Methodological issues in meta-analysis, benefit transfer and environmental valuation

Metodiske spørsmål i meta-analyse, verdioverføring og verdsetting av miljø

Philosophiae Doctor (PhD) Thesis

Henrik Lindhjem

Dept. of Economics and Resource Management
Norwegian University of Life Sciences

Ås 2008

Thesis number 2008: 18
ISSN 1503-1667
ISBN 978-82-575-0819-7

“If the day and the night are such that you greet them with joy, and life emits a fragrance like flowers and sweet-scented herbs, is more elastic, more starry, more immortal – that is your success. All nature is your congratulation, and you have cause momentarily to bless yourself. The greatest gains and values are farthest from being appreciated. We easily come to doubt if they exist. We soon forget them. They are the highest reality. Perhaps the facts most astounding and most real are never communicated by man to man. The true harvest of my daily life is somewhat as intangible and indescribable as the tints of morning or evening. It is a little stardust caught, a segment of the rainbow which I have clutched”

Walden, Henry David Thoreau, 1854

“I think that each town should have a park, or rather a primitive forest, of five hundred or a thousand acres, either in one body or in several – where a stick should never be cut for fuel – nor for the navy, nor to make wagons, but stand and decay for higher uses – a common possession forever, for instruction and recreation”.

Huckleberries, Henry David Thoreau, 1862

Abstract

This thesis consists of four papers on methodological issues in meta-analysis (MA), benefit transfer (BT) and environmental valuation. The first paper presents a MA of stated preference studies valuing non-timber benefits in Norway, Sweden and Finland over the last 20 years. It investigates using different meta-regression models to what extent willingness to pay (WTP) estimates conform with standard expectations, tests a number of novel hypotheses and identifies gaps in the literature. Papers 2-4 then each pick up an important research theme following from the first paper. The second paper utilises the same data to investigate the precision in using MA for international BT, as compared with simpler and more common BT techniques. The third paper investigates, using a different and more extensive dataset of biodiversity and nature conservation values from Asia and Oceania, how the heterogeneity or scope of the MA data influences the results of different meta-regression models and their precision when used for BT. The fourth paper tests using a primary contingent valuation data set of WTP for forest protection in Norway, whether people state different WTP if asked as individuals or on behalf of their household. This paper tests in a more controlled way the question also investigated in the first paper. Results from the four papers are encouraging in contributing to our understanding of people's preferences for complex environmental goods. However, more research is required to determine the conditions under which MA may be reliably used for BT.

Acknowledgements

This PhD work started in February 2005 and was carried out at the Department of Economics and Resource Management while working part time with the environmental economics group of Econ Pöyry (formerly Econ Analysis). I am indebted to many people for making the thesis possible.

First, I want to thank my supervisor Ståle Navrud for his enthusiasm, his creativity and his willingness to provide assistance almost at any time. Further, I am indebted to Olvar Bergland for sharing his knowledge of econometrics, Eirik Romstad, co-supervisor, for writing the original PhD scholarship proposal, and Reidun Aasheim and Rita Ovesen for helping with practical things. I also want to thank my PhD colleagues, especially Tran Huu Tuan (who is also my co-author) and Ronnie Babigumira, and other staff in the Department, for interesting discussions.

Special thanks also go to the following: Haakon Vennemo (Econ Pöyry) to grant me full flexibility in pursuing PhD research while working part-time with Econ Pöyry; Rolf Røtnes (Econ Pöyry) for essential help with a proposal; Birger Vennesland, Erlend Nybakk (both The Norwegian Forest and Landscape Institute – NFLI), Erling Bergsaker (NORSKOG), Even Bergseng, Hans Fredrik Hoen, Anders Lunnan, Mikael Ohlson, Birger Solberg (all Norwegian University of Life Sciences) for comments/discussions of my surveys; Barbi Nilsen (NFLI) for designing electronic maps; Ole Fredrik Ugland, Lars Hansen and Haakon Korsgaard (all TNS Gallup) for the administration of the survey data collection. I have also benefited from comments to my work from Anna Bartczak, Shelby Gerking, Michael W. Hanemann, Jette B. Jacobsen, Bengt Kriström, Mark Morrison, Randall Rosenberger, Jon Strand and Laura O. Taylor to whom I am grateful.

Funding for my research is greatly acknowledged from "Letterstedtske Forening" (Paper 1), Norwegian Research Council (grant 173230/I10, "AREAL-programmet") and "Skogtiltaksfondet" (Paper 1, 2 and 4), and the Environment and Economy Programme for Southeast Asia (Paper 3). Further, support for my stay at University of California, Berkeley autumn 2005, is greatly acknowledged from "Eckbos Legater", "Professor Wilhelm Keilhaus Minnefond", "Norges Banks Fond til økonomisk forskning", and "Dosent Theodor Landmark og hustrus familie- og landbrukslegat".

The papers included in this thesis represent a subset of a wider research agenda I have pursued during my 3 year PhD scholarship. Other papers include[1]:

1. Vennemo, H., K. Aunan, H. Lindhjem and H. M Seip "Environmental pollution in China: Status and trends". By invitation. Forthcoming in *Review of Environmental Economics and Policy* **3**(2), Summer, 2009.

2. Bartczak, A., H. Lindhjem, S. Navrud, M. Zandersen and T. Zylicz "Valuing forest recreation on the national level in a transition economy: The case of Poland". *Forest Policy and Economics* **10**(7-8): 467-472, 2008.

3. Lindhjem, H., T. Hu, Z. Ma, J. M. Skjelvik, G. Song, H. Vennemo, J. Wu and S. Zhang, (2007) "Environmental Economic Impact Assessment in China: Problems and Prospects". *Environmental Impact Assessment Review* **27**(1): 1-27, 2007.

[1] Fulltext of these and other ongoing research can be found at www.lindhjem.info/phd.htm

4. Lindhjem, H. and S. Navrud “Internet CV surveys - a cheap, fast way to get large samples of biased values?”. Manuscript under review by European Association of Environmental and Resource Economists, Amsterdam, June 2009.

5. Bartczak, A, H. Lindhjem and A. Stenger (2008) "Review of benefit transfer studies in the forest context". Published in Bergseng, E., G. Delbeck and H. F. Hoen (eds) (2009) Proceedings *of the Biennial Meeting of the Scandinavian Society of Forest Economics,* Lom, Norway, 6th – 9th April, 2008: pp 276-305.

Finally, I want to thank my parents for encouraging me to be curious, to learn and to work towards maximising my potential – albeit as always in the dismal science – under constraints of various sorts.

Henrik Lindhjem

Oslo, February 2009

Table of contents

Introduction and summary

This thesis consists of four papers on methodological issues in meta-analysis (MA), benefit transfer (BT) and environmental valuation. The papers are independent and can be read separately, though they are closely linked thematically and empirically. The first paper presents a MA of stated preference studies valuing non-timber benefits in Norway, Sweden and Finland (Fennoscandia) over the last 20 years. It investigates using different meta-regression models to what extent willingness to pay (WTP) estimates conform with standard expectations, tests a number of novel hypotheses and identifies gaps in the literature. Papers 2-4 then each pick up an important research theme that follows and is closely related to the first paper. The second paper utilises the same data to investigate the precision in using MA for international BT, as compared with simpler and more common BT techniques. The third paper investigates, using a different and more extensive dataset of biodiversity and nature conservation values from Asia and Oceania, how the heterogeneity or scope of the MA data influences the results of different meta-regression models and their precision when used for BT. The fourth paper tests using a primary contingent valuation (CV) data set of WTP for forest protection in Norway, whether people state different WTP if asked as individuals or on behalf of their household. This paper tests in a more controlled way the question also investigated in the first paper, the answer to which has important implications for CV applications and for the use of such data both for MA and BT.

This introduction first provides the background to the research themes investigated in the thesis, and gives a condensed review of theory, methods, and applications of environmental valuation, MA and BT, in section two. Section three presents the research questions and the methods and datasets used to investigate them. The fourth section summarises the main contents, results and contributions of each paper and links them together. Section five concludes.

1. Background

No other subfield within environmental economics has grown at the same rate as economic valuation of environmental goods[2], i.e. the pursuit of measuring what we are or would be willing to give up of other goods and services to improve the environment (or avoid that it is damaged). One important reason for this trend is the increased demand from government departments, public agencies and other institutions for monetary estimates of environmental goods and services provided to society. Most of these goods and services have no prices and effectively no value in ordinary markets. The most common applications of such monetary estimates are in (Navrud and Pruckner 1997): (1) Cost-benefit analysis of public projects that have environmental impacts; (2) In reviewing effects of new regulations and policies; (3) Assessing compensation for damages to natural resources from oil spills etc; (4) Estimating environmental costs as a basis to set environmental taxes, and; (5) Measuring "green" Gross Domestic Product adjusted for use and degradation of environmental resources.

The environment is not just valuable in that we can use it to our immediate benefit, for example harvesting resources, for recreation, or for breathing clean air, termed use value. It may also be valuable to a great number of people that, for example, wilderness exists even if they never intend to visit, yielding non-use or existence values (Krutilla 1967). There are two main ways to measure environmental values: (1) Ask people in

[2] For example as illustrated by the growth in publications presented in Carson and Hanemann (2005).

surveys to *state* their preferences for a change in the provision of an environmental good (giving both use and non-use values) and (2) Observe people's behaviour and how it is linked with the environmental good we are interested in, i.e. people *reveal* their preferences (giving just use value). The main approaches under (1) are contingent valuation (CV) if people are asked their willingness to pay (WTP) and choice experiments (CE) if they are asked to choose between different combinations of attributes of the environmental good and the costs of providing them. The family of methods under (2) is more diverse, but includes at least five approaches: (i) measuring people's travel costs to visit a site (the travel cost method – TCM); (ii) measure how house prices (or other prices such as wages) vary with environmental (or risk) attributes to infer their value (hedonic pricing); (iii) measure the damage to people's health (cost of illness), property, agricultural crops etc. caused by pollution (the damage cost method); (iv) measure people's expenditures to avoid or protect themselves against environmental impacts (the defensive behaviour method); and finally (v) measure the cost of replacing an ecosystem and its services (the replacement cost method).

The growing stock of valuation research contains important, accumulated knowledge about both people's preferences for environmental goods and how valuation methods and different contexts influence valuation results. However, paraphrasing Glass et al. (1981: p11)[3], results of much of this work "are strewn among the scree of a hundred journals and lies in the unsightly rubble of a million dissertations." While traditional qualitative literature reviews are still common in the literature to synthesise empirical research, the tool of meta-analysis (MA) has increasingly been used in economics in general, and in environmental economics in particular. MA is defined as "..a body of statistical methods that have been found useful in reviewing and evaluating empirical research results" (Stanley 2001). MA combines the results of several studies of a similar phenomenon, normally done by identification of a common measure of "effect size", which is modelled using a form of meta-regression analysis[4]. MA is typically used for three main purposes in the analysis of empirical valuation research (Smith and Pattanayak 2002): (1) research synthesis (i.e. a quantitative type of literature review); (2) testing hypotheses; and (3) benefit (or value) transfer (BT). Due to time and budget constraints, and because not all situations may require a full, primary valuation study, benefits can instead be transferred from the literature to the policy situation where a value estimate is needed, i.e. a meta-analytical BT can be performed (MA-BT). The literature on MA in environmental economics is still relatively immature in all three areas above. Much research remains to understand how the tool of MA can be used to its full potential environmental valuation and BT research, and also ensure that MA is not "abused" (to paraphrase the title of Nelson and Kennedy 2009).

The next section reviews the use of MA in environmental economics, with particular focus on issues of relevance to this thesis.

2. Meta-analysis, benefit transfer and valuation in environmental economics

2.1 Theory

The first application of MA in environmental and natural resource economics was according to Florax (2002) the assessment of noise and property values by Nelson (1980). However, Nelson himself in a recent survey of 140 MA studies in the field

---

[3] Originally quoted in Stanley and Jarrel (2005).

[4] In the wider MA literature, meta-regression analysis, which is the technique mostly used in economics, is sometimes seen as only one type of MA. In the following the two terms are used interchangeably.

(Nelson and Kennedy 2009) credits instead a cluster of studies published ten years later: Smith and Kaoru (1990a, b) and Walsh et al. (1989, 1990). The early investigations of MA in economics more generally started around the same time (a key publication is Stanley and Jarrel 2005[5]). Hence, the use of MA in environmental economics is relatively recent, compared to other fields such as clinical psychology where MA has been applied for at least 10-15 years longer (Schulze 2004; Lipsey and Wilson 2001). In comparison, environmental valuation itself dates back to the 1950-60s USA (Hanemann 1992). Given the relatively recent introduction of MA in the environmental valuation field, both the theory and methods used are characterised by a certain degree of immaturity. There is for example no unified or generally agreed MA methodology applicable to different types of environmental valuation meta-data – for the three areas of MA applications mentioned above – though a recent study have made a first attempt at synthesising best practice (Nelson and Kennedy 2009). For example, there is still an ongoing debate in the USA about how and if MA can be used to analyse empirical estimates of the value of statistical life for research and policy use (USEPA 2006; 2007). Compared to the use of CV for example, there are no guidelines for MA such as those given by the NOAA panel (Arrow et al.1993) or SEPA (2006). Hence, the field is evolving along a steep learning curve.

We begin by linking MA to utility theory. A common theoretical model, based on Bergstrom and Taylor (2006), considers the WTP for some environmental non-market commodity (Q) or some set of services (S) provided by Q. The underlying indirect utility function is:

(1) $V = V_j(P_j, M_j; Q_j, QUAL_j, SUB_j, H_j, I_j)$

[5] This paper was originally published in 1989, and reprinted in 2005. It is the 2005 reference which is used here.

Where P_j is a price index of market goods faced by individual j; M_j is her income; Q_j and $QUAL_j$ are quantity and quality of the environmental commodity available to individual j, respectively; SUB_j is a measure of substitutes; H_j is non-income characteristics of individual j's household; and I_j is the information set available to the individual. Introducing an environmental change, the bid function for the representative individual j for this change derived from (1) is[6]:

(2) $WTP = f(P_j, M_j, Q_j^T - Q_j^R, QUAL_j^T - QUAL_j^R, SUB_j^T - SUB_j^R, H_j)$

Where R indicates the reference situation ("status quo") and T the target state-of-the-world. According to Bergstrom and Taylor (2006) there are three main ways to build a bridge between the standard bid function in (2) to an empirical specification "elastic" enough to be estimated using meta-regression models. The first way, termed "strong structural utility theoretic approach" involves specifying a structural form of the indirect utility function (1) and then derive (2), which inherits the restrictions following from (1) (an example is Smith et al. 2002). The second, and empirically more flexible way, is termed "weak structural utility theoretic approach", in which the underlying variables in the bid function are assumed to be derivable from some unknown utility function. However, flexibility is maintained to introduce other explanatory variables into the model, such as study and methodological characteristics, that do not necessarily follow from (1). This is the most common approach in current MA studies in environmental economics. The third approach, "non-structural utility theoretic approach" is at the other end of the spectrum compared to the first approach, where the link to the underlying utility function is not explicitly specified. Based on the weak structural utility theoretic

[6] Assuming constant prices, and suppressing the information variable for simplicity.

approach above, the following meta-regression equation with three main groups of explanatory variables, is typically specified[7]:

$$(3)\ WTP_{ms} = \beta_0 + \beta_X X^j_{ms} + \beta_M M^k_{ms} + \beta_S S^l_{ms} + e_{ms} + u_s$$

Where WTP_{ms}, the effect size in this meta-regression model, is mean WTP estimate *m* taken from valuation study *s* (for example WTP per household per year); variables *X* are characteristics (*j*) of the environmental good or site valued (i.e. as accurate as possible proxies to the *Q* and *QUAL* variables above); *M* are the methodological characteristics of the valuation study (*k)* (i.e. which is additional to variables in (2)), and *S* are the socio-economic characteristics (*l*) of the sample or population surveyed (i.e. covering the variables M and H in (2)). β_0, β are constant term and parameter vectors for the explanatory variables, and e_{ms} and u_s are random error terms for the measurement and study levels, respectively[8].

Equation (3) is the basis for most classical applications of MA in environmental economics. In the next section some challenges and research gaps related to research synthesis and particularly hypothesis testing in MA are discussed in connection with equation (3). Section 2.3 discusses some important issues in data collection, model estimation and the use of MA for BT.

[7] As Bergstrom and Taylor (2006) also acknowledge, one increasingly popular alternative to this classical MA approach, not considered in this thesis, is the use of Bayesian modelling techniques (e.g. Moeltner et al. 2007, Moeltner and Woodward 2009).

[8] Contrasting equation (3) to a single study, it is worth noting that this bid function is often specified as $WTP_i = a + bX_{ij} + cY_{ik} + e_i$, where WTP_i is WTP of respondent *i*, *X* site/good characteristics (*j*), *Y* respondent characteristics (*k*), e_j random error, and the number of observations is equal to the number of respondents (Brouwer 2000).

2.2 Research synthesis, hypotheses testing and environmental valuation research

The most common application of MA in environmental valuation is as a tool for quantitative literature review. This means trying to explain the variation in WTP observed in the literature using more or less standard variables from theory and empirical research known to influence WTP, and estimating different meta-regression models based on (3). Several MA studies reviewed in Nelson and Kennedy (2009) are of this sort. Such research synthesis can be very useful to give a clear overview of for example WTP for a particular environmental good and to detect and open new avenues of research. Typically, research synthesis using meta-regressions in this way is part of a check of the reliability of the data and a first step towards deriving values for policy use in BT (see next subsection).

Another growing and highly valuable use of MA is in the testing of new and old hypotheses in valuation research. This can imply introducing new variables from theory or empirical research into (3) that have not been investigated before in a systematic way, and test if their coefficients are statistically different from zero. This exercise often forms part of a more standard MA research synthesis. It can also imply collecting, coding and arranging meta-data for the main purpose of testing a specific hypothesis in environmental valuation. Some hypotheses explored through MA to date include "classical" questions in the non-market valuation literature such as the WTP vs. willingness to accept (WTA) compensation disparity and income effects (Horowitz and McConnell 2002, 2003, Sayman and Onculer 2005, Schlapfer 2006), WTP's (in)sensitivity to change in quality or quantity ("scope") of the good (Smith and Osborne 1996; Ojea and Loureiro 2008), WTP's sensitivity to income (Jacobsen and Hanley 2009), convergent validity of different valuation methodologies (Carson et al.

1996), the relationship between use values and non-use values (Johnston et al. 2003), and differences in real and hypothetical WTP ("hypothetical bias") (List and Gallet 2001; Murphy et al. 2005). These are just a subset of interesting hypotheses in environmental valuation research that could conceivably be tested using MA.

To date most MA studies have investigated WTP for fairly simple goods (such as air and water quality, noise, recreation days) often within a national setting (e.g. Desvouges et al. 1998; Rosenberger and Loomis 2000a; van Houtven et al. 2007). There is a recent trend towards using MA to study more complex goods with higher non-use values in international settings (e.g. wetlands, coral reefs, forests and biodiversity) (e.g. Brander et al 2006; 2007; Richardson and Loomis 2009). However, comparatively less is known about preferences for complex goods and differences in such preferences between countries and cultures[9]. More research is needed, especially when attempting to use such valuation data for BT.

Another area where MA may potentially be very valuable in increasing our understanding of people's preferences, is the introduction and testing of new variables not normally included in (3), for example from psychology, sociology or family economics. Three examples from psychology include: (1) the tendency people have to value avoiding a loss higher than an equal-sized gain (called loss aversion) (Kahneman and Tversky 2000); (2) the calculation and discounting errors people typically make when asked WTP using different payment formats (e.g. paying per month vs. per year or over several years) (Rabin 1998); (3) the tendency to let the time of year influence the WTP, respondents' displaying a "season illusion". An example from sociology and family economics is that stated WTP may depend on the resource allocation model and

[9] For example as documented in the cross-country comparison of hypothetical bias by Ehmke et al. (2008).

the level of economic integration within a respondent's household (e.g. Strand 2007). Such issues lend themselves well to investigation in MA, though to fully understand and pin down the reasons for the patterns detected, primary valuation research, e.g. CV, may be better suited. This thesis makes a contribution to testing new explanatory variables in MA, and investigates one of the hypotheses more in depth in a primary CV study.

2.3 Data and modelling issues in meta-analysis and benefit transfer

A host of challenges are involved in conducting MA, both in relation to the type of data typically available in the literature and the econometric methods used to estimate (3). How these challenges are dealt with have implications for research synthesis and hypothesis testing, and for the third use of MA, meta-analytical BT (MA-BT). MA-BT involves estimating (3) based on previous studies, and transferring the function to a policy context of interest where there is no value estimate available. More specifically, policy context values are inserted for the characteristics of the environmental good or site to be valued (variables X in (3))[10] and for the socio-economic characteristics of the relevant population (variables S in (3)). In addition, values for the methodological characteristics are set, for example at the average of the studies in the meta-dataset, at some "best practice" values (e.g. only studies using dichotomous choice question format) or drawn from a distribution (e.g. Johnston et al. 2006; Stapler and Johnston 2009).

MA-BT is typically assumed to be more accurate than using simpler BT techniques, such as a single (adjusted) WTP estimate from a study that values a similar good in a similar context (so-called unit value transfer) or the average WTP from several studies (Navrud and Ready 2007). The main reason put forward is that more information can be transferred to the policy context. Compared to unit value transfer or function transfer based on a single study MA utilizes information from several studies providing more rigorous measures of central tendency that are sensitive to the underlying distribution of the study values (Rosenberger and Loomis 2000a). To date, only a few studies have investigated the validity and reliability of MA-BT (Santos, 1998; Rosenberger and Loomis, 2000a; Shrestha and Loomis, 2001; 2003; Santos, 2007; Shrestha et al., 2007). Santos (2007) is the only study attempting a comprehensive comparison of two versions of a domestic MA-BT with simple BT techniques often used in practice. Much is still unknown about the conditions under which MA may be reliably used for BT. There has been little testing of how different MA models perform in BT and how heterogeneity of data (for example in terms of valuation methods included, definition of the effect size and environmental good valued), influence results. Given the exponential growth in MA studies in the literature, as documented by Nelson and Kennedy (2009), more research is clearly needed to understand how errors in BT can be controlled when MA is increasingly used for e.g. cost-benefit analysis.

MA data from environmental valuation studies have certain characteristics in common with other "standard" MA data from for example epidemiology and clinical psychology, but also have their own characteristics creating particular challenges for data collection and modelling. The first challenge is to define a relevant "effect size statistic capable of representing the quantitative findings of a set of research studies in a *standardized form* that permits *meaningful numerical comparison and analysis* across studies" (Lipsey and

[10] This is provided that the policy site characteristics are represented within the range of the meta-data. Otherwise the meta-model would be unsuitable for BT to that particular policy site.

Wilson 2001:5) (my italics). For clinical trials testing a new drug on experimental groups, the effect size is relatively easy to define, standardise and compare between studies. For environmental valuation studies heterogeneity in study designs, in types of goods and populations, and in the contexts of the valuation exercise reported in the literature is by definition much greater. Hence, from the start, meta-analysts may need to compromise on the strict criteria of MA in other disciplines which to a larger extent draw effect size results from (better) controlled experiments, e.g. as discussed in USEPA (2006).

The effect size from environmental valuation studies is typically defined as some consumer surplus measure, mostly WTP for a change in the provision of the quality or quality of an environmental good, as represented by equations (2)-(3). A first step in standardising the measure in the literature is to adjust estimates from studies conducted in different years by inflation, and estimates from different countries by purchasing power parity exchange rates (Ready and Navrud 2006). Further, some MA studies convert WTP estimates reported in different formats to per trip for recreation, per year, per household, per environmental change (e.g. WTP/hectare or WTP/water quality) or similar. These conversions use implicit assumptions about people's preferences (e.g. regarding discounting, how values of environmental goods change over time etc.) that may or may not be true. Some MA studies relax the strict standardisation of the effect size, instead trying to control for variation using explanatory variables in (3). Given some degree of standardisation, the effect sizes will be measured in the same unit between studies. However (referring back to the quote by Lipsey and Wilson above), whether this permits meaningful numerical comparison and analysis across studies, also depends on the acceptable level of heterogeneity of the data in terms of variation in the variables X, M and S in (3). In the MA literature there is very little guidance or consensus (or even discussion) on how to balance heterogeneity at the data entry level compared to controlling for such variation using moderator variables in the meta-regressions. The studies in the literature range from measuring WTP for changes in a relatively homogenous goods (e.g. changes in mortality risk, water quality) to more heterogeneous goods (e.g. nature conservation). However, the same MA studies which are careful in controlling the homogeneity of the good valued may include WTP estimates derived from a large diversity of valuation methods (for example hedonic pricing and CV)[11], a criticism for example levied against MA studies of value of statistical life by USEPA (2006). The question of acceptable "scope" of the MA when applied to environmental valuation data is a fundamental one, which to date has not been investigated much neither from a theoretical/conceptual or an empirical perspective[12]. And it has potentially important, and as of yet largely uncertain, implications for MA-BT.

The characteristics of environmental valuation meta-data pose particular challenges in the data collection and model estimation. As with other data, there is a trade-off between quality or completeness of the dataset and the size. Most valuation studies report many estimates of mean WTP from the same or different samples, but may not give a complete reporting for all variables the meta-analyst would like to code. In some cases the reason why estimates from the same study are different may not be explained in detail. Further, many studies do not report mean income, education level and age of their samples, i.e. the *S* variables in (3). Many of the problems the meta-analyst experiences with the data are related to the fact that at least most published studies are

[11] In some cases the good valued may also not be completely idependent from the method used, i.e. subtle differences in the good valued that are not easy to control for/detect may be introduced by the method (e.g. as often observed in CV applications).

[12] A notable exception is Moeltner and Rosenberger (2008) who investigate "optimal MA scope" using Bayesian modeling techniques.

designed to report results of methodological tests, rather than the welfare measures *per se* for use in MA (Loomis and Rosenberger 2006). Hence, expanding the number of variables in (3) will reduce the number of studies having complete reporting for all variables. To increase the meta-data set analysts sometimes increase the heterogeneity of the data (i.e. include more studies) and introduce "higher level" or cruder explanatory variables that fit a wider range of study types.

Some meta-analysts exclude studies on the basis of subjective or objective criteria of quality, e.g. unpublished studies. However, Stanley and Jarrel (2005) recommend "to err on the side of inclusion" of studies and estimates. Including estimates from unpublished studies may also reduce the problem of publications bias, the tendency that significant results are more likely to get published (also known as the "file drawer problem") (USEPA 2006; Rosenberger and Stanley 2007).

Given the challenge to find sufficiently many valuation studies for the same type of good, it is also a challenge to decide how many WTP estimates to include from the same sample or study. A mentioned, a defining feature of valuation data is that many estimates are reported from the same study for different methodological split-sample tests, use of different statistical estimation methods etc. While MA studies in other professions typically only include one effect size estimate drawn from each population (sample), environmental MA studies include many estimates from the same sample or study[13]. The advantages to include more than one estimate are that the full range of available information can be used in the modelling and that the sample size is boosted. The disadvantage is that including several estimates may introduce dependencies in the data that are hard to account for (USEPA 2006).

In some studies, WTP estimates are weighted down by the inverse of the number of estimates from each study, so that estimates count equally in the data. However, this procedure is not recommended by some authors (e.g. Bateman and Jones 2003). In some cases, there may be many estimates from a study that cannot be meaningfully coded and explained given the variables used in the meta-regression equation. There is no explicit consensus or guidance in the literature on how to deal which such observations in MA. Few studies report how they explicitly have collected and coded the data and which criteria have been applied for including or excluding observations. Analysis of the implications for MA results and MA-BT of different exclusion criteria and weighting procedures are also scarce in the literature.

In the wider MA literature, effect size estimates are typically weighted by their precision e.g. as measured by their standard error. This procedure is often difficult to follow in the environmental MA literature, as standard errors are often not reported. A more feasible procedure is to use sample sizes, which are more often reported in the valuation literature, as a proxy for precision.

The discussion above reflects some of the issues related to data collection, coding and preparation of a dataset for meta-regression analysis. The next step is the estimation of the meta-regression models in (3). There are several approaches to estimating the model depending on the assumptions regarding the error covariance matrix. A range of different modelling approaches have been used in the literature. Nelson and Kennedy (2009) provide the first review of meta-regression methods in environmental and natural resource economics. The simplest approach, which has been used in several MA studies (e.g. Loomis and White 1996; Rosenberger and Loomis 2000a), is to treat all

---

[13] The same valuation study can report mean WTP from several different samples, and can also estimate different

measurements as independent replications and hence assume that study level error (u_s) is zero. This model can be estimated using ordinary least squares (OLS). A more advanced approach commonly used in MA is to apply a Huber-White robust variance estimation procedure to adjust for potential heteroskedasticity and intercluster correlation (Smith and Osborne 1996). The cluster is typically defined as the observations coming from the same study (or valuation survey). If correlation exists between estimates within the same cluster, OLS regression will be inefficient and inconsistent in estimated parameters. The Huber-White procedure does not affect the parameter estimates of the model, but provides robust standard errors of the parameters, and therefore influences the hypothesis testing.

More advanced models, increasingly more common in the MA literature, to deal with the panel structure of meta-datasets include fixed and random effects panel models, and multi-level models. Standard statistical tests and procedures can be used to identify and define stratifications of the data (study level is one possibility, but there are others such as estimates from the same author) (Nelson and Kennedy 2009). One procedure to check for panel structures is proposed by Rosenberger and Loomis (2000b). Bateman and Jones (2003) use multilevel models in MA, allowing them to cluster the data in hierarchies accounting for residual variance of estimates for each level. With regards the functional specification of (3), different forms have been used in the literature, with the double log form perhaps the most commonly applied. Only a few studies test different MA regression models on the same data (two examples are Johnston et al. 2003 and Bateman and Jones 2003). Given the immaturity of the field, there is still much experimentation with different types of MA models in the literature, as reflected by the review by Nelson and Kennedy (2009). To date, the implications of the use of different

---

WTP estimates from the same sample (for example applying different statistical estimation procedures).

data, meta-regression models and specifications on MA results and MA-BT have rarely been investigated systematically.

3. Research questions, methodology and data

The previous section identified a number of gaps in current knowledge regarding the use of MA, BT and environmental valuation. On that basis, the current section formulates the research questions the thesis attempts to answer, and describes the methods and data used to answer the research questions.

3.1 Research questions

The research questions investigated in this thesis can be divided into two closely related main themes, corresponding to the interrelated discussions in sections 2.2. and 2.3 above, respectively. The first set of questions (see I below) aims to: (1) increase our understanding of people's preferences for complex environmental goods with high non-use values; and (2) investigate whether the values reflecting those preferences in the literature display a degree of regularity and validity giving us some confidence that such data may be used for BT purposes. The answers to the first set of research questions contribute to the wider environmental valuation and MA-BT literature. The main research questions asked under the first theme are:

I. *Meta-analysis and environmental valuation:*

a. Does WTP for a complex good such as forest protection, multiple use forestry, nature and biodiversity conservation vary in ways expected from theory and empirical research?

b. Do people's stated WTP depend on whether they are asked as individuals or behalf of their household?

c. Do people display a "season illusion" when asked WTP for forest protection, i.e. does WTP vary with seasons?

d. Are people willing to pay more to avoid a loss in forest protection or multiple use forestry than an increase?

e. Does the WTP for forest protection and multiple use forestry increase with the size of the forest?

f. Do people have higher WTP for protection of certain types of species or habitats?

As discussed in the previous section, MA has often been used as a tool for quantitative research synthesis for more homogenous environmental goods in a national setting. However, applications in international settings and for more complex goods are still fairly scarce, but growing. Hence, research question Ia attempts to investigate if valuation data in such contexts using MA, display a similar regularity as compared with simpler goods. Question Ib draws on a small, but fast growing literature connecting theories of family economics with environmental valuation research (e.g. Strand 2007). The consensus in the literature seems to be that household WTP is higher than individual WTP, although this has not been tested empirically in for environmental goods in CV. The answer to this question has implications for the aggregation of welfare estimates for policy use and for the coding and treatment of WTP estimates in MA. Questions Ic-d have their origin in psychology or behavioural economics. Standard neoclassical environmental economics would predict an individual's stated WTP should not depend on the time of the year or the season the person is asked, as WTP would take into account the (discounted) stream of benefits over all years and seasons (Jakus et al. 2006). The same should be valid for forest protection and multiple use forestry, something question Ic will test. Question Id investigates people's preferences for avoiding a loss vs. achieving a gain. The extensive literature in psychology shows that people tend to value losses higher than equal-sized gains (Kahneman and Tversky 2000). This question is also tested in the context of forest valuation. Question Ie is an old, outstanding issue in environmental valuation research, though not before investigated specifically for forest protection and multiple-use forestry in the MA literature. It is an important question both for the validity of the CV method and for the use of per hectare WTP measures for BT. Finally, question If attempts to understand better people's preferences for nature and biodiversity conservation, especially their WTP for certain types of charismatic species and valuable habitats.

The second set of research questions (see II below) probes into more specific methodological issues in MA and BT, investigating the performance of different MA models and datasets, and the implications for explained variation and MA-BT reliability. The answers to the second set of questions contribute to the growing literature on MA and BT in environmental economics, and to the broader MA literature. The main research questions asked under the second theme are:

II. *Meta-analysis and benefit transfer methods:*

a. How reliable is MA-BT compared to other simple and more common BT techniques?

b. How does the precision of MA-BT depend on the types of MA models used?

c. How sensitive are meta-regression results and the precision of MA-BT to the level of heterogeneity of the MA data in terms of the good valued and valuation methods used?

Question IIa attempts to test more systematically, what is often just assumed in the literature, that MA-BT is more reliable than using simpler BT techniques. Question IIb investigates the precision in the predictions of MA-BT models depending on different model specifications. Finally, question IIc investigates a question of fundamental importance to MA in environmental valuation: how homogenous is homogenous enough in MA and how sensitive are results and MA-BT precision to model choices and scope of the MA?

3.2 Methodology and data

This section briefly describes and explains the methodologies applied and the data used to answer the research questions. Content, results and contributions of each paper are presented in the next section.

Research questions Ia-f are answered within a classic meta-regression analysis framework, using two main types of datasets and different meta regression models. Questions Ia-e are first analysed based on a meta-dataset consisting of stated preference studies (mostly CV) which ask people their WTP for protection of forests or for increases in more environmentally cautious forestry practices (multiple use forestry) on a local, regional or national scale in Norway, Sweden and Finland. The studies from which the meta-data are drawn span two decades starting from the mid 1980s. The data were collected during 2005-2006, coded and effect sizes (annual WTP) standardised following common procedures in the MA literature. Many of the estimates reported from the same study varied only along dimensions of statistical modelling choices (especially for dichotomous choice data) in the source studies, which were often impossible to code (accurately) due to insufficient reporting. Instead of excluding the values from such studies or to include all unweighted, the average of such estimates from each study were included in the meta-regressions. This was done to retain the information contained in the study, while at the same time reducing the influence of the "statistical noise" the meta-regression model would not be able to explain. Sensitivity of this choice was briefly assessed, though more thorough analysis based on these data were left for further research. This issue is part of the wider discussion on how to define the scope of the MA, in terms of limiting datasets or trying to control for variation in meta-regressions using moderator variables, an issue investigated further for a different meta-dataset under research question IIc (see below). The final dataset had 72 observations. Explanatory variables were coded based on expectations from theory and other empirical studies, of particular relevance for forest valuation. Specific dummies were included to test the hypotheses Ib-e, based on whether coefficients were statistically different from zero. Four meta-regression models were run: (1) A standard linear OLS; (2) & (3) robust Huber-White estimations using linear and double-log specifications, respectively; and (4) a Huber-White linear model where variables not significant at the 20 percent level were left out (a common model used for MA-BT).

Sensitivity analysis was carried out to investigate effects of different standardisation procedures (e.g. base year, including all observations, current vs. PPP adjusted exchange rates etc).

Research questions Ia-b,f were also tested for a different meta-dataset. Around 100 studies were collected during 2006-2007 valuing nature and biodiversity conservation in Asia and Oceania, using the full range of environmental valuation methods. The overall dataset consisted of 550 observations (after 27 estimates larger than two standard deviations from the mean had been excluded). For the meta-regressions, the dataset was divided into two primary levels of scope, according to level of homogeneity of the good and methods used: (1) Endangered species; and (2) Biodiversity and nature conservation more generally. The endangered species data included 124 WTP estimates from 16 studies using CV to value the preservation of single or multiple species. The second level of the data, included the studies from Level 1 and all the rest of the studies that value nature conservation more generally, with different types of methods (though the majority also use CV here). All in all the Level 2 dataset contained between 67 to 95 studies and 390 to 550 estimates, depending on the cleaning procedures and the subsets of the data used in the meta-regressions. Information was gleaned from the studies, coded and standardised following roughly the same protocol as the meta-dataset from the Nordic countries. Explanatory variables were coded based on expectations from theory and previous empirical studies. Additional variables were included to test hypotheses Ib,f. The division of the dataset into two levels was done to investigate research question IIb-c. This in line with the recommendation in Nelson and Kennedy (2009) that "meta-regressions also are estimated on more homogenuous subsamples, especially if a policy application is involved". A Breusch-Pagan Lagrange multiplier statistic test was conducted confirming the presence of panel effects in the data. A

Hausman χ^2 test was conducted confirming a random effects panel model. Six random effects meta-regression models were conducted for each of the two levels of data, respectively. The models differed in the level of heterogeneity (in terms of diversity of valuation methods, diversity of the good valued, use of GDP as proxy for unreported income, and range of moderator variables included). Eight alternative models where variables not significant at the 20 percent level were excluded, were also conducted for MA-BT tests.

Research questions IIb and the second part of question IIc (i.e. precision in MA-BT) were tested on the two main meta-datasets described above. The performance of MA-BT can only be accurately assessed if we knew the "true value", or an estimate of this, for a range of sites of interest, and then used the MA models to predict the value at those sites (i.e. inserting values for the *X, M, S* variables in (3)), and calculate so-called transfer errors (TE)[14]. Brander et al. (2006; 2007) and a few other studies, use different "benchmark" values from within their sample or from new studies to "simulate" the true value to assess TE performance. The same method was followed for research questions IIb-c. A jackknife data splitting technique was used to estimate n-1 separate meta-regression equations to predict (or forecast) the value of the omitted observation in each case (i.e. "the policy site"). The percentage difference between observed and predicted values can then be calculated, the TE in simple exercise, as well as the overall median and mean TE for all observations in the data[15]. The level of (median and mean) TE for both datasets were calculated and compared for the range of meta-regression models

---

[14] $TE = \frac{| WTP_T - WTP_B |}{WTP_B}$, where T = Transferred (predicted) value from study site(s), B = Estimated (observed) true value ("benchmark") at policy site.

[15] The mean prediction error for the *n* predictions is often termed Mean Absolute Percentage Error (MAPE).

and specifications described above for both datasets to give a good basis to answer research questions IIb-c.

For the meta-dataset based on the Nordic forest valuation studies, a further test of MA-BT performance was conducted (research question IIa). The performance of the two models with the lowest TE overall from the first MA-BT exercise were compared with simple BT methods typically used in practice, i.e. transfer of WTP from a study from the most similar site or use mean WTP from studies of similar domestic or international sites. This comparison was conducted using a simulation technique more closely resembling how an actual BT exercise would be conducted. The protocol for choosing the most similar study, or "best study" was simply to choose from the studies of similar site characteristics to the policy site, the one study with the lowest TE (i.e. the mean WTP closest to the policy site WTP)[16]. The overall TE for these different BT methods are also compared with just taking the raw mean WTP of all studies in the dataset regardless of the similarity with the policy site in question, i.e. an upper ceiling to BT performance ("the worst you can do").

Finally, question Ib was tested more in depth using primary CV data and a 2x2 split sample design. In the first sample the respondent first got a household WTP question and then prompted to instead think about personal WTP she got an individual WTP question[17]. After the second WTP question, the respondent was automatically directed in the survey to a question offering 4-6 reasons for why household WTP was higher, lower or the same as individual WTP. The design was the same in the second sample, except the order of the WTP questions and the way the prompt was phrased were

---

[16] This can of course not be done in reality, since the benchmark value at the policy site is not known. Hence, this is the reason it is called the "the best one can do".

reversed. The data used for the split sample test were collected from an Internet survey as part of a large multi-mode CV survey of forest protection in Norway. A professional polling firm collected the data in the autumn of 2007 from a pre-recruited nation-wide panel of respondents, giving response rates of 72 and 69 percent for the two samples, respectively. Each sample had a gross number of observations of around 400. The survey was designed following similar forest protection surveys well tested and tried in the Nordic context and recent best practice guidelines in the CV field (e.g. Bateman et al. 2002; SEPA 2006). The instrument went through thorough testing in focus groups and two small pilots (using both internet and personal interviews). Mean WTP for the two samples and WTP questions were estimated assuming a log-normal distribution following standard procedures given in Cameron and Huppert (1989), after zero and protest responses had been removed. Bootstrap methods were used to calculate confidence intervals and standard statistical tests applied to compare mean WTP between and within samples. Further, analysis is made of the reasons respondents stated for choosing to go up, stay at the same level, or reduce their bid when confronted with the reverse response unit. Finally, a simple probit model is used to investigate whether household and respondent characteristics can explain the observed relationship between household and individual WTP.

4. Thesis content, results and contributions

This section presents summaries of content, results and contributions of the four papers making up this thesis. Paper 1 addresses research questions Ia-e, Paper 2 questions IIa-b, Paper 3 questions Ia-b,f and IIb-c, and finally, Paper 3 addresses question Ib. Papers

---

[17] The WTP questions were formulated either as ”what are you personally willing to pay?” or as ”what are you willing to pay on behalf of your household?”.

1, 2 and 4 are published in international, peer-reviewed journals. Paper 3 is currently under journal review.

Paper 1: 20 years of stated preference valuation of non-timber benefits from Fennoscandian forests: A meta-analysis[18]

CV, and to a lesser extent choice experiments, have been conducted to value non-timber benefits from forests in Norway, Sweden and Finland for about 20 years. The paper first reviews the literature and summarises methodological traditions in the three countries. Second, a meta-regression analysis is conducted explaining systematic variation in WTP for forest protection and multiple use forestry by differences in survey methodology, good characteristics, socio-economic and other variables. The meta-regression model results are promising in response to research question Ia, with regard to revealing systematic and expected variation in WTP along methodological variables, and to some extent along various characteristics of the forest good. This is an indication of validity of the valuation research. For example, WTP is shown to depend in largely predictable ways on the type of WTP question format (open-ended max WTP vs. dichotomous choice), mode of data collection (mail surveys with low, medium and high response rates vs. in-person interviews), payment vehicle (voluntary, forced), and payment format (annual, one time, monthly). A subset of the WTP data was not sensitive to socio-economic characteristics (education, income, age). Unpublished studies and master theses (two imperfect indicators of study quality) generally give lower WTP values, unknown for which reasons.

---

[18] This paper was written independently and is published in *Journal of Forest Economics*, Volume **12**, Issue 4, February 2007, Pages 251-277.

For non-methodological variables, it is shown in response to research question Ib that WTP is higher if stated individually compared to on behalf of the respondent's household. The reason for this result is uncertain, but may be explained by the fact that a person asked for household WTP automatically is forced to think about a restrictive family budget constraint (compared to a more generous individual, private consumption budget). However, this interpretation may be unlikely in light of the results (later) found in Paper 4. Further, to research question Ic, WTP is found to be higher during the spring/summer season compared to the autumn/winter season, demonstrating that people display a "season illusion". Testing different models and subsets of the data, WTP is found to be insensitive to the size of the forest valued, casting doubt on the use of simplified WTP/area measures for complex environmental goods (research question Id). The MA results also show that geography (urban; local; regional), year and policy type (full protection vs. multiple use forestry) are important, but WTP does not seem to be different between the three countries. The results are fairly robust to changes in model specification and meta-analysis scope (i.e. the number and type of observations included in the meta-data). It is acknowledged that the implications of the weighting procedure applied to deal with observations which varied mostly due to unexplained econometric assumptions applied in the source studies, could have been investigated more thoroughly. However, there is not much consensus or guidance in the literature to follow. However, given the issues experienced during the MA process and the sensitivity analyses that *were* conducted, the paper concludes that varying the scope of the MA as part of assessing sensitivity of results is a potentially important and so far largely overlooked area in MA research. This is the topic of Paper 3 of this thesis.

Paper 2: How reliable are meta-analyses for international benefit transfers?[19]

Paper 1 uses MA primarily for research synthesis and hypothesis testing, and does not assess the potential for using the results (i.e. the estimated MA functions) for BT purposes. This paper fills this latter gap, and thus answers research questions IIa-b. Only a limited number of studies have tested the use of MA-BT, and these are typically based on national studies only. However, MA of valuation studies across countries is a potentially powerful tool for BT, especially for environmental goods where the domestic literature is scarce. Based on the same dataset as for Paper 1, this paper tests the reliability of international MA-BT compared with simpler BT techniques often used in practice. The studies included in the MA are relatively homogenous in terms of valuation methodology and all three countries have similar cultural, institutional and economic conditions. Reliability of BT is measured in terms of transfer error (TE) – the percentage difference between a "true" benchmark value and the transferred estimate – and the paper compares TE across meta-model specifications and restrictions, and between alternative ways of conducting BT based on the same data. The initial check of within and out-of sample predictions of four meta-models shows substantial variation in performance. The best two models give median and mean TE of between 25-34 percent and 39-62 percent, respectively. The TE is lower for higher WTP estimates. The two best models are both double log specifications, where the first model has excluded one observation giving very high TE and the second model has excluded variables not significant on the 20 percent level.

In the comparison of BT techniques (research question IIa), testing the two best models above, MA-BT shows mean TE of between 47-126 percent (median 37-70 percent)

[19] This paper was co-authored with Associate Professor Ståle Navrud and was published in *Ecological Economics* Volume **66**, Issue 2-3, Pages 425-435, 2008.

depending on the model. The second model, with reduced number of variables, performs better. A simple transfer based on the mean of domestic studies with similar site characteristics to the policy site yields a mean TE of 86 percent (median 41 percent), as compared with 62 percent (median 7 percent) if a best study estimate could be chosen from a domestic study. Including international studies in the simple mean transfer increases the TE substantially to 166 percent (median 85 percent). The best simple BT technique yields TE in the middle of the range of the two international MA-BT models. Based on these results, it is questioned whether the use of MA for practical BT achieves reliability gains justifying the increased effort. The paper concludes that more MA-BT tests should be performed for other environmental goods and other countries before discarding international MA as a tool for BT. Paper 3 takes up this challenge for a more diverse and larger dataset.

Paper 3: Meta-analysis of nature conservation values in Asia and Oceania: Data heterogeneity and benefit transfer issues[20]

Paper 3 takes stock of studies estimating WTP for conservation of endangered species, biodiversity and nature more generally in Asia and Oceania. The MA shows that nature conservation is highly valued in the region. Dividing the dataset into two levels of heterogeneity in terms of good characteristics and valuation methods, the paper shows using six different meta-regression models for each of the two levels of data, that the degree of regularity and conformity with theory and empirical expectations as well as the explanatory power of the MA models is higher for the more homogenous dataset of

[20] This paper was co-authored with Dr Tran Huu Tuan (Hue University, Vietnam). It is submitted and under review by *Environmental and Resource Economics*, and European Association of Environmental and Resource Economists for the Annual meeting in Amsterdam, June, 2009.

endangered species values, as expected (research questions Ia, IIc). In fact, though the species are different, the values to preserve them generally follow predictable patterns. For example, in response to research question If, mammals are generally valued higher than other species, likely due to the “charismatic” nature of this family. In response to research question Ib, WTP stated on behalf of the household is found to be higher than individual WTP only in two models of the more homogenous data. This result does not carry over to the second level data. Further, WTP increases significantly with income (elasticity is around 0.8) (level 1 data).

The analysis of the endangered species data show that around half of the variation in the best model is due to non-study specific observable characteristics of the good and population surveyed, boding well for use of such data in BT applications. However, in response to research questions Ia and IIc, increasing the scope of the MA, i.e. gradually including more heterogeneous observations, generally preserves some of the regularity and the explanatory power of some the models is in the range of other MA studies of goods typically assumed to be more homogenous (such as national water quality, recreation days etc). Specific types of habitats are not consistently valued differently across the models used.

Subjecting both dataset levels to a simple check of benefit transfer error (TE), using the MA models to predict observations one-by-one when excluded from the datasets, show for the best models median (mean) TE of 23 (45) percent for the endangered species data and 46 (89) percent for the more heterogeneous nature and biodiversity data. This is in the low range compared to other MA studies. We also run models where variables not significant on the 20 percent are taken out, a common type of model for MA-BT, but do not find that these models change TE levels in a systematic way. This result contrasts with what was found in the MA-BT tests in Paper 2. Results from the standard models suggest that the levels of forecasting errors may approach acceptable levels for policy use. It is also clear from the results that for example including values estimated using a more heterogeneous set of methods for the second level of data, even a fairly broad range of covariates is unable to explain and control for the variation in a satisfactory way, translating into large mean TE. In other words, in response to research question IIc, while median TE are fairly robust, mean TE seem fairly sensitive to the type of meta-data and choice of meta regression model.

A more careful testing of explanatory variables and MA models than we have done (for example including interaction effects) may be required to better understand if heterogeneous good and method characteristics can be controlled for using classical meta-regression analysis. Hence, we are still grappling with the question of how to strike the right balance between screening out studies from the analysis and coding them with the aim of later controlling for increased heterogeneity in regression models. How homogenous is homogenous enough? Fundamentally, there is still much we do not know about people's preferences and how to represent and interpret them in MA models. Increasing clarity and transparency of effect size definitions, data collection and screening protocols offering others the chance to replicate results, is one important way forward for MA (e.g. as pointed out by Nelson and Kennedy 2009 and USEPA 2006). Using sensitivity analysis to investigate the effects of important analyst choices related to the scope and heterogeneity of the MA dataset is another, as exemplified in this paper.

Paper 4: Asking for individual or household willingness to pay for environmental goods? Implication for aggregate welfare measures[21]

Paper 1 found using MA that WTP in CV surveys depends on whether it is stated individually or on behalf of the respondent's household. A similarly coded variable distinguishing between WTP stated individually or behalf of the household was also included in the MA in Paper 3, where no consistent and robust relationship was found. This paper attempts to explain differences between household and indiviudal WTP and test for them empirically in a CV survey administered on the Internet – investigating research question Ib more in depth. The aggregate welfare measure for a change in the provision of a public good derived from a CV survey will be much higher if the same elicited mean WTP is added up over individuals rather than households. A trivial fact, however, once respondents are part of multi-person households, it becomes almost impossible to elicit an "uncontaminated" WTP measure that with some degree of confidence can be aggregated over one or the other response unit. The literature is mostly silent about which response unit to use in WTP questions, and in some CV studies it is even unclear which type has actually been applied. The paper tests for differences between individual and household WTP in a novel, web-administered, split-sample CV survey asking WTP for preserving biodiversity in old-growth coniferous forests in Norway. Two samples are asked both types of questions, but in reverse order, followed by a question with an item battery trying to reveal why WTP may differ. Results show that in a between-sample test the WTP respondents state on behalf of their households is not significantly different from their individual WTP. However, within

[21] This paper was co-authored with Ståle Navrud, and was published 31. January 2009 "Online first" in *Environmental and Resource Economics*, for its "Special Issue: The household, gender, children and environmental economics". A previous version was presented at the 16th Annual Conference of the European Association of Environmental and Resource Economists, Gothenburg, Sweden, 25 - 28 June 2008.

the same sample, household WTP is significantly higher than individual WTP; in particular if respondents are asked to state individual before household WTP. 80 percent of respondents state as an important reason for this result that they have a larger budget at their disposal when asked household WTP. There are few indications that altruism, though imperfectly measured in this paper, may be important in explaining that household WTP is higher than individual WTP within samples – the commonly held view in the literature. Instead, degree of financial integration and relevant budget constraints seem to be more important.

