weights.[1] This issue can alternatively be addressed by estimating a model based on mean values for all experts from each country. We addressed the problem both ways, as reported below.

The first two columns in Table 5 show results for the weighted models based on WTP predictions by individual experts. They yielded very similar income elasticities for mean and median WTP, 0.729 and 0.775, respectively. Note that the income elasticities differ slightly from the coefficient estimates on ln(GDP per capita, PPP), due to the dependent variables being defined as ln(1 + WTP) instead of ln(WTP). The elasticity estimates changed little if the models were estimated without the weights; the main change was that the standard errors became much smaller. The magnitudes of the elasticities are similar to elasticity estimates for general environmental improvements and values of statistical life (e.g., OECD, 2012). As potentially important from the perspective of benefit transfer, neither elasticity differed significantly from unity. This suggests that it may be possible to use the ratio of (PPP-adjusted) incomes between two countries to scale results in the transfer exercise without too large an error for goods similar to the one considered here (Flores and Carson, 1997).

The income elasticities were higher in the models based on country means of the WTP predictions (shown in the last two columns of Table 5), 0.940 and 0.996, both of which are close to and not statistically different from unity.

As expected, WTP was higher for Plan A than for Plan B in all four models. WTP, controlling for ln(GDP per capita, PPP), is estimated to be higher in North America than in Europe (the omitted region) while the other two regions, Asia and Oceania, are not statistically different from Europe.

3.5. Effects of Experts' Characteristics and Attitudes on WTP Predictions

The models in Table 5 do not control for the characteristics of individual experts. Table 6 shows results for models that include such controls. In addition to PPP-adjusted GDP per capita and dummies for Plan A and the three non-European regions (as in Table 5), the models in the first two columns of Table 6 include controls based on the respondent characteristics in Table 2. These represent all the questions from the study that provided objective information on expert characteristics; we excluded questions that asked experts to evaluate their own expertise. The controls included the expert's perceived error in the Delphi estimate; the expert's perceived difficulty with doing such valuation; the expert's reported number of surveys, publications, reviewed papers, and benefit transfer exercises; the journals the expert reads regularly; the expert's highest academic degree; and the expert's gender.

Comparing the results in the two leftmost columns of Table 6 to the corresponding columns in Table 5, we find that inclusion of these additional controls raised income elasticities sharply, to 1.14 (mean WTP) and 1.11 (median WTP). Neither estimate though is significantly different from unity, which was also true of the Table 5 estimates. The only expert characteristics that are significantly (5%) correlated with their WTP predictions are the non-PhD dummy (a positive effect), the number of SP papers they had reviewed (a negative effect), and reading the four journals in PC1. While the association with highest degree likely represents a learning effect, the association with reading a set of journals likely represents both learning and selection effects.

Adding country fixed effects to the models makes the estimated effects of expert characteristics less prone to omitted variables bias caused by correlation of those characteristics with unobserved country characteristics. Results for such models are shown in the last two columns of Table 6. Inclusion of the fixed effects by construction drops out the income variable and the regional dummies, but it reveals that several additional expert characteristics have a significant effect on the mean or median WTP predictions. Publishing a larger number of SP papers now has a significant positive effect. By contrast, a larger number of RP studies and CV surveys affects stated WTP negatively, as does the number of CE surveys (although less significantly). These results suggest a cumulative effect of experience with multiple valuation studies. When combined with the positive effect of the number of SP papers published, they suggest that numbers of publications and numbers of studies conducted are qualitatively different dimensions of expertise. Experts who rated conducting a CV study like the one described in the Delphi study as more difficult had lower WTP predictions.

Results from the models in Table 6 could be used to formulate predictions of mean or median WTP that give more weight to the beliefs

Table 6
Effects of respondent characteristics on mean and median WTP. Dependent variables: ln(1 + mean WTP) and ln(1 + median WTP). Observations: individual respondents in all models. Dummies are used to control for regional differences in the models in the first two columns and country differences in the models in the last two columns.