The results suggest that using individual WTP as the response unit would overestimate aggregate WTP, and thus bias welfare estimates in benefit-cost analyses. Thus, the choice of response format needs to be explicitly and carefully addressed in CV questionnaire design in order to avoid the risk of unprofitable projects passing the benefit-cost test. This result, combined with the results from the meta-regression analyses in Papers 2 and 3, suggests that controlling for individual and household dimensions of WTP may also be potentially important in using MA for BT.

5. Conclusions

This thesis has presented four papers on methodological issues in meta-analysis (MA), benefit transfer (BT) and environmental valuation. The first paper presented a MA of stated preference studies valuing non-timber benefits (forest protection and multiple use forestry) in Norway, Sweden and Finland over the last 20 years. It investigated using different meta-regression models to what extent willingness to pay (WTP) estimates conform with standard expectations, tested a number of novel hypotheses and identified gaps in the literature. The paper revealed using four different meta-regression models systematic and expected variation in WTP along methodological variables and to some extent also along various characteristics of the forest good. Some of the key findings were that individually stated WTP tended to be higher than household WTP, that people state higher WTP for forest goods during the spring/summer seasons and for avoiding forest losses compared to gains. Finally, WTP was found generally to be insensitive to size of the forest valued (either for protection or multiple use forestry practices). Determining the right scope of the MA was identified as a potentially important area of research.

Papers 2-4 then each picked up an important research themes following from the first paper. The second paper utilised the same data to investigate the precision in using MA for international benefit transfer (BT), as compared with simpler and more common BT techniques. It was found that even under conditions of homogeneity in valuation methods, cultural and institutional conditions across countries, and a MA with large explanatory power, the transfer errors could still be large. Further, international meta-analytic benefit transfers (MA-BT) were found not on average to perform better than simple value transfers averaging over domestic studies. However, more research is required and it is too early to discard MA as a tool for BT.

The third paper picked up issues identified and investigated in both Papers 1 and 2. It analysed, using a different and more extensive dataset of biodiversity and nature conservation values from Asia and Oceania, how the heterogeneity or scope of the MA data influences the results of different meta-regression models and their precision when used for BT. It was found that the degree of regularity and conformity with theory and empirical expectations is higher for the more homogenous dataset of contingent valuation of endangered species, as expected. Further, WTP for preservation of mammals tended to be higher than other species and WTP for species preservation increased with income (elasticity around 0.8). Individual WTP was not found to be higher than individually stated WTP, as in Paper 1. Subjecting the best MA models to a simple BT test forecasting values for out-of-sample observations, showed higher transfer errors for the more heterogenous dataset, though both data levels had median BT precision approaching levels that may be acceptable in for policy use. However, as more heterogeneous observations were included in the MA, the meta-regression models were unable to control for the variation in a satisfactory way, resulting in high mean transfer errors. Hence, heterogeneity in MA data have important implications for BT precision.

The fourth paper tested in a more controlled and in-depth way a question investigated both in Papers 1 and 3 on the MA level, whether people state different WTP if asked as individuals or on behalf of their household. A primary contingent valuation data set of WTP for forest protection in Norway was utilised for a split sample test. Two samples were asked both types of questions, but in reverse order, followed by a question with an item battery trying to reveal why WTP may differ. Results showed that in a between-sample test the WTP respondents state on behalf of their households was not significantly different from their individual WTP. However, within the same sample, household WTP was significantly higher than individual WTP. An important reason for this result was found to be that people perceive that they have a larger budget at their disposal when asked the household question. Results suggest that using individual WTP as the response unit may overestimate aggregate WTP, and therefore that the response unit issue should be carefully addressed in future CV surveys. The results also have implications for the standardisation procedures – individual or housel level welfare estimates – used in MA of environmental valuation studies.

Results from the thesis are mildly encouraging in contributing to our understanding of people's preferences for complex environmental goods, along the limited number of variables investigated. People tend to answer and methods tend to give, values broadly conforming with expectations, though there is still much uncertainty about how preferences are formed and uncovered. In terms of using such environmental valuation results synthesised by MA for BT, more research is required to determine both the acceptable heterogeneity of MA data and appropriate meta-regression models, and the general conditions under which MA may be reliably used.

References

Arrow, K. J., R. Solow, E. Leamer, P. Portney, R. Radner and H. Schuman (1993), 'Report of the NOAA Panel on Contingent Valuation', *Federal Register* **58**: 4601-4614.

Bateman, I. J. and A. P. Jones (2003) 'Contrasting conventional with multi-level modeling approaches to meta-analysis: Expectation consistency in UK woodland recreation values', *Land Economics* **79**(2): 235-258.

Bateman, I. J., R. T. Carson, B. Day, W. M. Hanemann, N. Hanley, T. Hett, M. Jones-Lee, G. Loomes, S. Mourato, E. Ozdemiroglu, D. W. Pearce, R. Sugden and T. Swanson (2002) *Economic Valuation with Stated Preference Techniques: A Manual*. Edward Elgar Publishing, Cheltenham.

Bergstrom, J. C. and Taylor L O (2006) 'Using meta-analysis for benefits transfer: Theory and practice'. *Ecological Economics* **60**: 351-360.

Brander, L. M., P. van Beukering and H. Cesar (2007) 'The recreational value of coral reefs: a meta-analysis'. *Ecological Economics* **63**(1): 209-218

Brander, L. M., R. J. G. M. Florax and J. E. Verrmaat (2006) 'The Empirics of Wetland Valuation: A Comprehensive Summary and a Meta-Analysis of the Literature'. *Environmental and Resource economics* **33**: 223-250

Brouwer, R., (2000) 'Environmental value transfer: state of the art and future prospects'. *Ecological Economics*, **32**(1); 137-152.

Cameron, T. A. and D. D. Huppert (1989) 'OLS Versus ML Estimation Of Non-Market Resource Values With Payment Card Interval Data'. *Journal of Environmental Economics and Management* **17**(3): 230-246.

Carson, R. T. and M. Hanemann (2005), 'Contingent valuation', in K.-G. Mäler and J. R. Vincent, eds, *Handbook of Environmental Economics, Volume 2.* Elsevier.

Carson, R. T., N. E. Flores, K. M. Martin and J. L. Wright (1996), 'Contingent valuation and revealed preference methodologies: Comparing the estimates for quasi-public goods', *Land Economics* **72**(1): 80-99.

Desvousges, W. H., F. R. Johnson and H. S. Banzhaf (1998) Environmental policy analysis with limited information: Principles and applications of the transfer method. Cheltenham, Edward Elgar.

Ehmke, M. D., J. L. Lusk, J. A. List (2008) 'Is Hypothetical Bias a Universal Phenomenon? A Multinational Investigation'. *Land Economics* **84** (3): 489–500.

Florax, R. J. C. M. (2002) Accounting for dependence among study results in meta-analysis: Methodology and applications to the valuation and use of natural resources. Research Memorandum 2002:5. Vrije Universiteit, Amsterdam.

Glass, G. V., B. McGaw and M. L. Smith (1981) *Meta-analysis in Social Research*. Beverly Hills: Sage.

Hanemann, M. (1992) 'Preface', In S. Navrud (ed) *Pricing the European environment*, Scandinavian University Press.

Horowitz, J. K. and K. E. McConnell (2002), 'A Review of WTA / WTP Studies', *Journal of Environmental Economics and Management* **44**: 426-447.

Horowitz, J. K. and K. E. McConnell (2003), 'Willingness to accept, willingness to pay and the income effect', *Journal of Economic Behavior and Organization* **51**(4): 537-545.

Jacobsen, J. B. and N. Hanley (2009) 'Are there income effects on global willingness to pay for biodiversity conservation'. *Environmental and Resource Economics* (In press)

Jakus, P. M., B. Stephens and J. M. Fly (2006) 'Temporal reliability in contingent valuation (with a restrictive research budget)', in A. Alberini and J. R. Kahn (eds) *Handbook on Contingent Valuation*. Edward Elgar.

Johnston, R. J., E. Y. Besedin and M. H. Ranson (2006) 'Characterizing the effects of valuation methodology in function-based benefits transfer'. *Ecological Economics* **60**(2): 407-419

Johnston, R. J., E. Y. Besedin and R. F. Wardwell (2003), 'Modeling relationships between use and nonuse values for surface water quality: A meta-analysis', *Water Resources Research* **39**(12).

Kahneman, D. and A. Tversky (Eds) (2000) *Choices, Values and Frames*, Cambridge University Press.

Krutilla, J. (1967) 'Conservation Reconsidered', *American Economic Review* **57**: 777-86

Lipsey, M. W. and D. B. Wilson (2001) *Practical meta-analysis*. Sage, Thousands Oaks.

List, J. A. and C. A. Gallet (2001) 'What experimental protocol influence disparities between actual and hypothetical stated values?' *Environmental and Resource Economics* 20(3): 241-254.

Loomis, J. and R. Rosenberger (2006) 'Reducing barriers in future benefit transfers: Needed improvements in primary study design and reporting', *Ecological Economics* **60**(2): 343-350.

Loomis, J. B. and D. S. White (1996) 'Economic benefits of rare and endangered species: Summary and meta-analysis', *Ecological Economics* **18**(3): 197-206.

Moeltner, K., K. Boyle and R. W. Paterson (2007) 'Meta-analysis and benefit transfer for resource valuation - addressing classical challenges with Bayesian modeling'. *Journal of Environmental Economics and Management* **53**: 250-269.

Moeltner, K. and R. Rosenberger (2008) 'Predicting Resource Policy Outcomes via Meta-Regression: Data Space, Model Space, and the Quest for 'Optimal Scope''. *The B.E. Journal of Economic Analysis & Policy* **8**(1), Article 31.

Moeltner, K. and R. Woodward (2009) 'Meta-Functional Benefit Transfer for Wetland Valuation: Making the Most of Small Samples.' *Environmental and Resource Economics* **42**(1): 89-108.

Murphy, J. J., P. G. Allen, T. H. Stevens and D. Weatherhead (2005) 'A Meta-Analysis of Hypothetical Bias in Stated Preference Valuation', *Environmental and Resource Economics* 30: 313-325.

Navrud, S. and G. Pruckner (1997), 'Environmental Valuation - To Use or Not to Use? A Comparative Study of the United States and Europe', *Environmental and Resource Economics* **10**(1): 1-26.

Navrud, S. and R. Ready (Eds) (2007) *Environmental Value Transfer: Issues and Methods*. Springer.

Nelson, J. P. and P. E. Kennedy (2009) 'The Use (and Abuse) of Meta-Analysis in Environmental and Natural Resource Economics: An Assessment'. *Environmental and Resource Economics* (In press)

Nelson, J.P. (1980) 'Airports and property values: a survey of recent evidence', *Journal of Transport Economics and Policy* **19**: 37-52.

Ojea, E. and M. L. Loureiro (2008) 'Scope Tests in a Meta-Analysis of Contingent Valuation Studies'. Presented at EAERE Goteborg, June, 2008.

Rabin, M. (1998), 'Psychology and Economics', *Journal of Economic Literature* **XXXVI**: 11-46.

Ready, R. and S. Navrud (2006) 'International benefit transfer: Methods and validity tests'. *Ecological Economics* **60**: 429-434.

Richardson, L. and J. Loomis (2009) 'The total economic value of threatened, endangered and rare species: An updated meta-analysis.' *Ecological Economics* (In press)

Rosenberger, R. and J. Loomis (2000b), 'Panel stratification in meta-analysis of economic studies: an investigation of its effects in the recreation valuation literature', *Journal of Agricultural and Applied Economics* **32**(1): 131-149.

Rosenberger, R. and T. D. Stanley (2007) 'Publication Effects in the Recreation Use Values Literature'. Annual Meeting of the American Agricultural Research Association, July 29-August 1, 2007, Portland, Oregon

Rosenberger, R. S. and J. B. Loomis (2000a) 'Using meta-analysis for benefit transfer: In-sample convergent validity tests of an outdoor recreation database'. *Water Resources Research* **36**(4): 1097-1107.

Santos, J. M. L. (1998) The Economic Valuation of Landscape Change. Theory and Policies for Land Use and Conservation. Edward Elgar, Cheltenham.

Santos, J. M. L. (2007) 'Transferring landscape values: How and how accurately?' in S. Navrud and R. Ready (Eds) *Environmental Value Transfer: Issues and Methods*. Springer.

Sayman, S. and A. Onculer (2005), 'Effects of study design characteristics on the WTA-WTP disparity: A meta analytical framework', *Journal of Economic Psychology* **26**(2): 289-312.

Schlapfer, F. (2006), 'Survey protocol and income effects in the contingent valuation of public goods: A meta-analysis', *Ecological Economics* **57**(3): 415-429.

Schulze, R. (2004) *Meta-analysis: a comparison of approaches*. Hogrefe & Huber, Cambridge

SEPA (2006) An instrument for assessing the quality of environmental valuation studies. Swedish Environmental Protection Agency.

Shrestha, R. K. and J. B. Loomis (2001) 'Testing a meta-analysis model for benefit transfer in international outdoor recreation'. *Ecological Economics* **39**(1): 67-83.

Shrestha, R. K. and J. B. Loomis (2003) 'Meta-Analytic Benefit Transfer of Outdoor Recreation Economic Values: Testing Out-of-Sample Convergent Validity'. *Environmental and Resource Economics* 25: 79-100.

Shrestha, R. K., R. Rosenberger and J. Loomis (2007) 'Benefit transfer using meta-analysis in recreation economic valuation', in S. Navrud and R. Ready (Eds) *Environmental Value Transfer: Issues and Methods*. Springer.

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

Smith, V. K. and Y. Kaoru (1990a) 'What have we learned since Hotelling's letter? ameta-analysis'. *Economics Letters* **32**:267–272

Smith, V. K. and Y. Kaoru (1990b) 'Signals or noise? Explaining the variation in recreation benefit estimates'. *American Journal of Agricultural Economics* **72**:419–433

Smith, V.K., G. Van Houtven, S. K. Pattanayak (2002) 'Benefits transfer via preference calibration: "prudential algebra" for policy'. *Land Economics* **78** (1), 132–152.

Smith, V. K. and S. K. Pattanayak (2002), 'Is Meta-Analysis a Noah's Ark for Non-Market Valuation?' *Environmental and Resource Economics* **22**: 271-296.

Stanley, T. D. (2001), 'Wheat from chaff: Meta-analysis as quantitative literature review', *Journal of Economic Perspectives* **15**(3): 131-150.

Stanley, T. D. and S. D. Jarrel (2005) 'Meta-Regression Analysis: A Quantitative Method of Literature Surveys', *Journal of Economic Surveys* 19(3): 299-308.

Stapler, R. W. and R. J. Johnston (2009) 'Meta-Analysis, Benefit Transfer, and Methodological Covariates: Implications for Transfer Errors'. *Environmental and Resource Economics* (In press)

Strand, J. (2007) 'Public-good valuation and intrafamily allocation'. *Environmental and Resource Economics* **38**(4): 527-543.

U.S. Environmental Protection Agency (2006) *Report of the EPA work group on VSL meta-analysis*, report NCEE-0494. National Center for Environmental Economics, EPA, Washington, DC.

U.S. Environmental Protection Agency (2007) *SAB advisory on EPA's issues in valuing mortality risk reduction*, EPA-SAB-08-001. National Center for Environmental Economics, EPA, Washington, DC.

Van Houtven G, Powers J and Pattanayak S K (2007) 'Valuing water quality improvements using meta-analysis: Is the glass half-full or half-empty for national policy analysis?' *Resource and Energy Economics* **29**(3): 206-228.

Walsh, R. G., D. M. Johnson, J. R. McKean (1989) 'Issues in nonmarket valuation and policy applications: a retrospective glance'. *Western Journal of Agricultural Economics* **14**:178–188.

Walsh, R. G., D. M. Johnson, J. R. McKean (1990) 'Nonmarket values from two decades of research on recreation demand'. In: A. Link and V. K. Smith (eds) *Advances in applied microeconomics*. JAI Press, Greenwich.

Paper I

Available online at www.sciencedirect.com
ScienceDirect
ELSEVIER
Journal of Forest Economics 12 (2007) 251–277
Journal of FOREST ECONOMICS
www.elsevier.de/jfe

20 years of stated preference valuation of non-timber benefits from Fennoscandian forests: A meta-analysis ☆

Henrik Lindhjem[a,b,*]

[a]*Department of Economics and Resource Management, Norwegian University of Life Sciences, P.O. Box 5003, N-1432 Ås, Norway*
[b]*ECON, P.O. Box 5, N-0051 Oslo, Norway*

Received 28 June 2006; accepted 15 September 2006

Abstract

Stated preference (SP) surveys have been conducted to value non-timber benefits (NTBs) from forests in Norway, Sweden and Finland for about 20 years. The paper first reviews the literature and summarises methodological traditions in SP research in the three countries. Second, a meta-regression analysis is conducted explaining systematic variation in Willingness-to-Pay (WTP). Two important conclusions emerge, with relevance for future research: (1) WTP is found to be insensitive to the size of the forest, casting doubt on the use of simplified WTP/area measures for complex environmental goods; and (2) WTP tends to be higher if people are asked as individuals rather than on behalf of their household.
© 2006 Elsevier GmbH. All rights reserved.

JEL classification: Q23; Q26; Q51; H41

Keywords: Non-timber forest benefits; Meta-analysis; Stated preferences; Norway; Finland; Sweden

☆A previous version was presented at the 13th Ulvön Conference on Environmental Economics in Sweden, 19–21 June, 2006.

*Department of Economics and Resource Management, Norwegian University of Life Sciences, P.O. Box 5003, N-1432 Ås, Norway.
E-mail address: henrik.lindhjem@umb.no.

1104-6899/$ - see front matter © 2006 Elsevier GmbH. All rights reserved.
doi:10.1016/j.jfe.2006.09.003

Introduction

Stated preference (SP) surveys (contingent valuation (CV) and choice experiments (CE)) have been conducted to value non-timber benefits (NTBs) from forests for about 20 years in Norway, Sweden and Finland ("Fennoscandia"), the three largest of the Nordic countries. Time is ripe for taking stock and synthesising this body of research. In the economics literature the meta-analysis tool, more commonly used in other disciplines, is increasingly being put to such tasks (Stanley, 2001; Stanley and Jarrel, 2005). The non-market valuation branch of environmental economics has developed a rich but still immature meta-analysis literature since Smith and Kaoru's (1990) seminal study[1] of recreational benefits. Since then, meta-analyses have been conducted for the purposes of research synthesis, hypothesis testing and benefit transfer for a number of environmental goods (Smith and Pattanayak, 2002). Meta-analyses of recreational benefits for various outdoor activities are the most common, but other goods studied include for example endangered species (Loomis and White, 1996), wetlands (Brouwer et al., 1999; Woodward and Wui, 2001), noise (Button, 1995), aquatic resource improvements (Johnston et al., 2005), and air quality and visibility (Smith and Osborne, 1996; Desvousges et al., 1998).

The methodological hypotheses explored through meta-analysis include "classical" questions in the non-market valuation literature such as the Willingness-to-Pay (WTP) vs. Willingness-to-Accept (WTA) compensation disparity and income effects (Horowitz and McConnell, 2003; Sayman and Öncüler, 2005; Schläpfer, 2006), WTP's (in)sensitivity to change in quality or quantity ("scope") of the good (Smith and Osborne, 1996), convergent validity of benefit estimates from different valuation methodologies (Carson et al., 1996), the relationship between use values (UV) and non-use values (NUV) (Johnston et al., 2003), and differences in real and hypothetical WTP ("hypothetical bias") (List and Gallet, 2001; Murphy et al., 2005). More recently, meta-analysis has also been used to synthesise WTP estimates for the purpose of benefit transfer to new unstudied, policy sites (Shrestha and Loomis, 2001, 2003). Accurate benefit transfer with its savings in primary study costs enabling increased use of cost-benefit analysis (CBA) is one of the "holy grails" of environmental economics, though still some way from its promise (Florax et al., 2002; Navrud and Ready, 2006).

Although the use of meta-analysis in the non-market valuation literature has grown in recent years, no studies we are aware of have looked at SP surveys of forest protection or multiple use forestry (MUF). The existing studies on recreational benefits often include forests, but are typically focused on consumer surplus estimates for activity days (such as fishing, hunting, hiking, etc.), and not people's WTP for protection or change in forestry practices per se (Rosenberger and Loomis, 2000a; Shrestha and Loomis, 2001; Bateman and Jones, 2003). Further, this literature is dominated by the travel cost method, often pooling meta-datasets with a smaller number of SP surveys. This approach rules out an analysis of potentially

[1]Sometimes also credited to Walsh et al. (1989), for example by Shrestha and Loomis (2001), or to Walsh et al. (1990) by Smith and Pattanayak (2002).

important NUV[2] of forests related to for example biodiversity protection, and often limits the analysis of important features of SP research. Several unanswered questions remain in understanding people's preferences and WTP for NTBs related both to the UV and NUV components.[3] This paper aims to begin to answer some of them based on a synthesis and meta-analysis of around 50 studies reporting results from 30 SP surveys of both urban[4] and non-urban forests in the Fennoscandia over the last 20 years. The paper first reviews this literature and summarises and categorises methodological traditions in SP valuation of forests in the three countries. Second, a selection of methodologically similar SP studies is included in a meta-regression analysis attempting to explain the variation in WTP for protection and/or MUF by differences in survey methodology, good characteristics, study quality, socio-economic variables and other variables. In addition to investigating WTP's conformity with standard theoretically and empirically derived expectations, the paper attempts to answer novel questions about WTP's seasonal variability, country differences, WTP for MUF vs. full protection, household vs individual valuation, sensitivity to scope, WTP's development over time, and differences between WTP for avoiding a loss and achieving a gain. Finally, the paper concludes and suggests future research directions – not losing sight of the main goal of meta-analysis and SP research in the flood of WTP estimates, studies and methodological twists: a better understanding of individual preferences for forest protection and management.[5]

Valuation of NTBs from Fennoscandian forests

Norway, Finland and Sweden are very similar countries in many respects and there are good reasons to study them together.[6] Their location on the Fennoscandian Shield yields similar climatic, geological and ecological conditions, resulting in a large cover of boreal forests. Second only to Russia, Sweden is the most forested country in Europe with its 22.7 million hectares of productive forest. Finland and Norway have approximately 20 and 7.5 million hectares (Framstad et al., 2002), respectively. The similarities between the three countries extend to the judicial, economic and cultural dimensions of recreation, forest conservation and forestry. The countries are on roughly the same level of economic and human development measured by GDP/capita and UN's human development index. Forestry is an important industry and seen together with agriculture as the backbone of local economies and the key to retaining dwindling populations in rural areas. All three

[2] Asking people to state their preferences is the only method that can capture NUV.

[3] The sum of UV, including the timber values, and NUV are often termed the total economic value (TEV).

[4] The differences between urban and non-urban forests are often not clear-cut in the Nordic countries, as even the capital cities have patches of forests (rather than parks) within their city zones.

[5] Paraphrased after Smith and Pattanayak (2002).

[6] Denmark, Iceland and the Faroe Islands were left out of the analysis as their forests can be considered to be different goods (both in terms of size, ecology and use) to the Fennoscandian.

countries have a large number of small, private forest owners. The everyman's right to access and harvesting of certain resources (for example mushrooms and berries) regardless of land ownership is an important and age-old traditional basis for the forest activities carried out by the public. Semi-private markets for fishing (for example salmon) and hunting permits are allowed and broadly accepted, while markets for other NTBs are generally not, among others due to the everyman's right. Growing wealth, and with it increasing demand for environmental goods, has resulted in high conflict levels between timber production and the supply of NTBs (Vatn et al., 2005). As a response to this, forest protection has increased and MUF, in accordance with various certification schemes, have been taken up by the industry. Sweden, Finland and Norway have protected about 4%, 5% and 1% of their productive forests, respectively, and the distribution of protected areas tends to a large degree to reflect economic rather than ecological considerations (Framstad et al., 2002; Lehtonen et al., 2003).

In parallel with the growing tension in the forestry sector a substantial literature has developed in Fennoscandia to value NTBs to compare with timber values. Navrud (1992) sums up some of the early literature. Some SP studies value single forest species such as large carnivores (Boman, 1995) or birds (Fredman, 1995), recreation activities such as fishing (Laitila and Paulrud, 2006), or hunting activities (Johansson et al., 1988). Our focus here is on those primary studies that value forest protection and/or MUF only. In addition, we include two studies that look at the value of forest biodiversity in general (which would directly require increased forest protection and/or MUF practices). The values from these studies can be interpreted as the WTP to obtain a positive change (or WTP to avoid a loss) in at least one element in an attribute vector describing the forest environment, i.e. level of biodiversity, forest density, forest size, scenic beauty, etc. A broad search for published (peer-reviewed papers and book chapters) and unpublished studies (Master and Ph.D. Theses, working papers, research reports) was conducted in the three countries. The starting point for Sweden was a recent and comprehensive database of valuation studies (Sundberg and Söderqvist, 2004), and for Finland a recent meta-analysis (in Finnish) (Pouta and Rekola, 2005). The relevant references in these studies were supplemented by a few more recent studies.[7] The search was limited to studies written in Swedish, Norwegian or English, which most likely has not skewed the selection unduly.[8] The search turned up about 50 studies reporting different aspects of the results from around 30 different SP surveys (see Table 1).

As can be seen from the table, the number of surveys is distributed fairly evenly between countries, though more surveys have been conducted in Finland in recent years. There is a mix between valuing full forest protection and MUF at the local, regional and national levels. Many of the Norwegian studies focus on the forest area

[7] Olsson (1993) was excluded since the study valued a cableway entry to a forest, while Johansson and Zavisic (1989) was excluded due to insufficient reporting.

[8] Swedes and Norwegians generally understand the two languages, but not Finnish. A large number of Fins, on the other hand, also understand Swedish and Norwegian. Only two studies referenced in Pouta and Rekola (2005) are known to be excluded by this rule, as most of the relevant studies from Finland are in English.

Table 1. Stated preference valuation surveys of Fennoscandian forests, 1985–2005

Main references[a]	Year[b]	Good[c]	Gain/loss[d]	Mode	Scope[e]	Method[f]	#[g]	WTP (USD)[h]
Finland								
Kniivilä (2004)[1]	2000	P	L	Mail	R, L	CV: DC	2	61–107
Lehtonen et al. (2003)[2]	2002	P	G	Mail	R	CV: DC	5	190–342
Pouta et al. (2000, 2002)[3]	1997	P	G/L	Mail	N	CV: DC	4	154–227
Pouta (2003–2005)	1998	M	G	Mail	N	CV: DC	2	287–299
Rekola and Pouta (2005)	1995	M	G	Mail	L	CV: DC	1	20
Siikamäki and Layton (in press)	1999	P	G	Mail	N	CV: DC, CE	3	79–134
Mäntymaa et al. (2002)	1999	P, B	G	Mail	N	CV: OE	4	224–380
Horne et al. (2005)[i]	1998	P, M	G/L	Interv.	L	CE	1	−16
Tyrväinen and Väänänen (1998)	1995	P, O	L	Mail	L	CV: OEPC	5	31–124
Tyrväinen (2001)	1996	P, O	L/G	Mail	L	CV: OEPC	6	22–248
Norway								
Simensen and Wind (1990)	1989	P, M	G	Interv.	L	CV: OE	3	21–159
Hoen and Winther (1993)[4]	1990	P, M	G	Interv.	N	CV: OEPC	6	14–65
Veisten et al. (2004a, b)[5]	1992	B	L	Interv.	N	CV: OE/OEPC	3	138–210
Sandsbråten (1997)	1997	M	L/G	Interv.	L	CV: DC	2	43–45
Leidal (1996)	1996	P	L	Interv.	L	CV: DC/OE	3	455–504
Skagestad (1996)	1996	P, M	G	Interv.	L	CV: OEPC	1	15
Veisten and Navrud (2006)	1995	P	L	Mail	R	CV: DC/OE	4	3–104
Hoen and Veisten (1994)	1992	M	G	Interv.	L	CV: OE	1	50
Hoen and Veisten (1994)	1993	M	G	Interv.	L	CV: DC	1	48
Strand and Wahl (1997)	1997	P	L	Interv.	L	CV: OE/DC	2	172–243
Sweden								
Bojö (1985)	1985	P	G	Interv.	L	CV: DC	1	58
Bostedt and Mattson (1991)	1991	M, O	L	Mail	L	CV: OE	1	385
Mattsson and Li (1993)	1991	M, O	L	Mail	R	CV: OE/DC	2	469–907
Mattsson and Li (1994)[6,j]	1992	M, O	L/G	Mail	R	CV: DC, CE	2	440–1280

Table 1. (*continued*)

Main references[a]	Year[b]	Good[c]	Gain/loss[d]	Mode	Scope[e]	Method[f]	#[g]	WTP (USD)[h]
Kriström (1990a, b)[7]	1987	P	G	Mail	N	CV: DC/OE	4	275–725
Johansson (1989)	1987	B	L	Mail	N	CV: OE	1	254
Bostedt and Mattsson (1995)[8]	1992	M, O	G	Mail	L	CV: OE	2	78–84
Fredman and Emmelin (2001)	1998	M, O	G	Mail	R	CV: OE	1	92
Total number of estimates							72	

[a] Also reporting WTP estimates from the same survey:1 = Ovaskainen and Kniivilä (2005), Kniivilä et al. (2002); 2 = Lehtonen et al. (2005a, b); 3 = Rekola et al. (2000), Li et al. (2004), Pouta (2003); 4 = Garnes and Winther (1991), Veisten and Hoen (1994); 5 = Veisten et al. (1993), Veisten (1993), Veisten and Hoen (1994); 6 = Li and Mattsson (1995), Li (1996), Holgen et al. (2000); 7 = Kriström (1989); and 8 = Bostedt (1997).

[b] Year of survey, rather than study publication year.

[c] Good type: P = forest protection, M = multiple use forestry (MUF), B = forest biodiversity specifically, O = other (e.g. tourism WTP attributed to forests in an area).

[d] WTP for proposed improvement (gain) or to avoid a proposed negative change (loss).

[e] Geographical scope: National (N), regional (R), local (L) forest good.

[f] Methodology: OE = open ended WTP format, OEPC = OE with the aid of a payment card (a range of values presented to the respondent to choose from), DC = dichotomous choice format.

[g] #: Number of estimates included in the final meta-regression analysis.

[h] WTP estimates converted from NOK used in the meta-regression analysis to USD 2005 using OECD purchase power parity (PPP) and Norwegian Consumer Price Index (CPI), and may therefore not correspond exactly to the WTP estimates as they are reported in the studies. The WTP formats are given as reported (i.e. lump sum, per month, per household or individual, long-term annual, etc.), and are therefore not directly comparable.

[i] This study, which uses a CE approach that is not directly comparable to CV, was judged too different from the other studies and taken out of the final meta-regression analysis. The WTP is negative here since people preferred open scenery (and less biodiversity) to a more closed forest (with more biodiversity).

[j] One extreme WTP value from Li (1996) of SEK 158 116 was excluded.

just north of the capital Oslo (''Oslomarka''), an area of significant friction between forestry and environmental and recreational interests. All the national (and to some extent regional) level surveys cover both users and non-users. It is interesting to note that Norway has a tradition of conducting more in-person interview surveys, perhaps reflecting the funding situation for such research. There is a mix of dichotomous choice (DC) and open-ended (OE) CV WTP question formats used, while the CE approach, which has come into fashion internationally in recent years (Hanley et al., 1998), has been tried only once or twice. Other interesting features of the Fennoscandian forest SP research, not displayed in the table, is that a range of payment mechanisms are used (from voluntary contributions to tax and user fees) and that only one study use an actual payment mechanism (Veisten and Navrud, 2006). All studies but one ask for WTP (either for a gain or to avoid a loss), and not WTA. No studies use more advanced WTP question formats, such as the double-bounded DC or iterative bidding. Econometric approaches to estimate the data vary widely (see next section), and cover a range of parametric and non-parametric approaches especially in the Swedish and Finnish studies. Many studies specifically test for WTP's sensitivity to scope, most often presented as size of forest (in percentage or hectare – ha). As we shall see, this simple approach is fraught with difficulties for forest goods. Few surveys remind the respondents of substitutes and budget constraints. While it is difficult to discern a trend in the research judging from an overview like this, it seems that the DC approach has become more common in the years after the NOAA panel report (Arrow et al., 1993), but that the in-person interview mode has not. Another trend, as can be expected, is that survey instruments have gradually become more realistic, informative and sophisticated. The selection procedure for estimates included in the meta-regression analysis and explanation of the WTP estimate format and variation (the last two columns in the table) are left for the next section.

Metadata and hypotheses

Data

SP studies typically explore impacts of different methodological assumptions and to a lesser extent conduct the survey to obtain one single WTP estimate of the environmental change in question, for example for use in CBA. This practice, driven by which studies tend to get published, makes the reporting very diverse and the metadata coding process complex. Some of the studies report extensively and also annex the full survey instrument while others are silent or very brief on important dimensions of the survey design and results. Even the average WTP for the sample is sometimes not reported. The collected SP studies in Table 1 were coded in a spreadsheet for variables hypothesised to have explanatory power for the variation in WTP. The coding procedure is an iterative process as new studies added to the spreadsheet may require recoding of previously recorded ones. The explanatory variables were chosen based on theory and previous empirical meta-studies, and the availability of information in the SP studies. It is a challenge and a judgment call for the meta-analyst to make the trade-off between the number of potentially interesting explanatory variables to include in the analysis and exclusion of relevant studies due to limited reporting. The more explanatory variables that are included, the fewer studies will have complete reporting for all variables. Further, too many variables will lead to over specification of the model, while too few will fail to capture important variation in the data. There is no consensus in the literature on how to resolve this meta-analysis scope problem, or other judgments required by the meta-analyst, other than to state clearly which choices have been made in the analysis and to conduct sensitivity analysis. Some meta-analysis applications may require a narrower scope, for example including studies using certain specific methods only. Generally, many meta-analysts recommend "to err on the side of inclusion" of studies and estimates (Stanley and Jarrel, 2005). This is the principle we abide by here, but we also investigate the sensitivity of our results to changes in the scope of the meta-analysis. We began by recording all raw WTP estimates reported in the 50 studies, which amounted to some 250 observations from the 29 surveys.[9,10] The number of observations ranged from around 35 in Strand and Wahl (1997) to one observation from several studies. Many of the estimates reported from the same study varied only along dimensions of statistical modelling choices (especially for DC or OEPC data), which were often impossible to code accurately due to insufficient reporting. Instead of including all of these as study-to-study level background, as recommended for variation due to "*minor modelling choices*" by Stanley and Jarrel (2005, p. 137), we averaged them into one or more observations. For example, if a study reported nine WTP estimates for probit, logit and non-parametric statistical models, respectively, for three different sized forest protection plans and all other variable values are the same, we include one average WTP estimate for each protection plan in the meta-data (i.e. three measurements). In this way all estimates were included but weighted down. This was done to reduce substantial variation and noise in the data due to statistical modelling choices our statistical model was not intended to explain. Averaging was also done with WTP estimates reported for different subsamples that could not be distinguished by our explanatory variables (for example samples split depending on attitudes to conservation or ethical dimensions, various trimming procedures, etc.). Further, we excluded overall sample averages if sub-sample averages for the same survey had also been reported. Some of the studies reported income, age, and education levels for their sample. Preliminary analysis with some 40–50 observations showed that these variables were generally insensitive to differences in WTP estimates.[11] This is a

[9]Two or three studies reported WTP estimates by socioeconomic categories (age groups, income and education levels) within the sample in addition to sample WTP estimate with average socioeconomic variable values. In these cases, only the sample average WTP estimates were included in the meta-data set.

[10]The full data set is available in an Excel spreadsheet on request from the author.

[11]Some meta-analyses supplement lacking primary socioeconomic data with official statistics. This is not likely to make WTP *more* sensitive to variation than for the primary data, so this approach was not followed here.

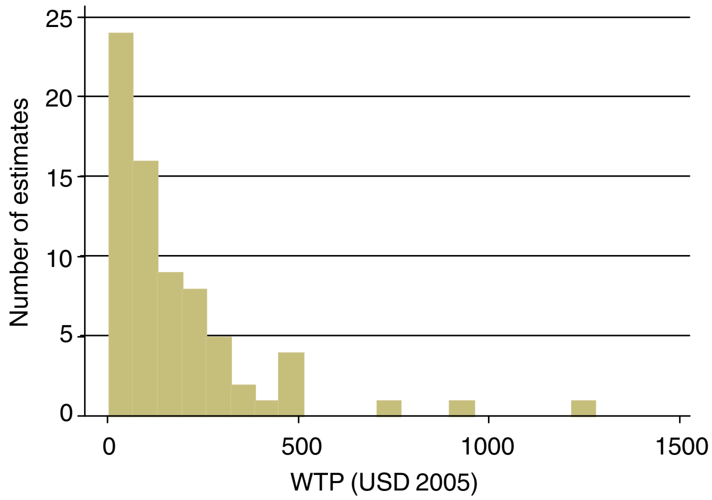


Fig. 1. Distribution of the dependant variable, WTP in 2005 USD, $N = 72$.

very common result in SP meta-regression analyses (Rosenberger and Loomis, 2000a; Johnston et al., 2003, 2005). Finally, about 10 observations that only varied along socioeconomic dimensions, for example WTP for different education or age segments of the sample, were taken out leaving a final meta-data set of 72 observations. The density of the WTP estimates are given in Fig. 1, and has a similar shape as in other meta-studies (for example, Rosenberger and Loomis, 2000a):

The variables that were eventually retained and fully coded are given in Table 2, and explained and justified in the next section. We chose long-term[12] average annual WTP per household as the base format (as this is most commonly asked) and coded WTP given in other formats (such as WTP per individual, per month, lump sum contribution, etc.), using dummy variables. An alternative would have been to adjust all reported estimates into an annual household WTP. Since respondents' discount rates are not known, we felt it was more prudent to use dummies. To make WTP from different countries comparable, estimates from Sweden and Finland were converted to NOK at the year of the survey[13] (rather than the publication year) using annual average OECD purchase power parity (PPP) rates, and then adjusted to 2005 figure by use of the standard Norwegian consumer price index (CPI).[14] The reason for using PPP to adjust for differences in actual purchasing power is that nominal exchange rates may not accurately measure differences in income and consumption (and therefore WTP) between countries. Using the Norwegian CPI implicitly assumes that WTP for NTBs increases at the same rate as market goods.

[12]WTP asked for 10 years was included in this category as people most likely do not distinguish between 10 years and an indefinite horizon.

[13]In some rare cases the year of reported WTP estimates was unclear. We have assumed reported in current values of the year of the survey (rather than for instance year of submission to a journal, etc.).

[14]This procedure is also recommended in the health benefit transfer literature, where international comparisons and transfer are more common (Eiswerth and Shaw, 1997; Pattanayak et al., 2002).

Table 2. Meta-analysis variables and descriptive statistics

Variable	Description	Sign	Mean (SD)
Dependent variable			
WTP2005	WTP in 2005 NOK	...	1192 (1374)
Methodological variables			
CE	Binary: 1 if choice experiment, 0 if CV	±	0.08 (0.25)
OE	Binary: 1 if OE without payment card, 0 if dichotomous choice	–	0.36 (0.48)
OEPC	Binary: 1 if OE with payment card, 0 if dichotomous choice	–	0.26 (0.44)
Volunpv	Binary: 1 if payment vehicle is described as a voluntary (unrelated to use) (e.g. donation to a fund), 0 if otherwise (e.g. tax)	+	0.18 (0.39)
Userpv	Binary: 1 if payment vehicle is related to recreational use or access (e.g. entrance fee, etc.), 0 if otherwise (e.g. tax)	–	0.19 (0.4)
Otherpay	Binary: 1 if payments were to occur on something other than an annual long-term basis, for example as a lump-sum, annual for a limited period, monthly or per season	+	0.5 (0.5)
Actualpay	Binary: 1 if payments were actually made, 0 if hypothetical WTP	–	0.03 (0.17)
Individual	Binary: 1 if individual WTP, 0 if household	±	0.32 (0.47)
Mailhigh	Binary: 1 if mail survey with high (more than 65% useable questionnaires), 0 if in-person interview	–	0.13 (0.33)
Mailmed	Binary: 1 if mail survey with medium (between 50% and 65% useable questionnaires), 0 if in-person interview	–	0.25 (0.44)
Maillow	Binary: 1 if mail survey with low (below 50% useable questionnaires), 0 if in-person interview	–	0.31 (0.46)
Study quality variables			
Unpub	Binary: 1 if WTP estimate unpublished, 0 if published	±	0.38 (0.47)
Mscthesis	Binary: 1 if primarily a Master thesis, 0 if otherwise	±	0.15 (0.36)
Good characteristics variables			
Forestpract	Binary: 1 if more cautious forestry practices; 0 if full protection	±	0.32 (0.47)
Protmix	Binary: 1 if mix of protection and forestry practices; 0 if full protection	±	0.07 (0.26)
Forestarea	Continuous: Total forest area of proposed change (ha)	+	See text
Impl	Binary: 1 if neither percentage of total land area nor forest area (ha) are mentioned in the survey, 0 if otherwise	±	0.78 (0.42)

Table 2. (*continued*)

Variable	Description	Sign	Mean (SD)
Hafrerc	Continuous: Area percentage of total productive forest area in the country (estimated in year 2005, or based on info provided in study)	+	See text
Haperc	Continuous: Area percentage of total land area	+	See text
Localgood	Binary: 1 if local good, 0 if nationwide	+	0.42 (0.5)
Reggood	Binary: 1 if regional good, 0 if nationwide	+	0.21 (0.41)
Sweden	Binary: 1 if study conducted in Sweden, 0 if Norway or Finland	±	0.19 (0.4)
Finland	Binary: 1 if study conducted in Finland, 0 if Norway or Sweden	±	0.44 (0.5)
Urban	Binary: 1 if primarily urban forest (major town), 0 if otherwise	±	0.33 (0.47)
Season	Binary: 1 if surveyed in autumn/winter (i.e. Sept.–March), 0 if spring/summer (i.e. April–August)	–	0.6 (0.49)
Avoidloss	Binary: 1 if it is WTP for avoiding a loss, 0 if it is for an improvement	+	0.4 (0.49)
Use	Binary: 1 if primarily use/users, 0 otherwise (i.e. users and non-users are incl.)	+	0.36 (0.48)
Other variables			
Year	Continuous: Range 1 (1985, year of first survey) to 16 (2002).	+	10.6 (4.2)

We included a set of methodological variables that are often used in SP meta-analyses, such as WTP question and reporting formats, survey mode and response rates for mail surveys, payment vehicles, and whether the WTP is asked from an individual or on behalf of a household. Instead of excluding observations on the basis of subjective judgement of study quality, a procedure that is generally not recommended in the meta-analysis literature (Woodward and Wui, 2001), we include proxy variables for quality; whether a study is a master thesis or otherwise unpublished (i.e. a research report or working paper). It is, however, difficult to capture the quality dimension with the ‘‘unpublished’’ variable as some of the studies (especially working papers) at some point may be published. In addition, SP meta-analyses sometimes use the year of the survey as a proxy for methodological quality assuming that advancements in SP methodology over time introduce prudence in survey design resulting in lower WTP estimates (Johnston et al., 2005). We favour a different interpretation (see next section). We also separate whether a mail survey has high, medium or low response rates. Finally, the last set of variables tries to capture the variation in good characteristics, along dimensions of geography (local or regional), country, time of the year (autumn/winter vs. spring/summer), scope (forest area percentage or ha) and other characteristics (use vs. non-use, urban forest). Our expectations and hypotheses regarding the signs of the model parameters are provided in the next section.

Hypotheses and expectations

The large body of theoretical and empirical SP research, for example as summed up by Carson (2004) and Carson et al. (2001) provide a rich set of expectations regarding the signs of our model parameters. We have indicated these in the third column of Table 2 and provide justification in the following. Regarding the methodological variables and WTP question formats first ("OE" and "OEPC"), with some exceptions, most comparisons of OE and DC question formats suggest that the DC format produces estimates that tend to be larger (for example, Cameron et al., 2002). The reasons are that OE ends to give a high number of zero bids due to free-riding behaviour and protest responses, and DC higher bids due to biases related to "yeah-saying" and different starting points, and distributional assumptions in the statistical analysis.[15] Comparisons between OE with the use of payment card and DC tend to show the same pattern (Cameron et al., 2002), though the results are more mixed. WTP estimates from OEPC surveys tend to be higher than those from OE surveys, among others since PC tend to reduce zero-responses (Mitchell and Carson, 1989). We therefore expect both the OE and OEPC variable parameters to be negative (though less so for the OEPC) as compared with the DC base case.

Empirical comparisons of WTP estimates from mail surveys with in-person interviews are few and results mixed, though it is clear that survey modes do affect value estimates (Boyle, 2003). There are forces at work in both directions. While an in-person interview may be better able to convey information about goods, it is not clear in general if this would lead the respondent to state a higher or lower WTP. However, in the case of highly complex goods such as forest protection and management, we hypothesise that interviews will lead to higher WTP simply due to a better understanding of the good. Further, the reporting in in-person survey studies is often silent on the number of houses visited or people asked, before someone accepted to take the time for an interview. The people included in the data are of course the ones who accepted, who are likely to have higher WTP than the average person. This is the real response rate that should be compared with mail surveys. As is generally assumed, the higher the response rates, the lower the average WTP, since the survey has managed to capture more of the less-interested, low-WTP respondents. We use three mail survey dummies depending on reported response rates for the mail surveys, "Mailhigh", "Mailmed" and "Maillow". For the reasons above we expect them to be negative, and for the "Mailhigh" to be more negative

[15] As pointed out by an anonymous referee, the log-logistic type of models tend to give very high WTP estimates, and further, if the DC model does not allow for zero responses (the distribution does not include a spike), the WTP difference caused by a higher number of zero responses in the OE data may be spurious. We believe this problem is relevant to relatively few estimates in our data, and not significant enough to cloud the overall question mode comparison.

than the "Maillow" coefficient. Response rates can also be interpreted as a proxy for study quality. As for the other proxy variables for quality, "Mscthesis" and "Unpub", it is unclear a priori how these variables relate to WTP.

To distinguish between different payment vehicles we use dummy variables for hypothetical voluntary contributions and actual payments ("Volunpv" and "Actualpay"). It can be expected that surveys requiring actual payments yield (much) lower WTP (Murphy et al., 2005). Research on voluntary contributions is more limited but it is likely that the voluntary payment vehicle may induce statements of higher WTP, since people do not expect to be charged the amount they stated if the project goes ahead (Boyle, 2003). As pointed out by Mitchell and Carson (1989), choice of payment vehicle is about balancing realism with payment vehicle rejection and protest responses. Whether WTP is stated on other than a long-term annual basis, related to use or by (or on behalf of an) individual rather than a household, is captured by the three dummies "Otherpay", "Userpv", and "Individual". The "Otherpay" coefficient can be expected to be positive a priori. This is because WTP estimates stated for a limited time period, as once for all lump sum contributions, per month or per season would be higher, the latter two simply due to human calculation and discounting errors (for example Rabin, 1998). The "Userpv" dummy can be expected to be negative, as WTP related to use does not include potentially important NUV. Very limited research we are aware of has studied whether WTP for environmental goods tend to be different if stated by an individual or by (or behalf of) a household.[16] Quiggin (1998) finds that under certain conditions household WTP will be higher than individual WTP. On the other hand, there are also reasons why individuals may state higher WTP. For instance, as is known in marketing, the individual may in practice invoke one (and a higher) budget for personal consumption goods and one (and lower) budget when "forced" to take the whole household into account. On balance the a priori sign of the "Individual" parameter is not clear.

The next set of variables describing the good are included to investigate how peoples' preferences for forest protection and/or MUF are related to time, scope, geographical dimensions and certain other characteristics.[17] Many of the hypothesised relationships are largely of an exploratory kind, as the literature on forest valuation (or indeed SP research in general) is relatively silent and give limited theoretical or empirical guidance. Starting with the time dimension, standard neoclassical environmental economics would state that whether an individual is asked to value the same good at different times of the year should not matter to her valuation (Jakus et al., 2006),[18] as WTP would take into account the (discounted) stream of benefits to her over all years and seasons from the proposed scenario. However, if asked specifically about WTP for forests activities in the winter season as

[16]Bateman and Munro (2005) and Strand (2005) compare household and individual valuation related to risk reductions, but the results are not immediately relevant for forests goods.

[17]Some of the variables included under the good description heading could also be called "methodological", but are included here as they relate specifically to the good valued.

[18]Given constant utility function, budget constraint and supply of other unpriced, environmental goods.

compared to the summer, the WTP can of course be different as two essentially different goods are valued. As literally all of the SP forest surveys in the meta-analysis ask for WTP for protection or MUF unrelated to different seasons as such, we thought it would be interesting to study whether people would see through this "season illusion". If they do not, we hypothesise that they have a lower WTP when asked during the darker and colder autumn and winter months,[19] i.e. the "Season" parameter would be negative.

Another interesting, and largely explorative, question we ask is whether people value scenarios that involve full protection more or differently from scenarios that only propose MUF or a mix (using the dummies "Forestpract" and "Protmix"). It is not clear which direction this relationship would go. NUV is higher for protection almost by definition, though some people may have a positive WTP to keep up "traditional" forestry rather than to leave forests "idle" even if they will never use the forest. Full protection may also increase UV for example related to certain recreation activities, but may also make forests dark and less accessible due to fallen trees and dense undergrowth (Horne et al., 2005). Another factor is that people may prefer alternatives to full protection due to the (perceived) economic and cultural importance of forestry and high conflict levels in the three countries.

It is difficult to capture the quality and/or quantity (scope) of a forest good to study whether peoples' WTP is sensitive to different provision levels of the good. Protection vs. MUF captures one quality dimension, while the size in hectares or share of total land or forest area is a crude measure of quantity (included as the dummies "Forestarea", "Hafperc" and "Haperc"). To probe deeper into the issue of scope sensitivity, we coded those surveys that explicitly mention as part of their good description the size and percentage of forest to be valued (dummy variable "Impl"). A complication is that when MUF is valued, the survey sometimes does not refer to a specific forest area (but implicitly, perhaps, means all the productive forests in the country). We therefore do not include the forest area and percentage variables in our primary estimation models in the next section, but utilise them when we look closer at the issue of scope sensitivity. Another factor complicating the issue of scope sensitivity is the geographical dimension of the good (captured in the dummy variables "Localgood", "Reggood"). The protection of a local municipality forest may yield higher WTP per person than for a national forest protection plan although the size of the forest is marginal. An interpretation of this phenomenon and a common result in the literature is that WTP decays with distance. Multiplied with the relevant population around the municipality forest, however, the total WTP is of course much lower. Using a measure of per person WTP/area as dependant variable or relating average WTP per person blindly to the size of the forest would of course not be meaningful. We return to the issue of scope in the next section.

We further include dummy variables for forest environments that are primarily urban ("Urban") (in or adjacent to large cities). Urban forests have potentially high

[19]The much documented psychological effect of lighter seasons on happiness, and its potential effect on WTP, may be difficult to discern from other aspects related to differences in the perceived forest good being valued.

UV but arguably lower NUV for example related to biodiversity, which leaves an ambiguous sign for the parameter. A confounding factor is the higher incomes of populations in urban areas potentially pushing WTP estimates upwards, which we cannot easily control for. Without having strong a priori expectations related to country differences, we include dummies for Sweden and Finland (base case is Norway). Incomes in Norway are somewhat higher, the forest good somewhat scarcer (both in terms of percentage protected and total forest area), which would tend to generate higher Norwegian WTP estimates. On the other hand, the demonstrated willingness to protect forests in Sweden and Finland and the relatively lower levels of user conflicts (Vatn et al., 2005), could reflect a higher underlying WTP for forests in these countries.

We include a dummy for whether respondents are asked WTP to avoid a loss or achieve a gain (“Avoidloss”). In principle, these need not be equal for an equal size change in environmental quality, as the reference scenarios are different. The extensive literature on psychological economics show that people tend to value losses higher than equal-sized gains (Kahneman and Tversky, 2000), which would indicate a positive parameter value for this variable. However, in many studies it is not always very clear whether respondents are asked their WTP to avoid a loss or to achieve a gain. For example, if you ask for WTP for forest protection, and the baseline scenario is accelerating loss of biodiversity, the estimate should be interpreted as WTP to avoid a loss. However, in another survey, the default scenario may be status quo, and increased protection a genuine positive change. As we indicated in Table 1, both of these approaches are equally common.[20] These ambiguities are generally caused by unclear good definitions and fuzzy scenario descriptions.

Further, we separate those surveys that stress user respondents over a mix of use and non-use respondents (not necessarily related to user payment vehicles, “Userpv”) with the dummy “Use”.[21] We hypothesise that users generally have higher WTP than non-users, because users are likely also to have higher NUV, i.e. they are more likely to want to protect or better manage forests (over and beyond providing them with for example recreational areas). Finally, we include a dummy for the year of survey. Rather than interpreting year as variable indicating quality, as discussed in the previous section, we would rather interpret this variable as capturing trends over time in WTP, for example reflecting increasing relative value of forest goods compared to other goods measured in the CPI due to growing scarcity and higher interest in and use of environmental goods in Fennoscandia.[22] We would therefore expect a positive parameter for this variable.

[20] In principle, one could imagine three different cases: (1) *increasing* environmental quality over time compared to a constant path; (2) *decreasing* environmental quality compared to a constant path; and (3) *increasing* environmental quality compared to a *decreasing* path.

[21] It was impossible to classify estimates into UV and NUV, as most of the studies do not explicitly use this distinction. However, we were able to classify studies that were predominantly asking users or focusing on UV, while the rest would include a mix.

[22] As mentioned, we found that WTP is generally insensitive to income (in the subset of the studies that reported it). Hence higher incomes can in our case not explain the increase in WTP over time. However,

Model and results

Meta-regression model

To analyse the impact on WTP of the explanatory variables above, the following standard meta-regression model is applied. A number of m ($m = 1, \ldots, M_s$) WTP estimates are identified from each study s ($s = 1, \ldots, S$), and the total number of WTP estimates can then be denoted $M = \sum_{s=1}^{S} M_s$. The set of k ($k = k, \ldots, K$) explanatory variables or regressors are further denoted $x_{k,\text{ms}}$. Measurements from the same SP study may share many of the same values (for example year, geographical area, payment vehicle, etc.), while varying along other dimensions (for example WTP question format). Hence, generally the random error for both the study and measurement levels may have an impact on the measurement of WTP, and the metadata may display panel effects. A meta-regression model that captures these two levels of error can be formulated as follows (Bijmolt and Pieters, 2001):

$$\text{WTP}_{\text{ms}} = \beta_0 + \sum_{k=1}^{K} \beta_k x_{k,\text{ms}} + e_{\text{ms}} + u_{\text{s}}, \tag{1}$$

where β_0 is the constant, β_k the slope parameter, and e_{ms} and u_{s} the random error terms for the measurement and study levels, respectively. The error terms are assumed to be normally distributed with zero mean and variances σ_{e}^2 and σ_{u}^2. There are several approaches to estimating this model depending on assumptions regarding the error covariance matrix. The simplest approach to the data, which has been used in several meta-analyses (Loomis and White, 1996; Rosenberger and Loomis, 2000a), is to treat all measurements (regardless of the source study) as independent replications and hence assume that study level error is zero. This model can be estimated using simple ordinary least squares (OLS) and may in many cases work well (Rosenberger and Loomis, 2000b). A more advanced approach often used in meta-analysis is to apply a Huber–White robust variance estimation procedure to adjust for potential heteroskedasticity and intercluster correlation[23] (Smith and Osborne, 1996). If such correlation exists, the OLS regression will be inefficient and inconsistent in estimated parameters. The Huber–White procedure does not affect the parameter estimates of the model, but provides robust standard errors of the parameters. Several authors advise against weighing estimates from different studies so that each study counts equally in the data, on the grounds that the information from the data is not used optimally (Bateman and Jones, 2003). Regarding

(*footnote continued*)
the way income is measured in CV surveys may not adequately capture the growing wealth in Fennoscandia (for example in property values). A study in Sweden that pools several data-sets allowing a more comprehensive analysis finds positive elasticity of WTP for environmental services to income (Hökby and Söderqvist, 2003).

[23]Some meta-analysis studies use multilevel models, but often find little improvement on the standard models applied here (for example, Bateman and Jones, 2003; Rosenberger and Loomis, 2000b). We therefore do not pursue this approach here.

specification of the functional form of the regression equation, there is no clear consensus in the meta-analysis literature. The most common specifications are linear, double log, semi- and translog (Johnston et al., 2005). Given this empirical framework, we choose four different models. The first is a simple OLS, the second and third are Huber–White robust estimations for the untransformed variables and a double log specification,[24] respectively. The fourth, and final model is a version of model 2, where we following Rosenberger and Loomis (2000a), retain only those variables that are significant at an 80% level or better based on *t*-statistics.

Model results and discussion

Results

The regression results displayed in Table 3 show that the models fit the data well and that many of our empirical or theoretical expectations are confirmed. The four models explain more than three-quarters of the variation in the data, which is high compared to other meta-studies with R^2's sometimes as low as 0.25 (Rosenberger and Loomis, 2000a). Likelihood ratio tests further demonstrate that the parameters are jointly significant at $p<0.01$ in all models. Starting with the first model, it confirms several of our expectations to the methodological variables, where such prior expectations exist. Open ended WTP format (OE, though not OEPC), payment vehicles (''Voluntpv'' and ''Actualpv''), and the mail survey variables all have the expected signs and are highly significant (''Mailhigh'', ''Mailmed'', ''Maillow''). It is worth noting that the coefficients for the mail survey variables are ranked as expected: the higher response rates the lower WTP. OEPC shows, somewhat unexpectedly, a low positive coefficient, though statistically insignificant. The ''Otherpay'' variable has the expected sign, but is not significant.

The model further shows that people have significantly higher WTP when stated as an individual than for a household. This result is interesting, but there is little research, we are aware of, that study such differences. One possible explanation we have mentioned is that a person asked for household WTP automatically is forced to think about a more restrictive family budget constraint, than an individual considering her own private consumption budget only. There are very few observations for CE, and the model is unable to distinguish CE estimates from CV DC estimates. The study quality dummies related to whether the estimates have been published or not (''Unpub'') or are from Master theses, give significantly different (negative) WTPs as compared with the other studies in the meta-data. This result is not immediately easy to explain, as a normal assumption many analysts make (though likely not based on hard evidence) is that higher methodological prudence should lead to more conservative and lower WTP estimates. This is not the case here, and may raise questions about inclusion of such studies in meta-analysis and benefit transfer exercises.

[24]Only the continuous ''Year'' variable of the regressors is transformed, while the dummy variables are kept on a linear form.

Table 3. Meta-regression results for different models

Variable	Model 1: OLS	Model 2: Huber–White (linear)	Model 3: Huber–White (dbl log)	Model 4: model 2 restricted
Intercept	1549.256*	1549.256*	4.140617**	1342.252**
	(854.0126)	(875.5331)	(1.170449)	(627.3681)
CE	192.6951	192.6951	.3,297,439	
	(539.4353)	(378.0004)	(0.2406569)	
OE	−1334.071***	−1334.071**	−.495455	−1287.111**
	(349.4965)	(594.0914)	(0.3395935)	(468.0961)
OEPC	227.536	227.536	−0.3608809	
	(385.0719)	(382.0898)	(0.2204971)	
Volunpv	3799.7***	3799.7***	2.803627***	3044.605**
	(857.556)	(988.7608)	(0.7711909)	(687.6332)
Userpv	−2564.024***	−2564.024***	−0.3300177	−2106.395***
	(596.0903)	(424.8793)	(0.4289763)	(368.3456)
Otherpay	183.4371	183.4371	−0.066285	
	(554.1872)	(620.5135)	(0.4875653)	
Actualpay	−571.5364	−571.5364*	−2.099854***	−784.9491*
	(822.7707)	(320.3029)	(0.1061977)	(452.4138)
Individual	1834.944***	1834.944***	1.295294***	1887.775***
	(514.8866)	(471.8069)	(0.2941284)	(356.5278)
Mailhigh	−6477.973***	−6477.973***	−4.986712***	−5414.957***
	(1302.404)	(1032.545)	(0.7683036)	(1116.645)
Mailmed	−4864.702***	−4864.702***	−4.270923***	−3766.46***
	(1391.61)	(1043.229)	(0.9019158)	(993.9104)
Maillow	−2476.168**	−2476.168**	−3.009995***	−1777.548**
	(1160.55)	(970.375)	(0.9114381)	(710.3121)
Unpub	−791.1643*	−791.1643**	0.0190386	−649.8414**
	(422.7837)	(320.2655)	(0.3603327)	(276.8186)
Mscthesis	−1916.265**	−1916.265**	−1.730453***	−1377.964**
	(696.9262)	(754.8593)	(0.5586125)	(633.7852)
Forestpract	765.1689**	765.1689**	0.2771635	724.4589**
	(395.7878)	(320.39)	(0.3163496)	(334.9457)
Protmix	−1261.768*	−1261.768	−0.6688487	1277.131**
	(751.8913)	(808.1531)	(0.5322865)	(560.8567)
Impl	1276.517**	1276.517	1.279632**	1461.405***
	(625.134)	(934.0211)	(0.525085)	(511.5116)
Localgood	649.1225	649.1225	−0.4468539	
	(575.4894)	(536.0937)	(0.4902242)	
Reggood	2350.52***	2350.52***	0.821114*	1576.816***
	(859.5872)	(746.4256)	(0.471253)	(462.357)
Sweden	1111.561	1111.561	2.147048**	1032.856**
	(947.6924)	(822.4675)	(0.9714438)	(484.5947)
Finland	644.2306	644.2306	2.131236*	
	(1110.856)	(1046.65)	(0.6016583)	

Table 3. (*continued*)

Variable	Model 1: OLS	Model 2: Huber–White (linear)	Model 3: Huber–White (dbl log)	Model 4: model 2 restricted
Urban	−1551.158**	−1551.158***	−0.5718084	−950.6182**
	(612.1044)	(552.4695)	(0.4513243)	(350.0053)
Season	−1879.212***	−1879.212***	−0.784065**	−1683.18***
	(433.4073)	(496.1174)	(0.313954)	(473.2106)
Avoidloss	627.9457	627.9457	0.5853566*	585.4352
	(401.3341)	(415.2456)	(0.3072963)	(346.2406)
Use	451.9457	451.9457	0.0224779	
	(526.0146)	(721.9776)	(0.3540051)	
Year	130.3553*	130.3553	1.242805**	140.1013**
	(71.80079)	(82.63281)	(0.5555091)	(61.82823)
Log-likelihood χ^2	101.47***	101.47***	121.56***	97.45***
R^2	0.756	0.756	.815	0.742
N	72	72	72	72

Note: $*p<0.10$, $**p<0.05$, $***p<0.01$, number of survey clusters for models 2–4 = 27. Estimated using Stata version 9.2.

We included a range of good description variables of a more explorative kind, i.e. not much research has been conducted on which to base solid prior expectations. The geographical variables in the model show as expected that regional and local forest goods are valued higher than a forest on a national level (the base case), though the latter is not statistically significant. Further, Sweden and Finland do not have significantly different WTP than observed in Norway. Urban forests are valued lower than other forests, which may indicate that NUV of non-urban forests is important. As hypothesised, WTP to avoid a loss is higher (though not significantly so) than WTP for a gain. WTP from users or related primarily to use ("Use") is not statistically different than from a mixed group. We also hypothesised that respondents would consider protection, MUF or a mix of the two as different types goods. Our results here are somewhat puzzling, as it seems that respondents value full protection lower than MUF ("Forestpract" is significantly positive), but higher than a mix between the two ("Protmix" is significantly negative). Further, it also seems to be important to the stated WTP whether forest area and percentage have been explicitly mentioned in the survey ("Impl" is positive and significant). These results are of an exploratory kind, but shows at least that it is not immaterial to people whether it is question of full protection or just a change in existing forestry practices. More research is required to probe deeper into people's preferences for different types of forest regulation.

Finally, the results regarding the temporal dimension are interesting. We hypothesised that people may value forests lower in the autumn/winter as compared to the spring/summer, due to a "season illusion". Our model shows that the season variable is negative and highly significant. In trial runs of the OLS model we also coded a winter-variable (November–March) to see if the snowy season would be a better categorisation, but for this variable we found no significant effects. There is not much theory and empirical evidence we can rely on to explain the negative season parameter, so it should be interpreted with caution. We also find as expected that the year of the survey influences WTP positively, indicating increased relative value of forest amenities in Fennoscandia over the last 20 years. Also for the temporal dimension, more research is required to better understand which forces are at work.

Sensitivity analysis

If we look at the results of the other three models the significance of many of the parameter values is relatively robust. Contrasting the first with the second model, where potential study level correlation and heteroskedasticity have been adjusted for, the results show small changes. Most notably, "Actualpay" is now significant at the 5% level, while the variables "Protmix", "Forestpract" and "Year" are no longer significant. For the other parameters there are minor changes. This supports the findings of other meta-studies that the effects associated with systematic study (or author) level variance are often not significant (Rosenberger and Loomis, 2000b; Johnston et al., 2003, 2005). As pointed out by Johnston et al. (2005) this is an important result suggesting that systematic variation in WTP is not driven by unobservable attributes unique to particular studies. The double-log transformation in model 3 shows a slightly better fit to the data compared to models 1 and 2, due to the relative skewedness of the average WTP-distribution towards zero. However, postestimation commands comparing residuals between the models, show that this is of minor importance to the performance of models 1 and 2. The results also display some degree of robustness to the double log specification, though there are some changes. Most notably both the Swedish and Finnish WTP estimates are now significantly positive. Many of the parameters are significant across the three models. We also estimated semi- and translog model formulations, which were found not to perform as well as the models reported. The fourth model in Table 3 removes variables from model 2 whose parameters have $p < 0.20$, and is a first step towards making the model more suitable for benefit transfer applications (Shrestha and Loomis, 2001, 2003). The model loses some explanatory power by removing variables, but the model now contains variables where all, but one ("Avoidloss"), are significant.

Concerned that the price and exchange rate adjustments would cloud our results, we also reran the simple OLS model using 1998 (mean survey year) as the base year, varying between PPP, market exchange rates, and using a weighted price index for the three countries. Both practices of choosing the current and mean survey year as the base year for analysis are common in the meta-analysis literature, though we have seen no studies testing for potential effects of such choices. The results of the model runs under these alternative WTP adjustments, left out of the Table 3 for sake of brevity, did not indicate that choice of base year, currency rates or inflation index had significant impact on results. Recalling that we decided in our meta-analysis to average over reported WTP estimates from the same studies that varied across dimensions that could not be meaningfully coded (especially econometric model specifications, trimmed vs. untrimmed estimates, etc.), we decided to assess preliminary the effects of this procedure. The model runs with all unweighted observations show that R^2, not surprisingly, falls significantly (to around 0.4) and many of the parameters are no longer significant (though their signs are generally preserved). Since the variation is too large just to be included as study background, we think it is justifiable to apply our weighting procedure as long as it is carried through consistently for all estimates. In this way we are able to pick up important and significant relationships from the meta-data that would otherwise remain obscure. In this case we can identify the main source of the variation (DC modelling choices especially), but cannot control for it due to insufficient study reporting practices. As a final check of the robustness of our parameters, we excluded two high estimates (WTP > NOK 5000), and one low estimate (from the only study measuring *actual* WTP) (WTP < NOK 15) and reran model 2. There are changes to parameter significance for "OE", "Forestpract", "Unpub", "Impl" and "Protmix" (no longer significant), "Swe" (now significant), while the other variable parameters remain significant at $p < 0.1$ or better. We think including the three observations above is the most prudent approach, as none of them are unrealistically large or small. Our sensitivity considerations here can at least be seen as a preliminary assessment of robustness of the meta-analysis model.

Is WTP sensitive to scope?

A CV critique that has been hotly debated since it was first raised is the issue of embedding effects (Kahneman and Knetsch, 1992). Embedding has come to mean at least three different things (Hanemann, 1994), the most important being scope insensitivity, i.e. that WTP is not (sufficiently) sensitive to changes in the quantity or quality of the good being valued. Second, WTP is sometimes found to depend on which number the good is in a sequence of items to be valued (sequencing effect). Third, WTP of a change of a composite public good may be less than the sum of the WTP for individual changes separately (sub-additivity effect). If these phenomena cannot be explained by legitimate economic reasons, the theoretical validity of the CV method can be challenged. Since both convergent and criterion validity are hard to judge for CV of NUV, a presumably important component of forest values, the pillars of theoretical and content validity will need to be all the more solid (Mitchell and Carson, 1989). Many of the studies in the meta-analysis consider within sample (internal) or split-sample (external) scope tests, often offering two or three different sized forest protection plans as measured in hectares and/or as percentage protected. A smaller number of the studies that only consider MUF assess sensitivity to scope. Only one or two studies consider the two other elements of embedding (Veisten et al., 2004b). In all cases the results are mixed. As discussed previously, it is problematic to assess scope sensitivity in the meta-models above due to higher WTP for local/regional goods and because some studies valuing MUF sometimes do not specify area (neither to the respondent nor to the reader) forcing us to code the whole productive forest area in the country. Another complicating issue is that some surveys do not distinguish clearly enough between the *change* in forest area, which is the good that should be valued, and the existing area of forests under certain protection or forestry restriction regimes. Further, the surveys use both the terms "productive forest area" and "total forest area" (with and without for example lakes and marshland). In other words, the good and scenario descriptions become unclear.

To account for these problems, we ran several models for subsets of the data to try to detect sensitivity to forest size (area in hectares and as a percentage of total land area of productive forest size). We first ran model 2 only for those 64 observations that had indicated a relevant forest area in the study. Second, for these observations, we also ran the model for surveys that value protection only and forests on a national level (i.e. excluding local and regional forest goods), hypothesising that protection may be more sensitive to scope than MUF. Finally, we estimated a small-sample model for those estimates from surveys where size and percentage protected were explicitly given to respondents in survey instruments, and for those estimates that were considered especially related to use. The somewhat discouraging result is that neither of these model approaches was able to detect any significant scope effects, beyond a generally weak, near-zero positive relationship. On the other hand, forests are complex environmental goods which scope may not be easily captured by simplified indicators such as area size or percentage. While other meta-analyses detect sensitivity to scope (Smith and Osborne, 1996), our findings strongly suggest showing caution in using WTP/hectare or similar measures in meta-analysis and benefit transfer applications for complex goods, as done for example in Woodward and Wui (2001). Since value per hectare is also a format much sought after in policy applications, it is tempting to overlook the challenges involved in estimation and interpretation.

Conclusions

This paper has taken stock of 20 years of stated preference (SP) research valuing non-timber benefits (NTBs) in Norway, Finland and Sweden by the use of meta-analysis. The paper first reviewed the literature and summarised methodological traditions showing a rich and varied body of SP research. Second, a meta-regression analysis was conducted attempting to explain the variation in Willingness-to-Pay (WTP) for protection and multiple use forestry (MUF) by differences in survey methodology, good characteristics, socio-economic and other variables. The model results are promising with regard to revealing systematic and expected variation in WTP along methodological variables, and to some extent along various characteristics of the forest good. Most notably, it is shown that geography (urban, local, and regional), seasons (autumn/winter vs. spring/summer), year and institution (full protection vs. MUF) are important, but WTP does not seem to be different between the three countries. The results are fairly robust to changes in model specification and meta-analysis scope, but it is acknowledged that some of the included variables are of an explorative kind requiring further research.

Two key conclusions with relevance for future research can be drawn from the meta-analysis. First, analysing several subsets of the data, no sensitivity to scope of WTP to the size of the forest (in hectare or percentage) was detected. It is likely that this result stems from a combination of weaknesses in SP survey design (especially unclear scenario and good descriptions) and respondent difficulties in assessing a complex and multidimensional forest good. In any case, it is an important result casting doubt on the validity of using simplified WTP/area measures, at least at current state of knowledge, pointing towards more research to understand embedding effects for complex environmental goods. Second, we find that individuals tend to value forests higher than households do. This result may run counter to some of the limited research in this area, but suggests that much is still unknown about which budgets people invoke in their minds when asked as individuals rather than on behalf of a household.

A final point of relevance to the meta-analysis literature worth emphasising in closing is the importance of conducting sensitivity analysis, varying the scope of the meta-analysis in particular. Since the reporting in Fennoscandian and international SP research still leaves much to be desired, the meta-analyst is left with difficult choices about which variables and studies to include. Many of the meta-analyses in the environmental economics literature conduct sensitivity analysis by applying different econometric model specifications, but tend to overlook and/or underreport the potentially important effects of varying the scope of the meta-analysis.

Acknowledgements

The author would like to acknowledge valuable comments from Associate Professors Ståle Navrud and Olvar Bergland at the Department of Economics and Resource Management at the Norwegian University of Life Sciences, Jette Bredahl Jacobsen at the Department of Economics, Policy and Management, Danish Royal Veterinary and Agricultural University, and from participants at the 13th Ulvön Conference on Environmental Economics, Sweden, 19–21 June, 2006.

References

Arrow, K.J., Solow, R., Leamer, E., Portney, P., Radner, R., Schuman, H., 1993. Report of the NOAA panel on contingent valuation. Federal Register 58, 4601–4614.

Bateman, I.J., Jones, A.P., 2003. Contrasting conventional with multi-level modeling approaches to meta-analysis: expectation consistency in UK woodland recreation values. Land Economics 79 (2), 235–258.

Bateman, I., Munro, A., 2005. Household vs. Individual Valuation: What's the Difference? European Association of Environmental and Resource Economics, Bremen.

Bijmolt, T.H.A., Pieters, R.G.M., 2001. Meta-analysis in marketing when studies contain multiple measurements. Marketing Letters 12 (2), 157–169.

Bojö, J., 1985. Kostnadsnyttoanalys av fjällnära skogar: Fallet Vålådalen (Cost-benefit analysis of mountainous forests: the Vala Valley Case). Research Report, The Economic Research Institute, Stockholm School of Economics.

Boman, M., 1995. Estimating costs and genetic benefits of various sizes of predator populations: the case of Bear, Wolf, Wolverine and Lynx in Sweden. Journal of Environmental Management 43 (4), 349–357.

Bostedt, G., 1997. Public goods in Swedish forests: essays on nonmarket valuation and environmental policy. Ph.D. Thesis, Swedish University of Agricultural Sciences, Umeå.
Bostedt, G., Mattson, L., 1991. Skogens betydelse för turismen: En samhällsekonomisk pilotstudie (The importance of forests for tourism: a pilot cost-benefit analysis). Arbetsrapport 141, Department of Forest Economics, Swedish University of Agricultural Sciences, Umeå.
Bostedt, G., Mattsson, L., 1995. The value of forests for tourism in Sweden. Annals of Tourism Research 22 (3), 671–680.
Boyle, K.J., 2003. Contingent valuation in practice. In: Champ, P.A., Boyle, K.J., Brown, T.C. (Eds.), A Primer on Nonmarket Valuation. Kluwer Academic Publishers, Dordrecht.
Brouwer, R., Langford, I.H., Bateman, I., Turner, R.K., 1999. A meta-analysis of wetland contingent valuation studies. Regional Environmental Change 1 (1), 47–57.
Button, K., 1995. What can meta-analysis tell us about transport? Regional Studies 29, 507–517.
Cameron, T.A., Poe, G.L., Ethier, R.G., Schulze, W.D., 2002. Alternative non-market value-elicitation methods: are the underlying preferences the same? Journal of Environmental Economics and Management 44 (3), 391–425.
Carson, R.T., 2004. Contingent Valuation: A Comprehensive Bibliography and History. Edward Elgar Publishing, Cheltenham, UK.
Carson, R.T., Flores, N.E., Martin, K.M., Wright, J.L., 1996. Contingent valuation and revealed preference methodologies: comparing the estimates for quasi-public goods. Land Economics 72 (1), 80–99.
Carson, R.T., Flores, N.E., Meade, N.F., 2001. Contingent valuation: controversies and evidence. Environmental and Resource Economics 19 (2), 173–210.
Desvousges, W.H., Johnson, F.R., Banzhaf, H.S., 1998. Environmental Policy Analysis with Limited Information: Principles and Applications of the Transfer Method. Edward Elgar, Cheltenham, UK.
Eiswerth, M., Shaw, W., 1997. Adjusting benefits transfer values for inflation. Water Resources Research 33 (10), 2381–2385.
Florax, R.J.G.M., Nijkamp, P., Willis, K.G., 2002. Comparative Environmental Economic Assessment. Edward Elgar, Cheltenham, UK.
Framstad, E., Økland, B., Bendiksen, E., Bakkestuen, V., Blom, H., Branderud, T.E., 2002. Evaluering av barskogvernet i Norge (Evaluation of the conservation of coniferous forests in Norway). NINA Fagrapport 54.
Fredman, P., 1995. The existence of existence value: a study of the economic benefits of an endangered species. Journal of Forest Economics 1 (3), 307–327.
Fredman, P., Emmelin, L., 2001. Wilderness purism, willingness to pay and management preferences: a study of Swedish mountain tourists. Tourism Economics 7 (1), 5–20.
Garnes, A.S., Winther, G., 1991. Holdninger til og betalingsvillighet for flersidig skogbruk og barskogvern i Norge (Attitudes and WTO for multiple use forestry and forest preservation in Norway). Master Thesis, Norwegian University of Life Sciences.
Hanemann, W.M., 1994. Valuing the environment through contingent valuation. Journal of Economic Perspectives 8 (4), 19–43.
Hanley, N., Wright, R.E., Adamowicz, V., 1998. Using choice experiments to value the environment: design issues, current experience and future prospects. Environmental and Resource Economics 11 (3–4), 413–428.
Hoen, H.F., Veisten, K., 1994. En undersøkelse blant brukere av Oslomarka: synspunkter på skogtilstand og skogbehandling (A survey of the users of Oslomarka: attitudes towards forest scenary and forestry practices). Skogforsk 6/94.
Hoen, H.F., Winther, G., 1993. Multiple-use forestry and preservation of coniferous forests in Norway: a study of attitudes and Willingness-to-pay. Scandinavian Journal of Forest Research 8 (2), 266–280.
Holgen, P., Mattsson, L., Li, C.Z., 2000. Recreation values of boreal forest stand types and landscapes resulting from different silvicultural systems: an economic analysis. Journal of Environmental Management 60 (2), 173–180.
Horne, P., Boxall, P.C., Adamowicz, W.L., 2005. Multiple-use management of forest recreation sites: a spatially explicit choice experiment. Forest Ecology and Management 207 (1–2), 189–199.
Horowitz, J.K., McConnell, K.E., 2003. Willingness to accept, willingness to pay and the income effect. Journal of Economic Behavior & Organization 51 (4), 537–545.
Hökby, S., Söderqvist, T., 2003. Elasticities of demand and willingness to pay for environmental services in Sweden. Environmental and Resource Economics 26 (3), 361–383.
Jakus, P.M., Stephens, B., Fly, J.M., 2006. Temporal reliability in contingent valuation (with a restrictive research budget). In: Alberini, A., Kahn, J.R. (Eds.), Handbook on Contingent Valuation. Edward Elgar, Cheltenham, UK, pp. 249–262.

Johansson, P.O., 1989. Valuing public goods in a risky world: an experiment. In: Folmer, H., van Ierland, E.C. (Eds.), Evaluation Methods and Policy Making in Environmental Economics. North-Holland, Amsterdam, pp. 39–48.

Johansson, P.-O., Zavisic, S., 1989. Svenska folkets miljöbudget. Ekonomisk Debatt 6, 472–474.

Johansson, P.O., Kriström, B., Mattson, L., 1988. How is the willingness to pay for Moose hunting affected by the stock of Moose? An empirical study of Moose-hunters in the county of Västerbotten. Journal of Environmental Management (26), 163–171.

Johnston, R.J., Besedin, E.Y., Wardwell, R.F., 2003. Modeling relationships between use and nonuse values for surface water quality: a meta-analysis. Water Resources Research 39 (12).

Johnston, R.J., Besedin, E.Y., Iovanna, R., Miller, C.J., Wardwell, R.F., Ranson, M.H., 2005. Systematic variation in willingness to pay for aquatic resource improvements and implications for benefit transfer: a meta-analysis. Canadian Journal of Agricultural Economics 53 (2–3), 221–248.

Kahneman, D., Knetsch, J.L., 1992. Valuing public-goods – the purchase of moral satisfaction. Journal of Environmental Economics and Management 22 (1), 57–70.

Kahneman, D., Tversky, A. (Eds.), 2000. Choices, Values and Frames. Cambridge University Press, Cambridge.

Kniivilä, M., 2004. Contingent valuation and cost-benefit analysis of nature conservation: a case study in North Karelia, Finland. D.Sc. (Agriculture and Forestry) Thesis, University of Joensuu.

Kniivilä, M., Ovaskainen, V., Saastamoinen, O., 2002. Costs and benefits of forest conservation: regional and local comparisons in Eastern Finland. Journal of Forest Economics 8 (2), 131–150.

Kriström, B., 1989. On the benefits of preserving virgin forests in Sweden. Scandinavian Forest Economics (30), 141–166.

Kriström, B., 1990a. A nonparametric approach to the estimation of welfare measures in discrete response valuation studies. Land Economics 66 (2), 135–139.

Kriström, B., 1990b. Valuing environmental benefits using the contingent valuation method – an econometric analysis. Ph.D. Thesis, Umeå University.

Laitila, T., Paulrud, A., 2006. A multi-attribute extension of discrete-choice contingent valuation for valuation of angling site characteristics. Journal of Leisure Research 38 (2).

Lehtonen, E., Kuuluvainen, J., Pouta, E., Rekola, M., Li, C.Z., 2003. Non-market benefits of forest conservation in southern Finland. Environmental Science and Policy 6 (3), 195–204.

Lehtonen, E., Kuuluvainen, J., Li, C.-Z., Pouta, E., Rekola, M., 2005a. Preference uncertainty in contingent valuation: the case of forest conservation in southern Finland, University of Helsinki, Department of Forest Economics.

Lehtonen, E., Kuuluvainen, J., Ovaskainen, V., Pouta, E., Rekola, M., 2005b. Influence of Logit Model Assumptions on Estimated Willingness to Pay for Forest Conservation in Southern Finland, University of Helsinki, Department of Forest Economics.

Leidal, K., 1996. Verdsetting av et bynært rekreasjonsområde: en betinget verdsettingsundersøkelse av området rundt Eigevannet i Kristiansand kommune (Valuation of an urban recreation area: a contingent valuation study of the Eige Lake area in Kristiansand municipality). Master Thesis, Norwegian University of Life Sciences.

Li, C.Z., 1996. Semiparametric estimation of the binary choice model for contingent valuation. Land Economics 72 (4), 462–473.

Li, C.Z., Mattsson, L., 1995. Discrete-choice under preference uncertainty – an improved structural model for contingent valuation. Journal of Environmental Economics and Management 28 (2), 256–269.

Li, C.Z., Kuuluvainen, J., Pouta, E., Rekola, M., Tahvonen, O., 2004. Using choice experiments to value the Natura 2000 Nature conservation programs in Finland. Environmental and Resource Economics 29 (3), 361–374.

List, J.A., Gallet, C.A., 2001. What experimental protocol influence disparities between actual and hypothetical stated values? Environmental and Resource Economics 20 (3), 241–254.

Loomis, J.B., White, D.S., 1996. Economic benefits of rare and endangered species: summary and meta-analysis. Ecological Economics 18 (3), 197–206.

Mattsson, L., Li, C.Z., 1993. The non-timber value of Northern Swedish forests – an economic-analysis. Scandinavian Journal of Forest Research 8 (3), 426–434.

Mattsson, L., Li, C.Z., 1994. How do different forest management-practices affect the non-timber value of forests – an economic-analysis. Journal of Environmental Management 41 (1), 79–88.

Mitchell, R.C., Carson, R.T., 1989. Using Surveys to Value Public Goods: The Contingent Valuation Method. Resources for the Future, Washington, DC.

Murphy, J.J., Allen, P.G., Stevens, T.H., Weatherhead, D., 2005. A meta-analysis of hypothetical bias in stated preference valuation. Environmental and Resource Economics 30, 313–325.

Mäntymaa, E., Mönkkönen, M., Siikamäki, J., Svento, R., 2002. Estimating the demand for biodiversity – vagueness band and open-ended questions. In: van Ierland, E.C., Weikard, H.P., Wesseler, J. (Eds.), Proceedings: Risk and Uncertainty in Environmental and Resource Economics.
Navrud, S., 1992. Pricing the European Environment. Scandinavian University Press, Oslo.
Navrud, S., Ready, R. (Eds.), 2006. Environmental Value Transfer: Issues and Methods. Kluwer Academic Publishers, Dordrecht.
Olsson, C., 1993. Linbanan i Norsjö – En skogsinriktad turistattraktion. En undersökning av betalningsvilja för att uppleva skog och natur (The cable car in Norsjö – a forest based tourist attraction). Master, Handelshögskolan i Umeå.
Ovaskainen, V., Kniivilä, M., 2005. Consumer versus citizen preferences in contingent valuation: evidence on the role of question framing. Australian Journal of Agricultural and Resource Economics 49 (4), 379–394.
Pattanayak, S.K., Wing, J.M., Depro, B.M., van Houtven, G., de Civita, P., Stieb, D.M., Hubbel, B., 2002. International Health Benefits Transfer Application Tool: The Use of PPP and inflation indices. Prepared for Economic Analysis and Evaluation Division, Health Canada.
Pouta, E., 2003. Attitude-behavior framework in contingent valuation of forest conservation. Ph.D., University of Helsinki.
Pouta, E., 2004. Attitude and belief questions as a source of context effect in a contingent valuation survey. Journal of Economic Psychology 25, 229–242.
Pouta, E., 2005. Sensitivity to scope of environmental regulation in contingent valuation of forest cutting practices in Finland. Forest Policy and Economics 7, 539–550.
Pouta, E., Rekola, M., 2005. Meta analysis of forest valuation studies. Working paper (In Finnish).
Pouta, E., Rekola, M., Kuuluvainen, J., Tahvonen, O., Li, C.Z., 2000. Contingent valuation of the Natura 2000 nature conservation programme in Finland. Forestry 73 (2), 119–128.
Pouta, E., Rekola, M., Kuuluvainen, J., Li, C.Z., Tahvonen, I., 2002. Willingness to pay in different policy-planning methods: insights into respondents' decision-making processes. Ecological Economics 40 (2), 295–311.
Quiggin, J., 1998. Individual and household willingness to pay for public goods. American Journal of Agricultural Economics 80, 58–63.
Rabin, M., 1998. Psychology and Economics. Journal of Economic Literature XXXVI, 11–46.
Rekola, M., Pouta, E., 2005. Public preferences for uncertain regeneration cuttings: a contingent valuation experiment involving Finnish private forests. Forest Policy and Economics 7, 635–649.
Rekola, M., Pouta, E., Kuuluvainen, J., Tahvonen, O., Li, C.Z., 2000. Incommensurable preferences in contingent valuation: the case of Natura 2000 Network in Finland. Environmental Conservation 27 (3), 260–268.
Rosenberger, R.S., Loomis, J.B., 2000a. Using meta-analysis for benefit transfer: in-sample convergent validity tests of an outdoor recreation database. Water Resources Research 36 (4), 1097–1107.
Rosenberger, R., Loomis, J., 2000b. Panel stratification in meta-analysis of economic studies: an investigation of its effects in the recreation valuation literature. Journal of Agricultural and Applied Economics 32 (1), 131–149.
Sandsbråten, L., 1997. Verdsetting av miljøgoder i Oslomarka: en betinget verdsettingsundersøkelse i privat og kommunal skog i Indre Oslomarka (Valuation of environmental goods in Oslomarka: a contingent valuation survey of private and municipality owned forest in inner Oslomarka). Master Thesis, Norwegian University of Life Sciences.
Sayman, S., Öncüler, A., 2005. Effects of study design characteristics on the WTA–WTP disparity: a meta analytical framework. Journal of Economic Psychology 26 (2), 289–312.
Schläpfer, F., 2006. Survey protocol and income effects in the contingent valuation of public goods: a meta-analysis. Ecological Economics 57 (3), 415–429.
Shrestha, R.K., Loomis, J.B., 2001. Testing a meta-analysis model for benefit transfer in international outdoor recreation. Ecological Economics 39 (1), 67–83.
Shrestha, R.K., Loomis, J.B., 2003. Meta-analytic benefit transfer of outdoor recreation economic values: testing out-of-sample convergent validity. Environmental & Resource Economics 25, 79–100.
Siikamäki, J., Layton, D., in press. Discrete choice survey experiments: a comparison using flexible methods. Journal of Environmental Economics and Management, forthcoming.
Simensen, K., Wind, M., 1990. Holdninger til og betalingsvillighet for ulike skogbehandlinger i fjellskog: En empirisk undersøkelse av hirkjølen statsallmenning (attitudes and wtp for different forestry practices in mountainous forests: a survey of the hirkjolen common). Master Thesis, Norwegian University of Life Sciences.

Skagestad, E., 1996. Friluftsliv og Skogbruk – En spørreundersøkelse blant turgjengere i ytre Oslomarka, Romeriksåsen, på vinterstid (Recreation and forestry – a survey of hikers in the outer Oslomarka, Romeriksåsen, in the winter time). Master Thesis, Norwegian University of Life Sciences.
Smith, V.K., Kaoru, Y., 1990. Signals or noise – explaining the variation in recreation benefit estimates. American Journal of Agricultural Economics 72 (2), 419–433.
Smith, V.K., Osborne, L.L., 1996. Do contingent valuation estimates pass a "scope" test? A meta-analysis. Journal of Environmental Economics and Management 31 (3), 287–301.
Smith, V.K., Pattanayak, S.K., 2002. Is meta-analysis a Noah's Ark for non-market valuation? Environmental and Resource Economics 22, 271–296.
Stanley, T.D., 2001. Wheat from chaff: meta-analysis as quantitative literature review. Journal of Economic Perspectives 15 (3), 131–150.
Stanley, T.D., Jarrel, S.D., 2005. Meta-regression analysis: a quantitative method of literature surveys. Journal of Economic Surveys 19 (3), 299–308.
Strand, J., 2005. Individual and household values of mortality reductions with intrahousehold bargaining. Journal of Risk and Uncertainty 31 (2), 217–236.
Strand, J., Wahl, T.S., 1997. Verdsetting av kommunale friområder i Oslo: en betinget verdsettingsstudie (Valuation of Municipality Recreation Areas in Oslo: A Contingent Valuation Study). SNF Report 82/97.
Sundberg, S., Söderqvist, T., 2004. The economic value of environmental change in Sweden: a survey of studies. Report 5360, Naturvårdsverket.
Tyrväinen, L., 2001. Economic valuation of urban forest benefits in Finland. Journal of Environmental Management 62 (1), 75–92.
Tyrväinen, L., Väänänen, H., 1998. The economic value of urban forest amenities: an application of the contingent valuation method. Landscape and Urban Planning 43 (1–3), 105–118.
Vatn, A., Framstad, E., Solberg, B., 2005. Tiltak og virkemidler for vern av biodiversitet i skog og våtmarker. Rapport TemaNord 2005, Nordisk Ministerråd, København.
Veisten, K., 1993. Samfunnsøkonomisk verdsetting av å oppretthalde artsmangfaldet i norske skogar: med testing for feilkilder ved betinga verdsetting (Social valuation of the preserving biodiversity in Norwegian forests: with multiple bias testing). Master Thesis, University of Oslo.
Veisten, K., Hoen, H.F., 1994. Om haldningar til – og betalingsvilje for endringar i skogbehandlinga i Noreg: resultat frå to nasjonale undersøkingar (On attitudes and WTP for changes in forest practices in Norway: results from two national surveys). Skogforsk 4/94, Skogforsk.
Veisten, K., Navrud, S., 2006. Contingent valuation and actual payment for voluntarily provided passive-use values: assessing the effect of an induced truth-telling mechanism and elicitation formats. Applied Economics 38 (7), 735–756.
Veisten, K., Hoen, H.F., Navrud, S., Strand, J., 1993. Valuing biodiversity in Norwegian forests: a contingent valuation study with multiple bias testing. Memo No. 7, Department of Economics, University of Oslo.
Veisten, K., Hoen, H.F., Navrud, S., Strand, J., 2004a. Scope insensitivity in contingent valuation of complex environmental amenities. Journal of Environmental Management 73 (4), 317–331.
Veisten, K., Hoen, H.F., Strand, J., 2004b. Sequencing and the adding-up property in contingent valuation of endangered species: are contingent non-use values economic values? Environmental and Resource Economics 29 (4), 419–433.
Walsh, R.G., Johnson, D.M., McKean, J.R., 1989. Issues in non-market valuation and policy application: a retrospective glance. Western Journal of Agricultural Economics 1, 178–188.
Walsh, R.G., Johnson, D.M., McKean, J.R., 1990. Nonmarket values from two decades of research on recreation demand. In: Link, A., Smith, V.K. (Eds.), Advances in Applied Micro-Economics, vol. 5. JAI Press, Greenwich, CT, pp. 167–193.
Woodward, R.T., Wui, Y.-S., 2001. The economic value of wetland services: a meta-analysis. Ecological Economics 37, 257–270.

Paper II

ECOLOGICAL ECONOMICS 66 (2008) 425–435

available at www.sciencedirect.com

ScienceDirect

www.elsevier.com/locate/ecolecon

ANALYSIS

How reliable are meta-analyses for international benefit transfers?☆

Henrik Lindhjem[a,b,*], *Ståle Navrud*[a]

[a]*Department of Economics and Resource Management, Norwegian University of life Sciences, P.O. Box 5003, N-1432 Ås, Norway*
[b]*ECON, P.O. Box 5, N-0051, Oslo, Norway*

ARTICLE INFO

Article history:
Received 6 July 2007
Received in revised form
25 September 2007
Accepted 2 October 2007
Available online 31 October 2007

Keywords:
Benefit transfer
Environmental valuation
Meta-analysis
Forest

JEL:
Q51; H41

ABSTRACT

Meta-analysis has increasingly been used to synthesise the environmental valuation literature, but only a few test the use of these analyses for benefit transfer. These are typically based on national studies only. However, meta-analyses of valuation studies across countries are a potentially powerful tool for benefit transfer, especially for environmental goods where the domestic literature is scarce. We test the reliability of such international meta-analytic transfers, and find that even under conditions of homogeneity in valuation methods, cultural and institutional conditions across countries, and a meta-analysis with large explanatory power, the transfer errors could still be large. Further, international meta-analytic transfers do not on average perform better than simple value transfers averaging over domestic studies. Thus, we question whether the use of meta-analysis for practical benefit transfer achieves reliability gains justifying the increased effort. However, more meta-analytic benefit transfer tests should be performed for other environmental goods and other countries before discarding international meta-analysis as a tool for benefit transfer.

© 2007 Elsevier B.V. All rights reserved.

1. Introduction

Meta-analysis (MA) is now common in environmental economics and non-market valuation. Since Smith and Kaoru's (1990) seminal study of recreational benefits, MA has been conducted for a wide range of environmental goods, from wetlands (Woodward and Wui, 2001) to visibility (Smith and Osborne, 1996). Common to all of these studies is the focus on research synthesis and hypothesis testing, rather than on the more interesting policy question of how MA can be used to improve benefit transfer (BT) practices (Smith and Pattanayak, 2002). Meta-analytic benefit transfer (MA-BT) to unstudied sites ("policy sites") has only been cursory treated in the literature, typically a few pages add-ons at the end of lengthy MA papers, although authors emphasise its potential importance for future research and applications, for example in cost–benefit analysis (see the special issue on BT in Ecological Economics, 2006; Van Houtven et al., 2007). While there is some knowledge of how unit value and value function-based BT from single studies perform (Rosenberger and Phipps, 2007), Bergstrom and Taylor (2006, pp. 359) point out that "before widespread application of MA-BT models, there is a need for additional MA-BT validity tests across different types of natural resources and environmental commodities." Only a few studies have, to our knowledge, investigated the validity and reliability of

☆ A previous version was presented at the conferences of the US Society of Ecological Economics (New York 26. June 2007) and the European Association of Environmental and Resource Economists (Thessaloniki, Greece 30. June 2007).

* *Corresponding author.* ECON, P.O. Box 5, N-0051, Oslo, Norway. Tel.: +4798263957; fax: +4764965701.
E-mail address: henrik.lindhjem@umb.no (H. Lindhjem).

0921-8009/$ – see front matter © 2007 Elsevier B.V. All rights reserved.
doi:10.1016/j.ecolecon.2007.10.005

MA-BT (Santos, 1998; Rosenberger and Loomis, 2000; Shrestha and Loomis, 2001, 2003; Santos, 2007; Shrestha et al., 2007). Four of the studies, however, are based on the same large dataset of use values for different recreational activities in the USA, and are unable to cover the breadth of issues involved in more typical MA-BT exercises, i.e. limited datasets, complex goods with significant non-use values, different level of methodological heterogeneity and mix of international studies to mention a few. Santos (2007) is the only study attempting a comprehensive comparison of two versions of a domestic MA-BT with simple BT techniques often used in practice. Further, all the above studies can be said to under-appreciate the potential impacts on the MA-BT performance of model specifications, values of methodological variables (Johnston et al., 2006) and other choices the meta-analyst needs to make (Hoehn, 2006)[1].

This paper aims to investigate the validity and reliability of international MA-BT of non-timber benefits based on a recently published MA of contingent valuation (CV) studies in Norway, Sweden and Finland (Lindhjem, 2007). Compared to previous research on MA-BT, our paper adds several new and interesting dimensions: (i) a more systematic and diverse testing of different MA-BT models, including comparisons with simple BT techniques, (ii) the good we investigate is complex and has substantial non-use values related to biodiversity (rather than mainly use values), (iii) data from three countries, which are similar culturally, economically, institutionally (e.g. people's right to walk in private forests), and in the way the good is perceived and used, and (iv) data are generally more homogenous methodologically since only CV studies are included. We investigate the transfer error (TE) of four different meta-regression model specifications, and use the best two models to compare MA-BT with simple unit value transfer techniques. A key question is whether MA-BT achieves reliability gains justifying the increased effort. As pointed out by Navrud and Ready (2007a, pp. 288): "Simple approaches should not be cast aside until we are confident that more complex approaches do perform better".

2. Validity and reliability of meta-analytic benefit transfer

2.1. *Underlying theory of MA-BT*

The simple underlying indirect utility function for a change from Q_0 to Q_1 in the quality/quantity vector describing an environmental good available to individual i is:

$$V_i(p_i, I_i, Q_0) = V_i(p_i, I_i - \text{WTP}, Q_1) \qquad (1)$$

where P_i, I_i are a market price vector and income, respectively, and WTP is Willingness-to-Pay. Eq. (1) solved for WTP, yields the bid-function that forms the (often implicit) basis for any MA-BT exercise. Following Bergstrom and Taylor (2006), we further assume what they call a "weak structural utility theoretic" approach[2], i.e. that the underlying variables in the bid-function is assumed to be derivable from some unknown utility function, but that flexibility is maintained to introduce explanatory variables, such as study characteristics, into the WTP model that do not necessarily follow from (1). This is the approach used in most previous MA-BT exercises (for example Rosenberger and Loomis 2000, Shrestha and Loomis 2003). We specify a meta-model that captures j site characteristics X, k study or methodological characteristics M, l programme characteristics P, and q socio-economic characteristics S. Mean WTP estimate (long term, per household in Norwegian Kroner 2005) m from study s, WTP_{ms}, can then be defined as:

$$\text{WTP}_{ms} = \beta_0 + \beta_X X^j_{ms} + \beta_M M^k_{ms} + \beta_P P^1_{ms} + \beta_S S^q_{ms} + e_{ms} + u_s. \qquad (2)$$

Where, β_0, β are constant term and parameter vectors for the explanatory variables, and e_{ms} and u_s are random error terms for the measurement and study levels, respectively. MA-BT involves estimating (2) based on previous studies, inserting values for X, P and S for the policy site under investigation, and choosing values for M (typically average of the meta-data, "best-practice" values or sample from a distribution — see e.g. Johnston et al. (2006)). The meta-model has several potential advantages for BT, compared to unit value transfer or function transfer based on a single study[3]. MA utilizes information from several studies providing more rigorous measures of central tendency that are sensitive to the underlying distribution of the study values (Rosenberger and Loomis 2000). Further, as specified in the model above, MA can control for study-specific choices of methodology, and finally it is possible to account for differences in site and programme characteristics between the policy site and the study sites in the meta-data, by setting these variable values equal to the policy site.

2.2. *Validity and reliability of BT*

Validity and reliability of BT can be explained using the concept of transfer error (TE), defined as:

$$\text{TE} = \frac{|\text{WTP}_T - \text{WTP}_B|}{\text{WTP}_B}, \qquad (3)$$

where T = transferred (predicted) value from study site(s), B = estimated true value ("benchmark") at policy site. Validity[4] has traditionally required "that the values, or the value functions generated from the study site, be statistically identical to those estimated at the policy site" (Navrud and Ready, 2007b, pp. 7), i.e. that TE is statistically indistinguishable from zero. Most of the studies testing BT validity have used the same valuation methodology for similar goods nationally or internationally often resulting in high TE levels and rejection of the hypothesis of TE = 0 (see Rosenberger and Phipps (2007, Table 1) for an overview of results). For MA-BT, such tests are scarce. Rosenberger and Loomis (2000) use raw

[1] An alternative to the classical MA approach, not considered here, uses Bayesian modelling techniques to address some of these challenges (Moeltner et al., 2007).

[2] Bergstrom and Taylor (2006) categorise three main utility theoretic MA-BT approaches (of which only the first two are recommended): Strong, weak, and non-structural.

[3] The BT function from a single study, for individual i, is often specified as $\text{WTP}_i = a + bX_{ij} + cY_{ik} + e_i$, where X is site/good characteristics (j), Y respondent characteristics (k), e_j random error, and the number of observations is equal to the number of respondents (Brouwer, 2000).

[4] In the BT literature the term "convergent validity" is sometimes used.

Table 1 – Objectives and transfer error calculation procedure for reliability check of meta-analytic benefit transfer

Objective	Transfer error calculation procedure
1. Initial reliability check of MA-BT model	
Analyse within-sample TE	Compare model predictions with WTP estimates and calculate overall mean and median TE for all estimates in the data.
Analyse out-of-sample TE	Compare N model predictions with WTP estimates for N-1 of the data for each prediction, and calculate overall mean and median TE for each model run.
2. Reliability comparison of different BT techniques	
Compare reliability of MA-BT and simple BT techniques	Domestic or international transfers of “single best” or mean estimate from studies of similar site and programme characteristics are compared with MA-BT. Single WTP estimates drawn from each study as benchmark policy site value for TE calculation.
3. Robustness of TE to methodological choices	
Analyse TE across model specifications and restrictions	Two different model specifications (linear, and double-log) and two restricted models are used for TE calculations under 1. The two specifications with the lowest overall TE are used in 2.

estimates from within their sample of recreation studies from the USA as benchmarks, compare them with the transferred estimates from their national and regional MA-BT models and calculate TE. Shrestha and Loomis (2003, 2001) and Shrestha et al. (2007) use values from a number of additional domestic and international studies as benchmarks, respectively (i.e. out-of-sample comparisons), and conduct several tests of validity. More recently BT validity assessment has shifted focus somewhat to the concept of reliability for policy use, which requires that TE is relatively small (but not necessarily zero). This shift comes from the realisation that BT can be considered valid even if the standard hypothesis of TE=0 is rejected — in fact the most appropriate null hypothesis is that TE>0 since environmental benefits from theory should be assumed to vary between contexts (see Kristofersson and Navrud (2005)). However, there is no agreement on maximum TE levels for BT to be reliable for different policy applications, though 20 and 40% have been suggested (Kristofersson and Navrud, 2007).