	Include income, exclude FE		*Exclude income, include FE*	
Variables	*Mean*	*Median*	*Mean*	*Median*
ln(GDP per capita, PPP)	1.104***	1.055***		
	(0.202)	(0.169)		
Dummy: plan A	0.290***	0.294***	0.348***	0.333***
	(0.0284)	(0.0302)	(0.0242)	(0.0190)
Dummy: Asia	0.570	0.901**		
	(0.461)	(0.405)		
Dummy: North America	0.426	0.682**		
	(0.292)	(0.265)		
Dummy: Oceania	−0.217	−0.194		
	(0.349)	(0.427)		
Dummy: female	−0.0933	−0.132	−0.278	−0.256
	(0.229)	(0.249)	(0.257)	(0.277)
Dummy: non-PhD	0.529**	0.622**	1.335***	1.316***
	(0.236)	(0.312)	(0.262)	(0.288)
No. CV surveys	−0.00829	−0.0170	−0.00569	−0.0163**
	(0.0120)	(0.0123)	(0.00705)	(0.00726)
No. CE surveys	−0.0187	−0.0187	−0.0410*	−0.0414*
	(0.0269)	(0.0247)	(0.0221)	(0.0225)
No. SP surveys (biodiversity)	−0.0112	0.00200	−0.00826	0.00120
	(0.0179)	(0.0193)	(0.0233)	(0.0247)
No. RP studies	0.00151	−0.00310	−0.0134***	−0.0121**
	(0.0170)	(0.0171)	(0.00438)	(0.00589)
No. benefit transfer exercises	0.00467	0.00826	−0.00359	2.41e-05
	(0.0114)	(0.00984)	(0.0206)	(0.0173)
No. SP papers published	0.0376	0.0590*	0.0309*	0.0448***
	(0.0338)	(0.0335)	(0.0174)	(0.0160)
No. SP papers reviewed	−0.0156*	−0.0180**	−0.000618	−0.00158
	(0.00928)	(0.00856)	(0.00811)	(0.00873)
Journals: PC 1	0.165**	0.165**	0.137***	0.133***
	(0.0681)	(0.0735)	(0.0456)	(0.0394)
Journals: PC 2	0.0247	0.0426	−0.0314	−0.0261
	(0.113)	(0.119)	(0.114)	(0.0946)
Journals: PC 3	−0.0801	−0.101	−0.112*	−0.0633
	(0.113)	(0.121)	(0.0620)	(0.0640)
Journals: PC 4	−0.0774	−0.0956	0.0487	0.0635
	(0.107)	(0.104)	(0.0685)	(0.0747)
Accuracy of Delphi prediction	0.00533	0.0273	0.0471	0.0692**
	(0.0452)	(0.0472)	(0.0309)	(0.0336)
Difficulty of CV study	−0.0448	−0.0498	−0.0607***	−0.0599***
	(0.0462)	(0.0456)	(0.0223)	(0.0223)
Country FEs?	No	No	Yes	Yes
Weighted regression?	Yes	Yes	No	No
Income elasticity	1.14	1.11	-	-
P-value for H_0: elasticity = 1	0.505	0.527	-	-
R^2	0.470	0.400	0.629	0.567
F statistic	16.9	12.6	11.6	8.94
Observations	401	400	401	400

*** p < 0.01, ** p < 0.05, * p < 0.1. Standard errors are in parentheses and were estimated jointly for the pair of equations (mean, median), with clustering by country.
Weights: inverse of number of respondents from a given country.

[1] Note that the reason for weighting is not to address heteroskedasticity; we used robust standard errors to address that issue. Additional weighting could arguably be done to account for differences in population across the countries, but weighting by population makes more sense when individuals are implicitly the units of observation (e.g., in studies on the effects of economic growth on poverty alleviation), not countries as in our study.

of particular experts. If expert characteristics do not systematically influence on estimates of WTP summary statistics, no issues arise. If, however, there are systematic differences based on observable expert characteristics, then an argument can be made to give greater weight to experts with characteristics thought to be associated with better quality forecasts. This can be done by estimating WTP as a function of expert characteristics, as in Table 6, and then obtain predictions from the model by setting all expert characteristics at their ideal ("most expert") level. Some of the variables in Table 6 have ideal levels, such as having a PhD instead of a lower degree, and having considerable experience doing SP studies (many surveys, many publications). Other variables, such as gender, do not, and these can be set equal to the mean across the experts from a given country.