Our objective is not to judge which levels of TE should be considered acceptable or to conduct traditional statistical tests of BT validity. Instead, we measure reliability in terms of TE and compare TE across meta-model specifications and restrictions, and between alternative ways of conducting BT based on the same data. Our approach is summarised in Table 1, and explained in the following.

A first check for our meta-model specified in (2) is to compare the in-sample model predictions with the WTP estimates and calculate TE for each estimate and the overall mean and median TE for the whole dataset[5] (as suggested by Brander et al. (2006) (Objective 1 in Table 1). Second, we estimate N different MA-BT functions using N-1 of the data for each run, since the WTP estimate we predict is taken out. Then we calculate the overall mean and median TE for all the N models taken together[6]. Third, to more closely resemble an actual BT situation, we draw randomly a single WTP estimate from each study to represent the benchmark, unknown policy site value (Objective 2 in Table 1). The next step is to use the other studies to transfer a best estimate to that policy site based on different BT techniques, including “single best” or mean WTP estimates from domestic or international studies that have similar site and programme characteristics. We compare TE from these techniques with the use of the two most promising MA-BT models, judged on the basis of the initial TE assessment. Finally, we assess the impact of the choice of MA model specifications and restrictions on TE (Objective 3 in Table 1). There are many different types of meta-model specifications in use, and there is little guidance as to which to choose (linear, semi-log, double-log etc) (Johnston et al., 2005). A restricted model frequently used (though rarely convincingly justified) in MA-BT is one where variables that are not significant at the 20% level are left out of the model. To investigate the implications of these choices, we decided to test linear and double-log specifications, and a fully specified and a restricted version of the meta-model.

3. Meta-data sources and regression results

A literature of around 50 studies reporting from more than 25 surveys valuing non-timber benefits has developed in Norway, Sweden and Finland over the last 20 years. The studies typically ask respondents' WTP for either full forest protection plans or for programmes introducing more environmentally and/or recreationally sensitive forestry practices — called multiple use forestry (MUF). The values from these studies can be interpreted as the WTP to obtain a positive change in at least one element in an attribute vector describing the forest environment, Q in the utility function (1), i.e. level of biodiversity, forest density, forest size, scenic beauty etc. A substantial portion of WTP can be assumed to be non-use values. We compiled a meta-dataset consisting of 72 WTP estimates, where 1–7 estimates were gleaned from each study. A more specific description of the base data and the coding process for the MA is given in Lindhjem (2007). All but one study use the CV approach, and the number of studies is about evenly distributed between countries. To make WTP from different countries comparable, estimates from Sweden and Finland were converted to NOK at the year of the CV survey using annual average OECD PPP rates, and then adjusted to 2005 by the use of the Norwegian consumer price index (Ready and Navrud, 2006). For each WTP estimate from a study, we coded explanatory variables according to the meta-model specified in

[5] The TE for each estimate, is often more appropriately (as a transfer as such is not performed) called Mean Absolute Percentage Error, e.g. in Brander et al. (2006).

[6] As pointed out by Brander et al. (2006) this is similar to a jacknife resampling technique.

(2) (see first two columns of Table 2). The year is the only continuous variable, the rest are dummies. We chose long-term average annual WTP per household as the base format, coding other formats using a dummy ("Otherpay") since respondents' discount rates are not known. Preliminary analysis showed that socio-economic variables, S (income, age and

Table 2 – Variable definitions and meta-regression results

		Full models		Restricted models (dbl log)	
Variable	Variable definition	I. Linear	II. Double-log	III. Trimmed (one obs. Excl.)	IV. Restricted (p>0.2 excl.)
Intercept		1549.25*	4.141**	1.943**	1.721***
		(875.53)	(1.170)	(.925)	(.594)
CE	1 if choice experiment, 0 if CV	192.69	.329	.096	
		(378.00)	(.240)	(.154)	
OE	1 if open ended question, 0 if dichotomous choice	−1334.07**	−.495	−.348	
		(594.09)	(.339)	(.340)	
OEPC	1 if OE with payment card, 0 if dichotomous. choice	227.53	−.360	−.279	
		(382.08)	(.220)	(.172)	
Volunpv	1 if voluntary payment vehicle, 0 if otherwise	3799.7***	2.803***	1.716***	1.845***
		(988.76)	(.771)	(.591)	(.381)
Userpv	1 if payment vehicle related to use, 0 if otherw.	−2564.02***	−.330	.296	
		(424.87)	(.428)	(.288)	
Otherpay	1 if WTP stated other than long-term, i.e. lump-sum, monthly, per	183.43	−.066	.419	
	season etc, 0 if otherwise	(620.51)	(.487)	(.435)	
Actualpay	1 if payments were actually made, 0 if otherwise	−571.53*	−2.099***	−1.974***	−1.715***
		(320.30)	(.106)	(.157)	(.367)
Individual	1 if individual WTP, 0 if per household	1834.94***	1.295***	1.581***	1.410***
		(471.80)	(.294)	(.176)	(.169)
Mailhigh	1 if mail survey w.>65% useable responses, 0 if in-person	−6477.97***	−4.986***	−5.232***	−4.10***
	interview	(1032.54)	(.768)	(.753)	(.695)
Mailmed	1 if mail survey w. 50–65% useable responses, 0 if in-person	−4864.70***	−4.270***	−4.919***	−3.735***
	interview	(1043.22)	(.901)	(.858)	(.721)
Maillow	1 if mail survey w.< 50% useable responses, 0 if in-person	−2476.16**	−3.009***	−4.184***	−3.328***
	interview	(970.37)	(.911)	(.765)	(.694)
Unpub	1 if WTP estimate is unpublished, 0 if published	−791.16**	.019	.084	
			(320.26)	(.360)	(.306)
Mscthesis	1 if MSc thesis, 0 if otherwise	−1916.26**	−1.730***	−1.299***	−1.121**
		(754.85)	(.558)	(.399)	(.414)
Forestpract	1 if more sensitive forestry practices, 0 if full protection plan	765.16**	.277	−.154	
		(320.39)	(.316)	(.222)	
Protmix	1 if mix of protection and forestry practices, 0 if full protection	−1261.76	−.668	−.474	
		(808.15)	(.532)	(.433)	
Impl	1 if percentage/hectare forest not given in survey, 0 if otherwise	1276.51	1.279**	1.246***	1.168***
		(934.02)	(.525)	(.404)	(.194)
Localgood	1 if local good, 0 if nationwide	649.12	−.446	−1.327***	−1.088***
		(536.09)	(.490)	(.322)	(.161)
Reggood	1 if regional good, 0 if nationwide	2350.52***	.821*	.538	
		(746.42)	(.471)	(.603)	
Sweden	1 if study conducted in Sweden, 0 if in Norway	1111.56	2.147**	3.889***	3.370***
		(822.46)	(.971)	(.768)	(.567)
Finland	1 if study conducted in Finland, 0 if in Norway	644.23	2.131*	2.254***	1.932***
		(1046.65)	(.601)	(.707)	(.607)
Urban	1 if primarily urban forest, 0 if otherwise	−1551.15***	−.571	.159	
		(552.46)	(.451)	(.306)	
Season	1 if surveyed in autumn/winter, 0 if spring/summer	−1879.21***	−.784**	−.689**	−.447**
		(496.11)	(.313)	(.275)	(.182)
Avoidloss	1 if WTP for avoiding a loss, 0 if WTP for a gain	627.94	.585*	.190	
		(415.24)	(.307)	(.156)	
Use	1 if primarily use, 0 if otherwise	451.94	.0224	−.316	
		(721.97)	(.354)	(.261)	
Year	Year/LnYear, continuous range 1–16 (1985–02)	130.35	1.242**	2.380***	2.246***
		(82.63)	(.555)	(.286)	(.342)
R^2		0.756	0.815	0.886	0.814
N		72	72	71	71

Note: *p<0.10, **p<0.05, ***p<0.01. Number of survey clusters for models = 27. STATA Version 9.1 used. Models I and II are identical with Models II and III in Lindhjem (2007).

education level), did not have a significant effect on WTP, and were therefore excluded from the subsequent analysis. This is a very common result in MA (Rosenberger and Loomis, 2000; Johnston et al., 2003, 2005).

Programme and site characteristics variables (*P* and *X* in Eq. (2)) try to capture the variation in the forest good valued and are of particular relevance for MA-BT. The size of the forest can a priori be assumed to capture an important dimension of the good. In preliminary analyses we used different measures, for example the size of the forest in hectares or as percentage of productive forestland in the country, to capture this scope dimension. This analysis is conceptually difficult for several reasons. Some surveys ask WTP for national changes in forest practices, which basically would involve all forest areas in the country. Further, the dataset included both surveys of local and national plans, with high non-use values at the national level and higher degree of resource conflicts at the local level[7]. We did not find any significant increase in WTP with simple measures of forest size, which in our opinion is *not* evidence against valid stated preference research. The complexity of the good, the high share of non-use values, the relatively small changes proposed, the geographical dimensions, may just mean that the area of the forest is too crude a measure to capture people's sense of scope in a MA[8]. It may also be that two forest plans that only differ marginally in size are seen as no different as long as people know that a minimum level of biodiversity is protected. The existing MA literature, with few exceptions such as Smith and Osborne (1996), can be said to have under appreciated the potential conceptual and practical problems involved in capturing scope sensitivity across very heterogeneous international studies of complex goods such as wetlands — where WTP/hectare often is used uncritically as the dependent variable (Woodward and Wui, 2001; Brander et al., 2006)[9]. Instead, we included other dimensions of the good that may be important to people; geography (local, regional, national, country levels), primarily use, and type of plan (forest protection, MUF or a mix, urban forests). We also included a dummy for the season of the survey, checking if people display "season illusion". In addition, a range of study specific or methodological dummy variables are included (*M* in Eq. (2)), such as survey mode, response rates for postal surveys, type of WTP question, type of publication etc. The full variable list is defined in Table 2, and expectations regarding their relationship with WTP are discussed in detail in Lindhjem (2007).

To estimate the meta-model in (2) we used a Huber–White robust variance estimation procedure to adjust for potential heteroskedasticity and intrastudy correlation, a common approach in the MA literature (Smith and Osborne, 1996). We chose four different models. The first two are linear and double-log specifications, while the third model is restricted in that one observation, which gave very high TE, was left out[10]. The fourth model is a version of the third where we following Rosenberger and Loomis (2000), retain only those variables that are significant at an 80 per cent level or better based on t-statistics. The regression results displayed in the third to sixth columns of Table 2 show that the models fit the data well, with adjusted R^2 between 75 and 88%. The models confirm several of our expectations about the methodological variables, for example related to open ended question formats, response rates of mail surveys, voluntary payment vehicles, actual payment etc (see Lindhjem (2007) for discussion). Methodological variables show a higher degree of significance than site and programme variables for explaining the variation in the data. This is a potential problem when using the meta-regressions for BT, and is common in the literature. Regional forests are valued higher than national (the base case) (though not statistically significant in model III), while local forests have lower WTP (except for the linear model). The resource use conflicts at local levels may explain the latter difference. Further, Sweden and Finland have significantly higher WTP than Norway for the last three models, suggesting that even if economic, cultural and institutional conditions are similar, WTP can still be different. Urban forests are valued lower than other forests, which may indicate that non-use values of non-urban forests are important. WTP to avoid a loss is higher (though not significantly so) than WTP for a gain. WTP of users or related primarily to use is not statistically different than from a mixed group. Regarding type of programme, our results are somewhat puzzling. It seems that respondents value full protection lower than MUF, but higher than a mix between the two (though not significant through the four models). It is worth noting that in model IV, the only site/programme description variables left are the local and country dimensions. Further, it also seems to be important to the stated WTP whether forest area and percentage have been explicitly mentioned in the survey. These results are of an exploratory kind, but show at least that it is not immaterial to people whether it is question of full protection or just a change in existing forestry practices. Finally, the models show that the season variable is negative and highly significant, while the year of the survey influences WTP positively. The discussion of meta-regression results is not elaborated further here since our intention is to use the estimated equations for BT analysis (see Lindhjem (2007) for details).

4. Transfer error results and comparison of benefit transfer techniques

4.1. *Within and out-of-sample overall mean TE*

The results from the initial check of the overall mean and median TE of the four MA-BT models, within and out of sample, i.e. Objective 1 in Table 1, are given in Table 3 below.

[7] Although there may also be a distance decay effect, i.e. that people value forests closer to where they live, higher.

[8] Some of the studies also had unclear scenario descriptions making it harder for people to judge differences between plans.

[9] Recent CV studies have moved beyond the relatively simplistic ("bird count") scope debate following the Exxon Valdez disaster in the early 1990s (see e.g. Bateman et al. (2004) and Heberlein et al. (2005)).

[10] In preliminary analysis we also ran several alternative models, e.g. following Shrestha and Loomis (2001), testing a trimming procedure of the data, leaving out WTP estimates larger or smaller than two standard deviations from the mean. This procedure did only marginally reduce the TE.

The first point to note is the relatively low overall *median* TE for all models, varying from 25–51%. Further, it is expected that TE will go up, more the smaller the dataset, when the observation we predict is left out. When considering means, the linear model I performs much worse with a TE of between 135 and 266 than the double-log models II–IV. The high TE of model II is considerably reduced when leaving out an extreme observation in model III and retained at around the same level when further restricted in that variables with $p>0.2$ are taken out in model IV. This increase in precision from model II to IV can be seen in the plots in Figs. 1 and 2, where estimates are sorted in ascending order. The figures also show that TE is higher for lower values of WTP, a similar result to Brander et al. (2006)[11].

Table 3 – Transfer errors for within-sample and out-of-sample runs of meta-models

	TE for different model specifications			
	Model I: linear	Model II[a]: Dbl log	Model III[a]:1 obs. excl.	Model IV[a]: $p>0.2$ excl.
Within-sample				
Overall mean TE	135	52	39	52
Overall median TE	37	26	25	30
0–25th percentile (1–18)[b]	390	71	77	76
25–50th percentile (19–36)	105	92	57	72
50–75th percentile (37–54)	24	17	25	24
75–100th percentile (55–71/2)	24	26	26	37
Out-of-sample				
Overall mean TE	266	222	62	63
Overall median TE	51	40	34	31
0–25th percentile (1–18)[b]	770	202	110	109
25–50th percentile (19–36)	213	592	70	53
50–75th percentile (37–54)	38	27	20	35
75–100th percentile (55–71/2)	42	67	50	54

[a] To account for econometric error in transforming ln(WTP) to WTP using antilog, we add standard deviation ($s^2/2$), which estimate varies when the sample changes, prior to transformation of ln (WTP) (Johnston et al., 2006). An alternative, or supplement, for brevity not considered here would be to replace s^2 with the variance of the prediction (Goldberger, 1968).

[b] Percentiles calculate the mean TE in four different segments of the data, when WTP estimates are sorted in ascending order (from estimate 1 to 72).

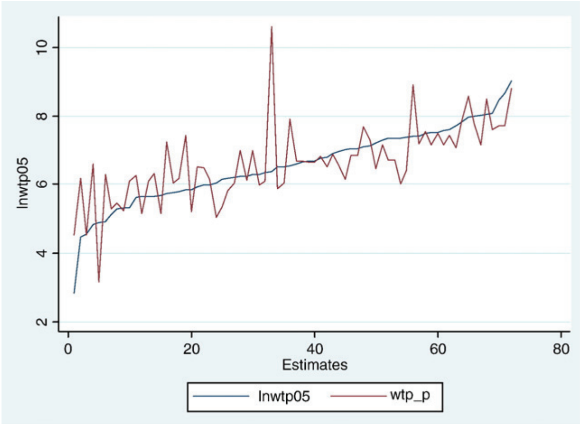


Fig. 1 – Plot of log WTP (lnwtp05) estimates and predicted values (wtp_p) for Model II (out-of-sample), sorted in ascending order.

Calculating average TE for different percentiles of the data, as shown in Table 3, when WTP is sorted in ascending order, also clearly shows the TE going down for higher WTP (though TE goes up again for the highest percentile). The predictions also seem to overshoot more often for lower WTP than for the ones above the median, which is an important consideration in making MA-BT conservative. The interpretation of TE for different levels of WTP is important also in terms of calculating a total welfare measure, i.e. summing WTP over the relevant population. In cost–benefit analysis it is the TE of the total welfare estimate that is important. If WTP per household from a local survey of a local protection plan is lower than a nationwide survey of a national plan (which is the case for models II–IV), the overall TE for the welfare measures of both plans may “even out” in the aggregate.

The overall mean TE of around 60% we find for out-of-sample models III and IV is somewhat lower than in the only two studies we have seen conducting this exercise (Brander et al., 2006, 2007). Their meta-analyses have 72 and 201 estimates, are based on more heterogeneous data, and use regression models with lower explanatory power. In their validity tests of MA-BT, Shrestha and Loomis (2001, 2003) find average TE ranging from a low 28% to 88%, respectively. The within-sample test results of Rosenberger and Loomis (2000) show mean TE ranging from 54 to 71% depending on whether a national or a region/activity specific model is used.

4.2. *TE for different BT techniques*

We compare the two models with the lowest overall mean TE above (i.e. models III and IV) with simple BT techniques using a more realistic simulation of actual BT. If we were faced with a policy site without sufficient time and resources to do a primary study, we could use a study from the most similar domestic or

[11] This is partly a result of the definition of TE, as the same absolute prediction error is higher in relative terms for low WTP values than for high.

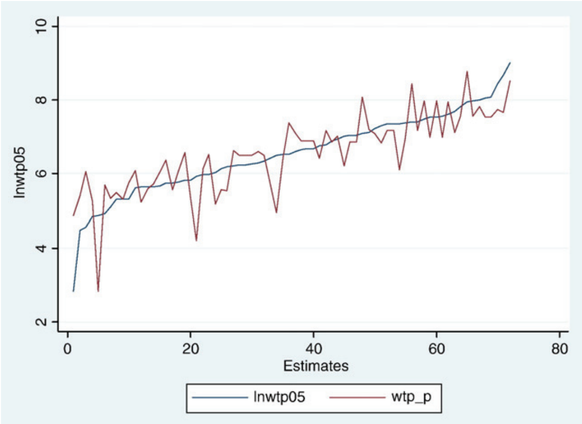


Fig. 2 – Plot of log WTP (lnwtp05) estimates and predicted values (wtp_p) for Model IV (out-of-sample), sorted in ascending order.

international site, use a mean from studies of similar domestic or international sites, or conduct an international MA-BT[12]. We compare these techniques in the following way. First we randomly draw one estimate from each of the 26 CV surveys included in the data, to represent a benchmark value for a policy site. All other WTP estimates from this survey are then excluded when the remaining data is used for BT. We then calculate TE for each site, and calculate the overall mean and median TE for each BT technique[13]. The benchmark value has of course its own error in measurement and is influenced by the survey methodology chosen. Nevertheless, a comparison of BT techniques for all sites represented by the data gives a valuable indication of the reliability and level of error that can be expected. Table 4 displays results.

The second column is the unknown benchmark value for a site to be predicted. This value can be seen as an estimate of long-term household WTP for a forest protection or MUF plan at a policy site, defined by certain site and programme characteristics[14]. Column three displays the raw mean WTP, regardless of site characteristics, for all 71 estimates in the data (except the benchmark estimate), representing an upper TE ceiling ("the worst you could do"). Column six displays the mean WTP for domestic surveys in the data that have the same site characteristics. The variables defining MUF, forest protection or a mix of the two, and local or national forests were used to assemble relevant value estimates.[15] Column seven is the mean WTP when estimates with the same site characteristics from the other two countries were also included. For both these mean value transfers study characteristics are ignored. In contrast with the raw mean WTP in column three, we picked the two estimates closest to the policy site benchmark from the set of domestic or international studies that have the same site characteristics (see columns four and five). This would not be possible in practice, but is a useful indication of the lower bound TE from choosing estimates from single, similar site studies ("the best you could do")[16]. Finally, the last two columns give the results from the use of the MA-BT models III and IV. Instead of setting the methodological dummy variables at average values, at 0.5 or at some best practice value as would have to be done in practical MA-BT (for example as investigated by Johnston et al., 2006), we set the values of these dummies to the same as for the benchmark estimate. This represents the lower bound TE for the MA-BT models. It would be unnecessary to introduce in our comparison the additional TE implied by the choice of methodological dummy values if the MA-BT models in our "best case" perform only marginally better than the simple BT techniques.

The last four rows in Table 4 sum up the mean and median TE for all BT techniques. First we ignore that some studies with matching policy site characteristics are not available (marked "na" in Table 4). Using the simplest of all techniques, just transferring raw mean WTP from the dataset of forest valuation studies would yield a mean TE of 217%. If it were possible to choose the closest value estimate with similar site characteristics, mean TE would be 62% if chosen from domestic studies and 71% if the set were expanded to include international studies. Taking means from domestic and the whole set of studies with similar site characteristics yields mean TE of 86% and 166%, respectively. Thus, expanding the dataset to include international studies in this case increases the TE substantially — close to "max" TE of 216%. In comparison, the two MA-BT models yield mean TE of 126 and 47%, a range that includes the TE from using mean from domestic studies. One reason why the MA-BT model IV gives a lower TE than model III is that simplified models often tend to give better predictions compared to fully specified models. Our BT testing procedure yields a lower number of observations for each model run, hence reinforcing this feature compared to the within and out-of-sample tests in the previous section. From comparing mean TE for all 26 sites, international MA-BT does not perform better on average than transferring mean WTP from domestic studies, though the best meta-model

[12] Most countries will not have enough domestic studies to conduct an MA, and would have to base their MA-BT on a mix of domestic and international studies, like in the present study.

[13] We realise that a fuller test could include a bootstrap to calculate TE for many random draws of single policy site benchmarks, and not just one draw.

[14] We do not distinguish between different formats of WTP in terms of long-term vs lump-sum and individual vs household etc, but assume that the value at the site and the simple transfer estimates roughly represent long-term household WTP (and as the regression results show, many of these dummies were also insignificant in the analysis).

[15] Using the whole set of site characteristics, i.e. also urban, regional and primarily use value etc have the disadvantage that there often are no estimates in the data with exact matching characteristics. A subset was therefore chosen.

[16] We first tried to use an objective rule to choose an estimate that would most closely resemble the policy site to mimic situation of simple BT. However, this is not straightforward as the set of studies with the full range of site and programme characteristics matching the policy site is often empty. In this case, secondary rules using a subset of the site characteristics need to be applied to end up with a unique, best estimate.

Table 4 – Comparison of transfer error between benefit transfer techniques

Main reference	Site benchmark value	Raw mean WTP all studies (-1)	Best similar domestic (*D*) or internat. (*I*) study		Mean of similar domestic (*D*) or international (*I*) studies				MA-BT models	
			D	*D*+*I*[a]	*D*		*D*+*I*		III	IV
					n	Mean	*n*	Mean		
Simensen and Wind (1990)	286	1225 (328%)	289 (1%)	289 (1%)	4	300 (5%)	14	756 (164%)	113 (60%)	272 (4%)
Hoen and Winther (1993)	340	1277 (275%)	na	1847 (443%)	0	na	6	3954 (1063%)	2367 (596%)	641 (88%)
Veisten et al. (2004a,b)	1355	1193 (11%)	na	1638 (20%)	0	na	1	1638 (20%)	572 (57%)	1256 (7%)
Sandsbråten (1997)	277	1218 (339%)	286 (3%)	286 (3%)	4	351 (27%)	14	771 (178%)	1175 (323%)	416 (49%)
Leidal (1996)	3248	1109 (65%)	1567 (51%)	1567 (51%)	3	1047 (67%)	10	519 (84%)	1985 (38%)	2258 (30%)
Skagestad (1996)	96	1207 (1157%)	na	na	0	na	0	na	278 (188%)	353 (266%)
Veisten and Navrud (2006)	204	1247 (511%)	201 (1%)	201 (1%)	6	131 (13%)	27	1100 (439%)	128 (37%)	282 (38%)
Hoen and Veisten (1994)	324	1204 (271%)	311 (4%)	311 (4%)	5	329 (1%)	15	736 (127%)	239 (26%)	456 (40%)
Hoen and Veisten (1994)	311	1204 (287%)	324 (4%)	324 (4%)	5	332 (6%)	15	736 (136%)	755 (142%)	547 (75%)
Strand and Wahl (1997)	1567	1187 (24%)	2930 (86%)	2930 (86%)	4	2438 (55%)	11	1072 (31%)	479 (69%)	660 (57%)
Kniivilä (2004)	393	1210 (208%)	342 (12%)	342 (12%)	5	256 (34%)	12	1173 (198%)	486 (23%)	422 (7%)
Lehtonen et al. (2003)	1534	1159 (24%)	1464 (4%)	1464 (4%)	12	791 (48%)	26	868 (43%)	1372 (10%)	1360 (11%)
Pouta et al. (2000, 2002)	1137	1192 (4%)	1226 (7%)	1226 (7%)	13	992 (12%)	27	962 (15%)	1433 (25%)	976 (14%)
Pouta (2003, 2004, 2005)	1847	1173 (36%)	na	2838 (53%)	0	na	5	4058 (119%)	873 (52%)	1153 (37%)
Rekola and Pouta (2005)	126	1207 (857%)	734 (482%)	734 (482%)	6	957 (659%)	15	749 (494%)	227 (79%)	173 (36%)
Siikamäki and Layton (2007)	531	1216 (129%)	512 (3%)	512 (3%)	14	1124 (111%)	28	1029 (93%)	1391 (161%)	652 (22%)
Mäntymaa et al. (2002)	569	1230 (116%)	531 (6%)	531 (6%)	13	1191 (109%)	27	1058 (85%)	307 (46%)	651 (14%)
Tyrväinen and Väänänen (1998)	796	1237 (55%)	734 (7%)	734 (7%)	4	875 (4%)	13	691 (13%)	2462 (208%)	959 (20%)
Tyrväinen (2001)	284	1238 (335%)	342 (20%)	342 (20%)	3	312 (9%)	10	1374 (383%)	525 (84%)	277 (2%)
Bojö (1985)	372	1203 (223%)	na	393 (5%)	0	na	12	1175 (216%)	252 (32%)	218 (41%)
Bostedt and Mattson (1991)	2478	1173 (52%)	393 (5%)	540 (78%)	2	519 (79%)	15	592 (76%)	3755 (51%)	2138 (13%)
Mattsson and Li (1993)	5843	1099 (81%)	8251 (41%)	8251 (41%)	2	5544 (5%)	5	3040 (47%)	1744 (70%)	2681 (54%)
Mattsson and Li (1994)	2838	1067 (62%)	3020 (6%)	3020 (6%)	2	4432 (56%)	5	2595 (8%)	10487 (269%)	5560 (95%)
Kriström (1990a, b)	1853	1110 (40%)	590 (68%)	590 (68%)	1	590 (68%)	27	756 (59%)	5291 (185%)	2645 (42%)
Johansson (1989)	1638	1185 (27%)	na	1355 (17%)	0	na	3	1151 (29%)	3636 (121%)	3660 (123%)
Bostedt and Mattsson (1995)	540	1211 (124%)	2478 (358%)	2478 (358%)	1	2478 (358%)	14	737 (36%)	2236 (313%)	688 (27%)
Mean TE		217%	62%	71%		86%		166%	126%	47%
Median		120%	7%	12%		41%		85%	70%	37%
Mean TE[b] (same obs.)		196%	62%	62%		86%		136%	111%	33%
Median[b] (same obs.)		120%	7%	7%		41%		85%	70%	29%

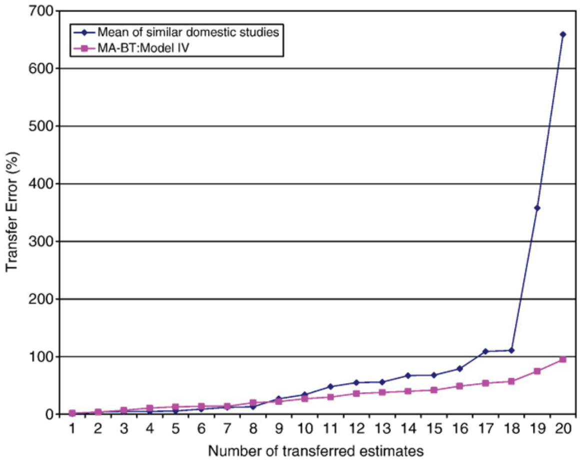


Fig. 3 – Transfer errors for MA-BT model 4 and mean of similar domestic studies arranged in ascending order of TE for each BT technique, respectively.

has lower TE. Considering the medians this conclusion is strengthened. It is clear from the data that the TE from using the simple BT techniques is pushed up by a number of high values compared with MA-BT. Medians of the best simple BT technique and MA-BT models are 41% and 37%, respectively. Comparing TE from all 26 sites is not entirely satisfactory as there are missing values for some of the simple techniques, while the MA-BT predicts values for all sites. Limiting the set for comparison to the sites where transferred estimates are available across all BT techniques does not change the general picture, though MA-BT comes out a little more favourably in this case (last two rows of Table 4). Plotting for the limited set the transferred estimates in ascending order of TE for the MA-BT model IV and the use of domestic mean, respectively, is instructive (see Fig. 3).

Fig. 3 clearly shows that the better performance of MA-BT model IV over using domestic means overall, is largely due to a few very high TE values for the latter. 50% of the domestic mean transfers and 70% of the MA transfers have TE below 40%, while 40% of the transfers for both techniques have TE below 20%. Excluding the two extreme transferred values from both sets brings the mean and median TE for both techniques down to around 35%. We also compared whether BT would work better to certain countries and it seems that there is no consistent pattern. Due to the already limited dataset it was not possible to investigate whether a subset of the data matching the policy context better, would improve the reliability of the MA-BT models. Santos (2007) investigates a subset of his meta-data and finds no improvement in MA-BT performance, though this result may not extend to our case. Another potentially relevant factor for our comparison that we were unable to investigate due to limited reporting in source studies is the different level of uncertainty in WTP estimates. A richer BT test could use confidence intervals for the benchmark at the policy site, as done by Santos (2007).

5. Concluding remarks

This paper has investigated the reliability of international meta-analytic benefit transfer (MA-BT) based on a data set of stated preference surveys of forest protection and multiple use forestry plans from Norway, Sweden and Finland. The studies included in the meta-analysis are relatively homogenous in terms of valuation methodology and all three countries have similar cultural, institutional and economic conditions. We measure reliability in terms of transfer error (TE) and compare TE across meta-model specifications and restrictions, and between alternative ways of conducting BT based on the same data. The initial check of within and out-of sample predictions of four meta-models shows substantial variation in performance. The best models give median and mean TE of between 25–34% and 39–62%, respectively. The TE is lower for higher WTP estimates. In the comparison of transfer techniques, MA-BT shows mean TE of between 47–126% (median 37–70%) depending on the model. A simple transfer based on the mean of domestic studies with similar site characteristics to the policy site yields a mean TE of 86% (median 41%), as compared with 62% (median 7%) if a best study estimate could be chosen from a domestic study. Including also international studies in the simple mean transfer increases the TE substantially to 166% (median 85%). The best simple BT technique yields TE in the middle of the range of the two international MA-BT models. It is worth emphasising that in practical BT applications, the TE for the MA-BT models would increase since values of methodological variables would have to be set. Finally, meta-model specifications and restrictions have substantial impact on TE. Our results suggest that MA-BT may not always yield reliability gains over simple BT techniques, as often claimed in the MA literature. However, more MA-BT tests should be performed for other environmental goods and other countries before discarding meta-analyses as a tool for international benefit transfer.

Acknowledgements

We would like to thank Olvar Bergland, Norwegian University of Life Sciences, Shelby Gerking, University of Central

Notes to Table 4

Norwegian Kroner 2005 (Transfer Error).
[a]If no domestic study with the right site characteristics was available, international studies were included.
[b]For a few of the benchmark values to be predicted there were no estimates in the data with matching site characteristics, indicated by “na”. Therefore, mean and median TE were also calculated only for those sites where WTP estimates were available across all BT techniques.

Florida, and John A. List, University of Chicago, for constructive comments.

REFERENCES

Bateman, I., Cole, M., Cooper, P., Georgiou, S., Hadley, D., Poe, G.L., 2004. On visible choice sets and scope sensitivity. Journal of Environmental Economics and Management 47, 71–93.

Bergstrom, J.C., Taylor, L.O., 2006. Using meta-analysis for benefits transfer: theory and practice. Ecological Economics 60, 351–360.

Bojö, J., 1985. Cost–benefit Analysis of Mountainous Forests: The Vala Valley Case (in Swedish). Research Report. Stockholm School of Economics.

Bostedt, G., Mattson, L., 1991. The Importance of Forests for Tourism: A Pilot Cost–benefit Analysis (In Swedish). Arbetsrapport 141, Swedish University of Agricultural Sciences, Umeå.

Bostedt, G., Mattsson, L., 1995. The value of forests for tourism in Sweden. Annals of Tourism Research 22 (3), 671–680.

Brander, L.M., Florax, R.J.G.M., Verrmaat, J.E., 2006. The empirics of wetland valuation: a comprehensive summary and a meta-analysis of the literature. Environmental and Resource economics 33, 223–250.

Brander, L.M., van Beukering, P., Cesar, H., 2007. The recreational value of coral reefs: a meta-analysis. Ecological Economics 63 (1), 209–218.

Brouwer, R., 2000. Environmental value transfer: state of the art and future prospects. Ecological Economics 32 (1), 137–152.

Ecological Economics, 2006. Environmental benefits transfer: methods, applications and new directions. Special Issue 60 (2), 335–482.

Goldberger, A.S., 1968. The interpretation and estimation of Cobb–Douglas functions. Econometrica 36, 464–472.

Heberlein, T.A., Wilson, M.A., Bishop, R.C., Schaeffer, N.C., 2005. Rethinking the scope test as a criterion for validity in contingent valuation. Journal of Environmental Economics and Management 50, 1–22.

Hoehn, J.P., 2006. Methods to address selection effects in the meta regression and transfer of ecosystem values. Ecological Economics 60 (2), 389–398.

Hoen, H.F., Veisten, K., 1994. A survey of the users of Oslomarka (In Norwegian). Skogforsk 6/94.

Hoen, H.F., Winther, G., 1993. Multiple-use forestry and preservation of coniferous forests in Norway: a study of attitudes and Willingness-to-pay. Scandinavian Journal of Forest Research 8 (2), 266–280.

Johansson, P.O., 1989. Valuing Public Goods in a Risky World: An Experiment, in H. Folmer and E. C. van Ierland, Eds, Evaluation methods and policy making in environmental economics. Amsterdam, North Holland. 39– 48.

Johnston, R.J., Besedin, E.Y., Iovanna, R., Miller, C.J., Wardwell, R.F., Ranson, M.H., 2005. Systematic variation in willingness to pay for aquatic resource improvements and implications for benefit transfer: a meta-analysis. Canadian Journal of Agricultural Economics 53 (2–3), 221–248.

Johnston, R.J., Besedin, E.Y., Ranson, M.H., 2006. Characterizing the effects of valuation methodology in function-based benefits transfer. Ecological Economics 60 (2), 407– 419.

Johnston, R.J., Besedin, E.Y., Wardwell, R.F., 2003. Modeling relationships between use and nonuse values for surface water quality: a meta-analysis. Water Resoures Research 39 (12), 1363–1372.

Kniivilä, M., 2004. Contingent valuation and cost–benefit analysis of nature conservation: a case study in North Karelia, Finland. PhD thesis, University of Joensuu.

Kristofersson, D., Navrud, S., 2005. Validity tests of benefit transfer — are we performing the wrong tests? Environmental and Resource Economics 30, 279–286.

Kristofersson, D., Navrud, S., 2007. Can use and non-use values be transferred across countries? In: Navrud, S., Ready, R. (Eds.), Environmental Value Transfer: Issues and Methods. Springer.

Kriström, B., 1990a. A nonparametric approach to the estimation of welfare measures in discrete response valuation studies. Land Economics 66 (2), 135–139.

Kriström, B.,1990b. Valuing Environmental Benefits Using the Contingent Valuation Method — An Econometric Analysis. PhD thesis, Swedish University of Agricultural Sciences, Umeå.

Lehtonen, E., Kuuluvainen, J., Pouta, E., Rekola, M., Li, C.Z., 2003. Non-market benefits of forest conservation in southern Finland. Environmental Science and Policy 6 (3), 195–204.

Leidal, K., 1996. Valuation of an urban recreation area (In Norwegian). Master Thesis, Norwegian University of Life Sciences.

Lindhjem, H., 2007. 20 Years of stated preference valuation of non-timber benefits from Fennoscandian forests: a meta-analysis. Journal of Forest Economics 12, 251–277.

Mäntymaa, E., Mönkkönen, M., Siikamäki, J., Svento, R., 2002. Estimating the demand for biodiversity — vagueness band and open-ended questions. In: van Ierland, E.C.M. Weikard, H.P., Wesseler, J. (Eds.), Proceedings: Risk and Uncertainty in Environmental and Resource Economics. Wageningen University: Environmental Economics and Natural Resources Group.

Mattsson, L., Li, C.Z., 1993. The non-timber value of Northern Swedish Forests — an economic analysis. Scandinavian Journal of Forest Research 8 (3), 426 – 434.

Mattsson, L., Li, C.Z., 1994. How do different forest management-practices affect the non-timber value of forests — an economic analysis. Journal of Environmental Management 41 (1), 79–88.

Moeltner, K., Boyle, K., Paterson, R.W., 2007. Meta-analysis and benefit transfer for resource valuation — addressing classical challenges with Bayesian modeling. Journal of Environmental Economics and Management 53, 250–269.

Navrud, S., Ready, R., 2007a. Lessons learned for environmental value transfer. In: Navrud, S., Ready, R. (Eds.), Environmental Value Transfer: Issues and Methods. Springer.

Navrud, S., Ready, R., 2007b. Review of methods for value transfer. In: Navrud, S., Ready, R. (Eds.), Environmental Value Transfer: Issues and Methods. Springer.

Pouta, E.,2003. Attitude–behavior framework in contingent valuation of forest conservation. PhD Thesis, University of Helsinki.

Pouta, E., 2004. Attitude and belief questions as a source of context effect in a contingent valuation survey. Journal of Economic Psychology 25, 229–242.

Pouta, E., 2005. Sensitivity to scope of environmental regulation in contingent valuation of forest cutting practices in Finland. Forest Policy and Economics 7, 539–550.

Pouta, E., Rekola, M., Kuuluvainen, J., Li, C.Z., Tahvonen, I., 2002. Willingness to pay in different policy-planning methods: insights into respondents' decision-making processes. Ecological Economics 40 (2), 295–311.

Pouta, E., Rekola, M., Kuuluvainen, J., Tahvonen, O., Li, C.Z., 2000. Contingent valuation of the Natura 2000 nature conservation programme in Finland. Forestry 73 (2), 119–128.

Ready, R., Navrud, S., 2006. International benefit transfer: methods and validity tests. Ecological Economics 60, 429–434.

Rekola, M., Pouta, E., 2005. Public preferences for uncertain regeneration cuttings: a contingent valuation experiment involving Finnish private forests. Forest Policy and Economics 7, 635–649.

Rosenberger, R., Phipps, T.T., 2007. Correspondence and convergence in benefit transfer accuracy: meta-analytic review of the

literature. In: Navrud, S., Ready, R. (Eds.), Environmental Value Transfer: Issues and Methods. Springer.
Rosenberger, R.S., Loomis, J.B., 2000. Using meta-analysis for benefit transfer: in-sample convergent validity tests of an outdoor recreation database. Water Resources Research 36 (4), 1097–1107.
Sandsbråten, L., 1997. Valuation of environmental goods in Oslomarka (In Norwegian). MSc Thesis, Norwegian University of Life Sciences.
Santos, J.M.L., 1998. The Economic Valuation of Landscape Change. Theory and Policies for Land Use and Conservation. Edward Elgar, Cheltenham.
Santos, J.M.L., 2007. Transferring landscape values: how and how accurately? In: Navrud, S., Ready, R. (Eds.), Environmental Value Transfer: Issues and Methods. Springer.
Shrestha, R.K., Loomis, J.B., 2001. Testing a meta-analysis model for benefit transfer in international outdoor recreation. Ecological Economics 39 (1), 67–83.
Shrestha, R.K., Loomis, J.B., 2003. Meta-analytic benefit transfer of outdoor recreation economic values: testing out-of-sample convergent validity. Environmental and Resource Economics 25, 79 –100.
Shrestha, R.K., Rosenberger, R., Loomis, J., 2007. Benefit transfer using meta-analysis in recreation economic valuation. In: Navrud, S., Ready, R. (Eds.), Environmental Value Transfer: Issues and Methods. Springer.
Siikamäki, J., Layton, D., 2007. Discrete choice survey experiments: a comparison using flexible methods. Journal of Environmental Economics and Management 53, 127–139.
Simensen, K., Wind, M.,1990. Attitudes and WTP for different forestry practices in mountainous forests. (In Norwegian). MSc Thesis, Norwegian University of Life Sciences.
Skagestad, E.,1996. Recreation and Forestry — A survey of hikers in the outer Oslomarka, Romeriksåsen, in the winter time (In Norwegian). MSc Thesis, Norwegian University of Life Sciences.
Smith, V.K., Kaoru, Y., 1990. Signals or noise — explaining the variation in recreation benefit estimates. American Journal of Agricultural Economics 72 (2), 419 – 433.
Smith, V.K., Osborne, L.L., 1996. Do contingent valuation estimates pass a “scope” test? A meta-analysis. Journal of Environmental Economics and Management 31 (3), 287–301.
Smith, V.K., Pattanayak, S.K., 2002. Is meta-analysis a Noah’s Ark for non-market valuation? Environmental and Resource Economics 22, 271–296.
Strand, J., Wahl, T.S., 1997. Valuation of municipality recreation areas in Oslo: A contingent valuation study (In Norwegian). SNF Report 82/97.
Tyrväinen, L., 2001. Economic valuation of urban forest benefits in Finland. Journal of Environmental Management 62 (1), 75–92.
Tyrväinen, L., Väänänen, H., 1998. The economic value of urban forest amenities: an application of the contingent valuation method. Landscape and Urban Planning 43 (1–3), 105–118.
Van Houtven, G., Powers, J., Pattanayak, S.K., 2007. Valuing water quality improvements in the United States using meta-analysis: Is the glass half-full or half-empty for national policy analysis? Resource and Energy Economics 29 (3), 206–228.
Veisten, K., Hoen, H.F., Navrud, S., Strand, J., 2004a. Scope insensitivity in contingent valuation of complex environmental amenities. Journal of Environmental Management 73 (4), 317–331.
Veisten, K., Hoen, H.F., Strand, J., 2004b. Sequencing and the adding-up property in contingent valuation of endangered species: are contingent non-use values economic values? Environmental and Resource Economics 29 (4), 419 – 433.
Veisten, K., Navrud, S., 2006. Contingent valuation and actual payment for voluntarily provided passive-use values: assessing the effect of an induced truth-telling mechanism and elicitation formats. Applied Economics 38 (7), 735–756.
Woodward, R.T., Wui, Y.S., 2001. The economic value of wetland services: a meta-analysis. Ecological Economics 37, 257–270.

Paper III

Meta-analysis of nature conservation values in Asia and Oceania:

Data heterogeneity and benefit transfer issues

Tran Huu Tuan[a] and Henrik Lindhjem[bc*]

[a] Hue College of Economics, Hue University, 100 Phung Hung Street, Hue City, Vietnam.

[b] Department of Economics and Resource Management, Norwegian University of Life Sciences, P.O. Box 5003, N-1432 Ås, Norway.

[c] Econ Pöyry, P.O. Box 5, N-0051, Oslo, Norway.

This paper has been submitted to:
Environmental and Resource Economics

* Corresponding author. E-mail: henrik.lindhjem@poyry.com.

Meta-analysis of nature conservation values in Asia and Oceania: Data heterogeneity and benefit transfer issues

Abstract

We conduct a meta-analysis (MA) of around 100 studies valuing nature conservation in Asia and Oceania. Dividing our dataset into two levels of heterogeneity in terms of good characteristics (endangered species vs. nature conservation more generally) and valuation methods, we show that the degree of regularity and conformity with theory and empirical expectations is higher for the more homogenous dataset of contingent valuation of endangered species. For example, we find that willingness to pay (WTP) for preservation of mammals tends to be higher than other species and that WTP for species preservation increases with income. Subjecting our best MA models to a simple benefit transfer test forecasting values for out-of-sample observations, shows median (mean) transfer errors of 23 (45) percent for endangered species and 46 (89) percent for nature conservation more generally, approaching levels that may be acceptable in benefit transfer for policy use. However, as more heterogeneous observations are included, our meta-regression models are unable to control for the variation in a satisfactory way, resulting in high mean transfer errors. Despite some encouraging results, more research is clearly required to answer the question of how homogenous is homogenous enough in meta-analysis and benefit transfer.

Keywords: Asia; benefit transfer; biodiversity; meta-analysis; Oceania; valuation

JEL Classification: Q26; Q51; Q57; H41

Introduction

According to the Millennium Ecosystem Assessment, more than 60 per cent of the world's ecosystems are being degraded or used unsustainably (MEC 2005). The pressure on nature is among the highest in the many rapidly growing economies of Asia and Oceania. The (neoclassical) economist's prescription to stemming this trend is to value changes in the provision of environmental goods in monetary terms, and create mechanisms to internalise their values in the billions of everyday decisions of consumers, producers and government officials. In response to this challenge, an enormous amount of primary valuation research has been produced using stated and revealed preference methods. However, paraphrasing Glass et al (1981: p11)[1], results of much of this work "are strewn among the scree of a hundred journals and lies in the unsightly rubble of a million dissertations." This valuation research could be much better utilised to demonstrate the social return to nature conservation, a key area where environmental economists need to do more in the future, as pointed out by the late David Pearce (2005). For a range of environmental goods meta-analysis (MA) techniques have been used to synthesize valuation research, test hypotheses, and facilitate the transfer of existing welfare estimates to new, unstudied policy sites ("benefit transfer" – BT) for use e.g. in cost-benefit analysis (Smith and Pattanayak 2002). Responding to Pearce's challenge, we use MA to review and take stock of the literature to date on environmental valuation of a complex and somewhat heterogeneous good: (changes in) conservation of habitat, biodiversity and endangered species, in a specific geographical region: Asia and Oceania. We attempt to answer the following two research questions; (1) To what extent do welfare estimates for this complex good

[1] Originally quoted in Stanley and Jarrel (2005).

conform with theoretically and empirically derived expectations regarding the good characteristics, valuation methods, study quality, socio-economic and other variables?; (2) How sensitive are the meta-regression results and the value forecasts for unstudied sites to; (a) the "scope of the MA", i.e. the level of heterogeneity of the good valued and the valuation methods used; and (b) the choice of meta-regression models?

The first question investigates whether the welfare estimates display the degree of validity and regularity more typically found for less complex environmental goods with higher share of use-values, and offers a first check of the potential for using such data for BT applications (Johnston et al 2005; Lindhjem 2007). The second question contributes to our understanding and refinement of MA methodology in environmental economics, where the meta-analyst typically is left to make a number of choices, potentially introducing various subjective biases (Hoehn 2006; Rosenberger and Johnston 2007). An important analyst choice both for the robustness of MA models and their suitability for use in BT applications, relates to the scope of the MA, i.e. the trade-off between the number of observations and the acceptable level of heterogeneity in the data, as pointed out by e.g. Engel (2002) and Nelson and Kennedy (2009) (Question 2a above). Another, related choice is which model to choose for BT, for example which covariates to include and how to treat insignificant variables (Question 2b)[2]. There are different practices and little is known of the empirical effects of these choices, though Lindhjem and Navrud (2008a) have shown that the precision of meta-analytical BT (MA-BT) depends on the model specifications, sometimes in unexpected ways.

---

[2] An alternative approach to dealing with classical MA challenges, not pursued here, is to use Bayesian techniques (e.g. Moeltner et al 2007 and Moeltner and Rosenberger 2008).

Previous MA studies have primarily analysed the values of more homogenous types of environmental goods (e.g. water and air quality, recreation days) often within the same country (Desvousges et al 1998; Rosenberger and Loomis 2000a; Van Houtven et al 2007). However, there is a trend towards using MA to study more complex goods in international settings (e.g. wetlands, coral reefs, forests, biodiversity, agricultural land preservation) (see e.g. Brander et al 2006; 2007; Jacobsen and Hanley 2009; Johnston et al 2008; Lindhjem 2007; Lindhjem and Navrud 2008a; Loomis and White 1996; Richardson and Loomis 2009; Stapler and Johnston 2009). With very few exceptions, these studies do not focus specifically on MA methodology or implications for BT, despite the growing focus on meta-analytic BT in the literature. Compared to previous work, we add several new and interesting dimensions; (1) To investigate the effect of MA scope, we divide our dataset into two levels of heterogeneity; endangered species (more similar good and methods used) and nature conservation more generally (more heterogeneity in good and methods used); (2) We then estimate a number of random effects meta-regression models for these two main datasets using different cleaning procedures and subsets of the data investigating conformity with expectations, explanatory power and the robustness of results, and finally; (3) We report the level of forecast (or transfer) errors for unstudied sites broken down by type of models, nature conservation habitat, geographic region and valuation method used, based on a jackknife resampling technique used in MA by e.g. Brander et al (2006) and Lindhjem and Navrud (2008a). This study is, to our knowledge, one of the first attempts to systematically investigate the issue of heterogeneity in MA and BT in environmental valuation

Conceptual framework and data

Conceptual framework

We start by defining "nature conservation" broadly as the protection or active management of any natural terrestrial or aquatic ecosystem, resource or amenity, Q. The economic value measure for an increase in the level of nature conservation (Q) is the change in the quantity and/or quality (QUAL) of Q, or some set of services provided by Q, and is referred to as consumers' surplus (CS) or Willingness to pay (WTP). From the standard indirect utility function, the bid function for a representative individual j for this change can be given by (Bergstrom and Taylor 2006)[3]:

$$(1)\ \ WTP = f(P_j, M_j, Q_j^T - Q_j^R, QUAL_j^T - QUAL_j^R, SUB_j^T - SUB_j^R, H_j)$$

Where P = a price index of market goods (assumed constant), M = (individual or household) income (assumed constant), Q^T-Q^R and $QUAL^T$-$QUAL^R$ are the changes in quantity and quality from a reference situation ("status quo") (R) to a target state-of-the-world (T), SUB = substitutes for Q available to individual j, H = non-income household or individual characteristics. Further to make (1) elastic enough for use in MA, we assume, following Bergstrom and Taylor (2006), a "weak structural utility theoretic approach" in which the underlying variables in the bid function are assumed to be derivable from some unknown utility function, but that flexibility is maintained to introduce explanatory variables into the model, such as study design and different valuation methods, that do not necessarily follow from (1). This is the most common approach in MA, where the meta-analyst records estimates of mean WTP from different

[3] For simplicity and brevity we do not elaborate the details of how nature conservation may increase utility e.g. related to market goods and household production, e.g. as done by Van Houtven et al (2007) for water quality.

studies and corresponding explanatory variables both informed by theory and empirical expectations. In this process the empirical specification chosen for (1) needs to trade off the availability of information reported in valuation studies with the range of potentially relevant explanatory variables. For example, information about substitute sites to a national park will mostly not be reported, even if important for WTP. In addition, if information *is* reported, for example about the exact change in nature conservation valued, this change may not be easily comparable across sites and studies. No MA studies are free of this problem. Some try to map changes to a common unit of measurement in terms of hectares or to a water quality ladder or similar, though such simplified common units may mask differences in other dimensions of the good important to individuals (see e.g. Lindhjem 2007). There are no easy solutions, and in our rather general case we interpret mean WTP from different studies as welfare estimates for a (small, though not marginal) change in Q and/or in one or more elements in an attribute vector of QUAL describing the quality of the nature site[4]. We then use dummy variables to detect differences in WTP depending on the type of habitat or change valued. For example, when considering studies that value preservation of biodiversity we use variables for types of species and other characteristics of the good to capture variation in this overall value category. Before discussing the empirical specification of (1), we first describe the data used for the MA.

Description of the meta-data from nature conservation studies

[4] The ecosystem services and functions and total economic value from nature and biodiversity conservation are discussed in depth elsewhere, and for sake of brevity not elaborated in detail here (see e.g. Fromm 2000).

Given this conceptual framework, we conducted a broad search for studies (published papers, reports, book chapters etc[5]) internationally available in English valuing nature conservation in the region drawn from various databases. The first studies were conducted in Australia in the 1980s. In the rest of Asia, valuation started much later, but has grown in number substantially since the 1990s. Based on the literature search we compiled a gross meta-dataset of 577 mean WTP estimates (i.e. observations) from 99 studies. A first crude screening of the studies was conducted by excluding the ones reporting negative mean WTP or very high or low estimates (2 standard deviations of the mean), leaving 550 estimates from 95 studies for detailed analysis. This procedure reduces the influence of outlier estimates in regressions. The resulting distribution of studies by region, by type of habitat or service valued, and valuation method used are given in Tables 1-3 below[6]. Most of the studies are from Southeast Asia, East Asia or Oceania (mostly Australia), with a smaller number of studies from South and Southwest Asia (Table 1). Australia has the largest number of studies (22), followed by the Philippines (10). Raw mean annual WTP is highest for Oceania at US$ 254, as expected, though also high for South Asia (US$ 206). The lowest WTP, all at around the same level, is found in Southeast Asia (US$ 83), East Asia (US$ 76) and Southwest Asia (US$ 66).

[5] We did not include Master degree theses for practical reasons (hard to find and/or to get hold of) and because many are written in the native language.

[6] We do not claim to have collected an exhaustive database of all studies in Asia and Oceania, but the extent of our search makes us confident that we cover the majority of such studies in the region. Further, it is unlikely that our search has been biased in any way (except for the focus on studies in English). Finally, to answer our research questions, completeness is also not strictly necessary.

Table 1 Regional distribution of valuation studies (WTP in US$ 2006)

Region	Mean WTP (St.dev.)	No. of obs.	No. of studies
Southeast Asia (SEA)	83 (212)	244	32
Oceania (O)	254 (914)	116	23
East Asia (EA)	76 (108)	99	23
South Asia (SA)	206 (286)	70	11
Southwest Asia (SWA)	66 (78)	21	6
Total		**550**	**95**

Note: O= Australia, Micronesia, Papua New Guinea, Vanuatu; SEA= Cambodia, Indonesia, Laos, Malaysia, the Philippines, Singapore, Thailand, Vietnam; EA= China, Japan, Korea, Taiwan; SA= India, Sri Lanka; SWA= Iran, Israel, Pakistan.

Table 2 Distribution of valuation studies by habitat types (WTP in US$ 2006)

Types of habitats/services	Mean WTP (St.dev.)	No. of obs.	No. of studies
Terrestrial habitats (reserves, national parks, forests)	116 (252)	176	33
Marine habitats (reefs, beaches, sea, watercourses)	80 (97)	162	27
Endangered species (single or multiple)	105 (220)	129	16
Wetlands (wetlands, mangroves)	514 (1503)	41	8
Other habitats/services (landscapes, eco.-services)	121 (182)	37	13
Total		**550**	**97***

Note: * Some studies have more split samples asking different types of good, and thus the number of studies is higher than reported in Table 1.

Table 3 Valuation studies by methods (WTP in US$ 2006)

Method	Mean WTP (St.dev.)	No. of obs	No. of studies
Contingent valuation method (CV)	124 (505)	417	77
Choice modelling/experiments (CM)	67 (41)	50	8
Travel cost method (TCM)	161 (162)	37	14
Others (market price, hedonic pricing)	269 (435)	46	5
Total		**550**	**104***

Note: * In some studies, there are more than one method used, and thus the number of studies is higher than reported in Table 1.

The most frequently valued habitat is terrestrial habitats (including forests, nature reserves and national parks), grouped together here for ease of exposition (Table 2). Marine and freshwater habitats (i.e. coral reefs, beaches, sea, rivers, watercourses) for simplicity termed “marine habitats”, follow second. Wetlands have the highest value at US$ 514, mostly due to the market price methods often used to value such habitats (see Table 3). Studies that value named and endangered species or groups of species are categorised as “endangered species”. Marine habitats provide the lowest value (US$ 80) compared to other types of habitats, while terrestrial habitats (US$ 116), endangered species (US$ 105), and other habitats (US$ 121) have values that are around 40-50 percent higher. The by far most frequently used method is contingent valuation (CV), with 77 studies, while the travel cost method (TCM) comes second with 14 studies (Table 3). A small number of studies (5) use other methods, such as the hedonic pricing method or calculating the value of wetlands and forests using the market price approach. These methods frequently calculate a different welfare measure than CV, CM and TCM studies, and also yield twice as high estimates as the other methods. Details of the individual studies (including references) are given in Appendix A.

Coding of the data for meta-regression analysis

Information from the studies was coded in a spreadsheet originally containing 30 of the likely most important variables chosen from a large universe of potential covariates, with between 1 and 36 observations drawn from each study (average 5.8). The same study typically has several sub-samples varying the methods used, scope and other aspects of the good being valued. Table 4 below give the variable names and definitions. Since there is no standardised way of reporting welfare estimates in the literature, a wide variety of units are typically used, e.g. WTP per individual or household, per unit of area[7], per visitor, for different time periods (e.g. per month, per visit, per year, one-time amount etc.), and in different currencies and reporting years. To deal with this, we standardized the values to a common metric, i.e. WTP (US$ in 2006 prices) per household per year as a default, and coded WTP per individual, WTP per month etc., using dummies. For WTP per visit from CV or TCM studies, we calculated per visit WTP per year (if the study had information about how many trips a person would make per year, we converted to WTP per year). Values from different years were converted to 2006 prices using GDP deflators from the World Bank World Development Indicators. Purchase Power Parity (PPP) exchange rates were used to correct for differences in price levels between countries, which is the recommended procedure in international BT and MA (Ready and Navrud 2006). Some theoretical models predict that WTP given per household is higher than individual WTP, though empirical evidence is mixed (Lindhjem 2007; Lindhjem and Navrud 2009). It can also be expected that WTP given per month multiplied by 12 to convert to an annual amount is higher than WTP originally stated on an annual basis (a well-known bias).

We also included other methodological variables that are often used in MA studies: whether the study was a stated preference study (i.e. CV or CM) or other methods, whether it used personal interviews, if the CV method applied a dichotomous choice (DC) question format (i.e. the respondent says yes or no to a given bid, rather than stating max WTP), and whether the CV data were analysed using non-parametric statistical methods. Some studies find that CV yields lower WTP than revealed preference studies (e.g. Carson et al 1996), which is also in line with results in Table 3 above. DC formats are often found to give higher mean WTP than open ended formats

[7] Studies that reported results with per unit of area were excluded, as the total size typically was not given.

(a main reason is so-called yea-saying), while non-parametric methods typically give a lower bound on WTP[8]. There is no clear prior for use of interviews vs. other modes, though type of survey mode is known to influence results (Lindhjem and Navrud 2008b).

Further, we include a set of geographic and good characteristics variables to control for differences in welfare estimates between types of species (mammals, turtles) and habitat types, between regions and countries, and between primarily non-use vs. use value. Larger and more charismatic or iconized species (for example elephants or pandas) are likely to yield higher welfare estimates than non-charismatic species or biodiversity/nature conservation in general (e.g. as found in Jacobsen et al 2008 and Richardson and Loomis 2009), though it is uncertain a priori if our MA will be able to detect such a pattern across several studies. Studies that primarily estimate non-use values are likely to give lower value estimates. There are no strong priors regarding other habitat types or regional/country dummies, though it is expected that these dimensions may influence WTP[9]. We considered including a dummy for the season of the study (e.g. rainy vs. dry season) similar to Lindhjem (2007), however in most cases such information was not reported.

The only socio-economic variable generally reported is income of the sample, which we include in our analysis. Around 78 percent of the studies report this. For those which do not, we follow common practice from other MA studies to use a proxy for income from

---

[8] Standard error of WTP estimates was generally not reported, making it impossible to weigh estimates by level of precision in the meta-regressions, a procedure recommended in the MA literature (e.g. USEPA 2007).

[9] We also considered using population density of the country of study as a variable, for example as done by Brander et al (2006) for wetlands. However, we think link between nature conservation and population density may be overly tenuous, and excluded this variable in our analysis.

other sources instead (we use GDP/capita for the country). It is expected that income will positively influence WTP, an often-found result in the literature for primary studies. However, in MA studies WTP is often relatively insensitive to income (see e.g. Johnston 2005; Jacobsen and Hanley 2009). One reason for this is the low variation in income levels in MA studies conducted within the same country or in Western countries with similar income levels. In our case we have a fairly large variation in income, so should expect that WTP may increase with income.

Finally, we include a proxy variable for study quality; whether a study is a published or unpublished paper (i.e. a journal article or research report/working paper). Though published studies may be expected to apply more stringent and perhaps conservative methods, it is not clear if this would result in lower WTP. There may also be publication bias with unknown influence on WTP (Rosenberger and Stanley 2007). A way to limit the potential impact of publication bias is also to include unpublished studies. To capture trends in WTP values over time not captured by income (or other coded variables), we include a trend variable for the year of the study (rather than publication year). Some MA studies find WTP to increase over time, reflecting, perhaps, both increased nature scarcity and "greener" preferences. Others argue that increased methodological prudence should result in lower WTP estimates in more recent studies. Since a portion of our studies is funded by the same institution, Environment and Economy Program for Southeast Asia, and may share similarities we have not otherwise coded, we include a dummy (EEPSEA) to control for that. This procedure is similar to Bateman and Jones (2003), who find indications of similarities in WTP estimates from the same authors.

Table 4 Definition of meta-analysis variables and descriptive statistics

Variables	Description	Mean (SD)*
Dependent variable:		
WTP 2006	WTP in 2006 prices (US$)	133 (461)
Methodological variables:		
SP	Binary: 1 if stated preference, 0 if otherwise	.84 (.35)
DC	Binary: 1 if SP using dichotomous choice, 0 if otherwise	.51 (.50)
TCM	Binary: 1 if travel cost method, 0 if otherwise	.07 (.25)
Hholdpay	Binary: 1 if household's WTP, 0 if individual	.67 (.46)
Month	Binary: 1 if payment is a monthly payment, 0 if otherwise	.35 (.47)
Nonpara	Binary: 1 if estimate is non-parametric (Turnbull), 0 otherwise	.07 (.25)
Interview	Binary: 1 if it is an in-person interview, 0 otherwise	.60 (.48)
Good characteristics variables:		
Mammal	Binary: 1 if it is a mammal, 0 otherwise	.04 (.20)
Turtle	Binary: 1 for sea turtle, 0 otherwise	.06 (.24)
Species	Binary: 1 for endangered species, 0 if other habitats/services	.23 (.42)
Terrestrial	Binary: 1 for terrestrial habitats, 0 if other habitats/services	.32 (.47)
Marine	Binary: 1 if marine habitat (beach, sea, watercourse, lake, river), 0 other habitats/services	.29 (.45)
Wetland	Binary: 1 for wetlands, 0 if other habitats/services	.07 (.26)
Nonuse	Binary: 1 for primarily non-use, 0 otherwise	.77 (.41)
Socio-economic variables:		
Income	Continuous: Mean household income from sample (PPP adjustment, 2006)	14,318 (17,258)
GDP	Continuous: GDP 2006 from country for survey. Inserted for household income in one model.	14,524 (12,191)
Geographic characteristics (countries and regions):		
Australia	Binary: 1 if the study in Australia, 0 otherwise	.19 (.39)
Philippin	Binary: 1 if a study in the Philippines, 0 otherwise	.22 (.42)
Oceania	Binary: 1 if a study in Oceania, 0 other region	.21 (.40)
East	Binary: 1 if a study in East Asia, 0 other region	.18 (.38)
Southeast	Binary: 1 if a study in Southeast Asia, 0 otherwise	.44 (.48)
Southwest	Binary: 1 if a study in Southwest Asia, 0 otherwise	.04 (.19)
South	Binary: 1 if a study in South Asia, 0 otherwise	.13 (.33)
Other variables:		
EEPSEA**	1 if the study is funded by EEPSEA, 0 otherwise	.39 (.48)
Journal	1 if it is a published paper, 0 otherwise	.47 (.49)
Year	Continuous: from 0 (2006) to 26 (1979)	6.36 (4.07)

Notes: *The Mean (SD) is for overview purposes given for the whole dataset. The scope of the dataset is limited in the model runs in the next section. Further, not all variables are used in all models.
**EEPSEA = Environment and Economy Program for Southeast Asia.

Meta-regression model and results

Meta-regression model

For our meta-regressions, we divided the dataset into two primary levels of scope, according to level of homogeneity of the good and methods used: Level 1: Endangered species; and Level 2: Biodiversity and nature conservation more generally. The endangered species data include WTP estimates from 16 studies using CV to value the preservation of single or multiple species. These CV studies typically ask how much local/domestic populations are willing to pay for various conservation programs for endangered species (e.g. WTP to conserve a viable population of sea turtles)[10].

10 of the studies are funded by EEPSEA (hence the importance of the control variable discussed above). The species valued in these studies include sea turtles (several countries), black-faced spoonbill (Macau), rhinos (Vietnam), eagles and whale shark (Phillipines), and various other species such as dugong dugong, elephants, rhinos, dolphins and tigers (Thailand). In addition we found six non-EEPSEA funded studies in the region using CV to value the preservation of the possum (a marsupial species native to Australia) and glider (the Mahogany Glider: a type of endangered possum), giant panda (China), and elephants (India, Sri Lanka). The 16 studies provide 124 estimates that will be used in the meta-regression analysis. Although the species are different, we consider the preservation of them as a good with many similar attributes in valuation (i.e. a larger degree of homogeneity of the good), as compared to nature and biodiversity conservation more generally. In addition, methodological heterogeneity is reduced since all the studies in this level use CV.

The second level of the data, include the studies from Level 1 and all the rest of the studies that value nature conservation more generally, with different types of methods (though the majority also use CV here). This dataset includes welfare estimates for a fairly heterogeneous good, however, not more so, it can be argued, than many other complex environmental goods studied in MA. Further, as almost all non-textbook goods in general (and environmental goods in particular) are heterogeneous to some degree, it is unclear from theory where to draw the line in practice. All in all the Level 2 dataset contains between 67 to 95 studies and 390 to 550 estimates, depending on the cleaning procedures and the subsets of the data used in the meta-regressions. The details of the Level 1 and 2 datasets are given in Tables A1 and A2, respectively, in Appendix A. We will conduct several meta-regression models based on these two levels of data to explain variation in welfare estimates and to investigate effects of different dimensions of heterogeneity.

As most studies provide more than one WTP estimate, the data should most prudently be treated as a panel to account for the correlation between the errors of estimates from the same study[11] (e.g. Nelson and Kennedy 2009). Thus we used the procedure proposed by Rosenberger and Loomis (2000b) to check for panel structures in the data. The panel structure model, our empirical specification of equation (1) above, can be written as:

---

[10] A small number of studies survey foreign populations, e.g. Bandara and Tisdell (2005) study OECD citizens' WTP for the preservation of the Giant Panda in China.

[11] We also tested two other stratifications of the data: by survey and by author. Results (available from the authors) show that in many model specifications of the two stratifications equal effects (and random effects) cannot be rejected.

$$(2)\ WTP_{ij} = \alpha + \sum_{i=1}^{n} \beta_i x_{ij} + \mu_{ij} + \varepsilon_i$$

where WTP is the *i*'th observation from the *j*'th strata (here study), α is a constant, x_{ij} is a vector of explanatory variables (as defined in Table 4), with a panel effect μ_{ij} and an error $\varepsilon_i \sim N(0, \sigma_\varepsilon^2)$. We also assume that μ_{ij}, ε_i and x_{ij} are uncorrelated within and across studies. A Breusch-Pagan Lagrange multiplier (LM) statistic test of whether panel effects are significant was conducted. The null hypothesis is that an equal effects model is correct ($H_0 : \mu_{ij} = 0$), and the alternative that a panel effects model is correct $(H_1 : \mu_{ij} \neq 0)$. If the hypothesis of fixed effects in the Breusch-Pagan LM test is rejected, the random effects model assuming heterogenous effect sizes across studies and within models should be more efficient in estimation. We chose a double-log specification of (2), common in the MA literature, which fitted our data better than linear or other specifications. For a model with income as the only explanatory variable[12], the Breusch-Pagan LM test showed that a model with equal effects was rejected, confirming the appropriateness of a panel estimation model (χ^2 = 274.90, p=0.000 with N=550 and j=95).

In order to test whether a random effects specification (which has a panel specific error component) is outperformed by a fixed effects model (which keeps the panel specific error component constant), a Hausman χ^2 test was performed for the whole dataset. The null hypothesis is that the random effect specification is correct, i.e. the panel effects are uncorrelated with other regressors, and the alternative that the fixed effect specification

---

[12] A comprehensive test would have included other explanatory variables with different model specifications, but for sake of simplicity and brevity, we only present the model with the income variable here

is correct (Zanderson and Tol 2009). The results in Table 5 show that the random effects model (B) cannot be rejected, and thus, is the one we use.

Table 5 Test of random vs. fixed effects panel structure (N=550, j=95)

	b Fixed effects model	**B Random effects model**	**b-B**	**S.E.**
Income variable	.0305127	-.0494427	.0799554	.2193994
$p > \chi^2$: 0.7155				

We also performed the Hausman test for all the models used in this study, i.e. for different subsets of the data and different explanatory variables included, and find that a random effects model is the best estimation approach for Level 1 and 2 of our data.

Results and discussion

First, we provide results in this section of different meta-regression models for Levels 1 and 2 of our data. Then, in the next section, we use the estimated models to investigate errors in predicted values for unstudied sites (i.e. a BT simulation). Results of six random effects GLS regression models for the Level 1 data (species) are reported in Table 6. Starting with Model 5, this is a model that includes all explanatory variables in Table 4 of relevance to the Level 1 data. Only one regional dummy and two species type dummies (instead of the full range of species types) are used, as estimates are thinly spread across categories. Model 5 shows how a fully specified model is able to deal with the heterogeneity of the data, for the most homogenous of the two datasets. Models 1-4 (and 6) are versions of Model 5 where adding different subcategories of variables illustrates changes in the explanatory power of the models. Model 1 contains methodological characteristics of the CV methods only, Model 2 adds good characteristics, Model 3 adds country variables (instead of region dummy in Model 5), and Model 4 includes income and the survey year variables. A range of models was tried using combinations of variables in Table 4. Models 1-4 presented here were chosen to avoid collinearity (excluding e.g. the EEPSEA variable), to include dummies reflecting a significant share of the data (i.e. excluding region dummies for Level 1 data), to obtain best fit and to enhance comparison between models and between Level 1 and Level 2 data.

Going from Model 1 to 4, the models gradually explain more of the variation in WTP for species preservation. The methodological variables in Model 1 explain around 40 percent of the variation ($R^2 = 0.398$), while adding characteristics of the species explain another 14 percent of the variation ($R^2 = 0.536$). Adding country specifics and income and year in Models 3 and 4 help explain another 22-27 percentage points of the variation. Model 4, the best fitting of the models, obtains an overall R^2 of 0.81, which is very high compared to other MA studies. Model 5 obtains nearly the same level of R^2. It is comforting for our belief in the validity of the data and for the potential use of such value estimates for BT that around half of the explained variation in the best model is due to non-study specific, observable characteristics related to the good, geographical area, year of study and income level of the population surveyed. Model 6, a version of Model 5 where all method variables are excluded, drives home the same point, with a R^2 of 61 percent. This model is particularly interesting for testing in the next section how ignoring methodological differences translates into BT errors predicting values for new sites. Note that the models are directly comparable since they all include the same observations.

Individual parameter estimates in the best Model 4 confirm well with expectations, where such priors exist. The DC format tends to provide higher estimates than other formats, as expected, as do monthly payments compared to other periods of payment. Non-parametric estimates are significantly lower than estimates using parametric methods, also as expected. Household payment is significantly higher than WTP from individual payment (Models 4 and 5), though theoretical and empirical expectations here are not clear. Personal interview is not significantly different than other survey modes in the more fully specified Models 2-5.

Table 6 Meta-regression models for Level I: Endangered species studies

Variable#	**Model 1**	**Model 2**	**Model 3**	**Model 4**	**Model 5**	**Model 6**
	Method variables	**+ Species types**	**+ Country variables**	**+ Income and year**	**All variables**	**No method variables**
Constant	1.298*	2.413***	1.493	-9.365***	(dropped)	-.455
	(.095)	(.002)	(.156)	(.001)		(3.861)
DC	1.517*	.695	1.187	1.555***	1.294**	
	(.064)	(.374)	(.102)	(.002)	(.566)	
Hholdpay	.855	.038	.563	1.722***	2.222**	
	(.295)	(.961)	(.438)	(.003)	(.922)	
Month	.168	.657	1.116*	.140	1.015**	
	(.810)	(.274)	(.092)	(.788)	(.419)	
Nonpara.	-.259**	-.278**	-.273**	-.281***	-.277**	
	(.032)	(.022)	(.016)	(.010)	(.109)	
Interview	1.525***	.113	.729	-.972	.064	
	(.004)	(.873)	(.375)	(.192)	(.571)	
Turtle		-.363	-.675	-.954***	-.830**	.004
		(.470)	(.151)	(.001)	(.375)	(.469)
Mammal		1.740**	.856	1.569***	1.664***	2.574***
		(.035)	(.277)	(.004)	(.578)	(.470)
Australia			.698	-2.048**		
			(.415)	(.019)		
Philippin.			-.982***	-.126		
			(.000)	(.699)		
Southeast					-.248	-.587**
					(.281)	(.288)
EEPSEA					-11.421***	-.441
					(3.908)	(1.720)
Journal					-14.870***	(dropped)
					(5.394)	
LnIncome				.854***	.798***	.505**
				(001)	(.225)	(.228)
LnYear				2.189***	4.424**	-.669
				(.000)	(2.003)	(1.445)
Summary statistics:						
R^2: within	.044	.044	.157	.231	.234	.168
R^2: betw.	.604	.785	.879	.961	.947	.727
R^2: overa.	**.398**	**.536**	**.757**	**.810**	**.804**	**.611**
Sigma_u	.841	.674	.614	.330	.420	.630
Sigma_e	.470	.470	.444	.425	.424	.435
Rho	.761	.672	.656	.376	.459	.677
N	124	124	124	124	124	124
# studies	16	16	16	16	16	16

Note: *p < 0.10, **p < 0.05, ***p < 0.01. STATA 9.2 used. # Blank space means variable not included in regression.

Valuation of turtle preservation is significantly lower than for other species (though insignificant in Models 2 and 3), while mammals are valued significantly higher[13]. Higher values for mammals can be explained by their higher degree of “charisma” than for other, lower-profile species. Australian studies provide lower WTP estimates than other countries, when controlling for income level in Models 4. This may be because Australian studies value species we have classified as “non-charismatic”, i.e. the possum (see Appendix A). Studies conducted in the Philippines are likely to give lower values (though only significant in Model 3) than studies conducted in other countries. The income parameter, i.e. the income elasticity of WTP in our double-log formulation, is around 0.8 and highly significant in the best Models 4 and 5. Income elasticity of WTP in the 0-1 range is commonly found in the CV literature (e.g. Kriström and Riera 1996). In Model 4 more recent studies yield significantly lower WTP estimates, reflecting perhaps increased prudence in the use of valuation methods over time.

In Table 7 we present results of six random effects GLS regression models instead using the more heterogeneous Level 2 data (nature and biodiversity conservation, i.e. Level 1 are a subset of the Level 2 data). In this case we include the fuller range of explanatory variables (e.g. covering different valuation methods) using different subsets of the data. We keep the same methodological variables (except we include the dummy for stated preference values) for the sake of comparing the robustness of the results with Level 1. Further, we include the habitat/good characteristics variables that are significant across at least one of our four models. Finally, geographic region dummies were included if significant or if data from these regions dominate our dataset. Similar to the models in

---

[13] We also tried other groupings or specifications of types of species, such as size, degree of ”charisma” across types of species etc, but found that using the biological classification ”mammal” worked best in our models. Adding dummies for each species is not feasible due to the limited number of observations for each.

Table 6, we first run a fully specified Model 5 using all variables in Tables 1-4 and then we exclude in Model 6 method variables. Model 1 investigates the full dataset of 550 observations, inserting GDP as proxy for unreported income information, while Model 2 excludes studies that did not report income information. These two models illustrates the difference between a "conservative" meta-analyst who would not accept the imprecision introduced by inserting proxies for unreported variables and a more "pragmatic" approach. Both practices are found in the MA literature. Model 3 contains the Model 2 observations, excluding values estimated using other methods than CV, CM, and TCM (i.e. market price and hedonic pricing methods[14]), as these methods typically estimate conceptually different welfare measures. Model 4 contains studies of endangered species only (the same observations as in Model 4 from Level 1), for sake of comparison. Given the heterogeneity of the good included in the Level 2 data, our fully specified Model 5 does not do very well in controlling for this using the covariates we have been able to code and include. The model explains only 13.5 percent of the variation. This is only slightly increased for Model 1, to a R^2 of 16 percent, which offers the best combination of covariates for the full dataset. However, it is comparable to the 25-26 percent obtained in two national level MA studies of an apparently more homogenous good; recreation activity days in the USA (see Rosenberger and Loomis 2000a and Shresta and Loomis 2003)[15]. Our R^2 for the full dataset is generally higher than the random effects MA models of international biodiversity studies in Jacobsen and Hanley (2009).

---

[14] The TCM variable is the "hidden" category in Model 3, now that other non-SP methods are excluded. In Models 1-2 the TCM variable is excluded as it is not significant across models.

[15] Since R^2 obtained from random effects models is not directly comparable to standard R^2 OLS, the comparison should be interpreted with caution.

Table 7 Meta-regression models for Level 2: Biodiversity and nature conservation

Variable#	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
	GDP inserted for income	**Income reported**	**Only SP and TCM**	**Level 1 data**	**All variables**	**No method variables**
Constant	3.455**	4.058***	3.448***	(dropped)	6.554***	5.522***
	(.022)	(.001)	(.002)		(1.800)	(1.664)
SP	-.450	-1.713***	-1.769***	(dropped)	-2.593***	
	(.149)	(.000)	(.000)		(.628)	
DC	.580***	.0114	-.065	-1.856	.760***	
	(.007)	(.950)	(.642)	(.163)	(.221)	
TCM					-2.657 ***	
					(.676)	
Hhldpay	.335	.025	.008	-2.270**	-.085	
	(.248)	(.923)	(.976)	(.032)	(.332)	
Month	.606	1.377***	1.448***	2.893***	1.021**	
	(.108)	(.000)	(.000)	(.000)	(.404)	
Nonpara	-.252	-.209	-.220*	-.307***	-.267	
	(.300)	(.229)	(.078)	(.006)	(.237)	
Interview	.080	-.009	.176	1.749**	.153	
	(.778)	(.970)	(.442)	(.049)	(.309)	
Turtle	-.026	-.117	-.275	-.912** (.014)		
	(.968)	(.811)	(.579)			
Mammal	1.666***	1.885***	1.715***	1.710***		
	(.007)	(.000)	(.001)	(.002)		
Marine	.888***	.562**	.554**	(dropped)	.046	.134
	(.004)	(.035)	(.042)		(.447)	(.437)
Wetland	-.991**	1.258	1.218***	(dropped)	-1.967***	-1.718***
	(.021)	(.003)***	(.003)		(.538)	(.528)
Species					-.942**	-.372
					(.439)	(.423)
Terrestrial					-1.143***	-.893
					(.446)	(.442)
Nonuse	.057	-.240	-.084	(dropped)	.175	.093
	(.809)	(.269)	(.639)		(.237)	(.210)
Oceania	.755*	.677 *	.588	(dropped)	.994	.513
	(.099)	(.095)	(.146)		(.647)	(.630)
East	-.204	.180	-.105	-3.825	-.421	-.646
	(.622)	(.612)	(.776)	(.108)	.638	(.632)
Southeast	-.766*	-.323	-.841**	-3.997*	-.879	-.975
	(.063)	(.364)	(.028)	(.080)	(.670)	(.665)
South					.131	.433
					(.751)	(.731)
EEPSEA	-.449**	-.561*	.188	(dropped)	-.357	-.266
	(.024)	(.070)	(.609)		(.403)	(.406)
Journal	-.318	-.263	-.017	-5.309	-.096	-.354
	(.351)	(.387)	(.956)	(.373)	(.371)	(.366)
LnIncome	-.022	.062	.103	.867***	-.027	-.068
	(.863)	(.558)	(.260)	(.000)	(.140)	(.136)
LnYear	.281	.213	.180	.791	.168	.020
	(.234)	(.270)	(.342)	(.818)	(.262)	(.256)
Summary statistics:						
R^2 within	.124	.124	.212	.227	.179	.103
R^2:betwen	.172	.550	.572	.953	.155	.074
R^2:overall	**.159**	**.337**	**.459**	**.790**	**.135**	**.095**
Sigma_u	.955	.708	.764	.396	1.032	1.037
Sigma_e	1.083	.809	.582	.440	1.066	1.108
Rho	.437	.434	.632	.447	.484	.466
N	550	431	390	124	550	550
# studies	95	70	67	15¤	95	95

Note: *p < 0.10, **p < 0.05, ***p < 0.01. STATA 9.2 used. # Blank space means variable not included in regression.
¤ Due to different st.dev. in Level 2 data, screening procedures reduce studies from 16 to 15 for Level 1 data, though number of observations turns out by coincidence to be the same.

Excluding the studies from Model 1 for which a crude GDP/Capita measure was substituted for missing income information, more than doubles the explained variation (Model 2, R^2 = 0.34). The coefficient on income turns positive, but is not significant. Enhancing methodological homogeneity in Model 3 increases the explained variation further to 46 percent, the same level as for example found in the MA of Brander et al (2006) of international wetland valuation studies. Finally, in Model 4, using the Level 1 dataset, with the more complete range of explanatory variables does not change R^2 much compared to Model 4 in Table 6. Despite a higher degree of heterogeneity than for the Level 1 dataset, the data show some degree of regularity, and many of the parameters have the expected signs. Stated preference (SP) methods tend to give lower estimates than revealed preference (RP) methods, as expected. It is also as expected that monthly payments yield higher estimates than other payment vehicles and that non-parametric estimates are lower than parametric ones, like for the Level 1 dataset. The other methodological parameter estimates (i.e. household WTP, personal interview) are not robust across models and there are no strong priors for their signs. The signs and significance of the turtle and mammal parameters are preserved from the Level 1 models.

Marine habitats are valued significantly higher than other habitats across Models 1-3, while wetlands tend to be valued higher (though not robust across models). Primarily non-use value estimates are not significant across models. Studies conducted in Oceania (mostly Australia) tend to yield significantly higher values (in Models 1-2). Studies from Southeast Asia give lower values, compared to other regions. Interestingly, studies funded by EEPSEA give lower values than studies funded by other institutions, though not robust across all models. The income parameter is positive for studies that have reported income information from their samples, but only significantly so in Model 4 for the endangered species data. Year is positive but not significant in any models. Removing the methodological variables from the fully specified Model 5, reduces the explanatory power to a dismal 9.5 percent in Model 6 – an aspect that may invalidate the model for BT purposes (see next section). In addition to the models in Table 6 and 7, we also ran Models 1-4 in reduced form, in which variables not significant at the 20 percent level are removed. This form is often used in MA-BT (see Rosenberger and Loomis 2000a and Lindhjem and Navrud 2008a). Detailed regression results are given in Appendix B. Finally, we also removed the method variables for Model 3 (Level 2), to test if that model may still be useable for BT, if choosing to ignore methodological differences (see regression results in Appendix C). The results are discussed in relation to the BT simulations in the next section.

Increasing the degree of homogeneity of our data in terms of good characteristics and methods, then, generally increases the conformity with theoretical and empirical expectations and explanatory power of the models, as expected. For the more homogenous Level 1 data, observable characteristics of the type of species, region and other variables (income, year) add significantly to the explanatory power of the models. Even with the fairly heterogeneous Level 2 dataset, two models are still able to explain a significant part of the variation of up to 34 and 46 percent, respectively, comparable with other MA studies. For example, the R^2 of 46 percent of the Level 2 Model 3 is only about 10-15 percentage points lower than what was found in the MA of van Houtven et al (2007) of water quality valuation studies in the USA. They screened 300 publications related to water quality valuation and found only 11 studies (96 observations) they considered "sufficiently comparable" to be included in the MA. However, for our most heterogeneous Models 1 and 5, the chosen covariates are not able to control for the heterogeneity in a satisfactory way, judged from the level of explained variation. Given the degree of confirmed validity of our data and explanatory power, the next, and directly policy relevant question, is how the MA models will perform forecasting values for unstudied sites, i.e. used for BT. This is the question we turn to in the next section.

A check of the transferability of nature conservation values

MA-BT involves transferring one or more estimated meta-regression equations (2) to an unstudied policy site, and insert values from this site for the geographic, socio-economic, good characteristics variables and relevant year, and predict or forecast annual WTP per household. The values of methodological variables would typically be set at some best practice level, at the average sample value (Stapler and Johnston 2009) or drawn from the MA sample distribution (Johnston et al 2006), since there is no such information for an unstudied policy site. To the extent observable characteristics of the habitats/good valued and the population explain a significant portion of the variation in WTP, and not only the methodological differences between studies, it gives us confidence that MA-BT may be a credible alternative to a new valuation study as input for example in cost-benefit analysis. The performance of MA-BT can only be accurately assessed if we knew the "true value", or an estimate of this, for a range of sites, and then used the MA models to predict the value at those sites, and calculate so-called transfer errors (TE)[16]. Lindhjem and Navrud (2008a) and a few other studies, use different "benchmark" values from within their sample or from new studies to "simulate" the true value to assess TE performance. We will not conduct a full such investigation, but only carry out a first check on how our MA models forecast nature conservation values for

---

[16] $TE = \frac{|WTP_T - WTP_B|}{WTP_B}$, where T = Transferred (predicted) value from study site(s), B = Estimated (observed) true value ("benchmark") at policy site.

our two datasets. We use a jackknife data splitting technique, used e.g. by Brander et al (2006), where we estimate n-1 separate meta-regression equations to predict (or forecast) the value of the omitted observation in each case (i.e. "the policy site"). We then calculate the percentage difference between observed and predicted values, the TE in our simple exercise, and the overall median and mean TE for all observations[17].

This measure gives a first indication of how far off our MA models would be in a real BT exercise. We start by reporting the results for the six models using the Level 1 and Level 2 data (Table 8 and 9 below, respectively)[18].

Table 8 Median and mean transfer error (percent) for models Level 1: Endangered species

	Model 1	**Model 2**	**Model 3**	**Model 4**	**Model 5**	**Model 6**
	Method variables	**+ Species types**	**+ Country variables**	**+ Income and year**	**All variables**	**No method variables**
Median	61	59	33	24	23	36
Mean	108	85	58	46	45	67
N	124	124	124	123	123	123

Table 9 Median and mean transfer error (percent) for models Level 2: Biodiversity and nature conservation

	Model 1	**Model 2**	**Model 3**	**Model 4**	**Model 5**	**Model 6**
	GDP inserted for income	**Income reported**	**Only SP and TCM**	**Level 1 data**	**All variables**	**No method variables**
Median	71	52	46	22	77	64
Mean	7344	377	89	45	5277	8363
N	547	428	387	121	547	547

First, using a fully specified model such as Models 4 and 5 for Level 1, yields fairly high precision at around median TE of 23-24 percent (mean 45-46 percent) (Table 8).

[17] The mean prediction error is often termed Mean Absolute Percentage Error (MAPE).

[18] To account for econometric error in transforming ln(WTP) to WTP using antilog, we add standard deviation ($s^2/2$), which estimate varies when the sample changes, prior to transformation of ln(WTP) (see e.g. Johnston et al 2006) Some of the observations were dropped by STATA performing the TE estimations in Tables 8-9 as compared to Tables 6-7.

Mean TE is low compared to other studies performing this check, e.g. Lindhjem and Navrud (2008a) (62-266 percent), Brander et al (2006; 2007) (74-186 percent), Stapler and Johnston (2009) (152 percent), Richardson and Loomis (2009) (34-45 percent for a within-sample test), indicating a level of precision that could be acceptable for policy use. A general level of 20-40 percent has been suggested (Kristofersson and Navrud 2007). Precision increases with the more fully specified models, as expected. Interestingly, introducing species dummies to Model 1 reduces median TE by only 2 percentage points in Model 2 compared to an almost halving of TE from introducing country variables in Model 3. Model 6, which uses only observable values at the "policy site" (i.e. no methodological controls included), still manages to predict values with a median precision of 36 percent. Stapler and Johnston (2009) find that using the hypothetical "ideal" values for the methodological variables instead of the means in the sample in a fully specified MA-BT model, gives a TE gain overall of only 26 percentage points (from 151.9 to 125.6 percent). In an earlier study Johnston et al (2006) find that the choice of values for methodological variables may have a large impact on forecasts. Although ignoring methodological differences between studies altogether may not generally be a sensible approach in MA-BT, our case illustrates at least that loss in precision is relatively low as long as the good and methods used are relatively similar. We also ran the BT tests using reduced versions of Models 1-4 (Level 1 & 2), excluding variables not significant at the 20 percent level, a common procedure in MA-BT studies. We found very similar TE levels and no clear relationship between TE and reduced vs. full versions of the models (see results in Tables B1-B4 in Appendix B)[19].

---

[19] We also ran the same TE simulations using a rule-of-thumb of p>0.1 instead of p>0.2 for the reduced models,

For the Level 2 data Models 1-4 median TE is, somewhat surprisingly, comparable to the Model 1-4 Level 1 results, despite lower explanatory power (Table 9). However, the Level 2 data produce more high TE values (i.e. the mean is much higher than the median). Keeping to the median TE, Models 5 and 6 for Level 2 do a poorer job at controlling heterogeneity as judged by the TEs of 77 and 64 percent, respectively, compared to Models 5 and 6 for Level 1. However, although the mean TEs are in the thousands, it is somewhat surprising that medians are not influenced more by the low explanatory power of Models 5 and 6 for Level 2. We also ran a version of Model 6 for Level 2 based on the more methodologically homogenous dataset used for Model 3. This produced median TE of 49 percent (mean 108 percent), down from 64 percent in Table 6, caused by only excluding hedonic and market price method observations (having no methodological covariates)[20] (see Model 6a in Appendix C for results). Reducing methodological heterogeneity for the Level 2 data from Model 2 to 3 reduces median TE from around 52 to 46 percent, while mean TE comes down from an unacceptably high level of 377 percent to a more reasonable 89 percent. For both Level 1 and 2 models there is generally an inverse relationship between the level of explained variation and TE, as expected. Hence, increasing degree of homogeneity of the data in terms of good characteristics (biodiversity and nature conservation in general to endangered species) increases the precision, as does the enhanced homogeneity of valuation methods used within Level 2. However, in median terms, the gain in precision is perhaps not as highly related to explanatory power or homogeneity, as expected. Even with a heterogeneous dataset, median TE may approach acceptable levels for policy use.

---

detecting no clear(er) relationship with TE.

[20] We acknowledge that reducing methodological heterogeneity may also implicitly reduce good heterogeneity, as some types of nature conservation values may be more likely to be estimated using particular methods.

But for mean TE the results are clear. The plot of observed WTP values (estimates sorted in ascending order) vs. predicted (zigzag line,) for Model 4 (Level 1 data) is illustrated in Figure 1. The forecasts follow the observed values well except at the extremities of the data, a characteristic of forecasting models. For comparison, Model 1 (the whole dataset, 550 observations) for Level 2 is plotted in Figure 2. This plot shows a lower level of precision than for Level 1 in Figure 1 (note that the scale is different).

Figure 1 Plot of predicted vs. observed lnWTP for Model 4 (Level 1)

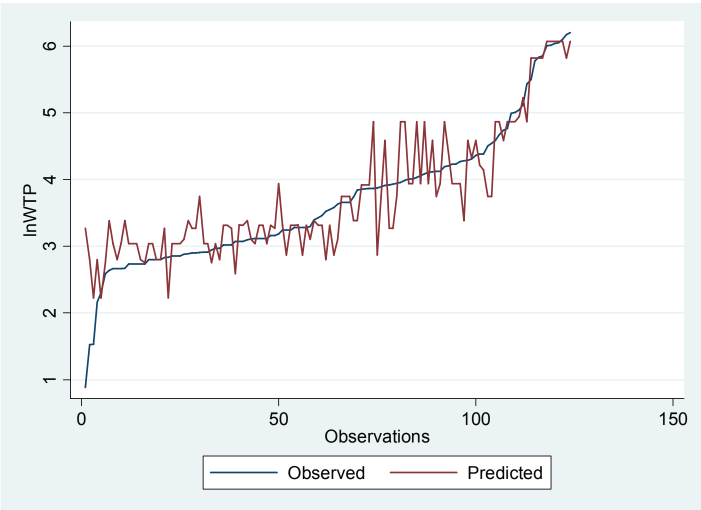
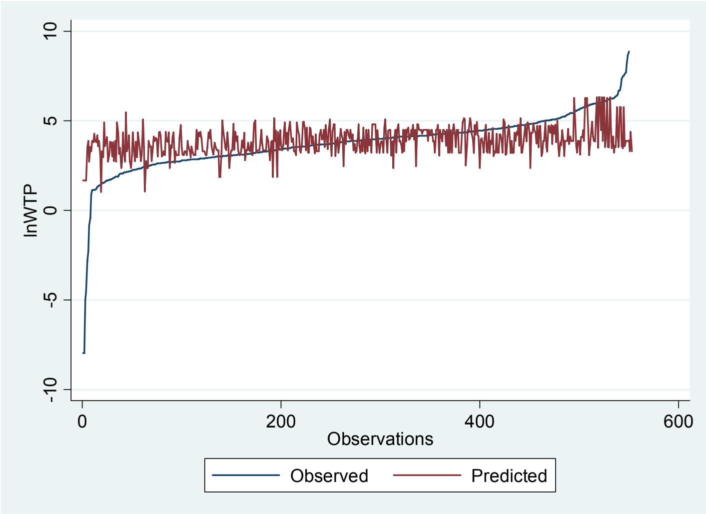


Figure 2 Plot of predicted vs observed lnWTP for Model 1 (Level 2)



We also break down estimated median and mean TE from Tables 8 and 9 for Models 1-4 only, for different subsets of the Level 1 and 2 data, i.e. by different characteristics of the good (Levels 1 and 2), valuation methods and geographical region (Level 2). First, TE for the four Level 1 models predicting values for preservation of turtles, mammals and other species, are given in Table 10. Precision increases from Model 1 through to 4 for all species types. WTP for mammal preservation is predicted with a median (mean) precision of 16 (17), percent, while for other species median TE doubles.

Table 10 Median (mean) transfer error in percent for different species, Models 1-4, Level 1 data

	Turtle	**Mammal**	**Other species**
Model 1	45 (169)	50 (120)	67 (75)
Model 2	52 (114)	43 (86)	65 (71)
Model 3	32 (97)	36 (61)	33 (39)
Model 4	24 (69)	16 (17)	32 (43)
# of obs.	**34**	**19/20**	**70**

In Table 11 we split the estimated TE for species and different types of habitats for the Level 2 data. The precision is generally higher for the endangered species (median TE of 36 percent for Model 3 and 22 percent for Model 4). Model 3 predicts WTP for terrestrial and marine habitats with the same median error of 40-46 percent (means at around 100 percent), while wetlands and other habitats have higher median errors. The similarity (homogeneity) in the change of environmental quality (ensure preservation) among the endangered species studies may be the reason that their values are predicted at a higher level of precision. Santos (1998) argues that the prediction errors he obtains in a MA of CV studies of landscape conservation are higher than those estimated in Loomis' (1992) study of rivers within the same US state, due to landscapes being a more heterogeneous good.

Table 11 Median (mean) transfer error in percent for different habitat types, Models 1-4, Level 2 data

	Terrestrial habitats	**Marine habitats**	**Endangered species**	**Wetlands**	**Other habitats**
Model 1	86 (545)	63 (605)	47 (85)	71 (92838)	77 (184)
Model 2	62 (1134)	44 (105)	36 (60)	77 (116)	78 (79)
Model 3	46 (104)	40 (106)	36 (57)	71 (119)	81 (75)
Model 4	-	-	22 (45)	-	-
# obs.	81- 173	129-162	121-129	31-41	17-37

Enhancing methodological homogeneity from Model 2 to 3 (i.e. removing estimates using market price or hedonic methods) reduces TE especially for terrestrial habitats. This is an indication that other valuation methods introduce substantial noise for terrestrial habitat valuation in the MA. In Table 12 we break down TE by valuation methods used. WTP estimates derived by CV has a median (mean) TE of 41 (71) percent in the most homogenous Model 3. Estimates derived by TCM or other valuation methods generally have higher TE than stated preference methods.

Table 12 Median (mean) transfer error in percent for different valuation methods, Models 1-4, Level 2 data

	CV	CM	TCM	Others
Model 1	64 (157)	70 (159)	73 (104785)	93 (1776)
Model 2	41 (81)	78 (167)	101 (141)	84 (3882)
Model 3	41 (71)	26 (149)	105 (145)	-
Model 4	22 (45)	-	-	-
# of obs.	121-423	50	17-37	37

Finally, breaking the TE estimates down by region in Table 13 shows that using the model to predict values in Southeast Asia produces the lowest TE, which is partly due to the larger number of estimates from this region. Except for some very high TE estimates pulling up the mean, median TE for transfers to all regions is below 80 percent and approaching acceptable levels for policy use.

Table 13 Median (mean) transfer error in percent for different regions, Models 1-4, Level 2 data

	Southeast Asia	Oceania	East Asia	South Asia	Southwest Asia
Model 1	59 (16000)	80 (209)	66 (163)	80 (1605)	45 (76)
Model 2	42 (102)	61 (81)	68 (304)	67 (3184)	44 (43)
Model 3	37 (90)	59 (76)	68 (110)	21 (27)	36 (36)
Model 4	23 (44)	31 (50)	28 (57)	15 (17)	-
# of obs.	69-244	16-116	26-99	10-41	12-21

Concluding remarks

Pushing the boundaries of meta-analysis (MA) in environmental economics, we have taken stock of studies estimating willingness to pay (WTP) for conservation of endangered species, biodiversity and nature more generally in Asia and Oceania. Our literature review shows that nature conservation is highly valued, probably more so in many cases than the opportunity costs of increasing conservation efforts in the region, though such a comparison is beyond the scope of our study. Dividing our dataset into two levels of heterogeneity in terms of good characteristics and valuation methods, we show that the degree of regularity and conformity with theory and empirical expectations as well as the explanatory power of our MA models is higher for the more homogenous dataset of endangered species values, as expected. In fact, though the species are different, the values to preserve them generally follow predictable patterns. For example, we find that mammals are generally valued higher than other species, likely due to the "charismatic" nature of this family. Further, WTP increases significantly with income (elasticity is around 0.8). The analysis of the endangered species data show that around half of the variation in the best model is due to non-study specific observable characteristics of the good and population surveyed, boding well for use of such data in benefit transfer (BT) applications. However, importantly, increasing the scope of the MA, i.e. gradually including more heterogeneous observations, generally preserves some of the regularity and the explanatory power of some of our models is in the range of other MA studies of goods typically assumed to be more homogenous (such as national water quality, recreation days etc).

Subjecting both our dataset levels to a simple check of benefit transfer error (TE), using the MA models to predict observations one-by-one when excluded from the datasets, show for the best models median (mean) TE of 23 (45) percent for the endangered species data and 46 (89) percent for the more heterogeneous nature and biodiversity data. This is in the low range compared to other MA studies. Results suggest that such levels of forecasting errors may approach acceptable levels for policy use. However, caution should be exercised in using values for single species for BT, as such estimates may include values of biodiversity or habitats more generally (see e.g. Veisten et al 2004). It is also clear from our results that for example including values estimated using a more heterogeneous set of methods for the Level 2 data, even a fairly broad range of covariates is unable to explain and control for the variation in a satisfactory way, translating into large mean transfer errors. A more careful testing of explanatory variables and MA models than we have done (for example including interaction effects) may be required to better understand if heterogeneous good and method characteristics can be controlled for using classical meta-regression analysis. However, even the most extensive list of explanatory variables in MA-BT we have seen to date in Stapler and Johnston (2009) is still unable to bring mean TE below 150 percent for recreational angling values in North America.

Hence, we are still grappling with the question of how to strike the right balance between screening out studies from the analysis and coding them with the aim of later controlling for increased heterogeneity in regression models. How homogenous is homogenous enough? Fundamentally, there is still much we do not know about people's preferences and how to represent and interpret them in MA models. Increasing clarity and transparency of effect size definitions, data collection and screening protocols offering others the chance to replicate results, is one important way forward for MA (e.g. as pointed out by Nelson and Kennedy 2009 and USEPA 2006). Using sensitivity analysis to investigate the effects of important analyst choices related to the scope and heterogeneity of the MA dataset is another, as exemplified in this study.

This paper is, to our knowledge, one of the first attempts to systematically investigate the issue of heterogeneity in MA and BT in environmental valuation. More research for other goods and geographical areas is needed to inform the development of a more consistent and generally applicable MA methodology, especially as MA is gradually being applied for BT to inform policy. Use of MA in economics is growing and the aim should be to move more of the methodological choices out of the black box.

Acknowledgements

We would like to thank Vic Adamowicz, Ståle Navrud and Randall Rosenberger for constructive comments. Funding from the Environment and Economy Programme for Southeast Asia (EEPSEA) is greatly appreciated.