There is a large and expanding literature (Bates and Granger, 1969) that shows that a forecast which is a weighted average of individual forecasters tends to have lower prediction error than any individual forecasts. However, even though the theoretical case for doing so is clear, empirical support for more complicated weighting schemes than giving each forecasts equal weight, is surprisingly weak (Smith and Wallis, 2009). There seems to be weak ability of in-sample information to precisely identify forecasters likely to make higher quality out-of-sample predictions, and few statistical degrees of freedom in trying to make this determination. While we find the results of the Table 6 regression models interesting from the perspective of what it suggests about different types of experience, we follow the conventional wisdom of using equal weights on all the forecasts. We also note that the results obtained doing this are not substantively different using Table 6 models variable values set to their ideal levels.

3.6. Predicting the WTP Distribution

Having estimates of both mean and median WTP for a particular region allows us to trace out the probability density function of the WTP distribution for that region if we are prepared to assume a two-parameter distribution. Assuming a log-normal distribution, commonly applied in stated-preference studies, the location parameter for this distribution equals the natural logarithm of the median, while the scale parameter equals the square root of twice the difference between the logs of the mean and median.[2] Fig. 3 displays the resulting probability density functions for WTP for Plans A and B, based on round 2 predictions by the North American experts. In countries where actual SP studies on Amazon protection were planned, functions such as these could be used as initial priors for experimental design where the variance in the estimated parameters could be used as an initial measure of parameter precision in a full Bayesian design (Choicemetrics, 2012; Rose and Bliemer, 2013).

An alternative approach is to estimate models similar to those in Table 5, which pool the data across countries while controlling for differences in GDP per capita and other regional and country characteristics. The difference compared to Table 5 is that the dependent variables are now the location and scale parameters implied by the experts' mean and median WTP predictions, instead of the mean and median WTP themselves. As in Fig. 3, we assumed a log-normal distribution, and set the variable for the location parameter, mu, equal to the natural logarithm of median WTP, and the variable for the scale parameter, sigma, equal to the square root of twice the difference between the logs of mean and median WTP. Again we added 1 to the predictions to avoid taking the log of zero when we constructed mu and sigma.

Results are shown in Table 7. Given that the location parameter in a log-normal distribution equals the log of the median, the results for mu

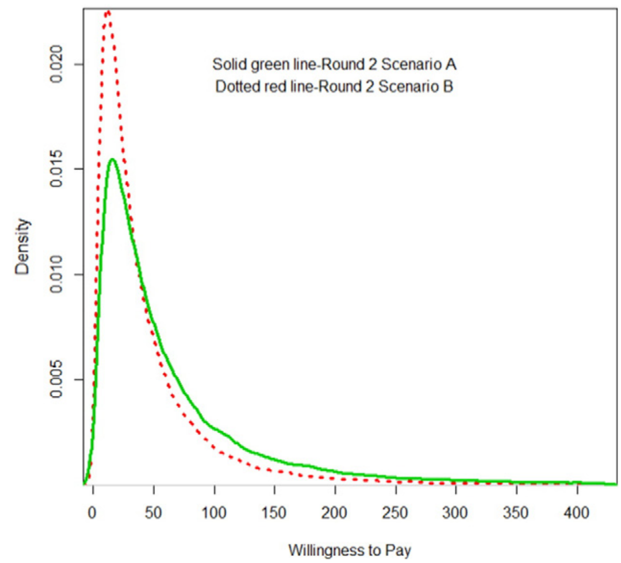

Fig. 3. North American Round 2 WTP Distributions for Plans A and B. Based on mean and median predictions for the regions and an assumed log-normal distribution.

in Table 7 are the same as the results for median WTP in Table 5. The results for the scale parameter, sigma, are more interesting, and indicate that income has no effect on it, neither in the model based on WTP predictions by individual experts, nor in the model based on country means of the WTP predictions. The only significant effect is for home region of experts, with a smaller scale parameter in Asia than in the other regions. These results could be used to generate parametric versions of Fig. 3 for any country in the sample, given information on its region and income level.

4. Conclusions and Final Discussion

Our Delphi study was intended to serve several objectives. First, it provided an initial estimate of WTP to protect the Amazon for a large share of the global population. Second, the distribution of the experts' WTP predictions may help us construct more efficient experimental designs (Ferrini and Scarpa, 2007) for the actual stated-preference (SP) valuation studies on Amazon protection that we intend to

Table 7
Effect of income on log-normal location and scale parameters (mu and sigma, respectively), with regional differences controlled by dummies. Dependent variables: mu and sigma. Observations: individual respondents in the models in the first two columns; countries in the models in the last two columns.

	Individual predictions		*Country means*	
Variables	*Mu*	*Sigma*	*Mu*	*Sigma*
ln(GDP per capita, PPP)	0.735***	−0.0422	0.932***	−0.0268
	(0.154)	(0.0800)	(0.157)	(0.0775)
Dummy: plan A	0.293***	−0.0114	0.317***	−0.000672
	(0.0367)	(0.0185)	(0.0315)	(0.0256)
Dummy: Asia	0.369	−0.343**	0.894**	−0.284**
	(0.321)	(0.140)	(0.368)	(0.131)
Dummy: North America	0.757**	−0.140**	0.661***	−0.104
	(0.314)	(0.0696)	(0.253)	(0.0764)
Dummy: Oceania	−0.298	−0.0876	−0.141	0.0145
	(0.475)	(0.165)	(0.251)	(0.112)
Country FEs?	No	No	No	No
Weighted regression?	Yes	Yes	No	No
R^2	0.283	0.084	0.611	0.211
F statistic	33.3	7.61	21.1	3.52
Observations	428	428	73	73

*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. Standard errors are in parentheses and were estimated jointly for the pair of equations (mu, sigma), with clustering by country.
Weights: inverse number of respondents from a given country.

[2] Alternative two parameter distributions that tend to have a shorter right tail and hence a smaller mean, such as the Weibull, can be examined in a similar way. The presence of zero estimates for the median by some experts suggests that it might be desirable to ask experts for three summary statistics, the mean, median and the percent of respondents having a WTP at or very near zero.

undertake. In addition, as part of the Delphi study we sought the experts' guidance on various other aspects of the design of population-based SP surveys involving Amazon protection plans where a key problem is that respondents may believe that if they pay to protect tropical forest for biodiversity purposes they must also be protecting the carbon from those forests from being released.[3] Third, we ultimately seek to compare the WTP estimates from our Delphi study to estimates from actual population SP studies. A large discrepancy between the Delphi and SP estimates, either absolutely or relatively across the countries for which we will have estimates from both, may provide useful information for analyzing and refining both approaches. Finally, the Delphi study provides WTP estimates for many countries in which we do not have the resources to conduct SP studies. Depending on the results of the comparison just mentioned, we may be able to use these estimates to augment the SP-based estimates in terms of either absolute WTP levels or, perhaps more likely (León et al., 2003), relative WTP differences across the countries with SP-based estimates and those without.

An objective was also to learn more about the Delphi method as such, as a tool for environmental and natural resource valuation. Hardly any such studies exist in the literature, and more applications are desired. We have through this study gained valuable insights into the processes by which initial expert predictions are formed, and later revised in response to updated information. It seems clear to us, in particular, that by applying this method one can go some way toward forming an initial consensus about natural resource values among experts; but a full consensus is neither possible nor desirable. Ultimately, ground truthing in the form of empirical SP studies is necessary. More research on how the Delphi approach works in applications to other global environmental goods would clearly be required if estimates from the approach are to be used for policy purposes. The distinction we see here between BT and the Delphi approaches is that the BT approach is most useful when there are policy decisions to be made for which the cost of undertaking original valuation studies is high and the quality of the BT estimate is likely to be good. The Delphi approach is likely to be most useful in assessing very large new programs, sometimes on a global scale, where little relevant valuation work has been done and for which the cost of doing original high quality valuation studies is obviously justified. The usefulness of the Delphi approach comes into play where having some reasonable economic estimates to work with until those original valuation studies are undertaken is important to the policy process. Our Delphi results on income elasticities show that the Delphi and BT approaches can be complementary as more original studies become available.