References

Bandara R and Tisdell C (2005) Changing Abundance of Elephants and Willingness to Pay for their Conservation. Journal of Environmental Management 76: 47-59

Bateman I J and Jones A P (2003) Contrasting conventional with multi-level modeling approaches to meta-analysis: Expectation consistency in UK woodland recreation values. Land Economics 79(2): 235-258

Bergstrom J C and Taylor L O (2006) Using meta-analysis for benefits transfer: Theory and practice. Ecological Economics 60: 351-360

Brander L M, Florax R J G M and Verrmaat J E (2006) The Empirics of Wetland Valuation: A Comprehensive Summary and a Meta-Analysis of the Literature. Environmental & Resource economics 33: 223-250

Brander L M, van Beukering P and Cesar H (2007) The recreational value of coral reefs: a meta-analysis. Ecological Economics 63(1): 209-218

Carson R T, Flores N E, Martin K M and Wright J L (1996) Contingent valuation and revealed preference methodologies: Comparing the estimates for quasi-public goods. Land Economics 72(1): 80-99

Desvousges W H, Johnson F R and Banzhaf H S (1998) Environmental policy analysis with limited information: Principles and applications of the transfer method. Cheltenham, Edward Elgar

Engel S (2002) Benefit Function Transfer versus Meta-analysis as Policy-Making Tools: A Comparison. In: R. J. G. M. Florax, P. Nijkamp and K. G. Willis (eds) Comparative environmental economic assessment, Cheltenham, Edward Elgar

Fromm O (2000) Ecological structure and functions of biodiversity as elements of its total economic value. Environmental and Resource Economics 16(3): 303-328

Glass G V, McGaw B and Smith M L (1981) Meta-analysis in Social Research. Sage

Hoehn J P (2006) Methods to address selection effects in the meta regression and transfer of ecosystem values. Ecological Economics 60(2): 389-398

Jacobsen J B, Boiesen J H, Thorsen B J and Strange N (2008) What's in a name? The use of quantitative measures versus 'Iconised' species when valuing biodiversity. Environmental & Resource economics 39(3): 247-263

Jacobsen J B and Hanley N (2009) Are there income effects on global willingness to pay for biodiversity conservation. Environmental and Resource Economics (In press)

Johnston R J, Besedin E Y, Iovanna R, Miller C J, Wardwell R F and Ranson M H (2005) Systematic variation in willingness to pay for aquatic resource improvements and implications for benefit transfer: a meta-analysis. Canadian Journal of Agricultural Economics 53(2-3): 221-248

Johnston R J, Besedin E Y and Ranson M H (2006) Characterizing the effects of valuation methodology in function-based benefits transfer. Ecological Economics 60(2): 407-419

Johnston R J, Kukielka J B and Duke J M (2008) Systematic Variation in Willingness to Pay for Agricultural Land Preservation. Paper presented at the 2008 Workshop on Meta-Analysis in Economics and Business, Nancy, France, October 17-18, 2008.

Kristofersson D and Navrud S (2007) Can Use and Non-Use Values be Transferred Across Countries? In: Navrud S and Ready R (ed) Environmental Value Transfer: Issues and Methods, Springer, Dordrecht, The Netherlands

Kriström B and Riera P (1996) Is the income elasticity of environmental improvements less than one? Environmental and Resource Economics 7: 45-55

Lindhjem H (2007) 20 Years of stated preference valuation of non-timber benefits from Fennoscandian forests: A meta-analysis. Journal of Forest Economics 12(4): 251-277

Lindhjem H and Navrud S (2008a) How reliable are meta-analyses for international benefit transfers? Ecological Economics 66(2-3): 425-435.

Lindhjem H and Navrud S (2008b) Internet CV surveys - a cheap, fast way to get large samples of biased values? Department of Economics and Resource Management, Norwegian University of Life Sciences

Lindhjem H and Navrud S (2009) Asking for Individual or Household Willingness to Pay for Environmental Goods. Implication for aggregate welfare measures, Environmental and Resource Economics (In press)

Loomis J B (1992) The Evolution Of A More Rigorous Approach To Benefit Transfer - Benefit Function Transfer. Water Resources Research 28(3): 701-705

Loomis J B and White D S (1996) Economic benefits of rare and endangered species: Summary and meta-analysis. Ecological Economics 18(3): 197-206

Millennium Ecosystem Assessment (MEC) (2005) Synthesis Report. Washington, Island Press

Moeltner K, Boyle K and Paterson R W (2007) Meta-analysis and benefit transfer for resource valuation - addressing classical challenges with Bayesian modeling. Journal of Environmental Economics and Management 53: 250-269

Moeltner K and Rosenberger R (2008) Predicting Resource Policy Outcomes via Meta-Regression: Data Space, Model Space, and the Quest for 'Optimal Scope'. The B.E. Journal of Economic Analysis & Policy 8(1), Article 31

Nelson J P and Kennedy P E (2009) The Use (and Abuse) of Meta-Analysis in Environmental and Natural Resource Economics: An Assessment. Environmental and Resource Economics (In press)

Pearce D (2005). Economists and biodiversity conservation: What can we contribute? European Association of Environmental and Resource Economists, Bremen, 2005

Ready, R and Navrud, S (2006) International benefit transfer: Methods and validity tests. Ecological Economics 60: 429-434

Richardson L and Loomis J (2009) The total economic value of threatened, endangered and rare species: An updated meta-analysis. Ecological Economics (In press)

Rosenberger R and Johnston R J (2007). Selection effects in meta-valuation function transfers. Aarhus Colloqium of Meta-Analysis in Economics. Sønderborg, Denmark

Rosenberger R S and Loomis J B (2000a) Using meta-analysis for benefit transfer: In-sample convergent validity tests of an outdoor recreation database. Water Resources Research 36(4): 1097-1107

Rosenberger R and Loomis J (2000b) Panel stratification in meta-analysis of economic studies: an investigation of its effects in the recreation valuation literature. Journal of Agricultural and Applied Economics 32(1): 131-149

Rosenberger R and Stanley T D (2007) Publication Effects in the Recreation Use Values Literature. Annual Meeting of the American Agricultural Research Association, July 29-August 1, 2007, Portland, Oregon

Santos J M L (1998) The Economic Valuation of Landscape Change. Theory and Policies for Land Use and Conservation. Cheltenham, Edward Elgar

Shrestha R K and Loomis J B (2003) Meta-Analytic Benefit Transfer of Outdoor Recreation Economic Values: Testing Out-of-Sample Convergent Validity. Environmental & Resource economics 25: 79-100

Smith V K and Pattanayak S K (2002) Is Meta-Analysis a Noah's Ark for Non-Market Valuation? Environmental and Resource Economics 22: 271-296

Stanley T D and Jarrel S D (2005) Meta-Regression Analysis: A Quantitative Method of Literature Surveys. Journal of Economic Surveys 19(3): 299-308

Stapler R W and Johnston R J (2009) Meta-Analysis, Benefit Transfer, and Methodological Covariates: Implications for Transfer Errors. Environmental and Resource Economics (In press)

U.S. Environmental Protection Agency (2006) Report of the EPA work group on VSL meta-analysis, report NCEE-0494. National Center for Environmental Economics

Van Houtven G, Powers J and Pattanayak S K (2007) Valuing water quality improvements using meta-analysis: Is the glass half-full or half-empty for national policy analysis? Resource and Energy Economics 29(3): 206-228

Veisten K, Hoen H F, Navrud S and Strand J (2004) Scope insensitivity in contingent valuation of complex environmental amenities. Journal of Environmental Management 73(4): 317-331

Zandersen M and Tol R S J (2009) A meta-analysis of forest recreation values in Europe. Journal of Forest Economics 15(1-2): 109-130.

Appendix A: Descriptions of and full references for meta-analysis datasets:

Table A1 Studies of endangered species used in meta-analysis (MA) – Level 1

Name of references	Country	Year[a]	Species	Method	Mode	Vehicle[b]	Payment format	# of values[c]	WTP (USD)[d]
Jakobsson and Dragun (2001)	Australia	2000	Possum	CV: DC, PC	Mail	M	Year	12	14-72
Tisdell et al (2005)	Australia	2002	Glider	CV: OE	Mail	V	One-off	3	13-19
Kontoleon and Swanson (2003)	China	1998	Panda	CV: DC	Interview	M&V	Visit	3	5-17
Jianjun et al (2006)[e]	China	2005	Turtles	CV: PL	Drop-off	M	Month	5	5-6
Jianjun (2006)	China	2005	Spoonbill	CV: DC	Drop-off	M&V	Month	17	4-19
Jianjun et al (2006)[e]	Philippines	2005	Turtles	CV: DC	Drop-off	M&V	Month	5	1-3
Harder et al (2006)	Philippines	2005	Eagles	CV: DC	Drop-off	M&V	Month	30	1-3
Indab (2006)	Philippines	2005	Shark	CV: DC	Drop-off	M	Month	5	2-4
Bandara and Tisdell (2004)	Sri Lanka	2002	Elephants	CV: DC	Drop-off	V	Month	4	20-40
Bandara and Tisdell (2005)	Sri Lanka	2001	Elephants[f]	CV: DC	Interview	V	Month	6	34-41
Jianjun et al (2006)[e]	Thailand	2005	Turtles	CV: DC	Drop-off	M&V	Month	5	3-8
Nabangchang (2006)	Thailand	2005	Multiple[g]	CV: DC	Interview	M&V	One-off	7	43-64
Jianjun et al (2006)[e]	Vietnam	2005	Turtles	CV: DC	Drop-off	M	Month	4	0.2-4
Thuy (2006)	Vietnam	2005	Rhino	CV: DC	Drop-off	M&V	Month	2	13-14
Tuan et al (2008)	Cn,Pp,Th,Vn[h]	2005	Turtles	CV: DC	Drop-off	M&V	Month	16	1-5
Ninan and Sathyapalan (2005)*	India	2000	Elephants	CV:DC	Interview	M	Year	5	341-1830
Total number of studies=16								**129**	

Notes: *4 of 5 observations from this study were excluded by the screening criterion (2x STD of mean) for the Level 1 data, but included in the Level 2 dataset (see Table B)

[a] Year of data.

[b] Payment vehicle: mandatory (M) or voluntary (V).

[c] Number of WTP values used in MA.

[d] WTP values in US$. The WTP formats are given as reported (i.e. lump sum, per month, per year, per visit, per individual or household). WTP values in local currencies are converted to US$ using PPP adjustments; and values from different years are converted to 2006 prices using CPI.

[e] Jianjun et al (2006) has four separate country case study components.

[f] Abundance of elephants.

[g] Multiple species: Dugong dugong, elephants, rhinos, Irawaddy dolphin, tigers.

Table A2 Studies of nature and biodiversity conservation used in meta-analysis (MA) – Level 2

Name of references	Country	Year[a]	Habitat/service type	Method	Mode	Vehicle	Payment format	# of values	WTP (USD)
Jakobsson and Dragun (2001)	Australia	2000	Flora & fauna	CV: DC, PC	Mail	M	year	7	24-175
Bennett et al (1998)	Australia	1996	Wetlands	CV: DC	Mail	M	One-time	2	122-187
Bennett (1984)	Australia	1979	Nature reserve	CV: OE	Interview	M&V	One-time	1	33
Blamey et al (1999)	Australia	1999	Water	CA	Interview	M	Year	4	29-116
Cameron and Quiggin (1994)	Australia	1991	Parks	CV: IB	Interview	M		4	228-664
Carr and Mendelsohn (2003)	Australia	2000	Reefs	TCM	Interview		Year	1	391
Carson et al (1994)	Australia	1990	parks	CV: DBDC	Interview	M	Year	4	30-129
Hundloe (1990)	Australia	1986	Reefs	TCM	Interview		Year	1	8
Kuosmanen et al (2003)	Australia		Parks	TCM	Mail		Year	6	54-418
Lockwood and Carberry (1998)	Australia	1997	Vegetation	CM, CV	Mail	M	One-time	8	35-90
Lockwood and Tracy (1995)	Australia	1993	Parks	CV: OE	Mail	V	One-time	1	21
Lockwood (1999)	Australia	1995	Parks	CV: OE	Computer	V		4	14-450
Lockwood (1996)	Australia		Natural environm.	CV: DC	Mail	V		9	5-123
Loomis et al (1993)	Australia		Forests	CV: OE, DC	Mail		Year	6	34-89
Morrison et al (2002)	Australia	1997	Wetlands	CM	Mail	M	One-time	18	25-117
Nillesen et al (2005)	Australia		Parks	TCM	Mail		Year	1	86
Streever et al (1998)	Australia	1996	Wetlands	CV: OE	Mail	M		1	151
Greiner and Rolfe (2004)	Australia	2000	Parks	CV: OE	Interview	M	Visit	3	23-39
Campbell and Reid (2000)	Australia	1996	Fisheries	CV: DC	Interview	M	Year	3	212-517
Flatley and Bennett (1995)	Vanuatu	1994	Forest	CV	Interview	V	One-time	2	33-36
Flatley and Bennett (1996)	Vanuatu	1994	Forest	CV	Interview	V	One-time	1	18
Chen et al (2004)	China	1999	Beaches	TCM	Interview	M	Visit	1	64
Day and Mourato (2002)	China	1997	Rivers	CV: DBDC	Interview	M		4	51-94
Gong (2004)	China	2002	National reserve	CV: BG	Interview	M		2	5-16
Guo et al. (2001)	China	1997	Ecosystem services	TCM	Interview		Visit	3	20-40
Jim and Chen (2006)	China	2003	Urban green spaces	CV: PC	Interview			1	15
Yaping (1998)	China	1996	Lakes	CV:OE& TCM	Interview	M	Visit Year	7	77-114 6-15
Zhongmin et al (2003)	China	2001	Ecosystem services	CV: PC	Interview	M&V	Year	3	8-87
Zhongmin et al (2006)	China	2003	Watershed	CV:DC, DBDC	Interview	M&V	Year	2	71
Wang et al (2007)	China	2006	Water	CV:DC	Interview	M	Month	2	
Xu et al (2007)	China	2002	Eco-services	CM	Interview	M	Year	7	51-134
Gundimeda and Kathuria (2003)	India	2003	Water	HPM	Interview			2	149-377
Hadker et al (1997)	India	1995	Parks	CV: DC	Interview	V	Month	2	6-8
Kohlin (2001)	India	1995	Woodlots	CV: DC, OE	Interview		Month	11	4-6
Maharana et al (2000)	India	1998	Lakes	CV: IB	Interview		Year	4	5-43

				& TCM					
Nallathiga (2004)	India	1995	Rivers	CV: PC	Interview	M	Year	2	22-25
Butry and Pattanayak (2001)	Indonesia	1996	Forests	CV:OE, PC & MP	Interview	M	Year	3	23-2006
Pattanayak (2001)	Indonesia	1996	Ecological services	CV: DC	Interview	M	Year	1	20
Pattanayak and Kramer (2001)	Indonesia	1996	Watershed	CV: DC	Interview	M	Year	10	7-21
Walpole et al (2001)	Indonesia	1995	Parks	CV: DBDC	Interview	M	Year	1	78
Amirnejad et al (2006)	Iran	2004	Forests	CV: DC	Interview	M	Month	1	9
Fleischer and Tsur (2000)	Israel	1997	Landscapes	TCM	Interview		Year	2	179-367
Shechter et al (1998)	Israel	1993	Parks	CV: OE, DC	Telephone	V	One-time	12	28-57
Tsgue and Washida (2003)	Japan	1998	Natural areas of the Sea.	CV: DC	Internet		One-time	6	132-159
Nishizawa et al (2007)	Japan	2003	Eco-services	CVM:DC	Mail	M		2	13-14
Kwak et al (2003)	Korea	2001	Forests	CV: DC	Interview	M	Year	4	3-6
Lee (1997)	Korea	1996	Nature-based tourism resources	CV: DC	Interview	M	Year	2	12-13
Lee and Han (2002)	Korea	1999	Parks	CV: DC	Interview	M	Year	10	8-23
Lee and Mjelde (2007)	Korea	2005	Eco-services	CV:DC	Interview	M	Year	2	22-26
Eom and Larson (2006)	Korea	2000	Water	CV	Interview	M	Year	2	35-62
Lee and Chun (1999)	Korea	1994	Forest recreation	CV:DC	Mail	V	Year	3	445-787
Othman et al (2004)	Malaysia	1999	Forests	CM	Interview		Year	5	0.5-8
Yeo (2002)	Malaysia	1998	Parks	CV: OE	Interview		Year	6	6-12
Mourato (2002)	Malaysia	1997	Water	CV: PL	Interview	M	Month	2	3
Naylor and Drew (1998)	Micronesia	1996	Mangroves	CV	Interview	M	Month	4	174-556
Khan (2004)	Pakistan	2003	Parks	TCM	Interview		Visit	2	13-18
Manoka (2001)	P.N.Guinea	1999	Forests	CV: OE, DC	Mail	V	Year	10	11-101
Arin and Kramer (2002)	Philippines	1997	Marine sanctuary	CV: PC	Interview	M	Year	3	21-34
Calderon et al (2005)	Philippines	2003	Watersheds	CV: DC	Interview	M	Month	36	2-6
Choe et al (1996)	Philippines	1992	Water	CV: DC, BG, OE	Interview		Month	18	0-13
Pattanayak and Mercer (1998)	Phillippines	1994	Soil	MP	Interview		Year	2	195-306
Subade (2005)	Philippines	2002	Reefs	CV: DC	Interview & drop-off	V	Year	12	15-83
Amponin et al (2007)	Philippines	2006	Watershed	CV:DC	Interview	M	month	9	3-6
Wei-Shiuen and Robert (2005)	Singapore	2002	Beaches	TCM & CV	Interview		Year	14	0.1-485
Bogahawatte (1999)	Sri Lanka	1997	Forests	MP	Interview	Year	Year	30	1-437
Ekanayake and Abeygunawardena (1994)	Sri Lanka	1992	Forests	CV: OE	Interview		Year	2	41-131
Chang and Ying (2005)	Taiwan	2001	Agri. lands	CV: DBDC	Telephone	M	Year	1	103
Chen (1998)	Taiwan		Agri. lands	CV:OE, DC	Mail		Month	6	0.3-7
Hammit et al (2001)	Taiwan	1993	Wetlands	CV: DC, OE	Interview	V	Year	4	46-173

Cushman (2004)	Thailand	2001	Beaches	CM	Interview	M	Year	5	17-526
Isangkura (1998)	Thailand	1996	Parks	CR & CV: OE	Interview	M	Year	9	2-28
Seenprachawong (2002)	Thailand	2002	Coastal ecosystem	CM	Interview	V	Year	5	9-188
Seenprachawong (2001)	Thailand	2000	Reefs	CV:DC &TCM	Interview	V	Year	3	31-555
Tapvong and Kruavan (1999)	Thailand	1998	Rivers	CV: DC	Interview	M	Month	2	9-10
Pham and Tran (2000)	Vietnam	2000	Reefs	CV:PC & TCM	Interview	M	Year	5	7-170
Pham et al (2000)	Vietnam	2003	Reefs	TCM CV	Interview	M	Year	4	17-390
Phuong and Gopalakrishnan (2004)	Vietnam	2001	Water	CV: PC	Interview	M	Year	7	4-40
Do (2007)	Vietnam	2006	Wetlands	CM	Interview		Year	2	4-12
Nam et al (2001)	Vietnam	1999	Forests	CV & MP	Interview	M	Year	9	24-1807
Total number of observations*								**421**	

Notes: * The total number of observations using the least strict screening criterion (WTP>0 and within 2x STD of the mean), i.e. 129 (Level 1) + 421 (Level 2) = 550 observations. Blank space means that information was not reported in the study. See also notes to Table A above.

References of studies included in the meta-analysis

Amirnejad H, Khalilian S, Assareh M H and Ahmadian M (2006) Estimating the existence value of north forests of Iran by using a contingent valuation method. Ecological Economics 58(4): 665-675

Amponin J A R, Bennagen M E C, Hess S and Cruz J D S d (2007) Willingness to pay for watershed protection by domestic water users in Tuguegarao City, Philippines. PREM 07/06 Working paper

Arin T and Kramer R A (2002) Divers' willingness to pay to visit marine sanctuaries: an exploratory study. Ocean Coastal Manage. 45(2-3): 171-183

Bandara D and Tisdell C (2004) The net benefit of saving the Asian elephant: a policy and contingent valuation study. Ecological Economics 48: 93- 107

Bennett J, Morrison M and Blamey R (1998) Testing the validity of responses to contingent valuation questioning. Australian Journal of Agricultural and Resource Economics 42(2): 131-148

Bennett J W (1984) Using Direct Questioning to Value the Existence Benefits of Preserved Natural Areas Australian Journal of Agricultural Economics 28(2-3): 136-152

Blamey R, Gordon J and Chapman R (1999) Choice Modelling: Assessing the Environmental Values of Water Supply Options. Australian Journal of Agricultural and Resource Economics 43(3): 337 - 357

Bogahawatte C (1999). Forestry Policy, Non-timber Forest Products and the Rural Economy in the Wet Zone Forests in Sri Lanka. Research report, EPPSEA.

Butry D T and Pattanayak S K (2001). Economic Impacts of Tropical Forest Conservation: The Case of Logger Households around Ruteng Park. Working Paper 01_01, Environment and Natural Resource Economics Program, Research Triangle Institute, North Carolina.

Calderon M M, Camacho L D, Carandang M G, Dizon J T, Rebugio L L and Tolentino N L (2005). A Water User Fee For Households in Metro Manila, Philippines. EEPSEA research report, No. 2005-RR2.

Cameron T A and Quiggin J (1994) Estimation Using Contingent Valuation Data From A Dichotomous Choice With Follow-Up Questionnaire. J.Environ.Econ.Manage. 27(3): 218-234

Campbell H P and Reid C R M (2000) Consumption Externalities in a Commercial Fishery: The Queensland Beam Trawl Fishery. Economic Record 76(232): 1-14

Carr L and Mendelsohn R (2003) Valuing Coral Reefs: A Travel Cost Analysis of the Great Barrier Reef. Ambio 32(5): 353-357

Carson R T, Wilks L and Imber D (1994) Valuing the Preservation Of Australia Kakadu Conservation Zone. Oxf. Econ. Pap.-New Ser. 46: 727-749

Chang K and Ying Y (2005) External Benefits of Preserving Agricultural Land: Taiwan's Rice Fields. The Social Science Journal 42: 285-293

Chen W, Hong H, Liu Y, Zhang L, Hou X and Raymond M (2004) Recreation Demand and Economic Value: An Application of Travel Cost Method for Xiamen Island. China Economic Review

Chen Y (1998). Biodiversity in the Framework of Ecological Economics, University of Wisconsin Milwaukee.

Choe K A, Whittington D and Lauria D T (1996) The economic benefits of surface water quality improvements in developing countries: A case study of Davao, Philippines. Land Econ. 72(4): 519-537

Cushman C (2004). External costs from increased island visitation: case study from Southern Thailand. Working paper, Department of Resource Economics, University of Massachusetts.

Day B and Mourato S (2002) Valuing River Water Quality in China In: Pearce D, Pearce C and Palmer C (eds) Valuing the Environment in Developing Countries: Case Studies, Edward Elgar

Do N T (2007). Impacts of dykes on wetland values in Vietnam’s mekong river delta: a case study in the plain of reeds. Research report submitted to EEPSEA, IDRC.

Ekanayake E R M and Abeygunawardena P (1994) Valuation of Conservation Commodity of the Sinharaja Forest: Towards Total Economic Value. Sri Lanka Journal of Agricultural Economics 2(1): 115-129

Eom Y-S and Larson D M (2006) Valuing housework time from willingness to spend time and money for environmental quality improvements. Review of the Economics of the Household 4: 205-227

Flatley G and Bennet J (1995). International Values of Tropical Forest Conservation: A Cross-cultural Contingent Valuation Experiment. Paper presented to the 39th Australian Agricultural Economic Society Annual Conference. Perth, Western Australia

Flatley G W and Bennet J W (1996) Using Contingent Valuation to Determine Australian Tourists' Values for Forest Conservation in Vanuatu. Economic Analysis and Policy 26(2): 111-127

Fleischer A and Tsur Y (2000) Measuring the recreational value of agricultural landscape. Eur Rev Agric Econ 27(3): 385-398

Gong Y (2004). Distribution of Benefits and Costs among Stakeholders of a Protected Area: An Empirical Study from China. EEPSEA Research Report 3, IDRC

Greiner R and Rolfe J (2004) Estimating consumer surplus and elasticity of demand of tourist visitation to a region in North Queensland using contingent valuation. Tourism Economics 10(3): 317-328

Gundimeda H and Kathuria V (2003) Estimation of Economic Value of Water scarcity and quality in Chennai, India: The Hedonic Approach. South Asian Network of Economic Initiatives Project

Guo Z, Xiao X, Gan Y and Zheng Y (2001) Ecosystem functions, services and their values - a case study in Xingshan County of China. Ecological Economics 38(1): 141-154

Hadker N, Sharma S, David A and Muraleedharan T R (1997) Willingness-to-pay for Borivli National Park: Evidence from a contingent valuation. Ecol. Econ. 21(2): 105-122

Hammit J K, Liu J-T and Liu J-L (2001) Contingent Valuation of a Taiwanese Wetland. Environment and Development Economics 6: 259-268

Harder D S, Labao R and Santos F I (2006). Saving the Philippine eagles: How much will this cost and are Filipinos willing to pay for it? Paper presented at the 25th EEPSEA biannual workshop. Singapore.

Hundloe T (1990) Measuring the Value of the Great Barrier Reef. Australian Parks and Recreation 26(3): 11-15

Indab A (2006). Rationalizing WTP for whale shark conservation in Sorsogon province. Paper presented at the 26th EEPSEA biannual workshop. Cebu, the Philippines.

Isangkura A (1998). Environmental Valuation: An Entrance Fee System for National Parks in Thailand. Economy and Environment Program for Southeast Asia, IDRC

Jakobsson K M and Dragun A K (2001) The Worth of a Possum: Valuing Species with the Contingent Valuation Method. Environmental and Resource Economics 21: 287-302

Jianjun J (2006). Economic valuation of the Black-faced Spoonbill conservation in Macao. Paper presented at the 25th EEPSEA biannual workshop. Singapore.

Jianjun J, Indab A, Nabangchang O, Thuy T D, Harder D and Subade R (2006). WTP for Marine Turtle Conservation: A Cross-country Comparison in Asia.. Paper presented at the Third World Congress of Environmental and Natural recourses economics, Kyoto, Japan

Jim C Y and Chen W Y (2006) Recreation - Amenity Use and Contingent Valuation of Urban Greenspaces in Guangzhou, China. Landscape and Urban Planning 75(1-2): 81-96

Khan H (2004). Demand of eco-tourism: estimating recreational benefits from the Margalla Hills national park in the northern Pakistan. SANDEE working paper, No. 5-04.

Kohlin G (2001) Contingent valuation in project planning and evaluation: the case of social forestry in Orissa, India. Environment and Development Economics 6(02): 237-258

Kontoleon A and Swanson T (2003) The Willingness to Pay for Property Rights for the Giant Panda: Can a Charismatic Species Be an Instrument for Nature Conservation? Land Economics 79(4): 483-499

Kuosmanen T, Nillesen E and Wesseler J (2003). Does Ignoring Multi-Destination Trips in the Travel Cost Method Cause a Systematic Downward Bias? Discussion Paper No. 9 2003, Mansholt Graduate School.

Kwak S J, Yoo S H and Han S Y (2003) Estimating the public's value for urban forest in the Seoul metropolitan area of Korea: A contingent valuation study. Urban Studies 40(11): 2207-2221

Lee C-K and Han S-Y (2002) Estimating the Use and Preservation Values of National Parks' Tourism Resources Using a Contingent Valuation Method. Tourism Management 23: 531-540

Lee C-K and Mjelde J W (2007) Valuation of ecotourism resources using a contingent valuation method: The case of the Korean DMZ. Ecological Economics 63: 511-520

Lee C K (1997) Valuation of nature-based tourism resources using dichotomous choice contingent valuation method. Tourism Management 18(8): 587-591

Lee H-C and Chun H-S (1999) Valuing environmental quality change on recreational hunting in Korea: A contingent valuation analysis. Journal of Environmental Management 57(1): 11-20

Lockwood M (1996) Non-compensatory preference structures in non-market valuation of natural area policy. Aust. J. Agric. Econ. 40(2): 85-101

Lockwood M (1999) Preference structures, property rights, and paired comparisons. Environmental and Resource Economics 13(1): 107-122

Lockwood M and Carberry D (1998). Stated Preference Surveys of Remnant Native Vegetation Conservation, Johnstone Centre, Albury, NSW.

Lockwood M and Tracy K (1995) Nonmarket Economic Valuation Of An Urban Recreation Park. Journal of Leisure Research 27(2): 155-167

Loomis J, Lockwood M and Delacy T (1993) Some Empirical-Evidence On Embedding Effects In Contingent Valuation Of Forest Protection. Journal of Environmental Economics and Management 25(1): 45-55

Maharana I, Rai S C and Sharma E (2000) Valuing ecotourism in a sacred lake of the Sikkim Himalaya, India. Environmental Conservation 27(3): 269-277

Manoka B (2001). Existence Value: A Re-appraisal and Cross-Cultural Comparison. Economy and Environment Program for Southeast Asia (EEPSEA) Research Report No. 2001-RR1, IDRC.

Morrison M, Bennett J, Blamey R and Louviere J (2002) Choice modeling and tests of benefit transfer. Am. J. Agr. Econ. 84(1): 161-170

Mourato S (2002) Valuing Improvements to Sanitation in Malaysia. In: D. Pearce, C. Pearce and C. Palmer (ed) Valuing the Environment in Developing Countries: Case Studies, Cheltenham, Edward Elgar

Nabangchang O (2006). Motivations for Charitable Behavior for Protection of Wildlife and Endangered Species of Thailand.

Nallathiga R and Paravastu R B (2004). Benefit Estimation of River Water Quality Conservation Using Contingent Valuation Survey: a Case-Study in the Yamuna River Sub-basin. European Association of Environmental and Resource Economics Thirteenth Annual Conference. Budapest, Hungary.

Nam M V, Nhan N T, Trihn B V and Thong P L (2001). Forest Management Systems in the Mekong River Delta, Vietnam. EEPSEA research report.

Naylor R and Drew M (1998) Policy Options: Valuing mangrove resources in Kosrae, Micronesia. Environment and Development Economics 3: 471-490

Nillesen E, Wesseler J and Cook A (2005) Estimating the Recreation-Use Value for Hiking in Bellenden Ker National Park, Australia. Environmental Management 36(2): 311-316

Ninan K N and Sathyapalan J (2005) The economics of biodiversity conservation: a study of a coffee growing region in the Western Ghats of India. Ecological Economics 55(1): 61-72

Nishizawa E, Kurokawa T and Yabe M (2007) Policies and resident's willingness to pay for restoring the ecosystem damaged by alien fish in Lake Biwa, Japan. Environmental Science & Policy 9: 448-456

Othman J, Bennet J and Blamey R (2004) Environmental values and resource management options: a choice modelling experience in Malaysia. Environment and Development Economics 9: 803–824

Pattanayak S K (2001). How Green Are These Valleys? Combining Revealed and Stated Preference Methods to Account for Ecosystem Costs of Deforestation, Environment and Natural Resource Economics Program, Research Triangle Institute

Pattanayak S K and Kramer R A (2001) Pricing ecological services: Willingness to pay for draught mitigation from watershed protection in eastern Indonesia. Water Resour. Res. 37(3): 771-778

Pattanayak S K and Mercer D E (1998) Valuing soil conservation benefits of agroforestry: contour hedgerows in the Eastern Visayas, Philippines. Agricultural Economics 18(1): 31-46

Pham K N and Tran V H S (2000). Recreation Value of the Coral-surrounded Hon Mun Islands in Vietnam. Economy and Environment Program for Southeast Asia, IDRC.

Phuong D M and Gopalakrishnan C (2004) Optimal Management of Water for Sustainable Fisheries and Aquaculture International Journal of Water Resources Development 20(4): 493-506

Seenprachawong U (2001). An Economic Analysis of Coral Reefs in the Andaman Sea of Thailand. EEPSEA research report.

Seenprachawong U (2002). An Economic Valuation of Coastal Ecosystems in Phang Nga Bay, Thailand. EEPSEA Research Report No. 2002-RR5, IDRC.

Shechter M, Reiser B and Zaitsev N (1998) Measuring Passive Use Value: Pledges, Donations and CV Responses in Connection with an Important Natural Resource. Environmental and Resource Economics 12: 457-478

Streever W, Callaghan-Perry J M, Searles A, Stevens T and Svoboda P (1998) Public attitudes and values for wetland conservation in New South Wales, Australia. Journal of Environmental Management 54: 1-14

Subade R F (2005). Valuing Biodiversity Conservation in a World Heritage Site: Citizens' Non-Use Values for Tubbataha Reefs National Marine Park, Philippines. Economy an Environment Program for Southeast Asia (EEPSEA), IDRC.

Tapvong C and Kruavan J (1999). Water Quality Improvements: A Contingent Valuation Study of the Chao Phraya River. EEPSEA, IDRC.

Thuy T D (2006). Willingness to Pay for Conservation of Vietnamese Rhinoceros. Paper presented at the 26th EEPSEA biannual workshop. Cebu, the Philippines.

Tisdell C, Wilson C and Nantha H S (2005) Policies for Saving a Rare Australian Glider: Economics and Ecology. Biological Conservation 123: 237-248

Tsuge T and Washida T (2003) Economic valuation of the Seto Inland Sea by using an Internet CV survey. Marine Pollution Bulletin 47: 230-236

Tuan T H, Thuy T D, Jianjun J, Indab A and Nabangchang O (2008). WTP for Marine Turtle Conservation in Four Countries: Benefit Transfer Testing. Final report submitted to EEPSEA, July, 2008.

Walpole M J, Goodwin H J and Ward K G R (2001) Pricing policy for tourism in protected areas: Lessons from Komodo National Park, Indonesia. Conserv. Biol. 15(1): 218-227

Wang X, Bennett J, Xie C, Zhang Z and Liang D (2007) Estimating non-market environmental benefits of the Conversion of Cropland to Forest and Grassland Program: A choice modeling approach. Ecological Economics 63(1): 114-125

Wei-Shiuen N and Robert M (2005) The impact of sea level rise on Singapore. Environment and Development Economics 1002: 201-215

Xu Z M, Cheng G D, Bennet J, Zhang Z Q, Long A H and Kunio H (2007) Choice modeling and its application to managing the Ejina Region, China. Journal of Arid Environments 69: 331-343

Yaping D (1998). The Value of Improved Water Quality for Recreation in East Lake, Wuhan, China: Application of Contingent Valuation and Travel Cost Methods. Economy and Environment Program for Southeast Asia, IDRC.

Yeo B H (2002) Valuing a Marine Park in Malaysia. In: Pearce D, Pearce C and Palmer C (eds) Valuing the Environment in Developing Countries: Case Studies, Edward Elgar,

Zhongmin X, Guodong C, Zhiqiang Z, Zhiyong S and Loomis J (2003) Applying Contingent Valuation in China to Measure the Total Economic Value of Restoring Ecosystem Services in Ejina Region. Ecological Economics 44: 345-358

Zhongmin X, Loomis J, Zhiqiang Z and Hamamura K (2006) Evaluating the performance of different willingness to pay question formats for valuing environmental restoration in rural China. Environment and Development Economics 11: 585-601

Appendix B: Reduced Models 1-4 and benefit transfer results

This appendix displays the results of reduced Models 1-4 for the Level 1 data (Table B1) and Level 2 data (Table B3), where variables not significant at the 20 percent level have been excluded. These models are then used to calculate transfer errors (Tables B2 & B4), which can be compared to results in Tables 8 and 9 in the main text.

Table B1 Meta-regression models for Level I: Endangered species studies

Variable[#]	Model 1	Model 2	Model 3	Model 4
Constant	1.779*** (.010)	3.347*** (.000)	2.488*** (.000)	-9.867*** (.000)
DC	1.992*** (.004)		1.797*** (.004)	1.567*** (.000)
Hholdpay				1.775*** (.000)
Month			.785 (.101)	
Nonpara.	-.264** (.028)	-.265** (.027)	-.300*** (.010)	-.286*** (.009)
Interview	1.419*** (.004)			-1.105** (.025)
Turtle			-1.347*** (.001)	-.919*** (.000)
Mammal		2.038*** (.000)		1.678*** (.000)
Australia				-2.221*** (.000)
Philippin.			-1.143*** (.000)	
SE Asia				
EEPSEA				
Journal				
LnIncome				.895*** (.000)
LnYear				2.309*** (.000)
Summary statistics:				
R^2: within	0.044	0.044	0.157	0.230
R^2: betw.	0.541	0.673	0.690	0.961
R^2: overa.	**0.391**	**0.438**	**0.548**	**0.810**
Sigma_u	.867	.701	.611	.248
Sigma_e	.470	.470	.444	.423
Rho	.772	.689	.654	.255
N	124	124	124	124
# studies	16	16	16	16

Note: *p < 0.10, **p < 0.05, ***p < 0.01. STATA 9.2 used. # Blank space means variable not included in regression.

Table B2 Median and mean transfer error (percent) for reduced Models 1-4, Level 1

	Model 1	Model 2	Model 3	Model 4
Median	69	44	68	25
Mean	108	77	103	44
N	124	124	124	123

Table B3 Meta-regression models for Level 2: Biodiversity and nature conservation

Variable#	Model 1	Model 2	Model 3	Model 4
Constant	3.672*** (.000)	4.854*** (.000)	4.798*** (.000)	(dropped)
SP	-.448* (.092)	-1.828*** (.000)	-1.779*** (.000)	-5.837*** (.000)
DC	.473** (.022)			
TCM				
Hhldpay				
Month	.549* (.066)	1.188*** (.000)	1.253*** (.000)	1.776*** (.000)
Nonpara			-.227* (.068)	-.315*** (.005)
Interview				
Turtle				-.878*** (.002)
Mammal	1.838*** (.001)	1.715*** (.000)	1.745*** (.000)	2.185*** (.000)
Marine	.963*** (.001)	.532** (.017)	.717*** (.001)	(dropped)
Wetland	-.822** (.036)	1.282*** (.001)	1.307*** (.000)	(dropped)
Species				
Terrestrial				
Nonuse				(dropped)
Oceania	.910*** (.003)	.755*** (.006)	.720*** (.006)	(dropped)
East				
Southeast	-.758*** (.004)		-.801*** (.000)	
South Asia				
EEPSEA		-.695*** (.001)		-.680*** (.010)
Journal				
LnIncome				.873*** (.000)
LnYear				
Summary statistics				
R^2 within	0.101	0.109	0.203	0.222
R^2:betwen	0.169	0.530	0.564	0.924
R^2:overall	**0.145**	**0.311**	**0.448**	**0.779**
Sigma_u	.909	.669	.710	.325
Sigma_e	1.111	.811	.583	.439
Rho	.401	.404	.596	.354
N	550	431	390	124
# studies	95	70	67	15

Note: *p < 0.10, **p < 0.05, ***p < 0.01. STATA 9.2 used. # Blank space means variable not included in regression.

Table B4 Median and mean transfer error (percent) for reduced Models 1-4, Level 2

	Model 1	Model 2	Model 3	Model 4
Median	67	58	46	26
Mean	10449	279	86	44
N	547	428	387	121

Appendix C: Model 6a, Level 2 and transfer error results

Table C1 displays a version of Model 6 for the Level 2 data in Table 7, based on Model 3 (instead of Model 5), where methodological variables have been taken out. This model is then used to calculate transfer errors (Table C2), comparable to Model 6 in Table 9 in the main text.

Table C1 Meta-regression Model 6a for Level 2 data

Variables	Model 6a
Constant	4.087*** (1.275)
SP	
DC	
TCM	
Hhldpay	
Month	
Nonpara	
Interview	
Turtle	
Mammal	
Marine	-.0203 (.429)
Wetland	.220 (.536)
Species	-.936*** (.284)
Terrestrial	-1.069** (.425)
Nonuse	-.577*** (.180)
Oceania	.639 (.476)
East	-.484 (.458)
Southeast	-.919* (.483)
South Asia	2.387*** (.724)
EEPSEA	.793* (.425)
Journal	-.006 (.383)
LnIncome	.076 (.102)
LnYear	-.036 (.229)
Summary statistics:	
R^2 within	.079
R^2:betwen	.383
R^2:overall	**.310**
Sigma_u	.897
Sigma_e	.626
Rho	.672
N	390
# studies	67

Note: *p < 0.10, **p < 0.05, ***p < 0.01. STATA 9.2 used. # Blank space means variable not included in regression.

Table C2 Median and mean transfer error for alternative version of Model 6, Table 7.

	Model 6a
Median	49
Mean	108
N	387

Paper IV

Environ Resource Econ
DOI 10.1007/s10640-009-9261-0

Asking for Individual or Household Willingness to Pay for Environmental Goods?

Implication for Aggregate Welfare Measures

Henrik Lindhjem · Ståle Navrud

Received: 31 January 2008 / Accepted: 29 December 2008
© Springer Science+Business Media B.V. 2009

Abstract The aggregate welfare measure for a change in the provision of a public good derived from a contingent valuation (CV) survey will be higher if the same elicited mean willingness to pay (WTP) is added up over individuals rather than households. A trivial fact, however, once respondents are part of multi-person households, it becomes almost impossible to elicit an "uncontaminated" WTP measure that with some degree of confidence can be aggregated over one or the other response unit. The literature is mostly silent about which response unit to use in WTP questions, and in some CV studies it is even unclear which type has actually been applied. We test for differences between individual and household WTP in a novel, web-administered, split-sample CV survey asking WTP for preserving biodiversity in old-growth forests in Norway. Two samples are asked both types of questions, but in reverse order, followed by a question with an item battery trying to reveal why WTP may differ. We find in a test between samples that the WTP respondents state on behalf of their households is not significantly different from their individual WTP. However, within the same sample, household WTP is significantly higher than individual WTP; in particular if respondents are asked to state individual before household WTP. Our results suggest that using individual WTP as the response unit may overestimate aggregate WTP. Thus, the choice of response format needs to be explicitly and carefully addressed in CV questionnaire design and further research in order to avoid the risk of unprofitable projects passing the benefit-cost test.

Keywords Contingent valuation · Household · Individual · WTP

JEL Classification Q51 · H41

H. Lindhjem (✉) · S. Navrud
Department of Economics and Resource Management, Norwegian University of Life Sciences,
P.O. Box 5003, 1432 Ås, Norway
e-mail: henrik.lindhjem@umb.no

H. Lindhjem
Econ Pöyry, P.O. Box 5, 0051 Oslo, Norway
e-mail: henrik.lindhjem@poyry.com

Published online: 31 January 2009

Abbreviations
WTP Willingness to Pay
PC Payment card

1 Introduction

The aggregate welfare measure for a change in the provision of a public good derived from a contingent valuation (CV) survey will be higher if the same elicited mean willingness to pay (WTP) is added up over (adult) individuals rather than households. A trivial fact, however, once respondents are part of multi-person households it becomes almost impossible to elicit an "uncontaminated" WTP measure that with some degree of confidence can be aggregated over one or the other response unit (e.g. Quiggin 1998; Bateman and Munro 2006). The correct unit will not only depend on how and to whom the WTP question is phrased, but on the respondent's self-perceived agency and the type of resource allocation model prevailing in her[1] household (Delaney and O'Toole 2006; Strand 2007). Failing to appreciate this problem has important implications for the credibility of welfare estimates from CV studies, and stated preference research more generally (as for example acknowledged by Boyle 2003).[2] The issue has received little attention in the extensive CV literature, though response unit distortions in welfare estimates could be higher than other more "high-profile" CV biases, discussed for example in Carson et al. (2001). The result is an ambiguous CV practice applying a mix of approaches asking respondents for their personal WTP, their WTP on behalf of the household, or even leaving the unit unspecified.

The aim of this paper is to investigate the empirical consequences of the choice of response unit—household or individual—and to inform the theoretical debate about household decision-making models in the context of CV of environmental goods. We attempt to answer the following questions: (1) What is the relationship between individual and household WTP?; (2) Do respondents within the same sample change their WTP when prompted instead to state individual or household WTP, and if so what are their stated reasons?; and finally; (3) Can household and respondent characteristics explain the observed relationship between household and individual WTP? The spill over and use of models from the large household decision-making literature to answer such questions has been limited to date within environmental valuation research. No generally agreed theoretical framework has been developed, though some attempts have been made (notably Quiggin 1998; Bergstrom 2003; Strand 2005, 2007; Smith and Van Houtven 1998, 2004; Munro 2005; Bateman and Munro 2003). We compare the collective household decision model of Strand (2007), which predicts that response unit bias may actually not be a problem in large samples, with other models deriving hypotheses we test within and between samples. This paper is, to our knowledge, the first empirical contribution investigating the relationship between household and individual WTP for CV of environmental goods. Our study also supplements the results from a few limited empirical studies for other types of goods (e.g. TV broadcasting, health risks) or valuation methods (choice experiments) (Bateman and Munro 2006; Beharry and Scarpa 2006; Delaney and O'Toole 2006; Hasler et al. 2008; Delaney and O'Toole 2008). Compared to previous research, we add several new dimensions: (1) Core elements of our CV question-

[1] Gender-neutral pronoun variation may be confusing in a paper on household preferences, where gender differences are well known. However, no such differences are intended unless explicitly stated.

[2] In fact, this is part of a wider problem as traditional microeconomics typically leave the definition of the "consumer" rather obscure—treating households and individual consumers the same (Vermeulen 2002).

naire (e.g. type of good, scenario, payment vehicle, budget reminders) are consistent with eliciting both household and individual WTP from individual respondents; (2) Within this framework, all respondents are given both household and individual WTP questions, but the order is varied between samples, offering a clean test of differences in WTP; finally (3) We utilise both respondents' stated reasons and a CV dataset merged with previously collected respondent web panel information to explain the observed differences in individual and household WTP. We find that people state a higher WTP on behalf of the household than as individuals, though this difference is not significant between samples at the 5% level. However, when people are prompted to answer using the other response unit, the WTP difference increases and becomes significant, especially if they have been asked individual WTP first. Results suggest that response unit uncertainty may continue to be a source of substantial noise in aggregate welfare estimates, unless the issue is much more carefully addressed in survey design and testing.

2 Theoretical Framework and Empirical Expectations

In a typical CV survey a random household member would be asked WTP for a change in the provision of an environmental good in one of two main ways[3]:

(I) What is your maximum individual (or personal) WTP (on your own behalf)?; or
(II) What is your maximum WTP on behalf of your household (or your household's maximum WTP)?[4]

In CV studies, WTP from question I would normally be aggregated over individuals, and from II—the most frequently used—over households. A commonly held view, as pointed out by Strand (2007), is that the answer to II is higher than to I, but only if the respondent shows interpersonal preferences, such as altruism, towards other household members. If there are no such preferences, aggregating WTP from question II over households would underestimate the total welfare change. In their landmark book on CV, Mitchell and Carson (1989, pp. 265–266), advise for pure public goods simply to "allow an adult who claims to be the household head" to answer question II.[5] In a footnote they refer to Becker (1981)'s unitary model, in which household choices can be described as if they were made by a single individual (Samuelson 1956; Becker 1973).[6] A feature of the model is income pooling among household members, implying that the source of income does not influence consumption decisions. The unitary model has increasingly come under fire (see e.g. Vermeulen 2002)

[3] The third way, not specifying the unit, as is frequently observed in CV would not be advisable. Further, in English "you", without reference to unit, introduces ambiguity since there is no difference between plural and singular interpretation.

[4] We use the open ended WTP question format here, but what we write would naturally extend to other formats, such as dichotomous choice.

[5] The two reasons Mitchell and Carson put forward for this advice were in our view both somewhat misguided. First, they claimed that most payments for pure public goods are made at the household level, using income tax as one example. But income tax was always measured out and paid individually (though at that time household income was more commonly derived from only one income earner). Second, they claimed that choosing a household head was the U.S. Census Bureau practice at the time, though this rather archaic practice seems to have been abandoned already in 1980 partly due to feminist critique (Presser 1998). For quasi-private goods such as hunting Mitchell and Carson recommend eliciting individual WTP.

[6] This outcome results either from imposing a structure on the household decision-making problem so that the household utility function reduces to one (Samuelson), or through an altruistic (benevolent) head optimally allocating household resources (Becker).

and is being replaced by collective models where resource allocations are determined by cooperative (Pareto efficient) or non-cooperative bargaining among household members.

The most advanced attempt to investigate household decision making in this context is Strand (2007). Introducing some notation is useful. Let: hwtp = the maximum amount a household would be WTP for the environmental good, so that all household members' utility levels are unchanged (which is typically unobserved in CV surveys); $\mathrm{iwtp}^{\mathrm{i}}$ = household member i's response to WTP question I; and $\mathrm{hwtp}^{\mathrm{i}}$ = household member i's response to WTP question II. Strand (2007) presents a collective model with no altruism assuming that the "true" hwtp is measured as the sum of the adult household members' individual WTP (suppressing the summation index: $\sum$ iwtp), i.e. each adult members' reply to WTP question I above[7]:

$$\mathrm{hwtp} \equiv \sum_{\mathrm{i}=1}^{\mathrm{m}} \mathrm{iwtp}^{\mathrm{i}} \equiv \sum \mathrm{iwtp}, \tag{1}$$

where the household has m (adult) members.[8]

The model of Strand (2007) assumes that the household allocates resources in efficient Nash bargains over a private and a household good (i.e. a good consumed commonly within the household). WTP question I is then interpreted as member i's willingness to give up units of the privately consumed good for the increase in the public good. Question II is member i's trade-off between the household good and the public good. A key result from the bargaining solution is that a member is generally willing to give up more of the household good than the private good:

$$\mathrm{hwtp}^{\mathrm{i}} > \mathrm{iwtp}^{\mathrm{i}} \tag{2}$$

The reasons for this is that only a share of a common budget increase can be spent on the private good and the marginal consumption values for both goods have to be equal in optimum for each member. Note that (2) does not arise from altruism, the commonly held view. Further, an implication of the model is that member i generally will misrepresent hwtp, in his answer to question II, i.e.

$$\mathrm{hwtp}^{\mathrm{i}} \neq \sum \mathrm{iwtp} \tag{3}$$

This result arises because the household members generally have different marginal valuations of the public good (in terms of the household good). Higher marginal valuation for the member asked than for the other members implies $\mathrm{hwtp}^{\mathrm{i}}$ higher than the sum of iwtp, which by definition (1) is hwtp. The converse is true for lower marginal valuation. However, importantly for practical CV research, Strand (2007, p. 541) argues that: "In a large random sample of households, such individual valuations should on average represent the respective

[7] In contrast to Strand, Quiggin (1998) and Bateman and Munro (2006) consider hwtp to be WTP obtained by consensus if adult household members are asked question II jointly. However, this interpretation makes implicit assumptions about the household bargaining structure allowing a consensus to be reached in a survey setting. Further, Bateman and Munro (2006, p. 3) state that "In one treatment a randomly selected individual is chosen from the couple [. . .] *providing responses on behalf of the household*" [our italics]. This is what they call "individual WTP", in our terminology $\mathrm{hwtp}^{\mathrm{i}}$, which they sum over the two partners to aggregate hwtp. In our setup, this would only make sense for $\mathrm{iwtp}^{\mathrm{i}}$ (Eq. 1). However, their imprecision may be due the income pooling model assumed in their study, in which the distinction between WTP questions I and II becomes immaterial (see Eq. 5).

[8] Strand assumes in his basic model two adult household members only, but his results extend to more than two members, so we take the general case here.

households correctly, only provided answers are truthful". In other words the true, unobservable mean household WTP in a (large) sample ($\overline{\mathrm{hwtp}}$) is equal to the observed mean response to question II ($\overline{\mathrm{hwtp}^{\mathrm{i}}}$):

$$\overline{\mathrm{hwtp}} = \overline{\mathrm{hwtp}^{\mathrm{i}}} = \overline{\mathrm{m}} \times \overline{\mathrm{iwtp}^{\mathrm{i}}} \tag{4}$$

Given the definition of hwtp in (1), the second equality in (4), where $\overline{\mathrm{m}}$ is average (adult) household size and $\overline{\mathrm{iwtp}^{\mathrm{i}}}$ average response to question I in the sample from one random household member from each household,[9] should also follow by approximation for large samples. If (4) is supported, the implication is that asking WTP questions I or II should be immaterial to the welfare estimate, as long as the aggregation is done according to the chosen unit.

Within a fairly general framework that does not depend on the type of household allocation model, Munro (2005) shows that if and only if the household members pool income, $\mathrm{hwtp}^{\mathrm{i}}$ and $\mathrm{iwtp}^{\mathrm{i}}$ will always be equal, i.e. the distinction between response units introduced with WTP questions I and II is unnecessary:

$$\mathrm{iwtp}^{\mathrm{i}} = \mathrm{hwtp}^{\mathrm{i}} = \mathrm{hwtp} = \mathrm{hwtp}^{\mathrm{j}} = \mathrm{iwtp}^{\mathrm{j}} \tag{5}$$

where members i and j are from the same household. The result follows directly from the properties of the standard indirect utility function once the functions depend on the sum of both individuals' income. In this case, summing over $\mathrm{iwtp}^{\mathrm{i}}$ for the individuals in the sample would grossly exaggerate hwtp, in contrast with (4), as these are really representations of hwtp.

Finally, in a meta-analysis of CV studies of non-timber benefits Lindhjem (2007) finds a counter-intuitive result from what we have discussed above:

$$\mathrm{hwtp}^{\mathrm{i}} < \mathrm{iwtp}^{\mathrm{i}} \tag{6}$$

He suggests that this result may be due to the fact that reference to individual or household in the WTP question triggers different "mental accounts" (Thaler 1999; Li et al. 2005) or "psychological purses" (Webley 1995) from which the payment for the environmental good is drawn. These frequently observed psychological phenomena make money non-fungible in practice. Further, this literature finds that the degree of financial integration within a household, from complete income pooling to separate finances, strongly influences household consumption decisions in general (Pahl 1995) and household versus individual WTP in particular (Delaney and O'Toole 2008). Being married, having children and being female are factors that suggest people respond more as households than as individuals. Further, bargaining strength between members in a household, typically measured as relative income of partners, has been shown to influence household consumption decisions in various ways (Dosman and Adamowicz 2006). Strand (2007) shows that the difference in (2) above will be larger for members with low bargaining power. The CV literature also suggests there are many respondent characteristics, often not directly derived from standard economic theory that may explain variation in WTP (e.g. difference between women and men, with or without children etc). Some of these variables may also help our understanding of differences between household and individual WTP.

[9] Equality should also hold if $\mathrm{iwtp}^{\mathrm{i}}$'s are added up with respondent-level household sizes instead of average household size in the sample. Note that only in two special cases will $1/\mathrm{m}\sum_{\mathrm{i}=1}^{\mathrm{m}} \mathrm{hwtp}^{\mathrm{i}} = \mathrm{hwtp}$, i.e. the error in representation of household WTP for all members will in general not average out within the same household.

3 Testing Procedure, Survey Design and Data

3.1 Testing Procedure

The theoretical predictions and empirically derived expectations discussed above were tested using a 2 × 2 split sample CV design. In sample A the respondent first got WTP question II ($hwtp^i$) and then prompted to instead think about personal WTP she got WTP question I ($iwtp^i$). After the second WTP question, the respondent was automatically directed in the survey to a question offering 4–6 reasons for why $hwtp^i$ was higher, lower or the same as $iwtp^i$ (see next section). The design was the same in sample B, except the order of the WTP questions and the way the prompt was phrased were reversed. The design allowed us to investigate the hypotheses discussed in the previous section (see Table 1 for summary of these) comparing mean WTP within and between the two samples. Note that we cannot test Eq. 3 directly with this design since we only collect responses from one random representative from each household. Giving both questions I and II to the same respondents allows us to model the relationship between $hwtp^i$ and $iwtp^i$ using explanatory variables for household and respondent characteristics.

3.2 CV Survey Design and Environmental Commodity

The data were collected from an Internet survey as part of a large multi-mode CV survey of forest protection in Norway. Currently ca. 1.4% of the productive forest area is protected, which according to biological assessments is too little to protect representative parts of forest habitats and endangered biodiversity. There are therefore plans to increase the level of protection, which can be assumed primarily to yield non-use values. A professional polling firm collected the data in the autumn of 2007 from a pre-recruited nation-wide panel of respondents. The panel is informed that surveys should be answered alone, so there is a higher degree of control over who actually answers than can be expected from mail surveys. The

Table 1 Summary of testable hypotheses of mean individual and household WTP[a]

	Within samples (k = A,B)	Between samples	Reference
H1	$\overline{hwtp^i_k} > \overline{iwtp^i_k}$	$\overline{hwtp^i_A} > \overline{iwtp^i_B}$	Strand (2007). Equation 2, Basic model
H2a[b]	$\overline{hwtp^i_k} = \overline{m_k} \times \overline{iwtp^i_k}$	$\overline{hwtp^i_A} = \overline{m_B} \times \overline{iwtp^i_B}$	Strand (2007). Equation 4, Response bias evens out in large samples
H2b[c]	$\overline{hwtp^i_k} = \frac{1}{n_k} \sum_{i=1}^{n_k} m^i_k \times iwtp^i_k$	$\overline{hwtp^i_A} = \frac{1}{n_B} \sum_{i=1}^{n_B} m^i_B \times iwtp^i_B$	
H3	$\overline{hwtp^i_k} = \overline{iwtp^i_k}$	$\overline{hwtp^i_A} = \overline{iwtp^i_B}$	Munro (2005). Equation 5, Income pooling makes units equal
H4	$\overline{hwtp^i_k} < \overline{iwtp^i_k}$	$\overline{hwtp^i_A} < \overline{iwtp^i_B}$	Lindhjem (2007). Equation 6, Explorative, based on mental accounting

Note: [a] Deriving the hypotheses it is reasonably assumed that the relationships discussed on the respondent level extend to means of samples. [b] It is somewhat artificial to test this hypothesis between samples for the second WTP question as people are expected to anchor their response to the first WTP question. [c] $iwtp^i$ is multiplied with respondent-level household size (m^i_k) and summed over the sample size (n_k)

survey was designed following similar forest protection surveys well tested and tried in the Nordic context (see Lindhjem 2007), and recent best-practice guidelines in the CV field (e.g. Bateman et al. 2002; SEPA 2006). The instrument went through thorough testing in focus groups and two small pilots (using both internet and personal interviews).[10]

The survey first included questions about use of government money for various ends to put the environmental good into a wider context, before focusing on the respondent's use of forests and attitudes towards their perceived biological and aesthetical state. Information was then presented about number and types of species, and the interplay between forestry practices, protection and evolution of ecosystem functions and biodiversity in forests. Six colour photos of endangered species and forest habitats were shown as well as pie and bar charts of number and percentage of species in different habitats, including forests. The information was broken up with questions to activate the respondent and encourage response. Hard-to-avoid technical terms in the text (such as "biodiversity", "nature reserve", etc.) were explained in boxes that would pop up when the cursor touched underlined words. Respondents were then presented current forest protection policy (status quo) and future plans. The environmental commodity was specified as two forest protection plans of either an increase to 2.8% (doubling) or to 10% (possible long-term target), presented together. The text was supplemented with digital, zoomable colour maps of current and future forest reserves, and a table giving information about the size of new reserves, location of reserves, and the improvements in the living conditions for main groups of species. The biological information was provided by a team of leading biologists in Norway, and checked by foresters to ensure a balanced presentation of the status quo and future plans.

3.3 Household and Individual WTP Elicitation and Follow-up Probes

The basis for the comparison of $hwtp^i$ and $iwtp^i$ was the 2.8% protection plan. After the information about the two plans, the respondent was given the following text (the text in italics was varied between Samples A and B):

"We ask you now to consider how much the two alternative plans are worth for *your household/you*. Think through carefully how much the 2.8% plan is worth compared to the current situation, before you give your final answer to the next question. Try to consider what would be a realistic annual amount given *your/your household's* budget. *You/Your household* must choose whether to spend the amount on the forest conservation plan, or on other things. What is the most *your household/you* almost certainly *is/are* willing to pay in an additional annual tax earmarked to a public fund for increased forest conservation from today's level of 1.4% to 2.8% of the productive forest area? Choose the highest amount, if anything, *your household/you* almost certainly will pay".

People could then indicate their maximum WTP in a payment card (PC) in the form of a drop-down menu with a non-linear scale containing 24 amounts (ranging from 0 first to NOK 15000[11]), including "don't know" (at the end). The amounts where chosen on the basis of previous CV studies (e.g. Lindhjem 2007). PC was chosen as response format over dichotomous choice, to avoid yea saying (at the expense of theoretical incentive compatibility) (Boyle 2003). PC also lends itself nicely to the drop-down menu format very familiar to internet-users. The payment vehicle (an earmarked tax to a forest protection fund) was chosen because it is response unit neutral, for example compared to an income tax (a potential

[10] The full survey questionnaire is available in English from the author upon request.

[11] There was also an option to choose "more than 15000", in which case a box would pop up where the exact amount could be specified.

problem in e.g. Hasler et al. 2008), is realistic and reduces people's scepticism that the money would not be spent on forest protection. The typical budget reminder, included in most modern CV studies, referred either to "your (personal)" or "your household's budget". Following the first WTP question, the respondent would get the following prompt, before getting the second WTP question with response unit changed: "We now ask you if it matters for your willingness to pay if you state it for yourself/on behalf of your household or on behalf of your household/for yourself."

After the second WTP question, respondents were automatically taken to a follow-up question asking whether a number of stated reasons were important, not important or not relevant for their response to the two questions. The suggested reasons allow the respondent to express her considerations regarding her (and her partner's) preferences (for herself, family), budget (individual, household, common), and role in the household (e.g. usually paying household expenses). The respondent could also state openly other reasons that may have been important. The details of these are given in the Appendix, and the descriptive statistics of peoples' responses given in the next section. The rest of the CV survey followed standard procedure, probing into why people answered zero or positive, checking their understanding and perceived realism of the scenario and WTP questions. The final part collected socio-economic background information, which was merged with existing web panel information about the respondents and their households.

4 Results and Analysis

4.1 Mean Individual versus Household WTP Between and Within Samples

The response rates were 72% and 69% for Samples A and B, respectively. Before estimating mean WTP the dataset was cleaned. Around 10% of respondents in both samples (and for both WTP questions) chose the "don't know" option in the drop-down menu for the PC, while between 20% and 25% chose zero. There were no significant differences in these answers across samples. Since our main aim here is to investigate the relationship between household and individual WTP and people's stated reasons, all zeros and "don't know" responses were therefore removed. Further, respondents from one-person households were taken out. This procedure reduced the samples from around 400 to 240 observations, each. A comparison of mean values of sample characteristics indicated no immediate reason for applying weighting procedures or using covariates in the estimation of mean WTP (see Table 6 in the next section). Since the stated WTP amounts had a skewed distribution with a long right tail, a log-transformation of WTP was applied. Mean WTP for the interval PC data for the two samples and WTP questions were estimated following standard procedures given in Cameron and Huppert (1989) (see Table 2).[12]

The response to household WTP (Euro 172) is higher than to the individual WTP (Euro 154) between the samples for the first question as expected from theory, confirming the common view in the CV literature. However, the difference seems not to be significant at the 5% level. We ran a likelihood-ratio test to check statistical significance, see (7):

$$q = -2\left[\log L_{\mathrm{PooledAB}} - (\log L_A + \log L_B)\right] \sim \chi^2(\mathrm{d.f.}) \quad (7)$$

[12] We compared a normal and lognormal model with a simple non-parametric survival function using the lower bound of the PC intervals. The lognormal model showed a better fit. Mean WTP from this model is given by $E(WTP) = \exp(a + \sigma^2/2)$, where a and σ are the estimated parameters from the lognormal model.

Table 2 Mean annual individual and household WTP (std. error), the two samples (Euros)

WTP Quest.	Sample A		Sample B	
	Mean	95% CI	Mean	95% CI
1st	$\overline{hwtp^i_A} = 172$ (16)	(141, 203)	$\overline{iwtp^i_B} = 154$ (17)	(121, 188)
2nd	$\overline{iwtp^i_A} = 147$ (13)	(121, 173)	$\overline{hwtp^i_B} = 237$ (28)	(182, 292)
N	239		234	

Note: Estimated using interval regression in STATA 9.2. Confidence intervals were calculated using 1000 bootstrap draws with replacement, following Efron (1997). 1 Euro = 8.07 Norwegian Kroner at time of study

Table 3 Respondents who answered higher, the same or lower on household WTP compared to individual WTP question (%)

	Sample A	Sample B
Higher ($hwtp^i > iwtp^i$)	32.6	52.9
Same ($hwtp^i = iwtp^i$)	59.4	44.4
Lower ($hwtp^i < iwtp^i$)	7.9	2.5
Total	100%	100%

where $logL_A$ and $logL_B$ refer to the log likelihood values of from the estimated models for individual samples, and $logL_{PooledAB}$ is the likelihood value for a pooled model. Running the pooled model without a sample dummy, yields a test static ($\hat{q}$) of 6.96, which allow us to reject that both parameters are equal at the 3% level. However, running the same model with a sample dummy yields $\hat{q} = 2.12$, which means we cannot reject that the standard errors are the same at the 10% level (i.e. the samples can therefore be pooled). The dummy is significant at the 2.8% level, indicating confirmation of the one-sided hypothesis that $\overline{hwtp^i_A} > \overline{iwtp^i_B}$. However, an extended bootstrap (10,000 draws with replacement) from each of the sample distributions combined with a simple non-parametric test of means indicates ca. 80% $\overline{hwtp^i_A} > \overline{iwtp^i_B}$ and 20% $\overline{hwtp^i_A} \leq \overline{iwtp^i_B}$. This means that we can reject the hypothesis that household WTP is higher than individual WTP between the samples for the first WTP question.[13] The confidence intervals estimated around the means in Table 2, also indicate that equality cannot be rejected at the 5% level. For the second WTP question where response units are reversed, respondents in Sample A generally reduce their bids (mean Euro 147) while respondents in sample B increase their bids (mean Euro 237), as expected. However, this difference is not symmetric, as can be seen from Table 3.

Around 53% of sample B increase their bid from $iwtp^i$ to $hwtp^i$, while only 33% reduce their bids from $hwtp^i$ to $iwtp^i$ in sample A. The reason for this stickiness downwards is not immediately clear. It is possible that some people in Sample A interpreted the $hwtp^i$ question as an $iwtp^i$ question (despite the unit being explicitly stated), and therefore saw no reason to reduce their bid in the second question (similar to what was found in Delaney and O'Toole 2008).[14] Drawing parallels to the extensive embedding debate in CV, the whole of a good

[13] This arises because sample B's distribution has a lower mean, but a higher standard error than sample A's distribution. The effects partly outweigh each other in the formula for mean WTP (see previous footnote).

[14] It was not possible directly to control whether respondents would take the time and trouble to go back a page in the Internet survey to change their response to the first WTP question. However, we included a control sample that only got one $hwtp^i$ question. Mean WTP from this sample was statistically identical with that of Sample A, indicating that this practice was not prevalent in the survey.

Table 4 Individual WTP scaled up by measures of household size (Euros)

Mean adjustment factor	Sample A	Sample B
Average household size in sample[a]	340	382
Adult members of each respondent's household	345	387
N[b]	224	218

Note: [a] Mean household sizes Sample A: 2.29, Sample B: 2.41 (excluding one-person households). [b] The sample sizes are lower than in Table 3 since the respondent database had a few missing household size values. This is also the reason why means from Table 3 scaled up with household sizes here are not exactly equal to 340 and 382

Table 5 Summary of empirical results by hypothesis (k = Samples A, B)

	Within samples	Test result	Between samples	Test result
H1	$\overline{\mathrm{hwtp}_k^i} > \overline{\mathrm{iwtp}_k^i}$	Supported A&B	$\overline{\mathrm{hwtp}_A^i} > \overline{\mathrm{iwtp}_B^i}$	Rejected
H2a	$\overline{\mathrm{hwtp}_k^i} = \overline{\mathrm{m}_k} \times \overline{\mathrm{iwtp}_k^i}$	Rejected A&B	$\overline{\mathrm{hwtp}_A^i} = \overline{\mathrm{m}_B} \times \overline{\mathrm{iwtp}_B^i}$	Rejected
H2b	$\overline{\mathrm{hwtp}_k^i} = \frac{1}{\mathrm{n}_k} \sum_{i=1}^{\mathrm{n}_k} \mathrm{m}_k^i \times \mathrm{iwtp}_k^i$	Rejected A&B	$\overline{\mathrm{hwtp}_A^i} = \frac{1}{\mathrm{n}_B} \sum_{i=1}^{\mathrm{n}_B} \mathrm{m}_B^i \times \mathrm{iwtp}_B^i$	Rejected
H3	$\overline{\mathrm{hwtp}_k^i} = \overline{\mathrm{iwtp}_k^i}$	Rejected A&B	$\overline{\mathrm{hwtp}_A^i} = \overline{\mathrm{iwtp}_B^i}$	Supported
H4	$\overline{\mathrm{hwtp}_k^i} < \overline{\mathrm{iwtp}_k^i}$	Rejected A&B	$\overline{\mathrm{hwtp}_A^i} < \overline{\mathrm{iwtp}_B^i}$	Rejected

(i.e. hwtp^i in our case) is typically valued more when valued after a smaller part of the good (i.e. iwtp^i) in a sequence than before (see e.g. Clark and Friesen 2008; Powe and Bateman 2003). This phenomenon is often termed sequencing or ordering effects. In our case it may offer an explanation why we observe a higher hwtp^i in sample B, but not why responses in sample A are sticky downwards. However, pairwise t-tests on the difference of bootstrapped mean WTP values between WTP questions I and II within each sample were conducted (for brevity not displayed here), strongly confirming $\overline{\mathrm{hwtp}_k^i} > \overline{\mathrm{iwtp}_k^i}$ for both samples at the 1% level.

To check hypotheses H2a & b, we also scaled individually stated WTPs with the number of adult household members (> 15 years) from each respondent's household and the average household size in the samples (see Table 4). For the former case, high and low limits from the PC data were multiplied by the household size, and interval regressions rerun. For the latter case, mean individual WTP was re-estimated and scaled up by the constant average household size for each sample.

Interestingly, the two methods to adjust individual WTP with household size yield very similar results. Compared to the household WTP for sample A in Table 2, adjusting individually stated WTP estimates to represent household WTP, yields significant overvaluation. $\overline{\mathrm{hwtp}_A^i}$ is about half of the individual WTP adjusted by household size from sample B, contrary to the expectation in Eq. 4. Also for the second WTP question where the majority end up answering $\mathrm{hwtp}^i > \mathrm{iwtp}^i$, there is no equality of mean hwtp^i and mean adjusted iwtp^i within samples. Finally, we can summarise our empirical results in Table 5.

Rejection of H1 and H2 between samples follows from the support to hypothesis H3. However, household WTP is significantly higher than individual WTP within each sample (i.e. H1 supported), where rejection of H3 and H4 (but not H2) within samples logically follows.

4.2 Explaining the Relationship Between Individual and Household WTP

We now turn to trying to explain the observed relationship between individual and household WTP within the two samples. Some explanations are given by the respondents themselves, when stating in the follow-up question for each proposed reason whether it was important, not important or not relevant to their choice. Figures 1 and 2 sum up the results for the pooled samples, for $hwtp^i > iwtp^i$ and $hwtp^i = iwtp^i$, respectively. Full versions of statements respondents considered are given in the Appendix.

The by far most important reason why people state higher household WTP is that they have a larger budget at their disposal and therefore can pay more (80% for reason A1 in Fig. 1). This means that individuals do not seem to consider the income of adult co-members as part of their own budget constraint, i.e. income is not pooled. The second most important reason (49%) is that the estimated WTP of a partner is added (A5), which may be consistent with both a separate and shared economy in practice. The third reason (A2) is that respondents think about the household also when answering individually (i.e. no generous individual "mental account" distorts the expected relationship between $hwtp^i$ and $iwtp^i$). The fourth most important reason is that respondents consider children especially when answering the household WTP question (A3). This is an indication of altruistic preferences, which is the traditional view of why household WTP may be higher (but as discussed not a necessary condition in Strand (2007)). Only 33% state that they have just doubled the individual amount since they are from a two-adult household (A6). Finally, it is interesting to note that very few respondents have answered higher household WTP because their partner has a stronger preference (higher marginal valuation) for forest conservation than themselves (A4). There were no substantial differences in answers to this follow-up question between samples A and B. However, reasons A1 and A5 seemed to be more important for sample B, and A6 for sample A (all by ca. 10-percentage points). Some of the open answers (A7) pointed to the

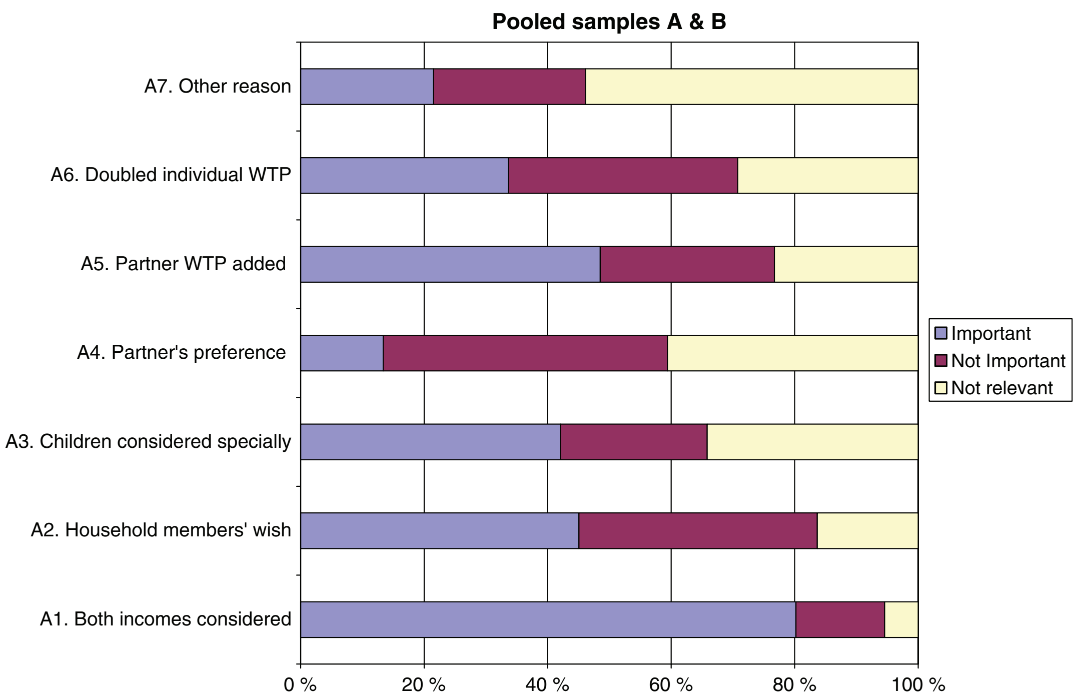


Fig. 1 Percentage of respondents rating given reasons for $hwtp^i > iwtp^i$ ($n = 202$)

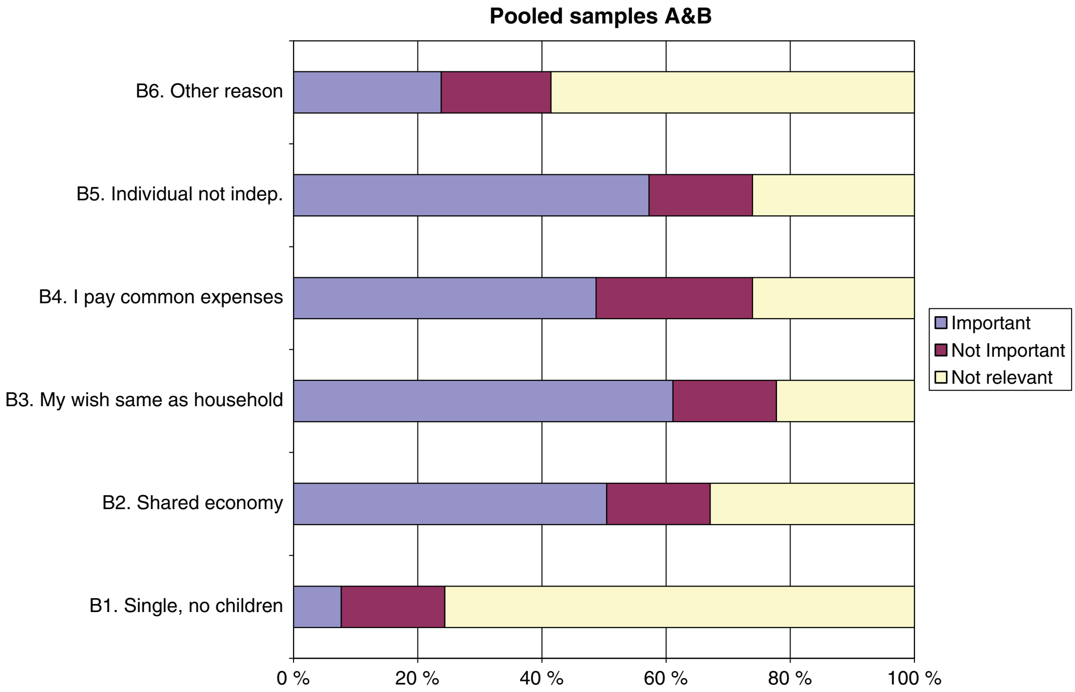


Fig. 2 Percentage of respondents rating given reasons for $hwtp^i = iwtp^i$ ($n = 234$) (*Note*: One-person households without children were taken out. Still, 7% had indicated that being single and having no children (B1) was important. This may be due to the fact that the CV data generally are more updated than the Internet panel information. Further, a few responses may be due to misunderstanding, in any case a low number giving us confidence in people's responses)

number of household members (with or without reference to use of forests) as important, that individual WTP was calculated as a percentage, that assumptions had been made about other members' WTP, and generally that a larger budget is available (supporting A1).

A similar battery of reasons where offered for $hwtp^i = iwtp^i$-respondents (see Fig. 2). The most important reason is that the respondent's wish concurs with what the household would collectively have decided (61%, reason B3 in Fig. 2). This is an indication of a unitary household model (as judged by the respondent), as is the reason almost as many states as important: taking household members into account even when answering as an individual (57%, B5). About 50% indicate as important that they have a shared economy with their partner (B2), 48% that they are responsible for paying household expenses and therefore that it does not matter for their WTP which response unit they are asked (B4). Reason B3 was indicated as more important for sample A than B (by 12-percentage points), while other reasons showed no big difference across samples. In their open statements, some respondents had noted that the amount they decided on the first question was "an appropriate amount" or "enough", i.e. that they saw no reason to increase or decrease from this level. This is a reason related to the embedding and scope insensitivity debate discussed above, however, in this case perhaps more a reflection that a "moral dump"—a suitable donation to a good cause—has been made. Hence, in our dataset there may be a few such responses, where "stickiness" ($hwtp^i = iwtp^i$) is not due to reasons related to household decision-making.

Some respondents also indicated that they made the decisions in their household, i.e. similar to the Mitchell and Carson's "household head", and therefore $hwtp^i = iwtp^i$. However, this stated reason may also be interpreted as an indication of high bargaining power. Only 7.9% in sample A and 2.5% in sample B (total of 25 respondents) chose $hwtp^i < iwtp^i$

(figure of stated reasons not displayed here). Around 50% stated as important that their personal budget was higher (reason C1 in Appendix), i.e. indicating "mental accounting" (or possibly separate finances). One respondent mentioned separate finances as important, another that it was easier to answer the individual question than assuming WTP for the other members.

We also posed the question of whether household and respondent characteristics, defined in Table 6, can explain the relationship between individual and household WTP. We do not rule out a priori that a range of respondent variables (e.g. sex, age, education, use, attitudes, etc.) often included in CV bid functions may also be important to explain the relationship between household and individual WTP. These variables may be considered explorative. Of household variables we included number of household members, presence of children, altruistic attitudes for WTP > 0 (latter two crude proxies for altruism), if children (>15 years) answered the survey, type of residence, whether grocery purchases are jointly planned, marital status (latter two crude proxies for economic integration), and the respondent's share of household income (common proxy for bargaining strength). The third and fourth columns in Table 6 indicate mean values and standard deviation, which are generally very similar between samples.

We chose a simple regression approach with a binary dependent variable of 1 if $\mathrm{hwtp}^{\mathrm{i}} > \mathrm{iwtp}^{\mathrm{i}}$ and 0 if $\mathrm{hwtp}^{\mathrm{i}} = \mathrm{iwtp}^{\mathrm{i}}$ estimated using a standard probit model, similar to the approach in Delaney and O'Toole (2004). The few respondents answering $\mathrm{hwtp}^{\mathrm{i}} < \mathrm{iwtp}^{\mathrm{i}}$ were excluded for simplicity. Models using the WTP ratio ($\mathrm{hwtp}^{\mathrm{i}}/\mathrm{iwtp}^{\mathrm{i}}$) or difference ($\mathrm{hwtp}^{\mathrm{i}} - \mathrm{iwtp}^{\mathrm{i}}$) as dependent variables instead of the binary variable, were specified and tested, but gave generally lower explanatory power. Being unfamiliar with the task, it is likely that respondents had a clearer idea about the direction than the exact magnitude of the difference between $\mathrm{hwtp}^{\mathrm{i}}$ and $\mathrm{iwtp}^{\mathrm{i}}$. Using a regression model we can control for the different characteristics that are underlying people's responses and hidden in Figs. 1 and 2. Results for the separate and pooled samples are displayed in Table 7. The models show reasonable fit to the data, but coefficient estimates should be interpreted with caution. For sample A, older people have significantly higher probability to state equality, indicating perhaps both for long relationships and for the older generation, the difference between the individual and the household gets increasingly blurred.

The other respondent variables are not significant for sample A. In sample B, men have a significantly higher probability than women of answering $\mathrm{hwtp}^{\mathrm{i}} > \mathrm{iwtp}^{\mathrm{i}}$, i.e. women may see income pooling as more natural than men. Individual income level has almost no effect on the probability. Interestingly, people who are certain not to visit the future forest reserves (variable "Nouse") have lower probability of separating between household and individual WTP (not significant for sample A). Non-users may be more likely to see the good in a broader family context (e.g. for bequest reasons) and therefore (mis)interpret the $\mathrm{iwtp}^{\mathrm{i}}$ question as a $\mathrm{hwtp}^{\mathrm{i}}$ question than users, in which case the WTP difference will be smaller (see also Delaney and O'Toole (2008)). Another explanation, if indeed respondents have clearly understood the response unit, is that a "moral dump" or donation has been made, which is less sensitive to unit of response (and, as is often found, to other key dimension of the survey instrument, such as scope of the good). People who favour a tax to pay for public goods ("Attax") also display smaller WTP differences.

For the household characteristics, only relative income is significantly negative through all three models. A higher share of household income reduces the probability of stating $\mathrm{hwtp}^{\mathrm{i}} > \mathrm{iwtp}^{\mathrm{i}}$, as expected. This result is consistent with the model of bargaining strength in Strand (2007), but can simply also be interpreted as an indication of separate finances as people's WTP is strongly correlated with their personal budget constraint. The variables on personal status ("Cohab" and "Married") do not pick up any consistent patterns across

Table 6 Explanatory variables and sample means (std. dev.) samples A and B

Variables	Definition	A	B
Respondent characteristics			
Gender[a]	Dummy: 1 if male, 0 if woman	0.504 (0.032)	0.512 (0.032)
Age[a]	Continuous: >15 years	41.2 (0.91)	41.7 (0.96)
Incomeind	Individual income 2006, Norwegian Kroner 25000 intervals	316912 (11903)	312948 (13597)
Eduhigh[a]	Dummy: 1 if >4 years university education; 0 if mid-education	0.100 (0.019)	0.119 (0.021)
Edulow[a]	Dummy: 1 if only primary education; 0 if middle education	0.075 (0.017)	0.085 (0.018)
Owner	Dummy: 1 if forest owner; 0 if not	0.306 (0.029)	0.235 (0.027)
Member	Dummy: 1 if member of nature organisation; 0 if otherwise	0.025 (0.010)	0.038 (0.012)
Use	Dummy: 1 if forest used for recreation last 12 months; 0 if not	0.924 (0.017)	0.935 (0.016)
Highuse	Interaction variable: 1 if >15 times in forest last month and "Use" = 1; 0 otherwise	0.063 (0.015)	0.136 (0.022)
Nouse	Dummy: 1 if sure not to use proposed forest reserves, 0 if otherwise	0.172 (0.024)	0.179 (0.025)
Attax[a]	Dummy: 1 if agree that high taxes ensure public goods; 0 if otherwise.	0.138 (0.02)	0.16 (0.02)
Altruism	Dummy: 1 if respondent indicated as reason for WTP > 0 that other people can enjoy old growth forests; 0 if otherwise	0.277 (0.029)	0.238 (0.027)
Household characteristics			
Relinc	Individual income as share of household income	0.570 (0.015)	0.545 (0.017)
Childdum[a]	Dummy: 1 if children <15 years of age in household; 0 if otherwise	0.231 (0.027)	0.227 (0.028)
Childresp[a]	Dummy: 1 if child (>15 years) answered survey; 0 otherwise	0.077 (0.017)	0.080 (0.018)
Married[a]	Dummy: 1 if married; 0 if previously married/single	0.596 (0.032)	0.607 (0.032)
Cohabit[a]	Dummy: 1 if cohabitants; 0 if previously married/single	0.236 (0.027)	0.209 (0.027)
Grocery[a]	Dummy: 1 if divided responsibility, grocery purchase; 0 if otherwise	0.454 (0.032)	0.446 (0.033)
House[a]	Dummy: 1 if detached house; 0 if otherwise	0.592 (0.032)	0.638 (0.032)
Hhldmem[a]	Number of adults and children (1–4, 5 or more)	2.98 (0.071)	3.03 (0.071)
N^{b}		239	234

Note: [a] Variable information taken from Internet panel of respondents. Other variables are from the CV survey. [b] Some averages based on reduced sample. No weighting was conducted between samples

models. Increasing number of household members (i.e. more than 2) reduces the probability of answering $hwtp^i > iwtp^i$ through all models (only significant for A).[15] This suggests that once children are involved, the household is more tightly integrated, resulting in smaller differences between household and individual WTP. This result runs contrary to the common argument that altruism drives a wedge between individual and household WTP. If the alleged effect of altruism is present, it may be outweighed by the higher degree of income pooling

[15] Mere presence of children in the household (i.e. "Childdum") did, however, not have a consistent and significant effect on response probability.

Table 7 Probit models on $hwtp^i > iwtp^i$ (Y = 1) or $hwtp^i = iwtp^i$ (Y = 0) for separate and pooled samples

Independent variables	Sample A		Sample B		Pooled sample (A + B)	
	Coefficient	$Z-$score	Coefficient	$Z-$score	Coefficient	$Z-$score
Dummy for sample (WTP question order)					0.482***	3.69
Respondent variables						
Gender	−0.158	−0.73	0.387*	1.73	0.126	0.85
Age	−0.025**	−2.34	0.004	0.44	−0.006	−0.94
Incomeind	0.000	1.48	0.000**	1.96	0.000***	2.59
Eduhigh	0.166	0.55	0.265	0.89	0.193	0.96
Edulow	0.412	0.96	0.670	1.36	0.534*	1.76
Owner	−0.370	−1.62	0.127	0.55	−0.013	−0.09
Member	−0.439	−0.51	0.249	0.44	0.214	0.51
Use	0.521	1.44	−0.040	−0.10	0.259	0.99
Highuse	−0.404	−0.97	−0.120	0.43	−0.145	−0.67
Nouse	−0.073	−0.29	−0.622**	−2.38	−0.405**	−2.34
Altruism	0.163	0.71	0.354*	1.66	0.198	1.32
Attax	−0.095	−0.34	−0.503*	−1.90	−0.375**	−2.00
Household variables						
Relinc	−1.362***	−2.16	−2.209***	−3.68	−1.653***	−4.01
Childdum	0.453	1.27	−0.251	−0.81	−0.064	−0.29
Childresp	1.150*	1.66	0.253	0.39	0.611	1.38
Married	0.842	1.59	−0.200	−0.48	0.202	0.65
Cohab	0.945*	1.86	−0.151	−0.36	0.348	1.13
Grocery	−0.173	−0.84	0.005	0.03	−0.048	−0.36
House	0.054	0.26	−0.242	−1.04	−0.103	−0.71
Hhldmem	−0.374**	−2.50	−0.022	−0.18	−0.104	−1.14
Constant	0.948	1.11	0.944	1.11	0.273	0.48
Log likelihood	−120.19		−129.82		−260.84	
Pseudo R^2	0.1329		0.1419		0.1228	
N	214		219		433	

Note: *, **, *** Significance at 10%, 5%, and 1% levels, respectively

that emerges once relationships mature. A few children (>15 years) answered the survey, and these generally had a higher probability of answering $hwtp^i > iwtp^i$ ("Childresp") as expected since they have lower income than their parents. A final point to note is that the dummy on sample in the pooled model is significantly positive indicating a higher probability of $hwtp^i > iwtp^i$ in sample B, confirming our earlier results. The results of the models give us some degree of confidence in the validity of the data, and supplement the insights provided by respondents in their stated reasons for differences in household and individual WTP.

5 Discussion and Conclusions

The practical implications for aggregate welfare estimates of the choice of response unit for WTP—household or individual—in CV surveys has been largely ignored in the literature to date. We demonstrate that the empirical consequence may be substantial noise or bias in welfare estimates. In our CV survey of forest protection in Norway, we find that people do not state a significantly different WTP when asked on behalf of the household than as individuals between two samples. Aggregating WTP over individuals in this case more than doubles the total welfare change compared to mean stated household WTP. This means that scaling individual WTP values with mean household size to get household WTP, as is commonly done, inflates welfare estimates. This result runs contrary to the collective household decision model of Strand (2007), which predicts equality between mean household WTP and scaled-up individual WTP in large samples. Since we exclude single-person households and true zero responses, the demonstrated distortion in welfare estimates is somewhat higher than in a full sample.

When people are prompted in the second WTP question to answer for the other response unit, an average of 43% decide to state higher household WTP than individual WTP, while 52% state the same WTP. More people state higher household WTP if they have been asked individual WTP first, i.e. people tend more easily to increase their bids than reduce them. Mean household WTP within the same samples is found to be significantly higher than individual WTP on the 1% level. 80% of respondents state as an important reason for this result that they have a larger budget at their disposal when asked household WTP. There are few indications that altruism, though imperfectly measured in this study, may be important in explaining that household WTP is higher than individual WTP within samples—the commonly held view in the literature. Instead, degree of financial integration and relevant budget constraints seem to be more important.

Our study is a first attempt to investigate the empirical differences between individual and household WTP for one type of environmental good in a particular CV setting—where both types of questions could meaningfully be asked. The degree to which our results can be generalised to other CV studies, or stated preference research more generally, types of goods, response formats, survey modes etc. is uncertain. For other environmental goods of a more quasi-private nature or where use values dominate, individual WTP questions may be the appropriate choice, as this may be what resonates best with respondents' interpretation. However, more empirical research is undoubtedly necessary within stated preference valuation, to advance the theory of intra-household resource allocation and to test it empirically. Recent research interviewing partners and households together in choice experiment settings (such as in Bateman and Munro 2005, 2006; Beharry and Scarpa 2006) are important contributions. However, since this approach will never really be a practical option due to excessive costs, empirical work should inform stated preference design, where random individuals typically are asked, with the aim to reduce response unit distortion in welfare estimates as much as possible. An important point made by Delaney and O'Toole (2006, 2008) is that people's self perceived agency—i.e. their interpretation of the unit of the WTP question notwithstanding explicit reference to "household" or "individual"—may vary depending on household and respondent characteristics. We think the risk of such misunderstandings to some extent may be alleviated by carefully designing the survey to be consistent with the chosen response unit. Not only wording of the actual WTP question, but the type of good (e.g. extent of non-use values), the payment vehicle (e.g. household tax versus income tax), budget reminders and scenario descriptions, and even the survey mode,[16] may give conflicting cues as to the intended agency of the respondent. Understanding respondent agency more broadly in the CV context, it is clear from the literature that when non-use values and

[16] People may for example view their agency differently depending on whether they are asked during an intercept at a forest site or in a shopping mall compared to filling in a paper questionnaire on their own, family kitchen table.

altruistic concerns are part of the valuation context—as is the case in our survey—individuals may also respond in accordance with moral or social norms or as citizens (Brekke et al. 2003; Nyborg 2000), rather than as neo-classical consumers (Sagoff 1988). Understanding how the individual and household roles are influenced by and influence these other roles of respondents, beyond the scope of this paper, would be a fruitful avenue for further research.

Our results indicate that people may need more information to state their WTP reliably for the household or as individuals. We think that our approach would have to be improved in different directions and thoroughly tested, before a clear recommendation on survey design is given. Some avenues of potential improvement are clear. Since there are parallels to the embedding debate in CV, some sort of advance disclosure of the fact that both individual and household questions will be asked (which we did not include) may help respondents think, improve consistency and reduce "surprise" effects (as recommended in scope tests—see e.g. Bateman et al. 2004). People may also be given a chance to revise WTP ex post. Since our findings show that people seem not to have a clear idea about how they should interpret the individual and household questions, "cheap talk" (e.g. explaining why there may be differences) and definition of "household" may be useful in clarifying the intended response unit. Further, it may prove useful to investigate whether our results are invariant to the use of a dichotomous response format rather than payment card (for an early test in this context see Bateman and Brouwer (2006)). These suggested approaches will have to be carefully tested not to introduce other, unintended biases, a common experience in the CV history. However, our results suggest that response unit distortions may be sufficiently problematic to need fixing. Though we have not uncovered the "true" (actual) household WTP in our survey and more research is needed, asking household WTP of random individuals for environmental goods seems to be the conservative approach that should be followed, even though it may lead to underestimation of welfare change in some situations.

Appendix

See Table 8.

Table 8 Given reasons in survey for why respondents chose higher, the same or lower household or individual WTP[a]

How important were the following reasons for you stating a higher/same/lower amount on behalf of your household than/as for yourself?
Answers: Cross "Important, Not Important, or Not Relevant"

Higher	Same	Lower
A1. I took both incomes into account when I was asked on behalf of the household	B1. I am single and have no children, so there is no difference	C1. I thought about my individual budget and can pay more than if I have to take my household into consideration
A2. I take the household members' wishes regarding increased forest conservation into account even if I consider willingness to pay for myself alone[b]	B2. My partner and I have a shared economy, so it does not matter if I am asked personally or on behalf of my household	C2. My partner is against more forest conservation, so I adjusted for that

Table 8 continued

How important were the following reasons for you stating a higher/same/lower amount on behalf of your household than/as for yourself?
Answers: Cross "Important, Not Important, or Not Relevant"

Higher	Same	Lower
A3. I especially consider the children when asked on behalf of the household	B3. What the household collectively would have decided concur with my wish	C3. I am normally not the one paying for our household expenses, so I chose a lower amount on behalf of my household
A4. My partner is more interested in forest conservation than I am, so I adjusted for that	B4. I am normally the one paying our household expenses, so in practice there is no difference if I am asked personally or behalf of my household	C4. We have a tight budget for household expenses, but my personal budget is more generous
A5. I added what I think my partner would be willing to pay	B5. I take my household members into account even if I consider willingness to pay for myself alone	
A6. I doubled my individual amount since we are two adults in the household		
Other reasons that were important? Specify:______		

Note: [a] Respondents would get automatically directed to one of these item batteries in the web-survey depending on their answers to the hwtp and iwtp questions. For brevity all three are compiled in one table here. [b] There was some ambiguity in the interpretation of this reason. The intended meaning is that the respondent does not focus on a more generous individual budget "mental account" for $iwtp^i$, but thinks about the whole household and therefore goes up from $iwtp^i$ to $hwtp^i$

Acknowledgements We would like to thank Jon Strand, Olvar Bergland, Anna Bartczak, Jette Bredahl Jacobsen, Mark Morrison and Laura O. Taylor for constructive comments. Funding from the Norwegian Research Council AREAL program grant 173230/I10 is gratefully acknowledged.

References

Bateman I, Munro A (2003) Non-cooperative decision-making and measures of household surplus. CSERGE Working Paper EDM 03-12

Bateman I, Munro A (2005) An experiment on risky choice amongst households. Econ J 115:C176–C189. doi:10.1111/j.0013-0133.2005.00986.x

Bateman I, Munro A (2006) Household versus individual valuation: what's the difference? Royal Holloway, University of London, Discussion Papers in Economics number 06/02.

Bateman IJ, Brouwer R (2006) Consistency and construction in stated WTP for health risk reductions: a novel scope-sensitivity test. Resour Energy Econ 28:199–214. doi:10.1016/j.reseneeco.2006.06.001

Bateman IJ, Carson RT, Day B, Hanemann WM, Hanley N, Hett T, Jones-Lee M, Loomes G, Mourato S, Ozdemiroglu E, Pearce DW, Sugden R, Swanson T (2002) Economic valuation with stated preference techniques: a manual. Edward Elgar Publishing, Cheltenham, 480 pp

Bateman I, Cole M, Cooper P, Georgiou S, Hadley D, Poe GL (2004) On visible choice sets and scope sensitivity. J Environ Econ Manage 47:71–93. doi:10.1016/S0095-0696(03)00057-3

Becker GS (1973) A theory of marriage: Part I. J Polit Econ 81:813–846. doi:10.1086/260084

Becker GS (1981) A treatise on the family (Enlarged edition). Harvard University Press, Cambridge

Beharry N, Scarpa R (2006) Joint vs. separate decisions by couples in choice experiments: the case of coastal water quality in Tobago. 3rd World Congress of Environmental and Resource Economists, Kyoto, Japan, July 2006

Bergstrom TC (2003) Benefit cost analysis and the entanglements of love. In Proceedings of session II: household decision making, USEPA Workshop, October 20–21. Washington DC

Boyle KJ (2003) Contingent valuation in practice. In: Champ PA, Boyle KJ, Brown TC (eds) A primer on nonmarket valuation. Kluwer Academic Publishers, Dordrecht
Brekke KA, Kverndokk S, Nyborg K (2003) An economic model of moral motivation. J Public Econ 87:1967–2003. doi:10.1016/S0047-2727(01)00222-5
Cameron TA, Huppert DD (1989) OLS versus ML estimation of non-market resource values with payment card interval data. J Environ Econ Manag 17(3):230–246. doi:10.1016/0095-0696(89)90018-1
Carson RT, Flores NE, Meade NF (2001) Contingent valuation: Controversies and evidence. Environ Resour Econ 19(2):173–210. doi:10.1023/A:1011128332243
Clark J, Friesen L (2008) The causes of order effects in contingent valuation surveys: an experimental investigation. J Environ Econ Manag 56:195–206. doi:10.1016/j.jeem.2007.12.005
Delaney L, O'Toole F (2004) Irish Public Service Broadcasting: a contingent valuation analysis. Econ Soc Rev (Irel) 35:321–350
Delaney L, O'Toole F (2006) Willingness to pay: individual or household? J Cult Econ 30:305–309. doi:10.1007/s10824-006-9021-8
Delaney L, O'Toole F (2008) Individual, household and gender preferences for social transfers. J Econ Psychol 29(3):348–359. doi:10.1016/j.joep.2007.09.003
Dosman D, Adamowicz W (2006) Combining stated and revealed preference data to construct an empirical examination of intrahousehold bargaining. Rev Econ Househ 4:15–34. doi:10.1007/s11150-005-6695-1
Efron B (1997) Bootstrap methods: another look at the jackknife. Ann Stat 7:1–26. doi:10.1214/aos/1176344552
Hasler B, Jacobsen JB, Lundhede TH, Martinsen L, Thorsen BJ (2008) Household versus individual willlingness to pay: evidence from a choice experiment on environmental valuation. Working Paper, University of Aarhus
Li H, Berrens RP, Bohara AK, Jenkins-Smith HC, Silva CL, Weimer DL (2005) Testing for budget constraint effects in a national advisory referendum survey on the Kyoto protocol. J Agric Resour Econ 30(2):350–366
Lindhjem H (2007) 20 Years of stated preference valuation of non-timber benefits from Fennoscandian forests: a meta-analysis. J For Econ 12:251–277
Mitchell RC, Carson RT (1989) Using surveys to value public goods: the contingent valuation method. Resources for the Future, Washington DC
Munro A (2005) Household willingness to pay equals individual willingness to pay if and only if the household income pools. Econ Lett 88(2):227–230. doi:10.1016/j.econlet.2005.02.011
Nyborg K (2000) Homo economicus and homo politicus: interpretation and aggregation of environmental values. J Econ Behav Organ 42:305–322. doi:10.1016/S0167-2681(00)00091-3
Pahl J (1995) His money, her money: recent research on financial organisation in marriage. J Econ Psychol 16:361–376. doi:10.1016/0167-4870(95)00015-G
Powe NA, Bateman IJ (2003) Ordering effects in nested 'top-down' and 'bottom-up' contingent valuation designs. Ecol Econ 45:255–270. doi:10.1016/S0921-8009(03)00084-3
Presser HB (1998) Decapitating the US Census Bureau's "head of household": feminist mobilization in the 1970s. Fem Econ 4(3):145–158. doi:10.1080/135457098338356
Quiggin J (1998) Individual and household willingness to pay for public goods. Am J Agric Econ 80:58–63. doi:10.2307/3180268
Sagoff M (1988) The economy of the Earth. Cambridge University Press, Cambridge
Samuelson PA (1956) Social indifference curves. Q J Econ 70(1):1–22. doi:10.2307/1884510
SEPA (2006) An instrument for assessing the quality of environmental valuation studies. Swedish Environmental Protection Agency
Smith VK, Van Houtven G (1998) Non-market valuation and the household, RFF Discussion Paper 98–31
Smith VK, Van Houtven G (2004) Recovering Hicksian consumer surplus within a collective model: Hausman's method for the household. Environ Resour Econ 28(2):153–167. doi:10.1023/B:EARE.0000029916.05630.89
Strand J (2005) Individual and household values of mortality reductions with intrahousehold bargaining. J Risk Uncertain 31(2):217–236. doi:10.1007/s11166-005-3555-6
Strand J (2007) Public-good valuation and intrafamily allocation. Environ Resour Econ 38(4):527–543. doi:10.1007/s10640-007-9098-3
Thaler RH (1999) Mental accounting matters. J Behav Decis Making 12(3):183–206. doi:10.1002/(SICI)1099-0771(199909)12:3<183::AID-BDM318>3.0.CO;2-F
Vermeulen F (2002) Collective household models: principles and main results. J Econ Surv 16(4):533–564. doi:10.1111/1467-6419.00177
Webley P (1995) Accounts of accounts: en route to an economic psychology of personal finance. J Econ Psychol 16:469–475. doi:10.1016/0167-4870(95)00021-F

Annex

Annex: Contingent valuation survey questionnaire

Developed by[1]:
Henrik Lindhjem (Norwegian University of Life Sciences & Econ Pöyry), and
Ståle Navrud (Norwegian University of Life Sciences).

This survey was administered to a webpanel of Norwegian respondents on the web, as personal interviews and as mail surveys in September-November 2007 by TNS Gallup. There were 10 different samples (two for the mail survey, one for in-person interviews, seven for web survey – that were all given different treatments – see Table 1 below. The sample treatments are explained in the survey below using comments inserted in brackets [..][2].