With respect to the Amazon, our study suggests, according to the almost 220 (overall, very highly qualified) international environmental valuation experts who participated in it, that there is considerable aggregate WTP among the global population outside of Latin America to avoid further forest losses in the Amazon region. Focusing on experts' predictions in Round 2 of the study for the more comprehensive rainforest protection plan, mean annual WTP per household varies from a high near $100 in Canada, Norway, and Germany, to intermediate levels closer to $50 in a broader set of OECD countries, to lower levels varying from $4 to $35 for Asian countries.

We recognize that the Delphi method represents a virtually untried technique for environmental valuation. Its usefulness for assessing average WTP levels in populations that are not covered by national population surveys is still an open question. Still, we believe that the results from our Delphi study can be helpful; perhaps less for providing accurate *valuation levels*, and more by indicating how WTP for Amazon forest protection can *vary across countries at different income levels*. The expert elicitation study by León et al. (2003) provides support to the idea that experts' WTP answers can be used constructively in this way. Such an assessment is based on their conclusion that experts' relative evaluations of alternative projects appear to be far more precise than their absolute-level valuations of individual projects; and the former is the crucial aspect in predicting how WTP is likely to vary with average per-capita incomes across countries. The elasticity estimates of WTP per household with respect to national income then represent a key, and interesting, set of results. These elasticities are not significantly different from unity when measured with respect to PPP-adjusted GDP levels per capita. This finding is consistent with similar results found for some other environmentally-related goods and features (e.g. for assessments of environmentally-related premature deaths using Value of Statistical Life estimates; see Lindhjem et al., 2011; OECD, 2012). In the (albeit, special) case where an elasticity estimate of 1 for PPP-adjusted per-capita GDP is an approximate global value, an approximate aggregate global WTP value can be obtained from a smaller set of national surveys, by scaling the aggregate, global WTP value up proportionately to the PPP-adjusted GDP values by country.

There are as noted few existing studies against which the numbers from our Delphi study can be compared. The only relevant earlier study, Horton et al. (2003), gave average annual per-household WTP of about $60 among Italian and U.K. households in 1999 for an Amazon forest protection plan that was less comprehensive than the ones presented to our experts. The equivalent expert-based figures for these two countries from our survey would be closer to $25–30. This might indicate that experts' assessments in our survey are on the low side.

A new study, Siikamäki et al. (2015), is far more relevant and points in a different direction. Using a survey instrument built off of the same basic scenarios used in our Delphi study administered to samples of the U.S. and Canadian population, results indicate that the average annual per-household WTP levels in the population survey for these two countries, are between twice and three times the average expert predictions for U.S and Canadian experts in the Delphi study. We consider this to be a remarkable similarity. Note that we in this Delphi study have explicitly instructed experts to ignore carbon values (which would likely constitute a substantial fraction of overall assessed value by experts), while no instructions, neither to include nor exclude carbon values, were given to those interviewed in the Siikamäki et al. survey. On the basis of our assessments made of answers from that survey, based on debrief questions about such possible motivations, up to about half of the stated WTP can be ascribed to carbon values. Note also that experts in the current study had few very concrete clues about likely outcomes of similar population surveys. No similar survey had been carried out in North America prior to that survey, and the two relevant existing surveys, Horton et al. (2003) and Sánchez (2008), covering households in the U.K., Italy and Spain, yielded far lower values, much closer to those in our current Delphi survey. Population surveys from more countries are here clearly required.

A very different and perhaps speculative calibration of our results can be obtained from comparisons with the Norwegian government's funding for its forest protection program, aimed at preventing both the release of carbon and protection of biodiversity, with Brazil, the most comprehensive such plan to date based on external funding. This program aims to provide $1 billion to Brazil over a 10-year period for "delivered" forest protection in that country (verified reductions in forest loss rates resulting from the funds being made available). With about 2 million households in Norway, this represents a total WTP per Norwegian household, reflected by this program, of about $500 to protect the Amazon rainforest over a 10-year period; thus $50 per household per year (or somewhat higher when applying the Norwegian government's risk-adjusted project discount rate of 4%). Evaluation by the Norwegian experts in our survey is closer to a mean WTP of $100 per household per year; thus higher than values directly embedded in the Norwegian program, but Norwegian expenditures are quite similar

[3] This difficulty is endemic to situations where the actual protection mechanism by its nature provides a bundle of services but policy makers want separate values for the individual services. Economic theory suggests that except under highly restrictive conditions that unique decomposition of values for the individual services does not exist and there is no reason why a respondent facing a consequential survey question about implementation of the protection plan should ignore some of its outputs.

to our median WTP estimate. Note that WTP revealed via actual payments through Norwegian government budgets must clearly be a lower bound of (generally, less than) the government's maximum WTP for the same service.

An issue, perhaps most important for Asian experts, is whether expert valuation figures can be interpreted as representing their entire homeland populations or only fractions of them. In several of the Asian countries, most households pay no income taxes, and many do not have formal utility services nor face utility bills. Although we asked experts to consider all households in their countries when predicting mean and median WTP, they might have felt it unreasonable to assume that payments could be collected from informal households that do not have formal relationships to taxing authorities nor to utilities. This represents an area ripe for future exploration.

References

Angrist, J.D., Pischke, S.J., 2009. Mostly Harmless Econometrics: An Empricist's Companion. Princeton University Press, Princeton, NJ.
Bates, J.M., Granger, C.W.J., 1969. The combination of forecasts. Oper. Res. Q. 20, 451–468.
Brookshire, D.S., Neill, H.R., 1992. Benefit transfers — conceptual and empirical issues. Water Resour. Res. 28, 651–655.
Carson, R.T., 1998. Valuation of tropical rainforests: philosophical and practical issues in the use of contingent valuation. Ecol. Econ. 24, 15–29.
Carson, R.T., Conaway, M.B., Navrud, S., 2013. Preliminary Valuation of a Cultural Heritage Site of Global Significance: A Delphi Contingent Valuation Study. In: Rizzo, I., Mignosa, A. (Eds.), Handbook on the Economics of Cultural Heritage. Edward Elgar, Cheltenham, UK.
Choicemetrics, 2012. Ngene 1.1.1 User Manual and Reference Guide. Choicemetrics, Sydney.
Curtis, I.A., 2004. Valuing ecosystem goods and services: a new approach using a surrogate market and the combination of a multiple criteria analysis and a Delphi panel to assign weights to the attributes. Ecol. Econ. 50, 163–194.
Dalkey, N.C., 1967. Delphi. Rand Paper P-3704. Rand Corporation, Santa Monica, CA.
Dalkey, N.C., 1969. The Delphi Method: An Experimental Study of Group Opinion. Rand Paper RM5888-PR. Rand Corporation, Santa Monica, CA.
Dalkey, N.C., Helmer, O., 1963. An experimental application of the Delphi method to the use of experts. Manag. Sci. 9, 458–467.
Evans, C., 1997. The use of consensus methods and expert panels in pharmacoeconomic studies. PharmacoEconomics 12, 121–129.
FAO, 2010. Global Forest Resources Assessment 2010: Main Report. United Nations Food and Agriculture Organization, Rome.
Ferrini, S., Scarpa, R.R., 2007. Designs with a priori information for nonmarket valuation with choice experiments: a Monte Carlo study. J. Environ. Econ. Manag. 53, 342–363.
Flores, N.E., Carson, R.T., 1997. The relationship between the income elasticities of demand and willingness to pay. J. Environ. Econ. Manag. 33, 287–295.
Holtsapple, C.W., Joshi, K.D., 2002. Knowledge manipulation activities: results of a Delphi study. Inf. Manag. 29, 477–490.
Horton, B., Colarullo, G., Bateman, I., Peres, C., 2003. Evaluating non-user willingness to pay for a large-scale conservation programme in Amazonia: a UK/Italian contingent valuation study. Environ. Conserv. 30, 139–146.
Hufschmidt, M.M., James, D.E., Meister, A.D., Bower, B.T., Dixon, J.A., 1983. Environment, Natural Systems, and Development: An Economic Valuation Guide. Johns Hopkins University Press, Baltimore.
Kramer, R.A., Mercer, D.E., 1997. Valuing a global environmental good: U.S. residents' willingness to pay to protect tropical rain forests. Land Econ. 73, 196–210.
Landeta, J., 2006. Current validity of the Delphi method in social sciences. Technol. Forecast. Soc. Chang. 73, 467–482.
León, C.J., Vázquez-Polo, F.J., González, R.L., 2003. Elicitation of expert opinion in benefit transfer of environmental goods. Environ. Resour. Econ. 26, 199–210.
Lindhjem, H., Navrud, S., Braathen, N.A., Biausque, V., 2011. Valuing lives saved from environment, transport and health policies: a meta analysis. Risk Anal. 31, 1381–1407.
Linstone, H.A., Turoff, M., 1975. The Delphi Method: Techniques and Applications. Addison-Wesley, Reading, MA.
Moulton, B.R., 1986. Random group effects and the precision of regression estimates. J. Econ. 32, 385–397.
Moulton, B.R., 1990. An illustration of the pitfalls in estimating the effects of aggregate variables on micro units. J. Econ. Stat. 72, 334–337.
Navrud, S., Strand, J., 2013. Valuing Global Public Goods: A European Delphi Stated Preference (SP) Survey of Population Willingness to Pay for Amazon Rainforest Preservation. Policy Research Working Paper no 6637. World Bank, Washington, DC.
Navrud, S., Strand, J., 2016. Valuing global ecosystem services - what do European experts say? Applying the Delphi method to contingent valuation of the Amazon rainforest. Environ. Resour. Econ. (forthcoming).
Nepstad, D., McGrath, D., Stickler, C., Alecar, A., Azevedo, A., Swette, B., Bezerra, T., DiGaino, M., Shimada, J., da Motta, R.S., Armijo, R., Castello, L., Brando, P., Hansen, M.C., McGrath-Horn, M., Carvalho, O., Hess, L., 2014. Slowing Amazon deforestation through public policy and interventions in beef and soy supply chains. Science 344 (6188), 1118–1123.
OECD, 2012. Mortality Risk Valuation in Environment, Health and Transport Policies. Organisation for Economic Co-operation and Development, Paris.
Okoli, C., Pawlowski, S.D., 2004. The Delphi method as a research tool: an example, design considerations and applications. Inf. Manag. 42, 15–29.
Ready, R., Navrud, S., 2006. International benefit transfer: methods and validity tests. Ecol. Econ. 60, 429–434.
Roman, H.A., Hammitt, J.K., Walsh, T.L., Stieb, D.M., 2012. Expert elicitation of the value per statistical life in an air pollution context. Risk Anal. 32, 2133–2151.
Rose, J.M., Bliemer, M.C.J., 2013. Sample size requirements for stated choice experiments. Transportation 40, 1021–1041.
Sackman, H., 1975. Delphi Critique: Expert Opinion. Forecasting and Group Process. D.C. Heath, Lexington, MA.
Sánchez, J., 2008. Disponibilidad a pagar por la conservación del bosque amazónico por parte de usuarios indirectos. Rev. Econ. Adm. 71, 57–82.
Scholl, W., König, C., Meyer, M., Heisig, P., 2004. The future of knowledge management: an international Delphi study. J. Knowl. Manag. 8, 19–35.
Scolozi, R., Morri, E., Santolini, R., 2012. Delphi-based change assessment in ecosystem service values to support strategic spatial planning in Italian landscapes. Ecol. Indic. 21, 134–144.
Siikamäki, J. Krupnick, A., Strand, J. and Vincent, J., 2015. Willingness to pay for Amazon rainforest protection. Unpublished manuscript, Resources for the Future/World Bank.
Smith, J., Wallis, K.F., 2009. A simple explanation of the forecast combination puzzle. Oxf. Bull. Econ. Stat. 71, 331–355.
Soares-Filho, B.S., Nepstad, D., Curran, L., Voll, E., Cerqueira, G., Garcia, R.A., Ramos, C.A., Mcdonald, A., Lefebvre, P., Schlesinger, P., 2006. Modeling conservation in the Amazon Basin. Nature 440, 520–523.
Weitzman, M.L., 2001. Gamma discounting. Am. Econ. Rev. 91, 260–271.
Wilson, M.A., Hoehn, J.P., 2006. Valuing environmental goods and services using benefit transfer: the state-of-the art and science. Ecol. Econ. 60, 335–342.
Ziglio, E., 1996. The Delphi Method and its Contribution to Decision-Making. In: Adler, M., Ziglio, E. (Eds.), Gazing into the Oracle: The Delphi Method and its Application to Social Policy and Public Health. Jessica Kingsley, London.