Table 1 *Survey design: Split (sub)samples and different treatments (i.e. different types of questionnaires or survey modes for different samples)*

	I (n=500)	**III (n=300)**	**IIa (n=450)**	**IIb (n=450)**	**IIc (n=450)**	**IId (n=450)**	**IIe (n=450)**	**IIf (n=450)**	**IIg (n=450)**
Mode type	Mail	PI	Web	Web	Web	Web	Web	Web	Web
Frame	Norw	Oslo	Oslo	Norw	Norw	Norw	Norw	Norw	Norw
Treatment 1: Modes									
Websurvey Oslo			X						
Pers.int Oslo		X							
Treatment 2: Scope									
Protection 2,8% & 4,5%	Ia: 250	X	X	X					
Protection 2,8% & 10%	Ib: 250				X	X	X	X	X
Cheap talk 1: Scope						X			
Info: # of species									X
Treatment 3: Hhld vs indiv.									
WTP household	X	X	X	X	X	X	X	X	X
WTP personally							X	X	
Follow up probe							X	X	
Treatment 4: Protest & High bids									
Cheap talk 2: High bids				X	X	X			X
Cheap talk 3: Protest 0				X	X	X	X	X	X
Websurvey features									
Time measure			X	X	X	X	X	X	X
Definitions of terms				X	X	X	X	X	X

[1] Several people have assisted in developing the survey questionnaires (see the acknowledgements section of the thesis).
[2] The samples IIe and IIf are the basis for paper 4 in this thesis.

Below is the generic form of the questionnaire translated from Norwegian. The specific lay-out is different depending on the mode of administration.

[In web survey, time starts counting]

q1 – Do you think the Norwegian Government should spend more, less, or the same amount of money as they do today, on the following public goods?
[Statements randomised in web survey and in-person interviews]

(Check one alternative per row)

	More	Same	Less	Don't know
Reduction of greenhouse gas emissions	☐	☐	☐	☐
Road construction	☐	☐	☐	☐
Better health care and care for the elderly	☐	☐	☐	☐
Burying power lines in mountains and fjord areas	☐	☐	☐	☐
Wind power development	☐	☐	☐	☐
Prevent oil spills along the coast	☐	☐	☐	☐
Protection of more forest areas and animal and plant species living there	☐	☐	☐	☐
Protection of cultural monuments	☐	☐	☐	☐

q2 – Have you used forests (in the lowlands or up to the mountains) for recreation over the past 12 months?

(Check one)

☐ Yes
☐ No [Filtered to the text after question 7]

q3 – How many times have you been to forests over the past month?

(Check one)

☐ 0
☐ 1-3
☐ 4-7
☐ 8-15
☐ More than 15

q4 – Have you been doing any of the following activities in forests over the past 12 months? You may provide several checks.

(Check all that apply)

☐ Hiking or biking on a road
☐ Hiking off road or on a trail
☐ Skiing
☐ Hunting or fishing
☐ Collecting berries, mushrooms or other things from the forests
☐ Collecting wood, Christmas trees or similar

q5 – Is there a specific forest that you frequently visit? If yes, how far from your home is it located?

(Check one)

- ☐ Yes. Write down the number of kilometres between your home and the forest area:_________________km
- ☐ No, I don't have a specific forest that I frequently visit [Filtered to text after q7]

q6 – What is the name of the forest that you frequently visit?
(Type text)

q7 – To what extent do you agree or disagree with the following statements about the forest area you most frequently visit? [Statements randomised in web survey and in-person interviews]

(Check one alternative per row)

	Totally agree	Agree to some extent	Disagree to some extent	Totally disagree	Don't know	Not applicable
"There are too many clear-cut areas"	☐	☐	☐	☐	☐	☐
"There are not enough views of lakes, streams and cultural landscapes"	☐	☐	☐	☐	☐	☐
"Planted forests are too homogenous (few tree species and age differences)"	☐	☐	☐	☐	☐	☐
"Forest roads make recreation easy for me"	☐	☐	☐	☐	☐	☐
"There are too many wheel tracks and other traces from logging machines outside the forest roads"	☐	☐	☐	☐	☐	☐
"Our ancestors' paths and roads and old settlements are getting overgrown"	☐	☐	☐	☐	☐	☐
"There are few clearings – the forest is too dense"	☐	☐	☐	☐	☐	☐
"There are too many visible traces of human encroachment, such as power lines"	☐	☐	☐	☐	☐	☐

[In web survey timing stops and starts]

Before you answer the next questions, we would like you to to read through the following information about Norwegian forests. Difficult words are underlined: by moving your mouse cursor over them, an explanation of the term will come up.

1830 vulnerable and threatened species in Norwegian forests

There are around 40 000 animal and plant species living in Norway, an important measure of what we call biological diversity. From around 18 500 species assessed by biologists, circa 3800 of them are classified as vulnerable or threatened by extinction over the next 100 years (pictures 2-7 below). Half of the vulnerable and threatened species live in forests (Click here for graph 1), and most of them are insects, fungi, lichens and mosses (Click here for graph 2). Many of them depend on large, old and dead trees for their survival. The species are part of the complicated ecological system in the forests, and there is much we do not know about this, since we do not know which function each species has. But we know that biological diversity in the long term makes the forest better able to endure and adapt to changes in the environment and in the climate.

Few forest areas are untouched by forestry

Forestry has been conducted in Norway for several hundred years, and has had large impact on the areas where animals and plants live. That is particularly true for modern, industrial forestry. This type of forestry include the cutting of larger areas using machines, planting of more uniform tree stands, change of tree species, use of fertiliser and construction of forest roads. The forestry sector also takes many environmental precautions in its activities, for example leaving old trees and areas rich in species (key biotopes), cutting smaller areas, leaving zones along streams forested and so on. Even so, there are many species that need larger and more natural forest areas to survive. There is very little virgin-like forest left in Norway (picture 1 below). Protection of virgin forest in nature reserves, where forestry is not permitted, can be important to secure the survival of species.

q8 – How well do you know how forestry is conducted today?

(Check one)

- ☐ Very well
- ☐ Fairly well
- ☐ Fairly badly
- ☐ Very badly

The following main functions are maintained through “environmentally cautious forestry” and “protection of forests in natural reserves”:

Environmentally cautious forestry maintains:

- Forest walks, skiing and bicycling
- Hunting and fishing
- Berry picking, mushroom collection, farm animal grazing
- Most of the functions of forests such as absorbing carbon dioxide, preventing erosion and leaching of nutrients to water bodies
- Protection of some vulnerable and threatened species

Protection of forests in nature reserves maintain:

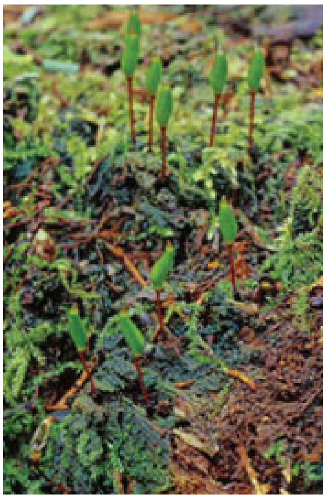
- Virgin-like forests as a segment of Norwegian nature for our descendants
- Virgin-like forests for research purposes
- Protection of many vulnerable and threatened species
- Possibility for education in natural sciences
- Norway’s international commitments regarding protection of biological diversity
- The intrinsic value of nature and species

Picture 1: Virgin-like forest

Distinctive features of virgin forest are large amounts of dead wood in all stages of decay, much standing dead wood, very high age of trees and good variation. Photo: Sigve Reiso/Naturarkivet

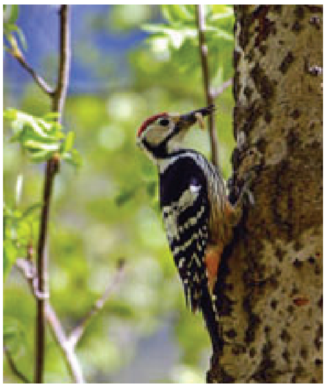
Pictures 2-7: Examples of vulnerable and threatened animal and plant species in forests in Eastern Norway.

The orchid Knottblom (Microstylis monophyllos) is classified as critically threatened. Photo: Kim Abel/Naturarkivet



The moss Grønnsko (Buxbaumia viridis) is classified as vulnerable. Photo: Kim Abel/Naturarkivet

The lichen Huldrestry (Usnea longissima) is classified as threatened. Photo: Bård Bredesen/Naturarkivet

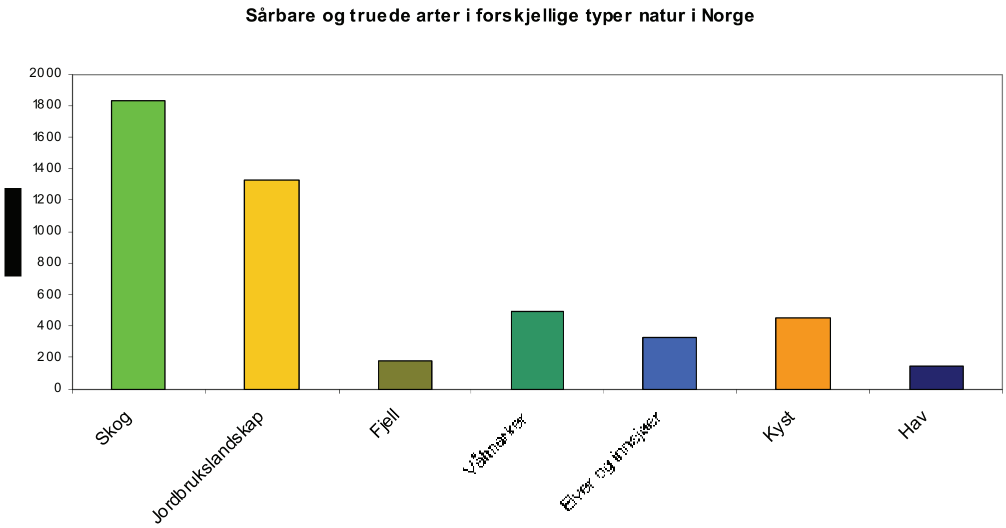


The White-backed woodpecker (Dendrocopos leucotos) is classified as near threatened. Photo: Sigve Reiso/Naturarkivet

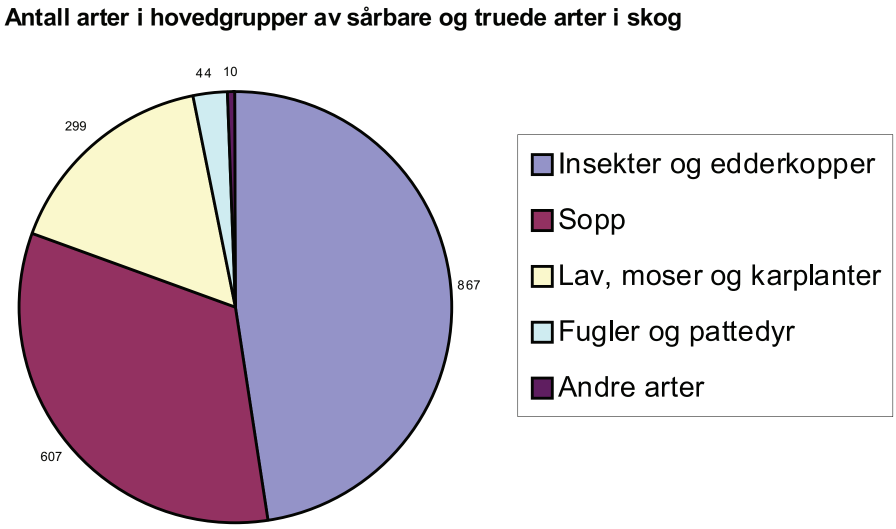
The assumed extinct fungus species Storporet flammekjuke (Pycnoporellus alboluteus) is classified as critically threatened. Photo: Kim Abel/Naturarkivet

Sinoberbillen (Cucujus cinnaberinus) has been found only in ca seven areas in Norway. Photo: Anne Sverdrup Thygeson/Naturarkivet

[Graph 1: Shows number of species by nature type in Norway. Bars from left to right: Forests, agricultural landscapes, mountains, wetlands, rivers and lakes, coast, ocean. In web version the two graphs below would appear by clicking on the links in the text above]



[Graph 2: Shows distribution of types of species in forests, i.e. breakdown of the left bar in graph 1 above. Categories: Insects and spiders (largest sector to the right), fungi (second largest sector), lichens, mosses and plants (third largest), birds and mammals (fourth largest), other species (smallest)]



q9 – Which of the following statements fits your view on protection of vulnerable and threatened animals and plants best? Check the statement which suits best

(Check one)

- ☐ I don't see extinction of plants and animals as a problem in Norway
- ☐ I am fairly concerned about the protection issue
- ☐ I am very concerned about the protection issue
- ☐ I work with nature protection in my job
- ☐ I don't know

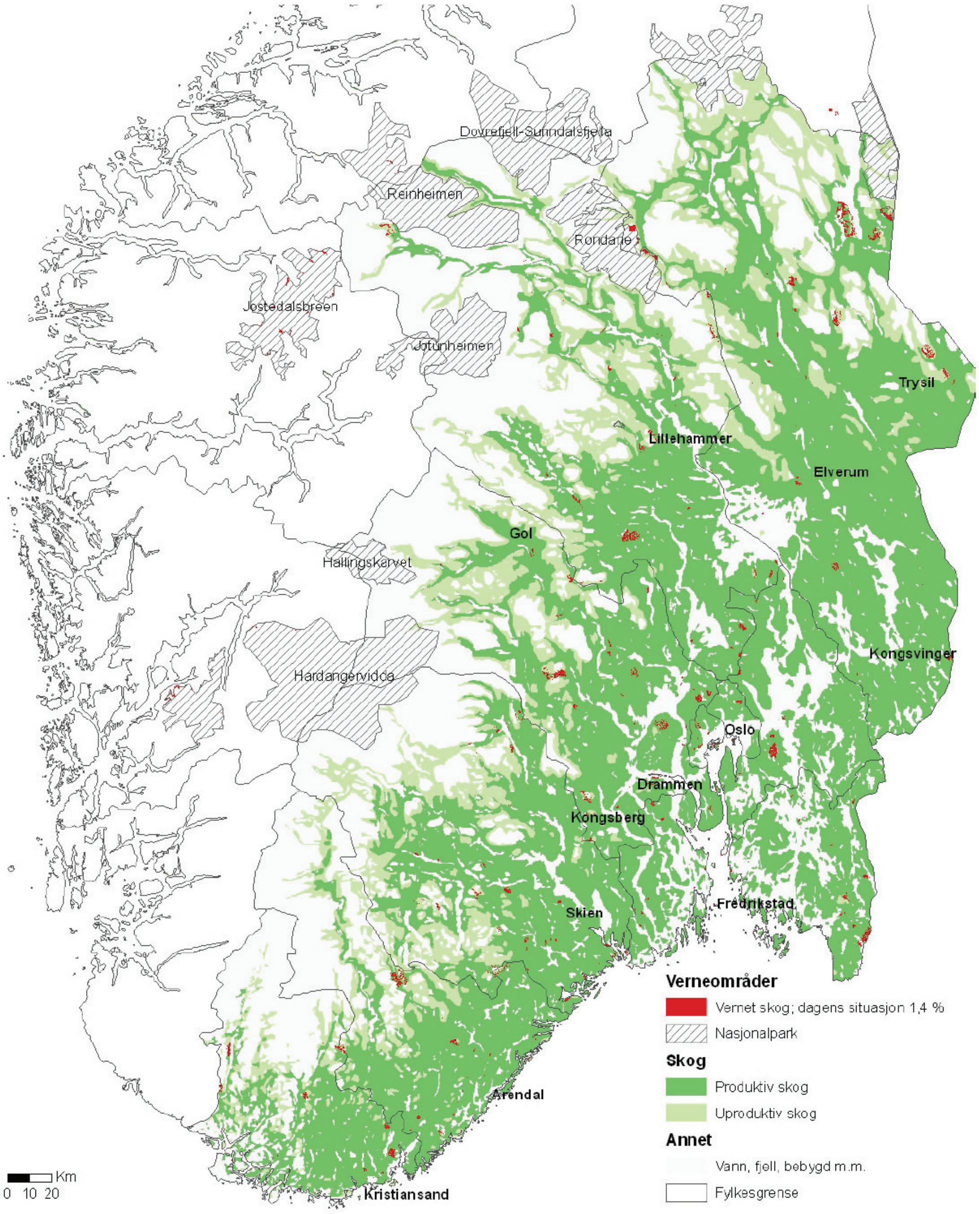
[In websurvey time for first part of survey measured. Time for second part started]

Current forest protection policy

One third of the land area in Norway is covered by forests and the majority of forests are located in Eastern Norway. The map below shows total forest area (marked as light green) and productive forest area (marked as dark green). "Productive forest" is forest that grows fast enough to be used for forestry. The "unproductive forest areas" consist of the areas up towards the mountains, along the coast and in the northern part of the country. Most of the forest is owned by private owners. The red spots on the map show the forest areas that are currently protected in nature reserves, that is where forestry is not permitted. Other activities such as hunting, fishing, berry and mushroom picking and grazing of farm animals are nomally permitted, like in other forests. Forest owners take environmental precautions in their forestry activities for most of the remaining forest areas. The forest owners receive compensation for loss of timber revenues if a forest they own is protected.

The nature reserves constitute today **1.4 percent (%)** of the procutive forest area, or approximately 1.05 million mål[3] (1050 square kilometres). Hardangervidda National Park (marked on the map) is three times as large as this area, for comparison. The National Parks in Norway cover mostly mountain areas, but also contain some forest.

[3] "Mål" is a common measure of land area in Norway = 1000 m^2. 10 mål equal 1 hectare (100m x 100m), 100 hectare equal 1 km^2.



Kart laget av Norsk institutt for skog og landskap

q10 – To what extent are you familiar with the ongoing debate about a planned nature reserve in Trillemarka in Buskerud?

(Check one)

- ☐ I have detailed knowledge about it
- ☐ I have heard about it
- ☐ I haven't heard about it

q11 – To what extent do you agree or disagree with the following statements about today's protection policy? [Statements randomised in web survey and in-person interviews]

(Check one alternative per row)

	Totally agree	Agree to some extent	Disagree to some extent	Totally disagree	Don't know	Not applicabl e
"The current protection level does not ensure the survival of vulnerable and threatened plant and animal species"	☐	☐	☐	☐	☐	☐
"There is too much forest in Norway not managed or harvested"	☐	☐	☐	☐	☐	☐
"Difficult terrain where forestry cannot be undertaken, ensures sufficient amount of virgin forests in Norway"	☐	☐	☐	☐	☐	☐
"It is positive if nature reserves are established in my community"	☐	☐	☐	☐	☐	☐
"Forestry should take more care of biological diversity"	☐	☐	☐	☐	☐	☐
"It is the quality of forest recreation that is important, not protection of species"	☐	☐	☐	☐	☐	☐

[In web survey, time is measured to here. Timing starts for next section]

Future protection plans

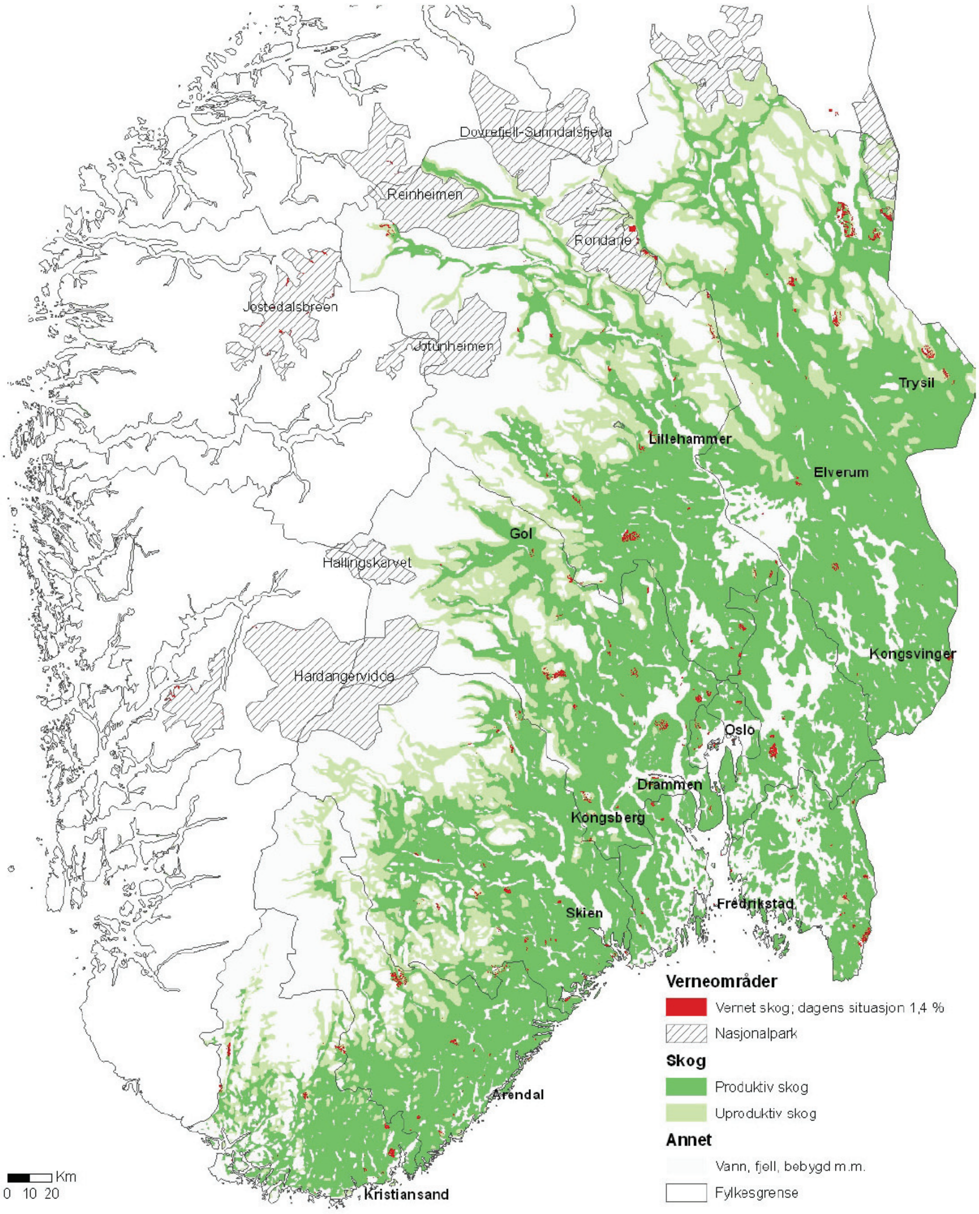
The government is currently considering two possible plans to increase forest protection in Norway from today's level of **1.4 %** of productive forest area to either **2.8 %** or **4.5 %** [Samples 2c-g: 10 %]. Most of the increased protection will be located in Eastern Norway. For the forest areas that are not protected, forest owners will take environmental precautions as described earlier. The more of the forest is protected, the more types of forests and types of animal and plant species will be conserved for the future. The new nature reserves will consist of voluntary protection in cooperation with private forest owners and increase in protection of government owned forests. Since the quality of the timber in the relevant nature reserves is not completely surveyed, it is not clear how much the two plans will cost.

The forest plan that may be chosen, would be financed through an annual extra tax for each household in Norway. The resources are earmarked a public fund for forest protection. The fund will compensate private forest owners for loss of timber revenues or give them new forest areas of equal value. Increased protection may mean that some forest owners can not continue their normal forestry activities, which has often been an important industry in the rural areas in Norway.

Look at the table below, where today's protection level (left column) is compared with the two alternative protection plans (the two columns to the right). Biologists recommend that two thirds (67 percent) of the increased protection should be located in Eastern Norway, since this part of the country contains forest types and species that are missing from today's protected areas. Also look at the three maps where we have marked today's protected areas in Eastern Norway and the possible new protected areas (red dots on the maps). It is not decided exactly where the new protected areas will be located. On the two maps of Eastern Norway we have increased the sizes of the current protected areas to illustrate the approcimate size of the protected areas under the two plans. Note that it may be difficult to see big differences between the maps.

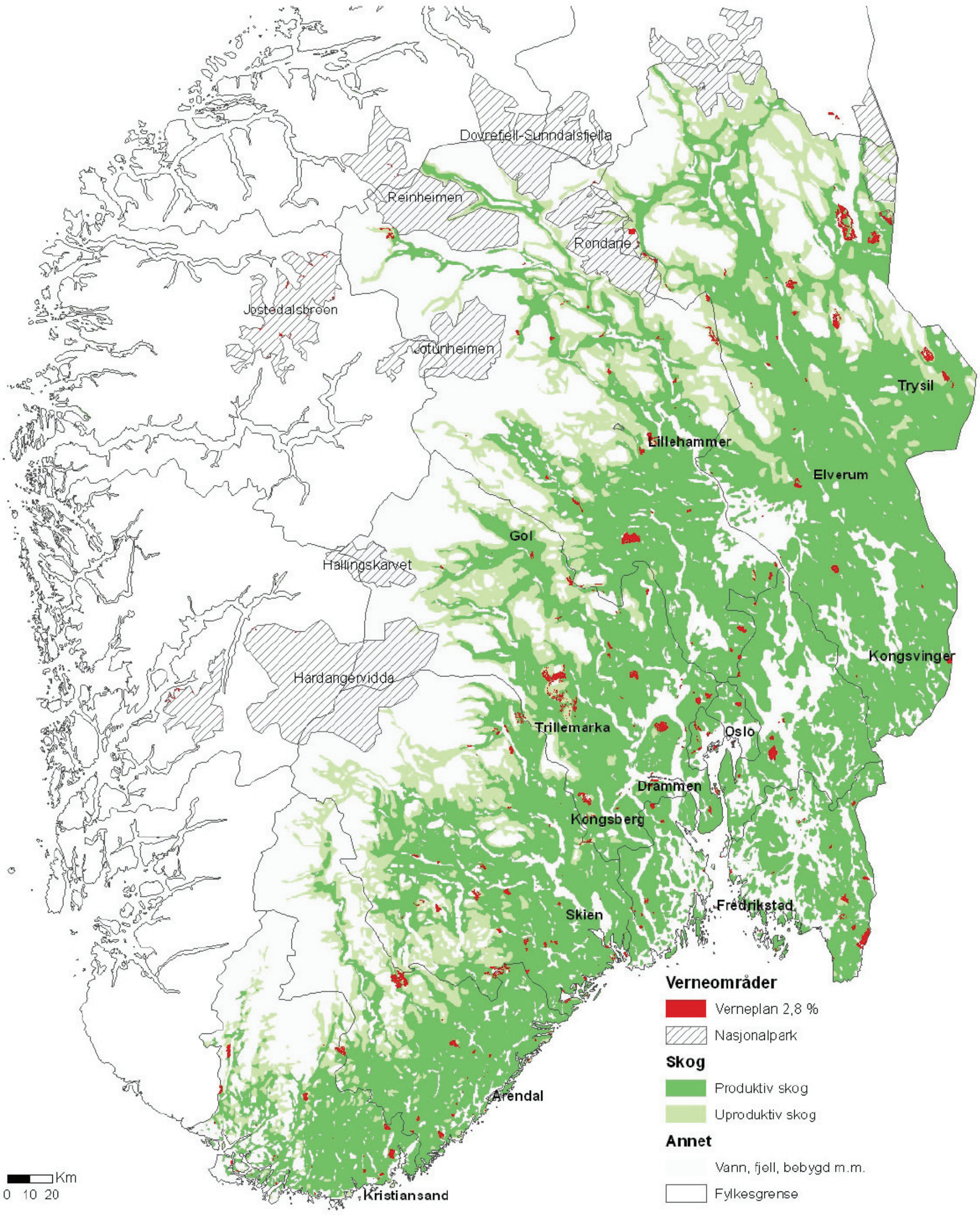
Today's situation: 1.4% of productive forest area	**Increase to 2.8% of productive forest area**	**Increase to 4.5 % [Samples 2c-g: 10 %] of productive forest area**
Total protected area: 1,05 million mål	**Total protected area:** 2,1 million mål	**Total protected area:** 3,3 million mål [10%: 7.4 million mål]
No new reserves	Twice today's protected area, 2/3 in Eastern Norway	Three times today's protected area, 2/3 in Eastern Norway. [10%: Seven times today's protected area]
Types of species: Same types of species as today remain vulnerable or threatened	**Types of species:** Large improvement for lichens, mosses, plants and fungi	**Types of species:** Very large improvement for lichens, mosses, plants and fungi. Large improvement for insects. [10%: Very large improvement for lichens, mosses, plants, fungi and insects. Some improvement for mammals and birds]
[Sample 2g , 10% level] **Number of vulnerable and threatened species:** Ca constant at today's level of 1827 species	[Sample 2g , 10% level] **Number of vulnerable and threatened species:** Ca 450 of the species in forests no longer vulnerable or threatened	[Sample 2g , 10% level] **Number of vulnerable and threatened species:** Ca 1000 of the species in forests no longer vulnerable or threatened

Today's situation [Clickable for larger map in web version]



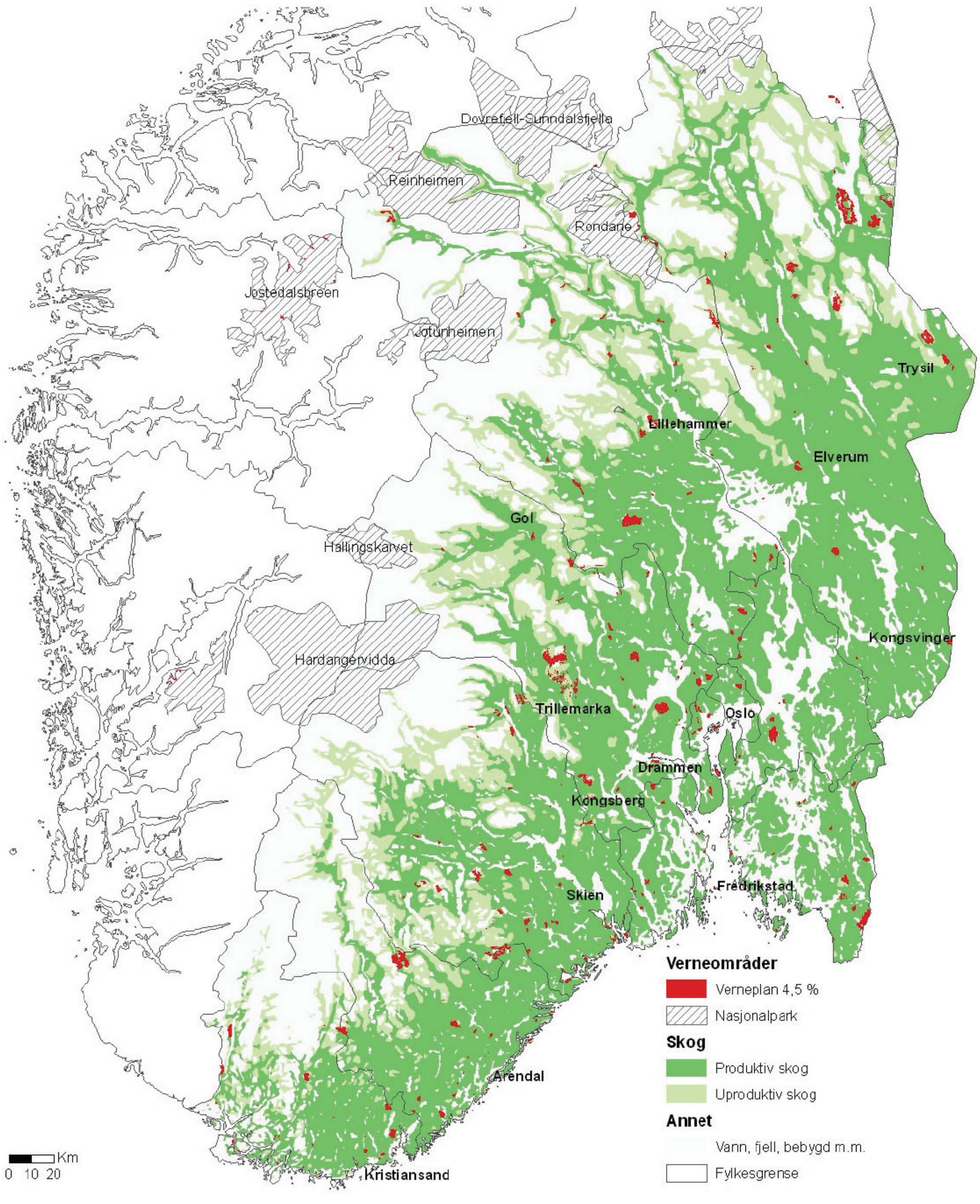
Kart laget av Norsk institutt for skog og landskap

Increase to 2.8% of productive forest area [Clickable for larger map in web version]



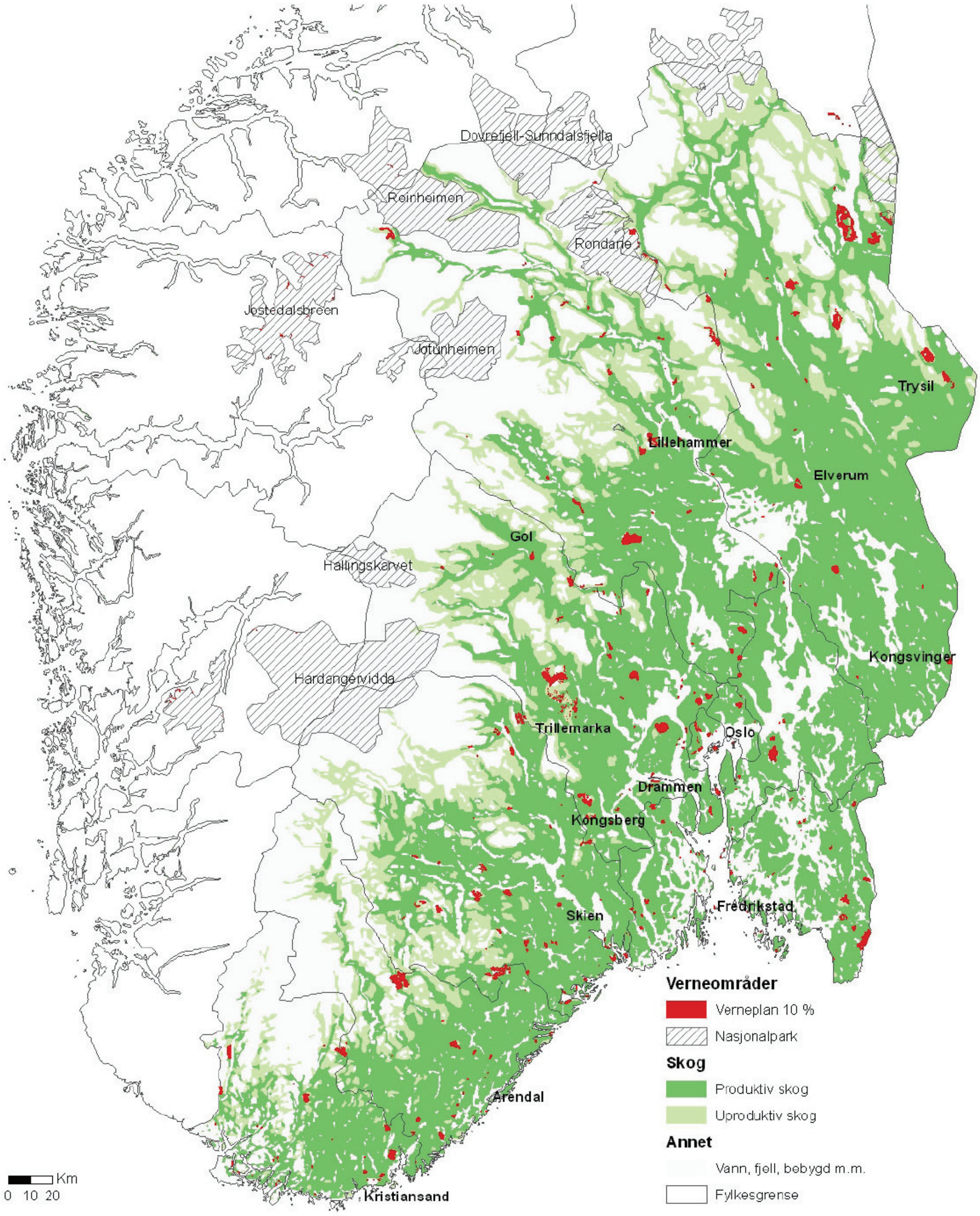
Kart laget av Norsk institutt for skog og landskap

Increase to 4.5% of productive forest area [Clickable for larger map in web version]



Kart laget av Norsk institutt for skog og landskap

[Samples 2c-g: Increase to 10% of productive forest area, instead of the 4.5 % plan [Clickable for larger map in web version]



Kart laget av Norsk institutt for skog og landskap

We now ask you to think about how much the two alternative plans are worth for your household [Sample 2f: you]. Think carefully through how much the **2.8 percent** plan is worth compared to the current situation, before you give your final answer to the next question. Try to think about how much will be a realistic annual amount to pay from your household's [Sample 2f: your] budget. Your household [Sample 2f: You] must choose to spend money on this plan, or on something else.

[Sample 2d: Cheap talk about scope. "It is often difficult for people to relate to two protection plans at the same time. When you compare the two protection plans with today's situation, try first to imagine that the changes in the forest will actually take place. Then consider whether the changes your household will get if 10 percent is protected is worth more than 2.8 percent protection, and if your household therefore is willing to pay more for the large plan. It is of course fully acceptable to think that it does not matter either way for your household which of the two protection levels are chosen".]

q12 – We ask you first to consider the smallest of the two protection plans, the one for 2.8%. Then you will be asked about the other plan.

What is the most **your household** [Sample 2f: **you**] almost certainly is willing to pay in an extra annual tax earmarked to a public fund for increased forest protection from today's level of **1.4%** to **2.8%** of productive forest area?

Choose the highest amount, if anything, your household [Sample 2f: you] almost certainly will pay.

(Check one) [the payment card below is designed as a drop-down menu in the web version]

- ☐ 0
- ☐ 25
- ☐ 50
- ☐ 100
- ☐ 300
- ☐ 500
- ☐ 700
- ☐ 900
- ☐ 1100
- ☐ 1400
- ☐ 1800
- ☐ 2200
- ☐ 2700
- ☐ 3200
- ☐ 3800
- ☐ 4400
- ☐ 5100
- ☐ 5800
- ☐ 7000
- ☐ 8500
- ☐ 10000
- ☐ 13000
- ☐ 15000
- ☐ More than 15000
- ☐ Don't know

q12b – Specify in kr per year [if more than 15000]
(Type number)

[Samples 2b-d, g: Cheap talk for high bids. If WTP > 5100 per year the following text would automatically come up: “People often state amounts that are higher than they actually would be willing to pay. This can be because they think the topic is important, and forget that money can only be used once. When you answer the question below try to imagine that the amount you state for forest protection no longer is available for other consumption”. The question below would follow:

q13 – Do you want to stick with the amount you stated in the last question or would you like to change it?

(Check one)

- ☐ I stick with my answer. I have given an amount I realistically would pay for increased forest protection.
- ☐ I stick with my answer. It is a hypothetical question and I don’t really have to pay what I have stated.
- ☐ No, I would like to change my answer, and think again how much the two protection plans are worth for me. [The payment card would come up again with this answer]

[q14: Sample 2e and 2f get another WTP question where reference to household and individual is reversed. Sample 2e gets the household WTP question first, sample 2f gets the individual WTP question first. Before the second WTP question, the following text is given: “We now ask you to think about whether it matters to your willingness to pay if you state it for yourself [Sample 2e: your household] instead of for your household [Sample 2e: yourself]”]

[Sample 2f and 2e get q15, 16 and 17 below depending on answers to q12 and q14]

q15 – How important were the following reasons for you stating a higher amount on behalf of your household than for yourself?

(Check one alternative per row)

	Important	Not important	Not applicable
"I took both incomes into account when I was asked on behalf of the household"	☐	☐	☐
"I take the household members' wishes regarding increased forest conservation into account even if I consider willingness to pay for myself alone"	☐	☐	☐
"I especially consider the children when asked on behalf of the household"	☐	☐	☐
"My partner is more interested in forest conservation than I am, so I adjusted for that"	☐	☐	☐
"I added what I think my partner would be willing to pay"	☐	☐	☐
"I doubled my individual amount since we are two adults in the household"	☐	☐	☐
Other reasons that were important? Specify:______________________________	☐	☐	☐

q16- How important were the following reasons for you stating a lower amount on behalf of your household than for yourself?

(Check one alternative per row)

	Important	Not important	Not applicable
"I thought about my individual budget and can pay more than if I have to take my household into consideration"	☐	☐	☐
"My partner is against more forest conservation, so I adjusted for that"	☐	☐	☐
"I am normally not the one paying for our household expenses, so I chose a lower amount on behalf of my household"	☐	☐	☐
"We have a tight budget for household expenses, but my personal budget is more generous"	☐	☐	☐
Other reasons that were important? Specify:______________________________	☐	☐	☐

q17 - How important were the following reasons for you stating the same amount on behalf of your household as for yourself?

(Check one alternative per row)

	Important	Not important	Not applicable
”I am single and have no children, so there is no difference”	☐	☐	☐
”My partner and I have a shared economy, so it does not matter if I am asked personally or on behalf of my household”	☐	☐	☐
”What the household collectively would have decided concur with my wish”	☐	☐	☐
”I am normally the one paying our household expenses, so in practice there is no difference if I am asked personally or behalf of my household”	☐	☐	☐
” I take my household members into account even if I consider willingness to pay for myself alone”	☐	☐	☐
Other reasons that were important? Specify:____________________________	☐	☐	☐

q18 – Imagine that the government chose the big plan of **4.5 percent** [Samples 2c-g: 10 percent] instead.

What is the most **your household** almost certainly is willing to pay in an extra annual tax earmarked to a public fund for increased forest protection from today's level of **1.4%** to **4.5%** [Samples 2c-g: **10%**] of productive forest area?

Choose the highest amount, if anything, your household almost certainly will pay.
(Check one)

- ☐ 0
- ☐ 25
- ☐ 50
- ☐ 100
- ☐ 300
- ☐ 500
- ☐ 700
- ☐ 900
- ☐ 1100
- ☐ 1400
- ☐ 1800
- ☐ 2200
- ☐ 2700
- ☐ 3200
- ☐ 3800
- ☐ 4400
- ☐ 5100
- ☐ 5800
- ☐ 7000
- ☐ 8500
- ☐ 10000
- ☐ 13000
- ☐ 15000
- ☐ More than 15000
- ☐ Don't know

q18b – Sepcify amount in kr per year [If higher than 15000]
(Type number)

[Samples 2b-d, g: Similar cheap talk for high bids as above]
[In web survey, time measured for the WTP questions]

q19 – How important were the different kinds of information that were given, when you compared the two alternative protection plans with today's situation?
(Check one alternative per row)

	Important	Not important
Increase in area of nature reserves measured in either mål or square kilometres	☐	☐
Increase in area of nature reserves measured in percent	☐	☐
The difference in which species types will get improved living conditions for the different plans	☐	☐
The maps showing the size of the nature reserves	☐	☐
The maps showing the distribution of nature reserves in relation to where I live	☐	☐
An overall evaluation of the differences between the plans	☐	☐
[Sample 2g: The number of species that will no longer be vulnerable and threatened]	☐	☐

[All samples get one of the next three questions, depending on the answer to the WTP questions for the two plans. Automatic filtering for the web survey]

q20 – How important are the reasons suggested below for you to state a higher amount for the large plan?
(Check one alternative per row)

	Important	Not important
I believe more forest types and species will be secured with a larger area	☐	☐
The large plan seams more realistic to implement	☐	☐
The large plan will be more expensive, and therefore I must pay my share	☐	☐
The large plan is better because it provides larger areas for recreation	☐	☐
Since our knowledge of vulnerable and threatened species is poor, we better be precautious	☐	☐
Other important reasons? Specify:____	☐	☐

q21 – How important are the reasons suggested below for you to state the same amount for the large plan?

(Check one alternative per row)

	Important	Not important
The large plan seems unrealistic to implement	☐	☐
For me, there are enough species and forest types that will be protected with the smaller plan	☐	☐
Both plans seem to give about the same improvement for different species types	☐	☐
I believe the large plan will be carried out anyway	☐	☐
I cannot afford to pay more for the large plan	☐	☐
It is the amount I am willing to pay for this purpose	☐	☐
Other important reasons? Specify:____	☐	☐

q22 – How important are the reasons suggested below for you to state a lower amount for the large plan?

(Check one alternative per row)

	Important	Not important
The large plan seems unrealistic to implement	☐	☐
For me, there are enough species and forest types that will be protected with the smaller plan	☐	☐
I believe the larger plan will be carried out anyway	☐	☐
I cannot afford to pay more for the large plan	☐	☐
The large plan will harm the forest industry and rural areas Norway too much	☐	☐
Other important reasons? Specify:____	☐	☐

[People answering “0” or “Don’t know” to both WTP questions are filtered directly to the question below]

q23 – Choose the statements below that best describe why you were not willing to pay anything to increase forest protection.

(Check no more than two statements)

☐ It is the government’s responsibility to protect forests
☐ I cannot afford to pay anything for this*
☐ The tax level is high enough as it is
☐ What I say will not affect whether the plans will be carried out or not
☐ Today’s level of forest protection is good enough*
☐ It is too difficult to give an amount
☐ I feel it is not right to value vulnerable and threatened species in monetary terms in this way
☐ Those who destroy the habitat where these plant and animal species live should pay
☐ The species can be preserved through environmentally cautious forestry - more protection is not necessary*
☐ We already pay enough to the farmers owning the forests
☐ I don’t know*

[Samples 2b-g: Cheap talk protest responses. Statements in the previous question not marked "*" are considered protest responses. The cheap talk script below would automatically come up if at least one such response was given.]

"In surveys like this people often protest against having to state how much they think environmental improvements are worth to them. But estimates of how much people are willing to pay give important information about how much of public resources we should spend on forest protection compared to other public goods.The answer you have given will be set equal to zero in our analysis, and this value of forest protection to you may not reflect your opinions in a good way. It is fully acceptable to answer what you have done, but it will make it harder to implement the protection plan based on the results from this survey."

q24 – After reading the above text, do you stick to your answer, or do you want to change it?

(Check one)

- ☐ I stick to my answer, and I am not willing to pay anything for increased forest protection
- ☐ No, I wish to change my answer and consider how much the protection plans are worth to me

[Samples 2b-g would be taken back to the first WTP question if no to q24].

q25 – What are the most important reasons why you wish to protect forests?

(Check up to three reasons)

- ☐ More untouched forest areas increase the value of my recreational experience in the forest
- ☐ The species must be preserved for scientific reasons
- ☐ All plant and animal species have the right to exist
- ☐ Norway is committed by international agreements to plant and animal protection
- ☐ I want the species to be preserved regardless of whether I will benefit from them
- ☐ If species disappear, it may disturb the natural system
- ☐ I enjoy the fact that other people have opportunities to visit virgin forests
- ☐ The species must be preserved as long as we don't have enough knowledge of their functions and which physical substances they consist of
- ☐ I don't know

q26 – Do you think you will visit any of today's protected areas marked on the Eastern Norway map during the next 12 months?

(Check one)

- ☐ Yes, definitely
- ☐ Yes, probably
- ☐ Maybe
- ☐ No, never

q27 – Protecting forests in nature reserves as done today may cause conflicts with forest owners, so alternatives to this are under consideration. How suitable do you think the following alternative measures to increase protection sound?

(Check one alternative per row)

	Very suitable	Fairly suitable	Fairly unsuitable	Very unsuitable	Don't know
Long-term lease of areas from forest owners (e.g. 30 years) instead of protecting the area forever	☐	☐	☐	☐	☐
The government buys several similar forest areas that can be exchanged with the areas to be protected	☐	☐	☐	☐	☐
Allow forest owners more freedom to develop business activities such as building cabins in the border areas of nature reserves	☐	☐	☐	☐	☐
The forest owners can offer to protect their forest against a stated compensation level. The government, private individuals or organisations can through an auction choose which areas to protect.	☐	☐	☐	☐	☐
The compensation level to forest owners is increased so that they will not loose income and future opportunities due to protection	☐	☐	☐	☐	☐
Forest owners are given the opportunity to use the nature reserves for tourism and other activities that may give economic development in rural parts of Norway as long as the forest is not harmed	☐	☐	☐	☐	☐

q28 – Did you find that the questions about your willingness to pay for increased forest protection were...?

(Check one)

☐ Very hard to answer
☐ Fairly hard to answer
☐ OK to answer
☐ Fairly easy to answer
☐ Very easy to answer
☐ I don't know

q29 - How thoroughly did you read the information about the proposed protection plans?

(Check one)

- ☐ Looked at the pictures and maps
- ☐ Read parts of the text thoroughly
- ☐ Read everything thoroughly

q30 – Do you or your closest family own forests?

(Check one)

- ☐ Yes
- ☐ No
- ☐ I don’t know

q31 –Are you a member of any of the following organisations?

(Check all that apply)

- ☐ Environmental conservation organisations such as the World Wide Fund for Nature (WWF) or Norwegian Nature Protection Association (Norges Naturvernforbund)
- ☐ The Norwegian trecking association (DNT)
- ☐ Norwegian Forest Owner Association (Norges Skogerieforbund) or other forest owner associations
- ☐ Hunter’s and Fishermen’s associations
- ☐ Norwegian farmer’s Union
- ☐ Other
- ☐ No organisation

q32– Approximately how much was your household's total gross income (before tax and deductibles) in 2006?

The answer will be used for statistical analysis only, and cannot be traced back to the individual respondent

(Check one)

- ☐ Up to 100000
- ☐ 125000
- ☐ 150000
- ☐ 175000
- ☐ 200000
- ☐ 225000
- ☐ 250000
- ☐ 275000
- ☐ 300000
- ☐ 325000
- ☐ 350000
- ☐ 375000
- ☐ 400000
- ☐ 425000
- ☐ 450000
- ☐ 475000
- ☐ 500000
- ☐ 525000
- ☐ 550000
- ☐ 575000
- ☐ 600000
- ☐ 625000
- ☐ 650000
- ☐ 675000
- ☐ 700000
- ☐ 725000
- ☐ 750000
- ☐ 775000
- ☐ 800000
- ☐ 825000
- ☐ 850000
- ☐ 875000
- ☐ 900000
- ☐ 925000
- ☐ 950000
- ☐ 975000
- ☐ 1000000
- ☐ 1025000
- ☐ 1050000
- ☐ 1075000
- ☐ 1100000
- ☐ 1125000
- ☐ 1150000
- ☐ 1175000

- ☐ 1200000
- ☐ 1225000
- ☐ 1250000
- ☐ 1275000
- ☐ 1300000
- ☐ 1325000
- ☐ 1350000
- ☐ 1375000
- ☐ 1400000
- ☐ 1425000
- ☐ 1450000
- ☐ 1475000
- ☐ 1500000
- ☐ More than 1500000

q32b – Specify income per year in NOK [if more than 1 500 000]
(Type numbert)

q33 – Approximately how much was your personal total gross income (before tax and deductibles) in 2006?

The answer will be used for statistical analysis only, and cannot be traced back to the individual respondent

(Check one)

- ☐ Up to 50000
- ☐ 75000
- ☐ 100000
- ☐ 125000
- ☐ 150000
- ☐ 175000
- ☐ 200000
- ☐ 225000
- ☐ 250000
- ☐ 275000
- ☐ 300000
- ☐ 325000
- ☐ 350000
- ☐ 375000
- ☐ 400000
- ☐ 425000
- ☐ 450000
- ☐ 475000
- ☐ 500000
- ☐ 525000
- ☐ 550000
- ☐ 575000
- ☐ 600000
- ☐ 625000
- ☐ 650000
- ☐ 675000
- ☐ 700000
- ☐ 725000
- ☐ 750000
- ☐ More than 750000

q33b – Specify income per year in NOK [if more than 750 000]
(Type number)

Comments – Do you have comments or views on the survey you just have completed?

(Type text)

[List of definitions: These would appear automatically in a box on the screen if the cursor in the web survey moved across an underlined word. A list of definitions were given at the end of the mail survey, and in the in-person interviews people were informed that they could have underlined words explained to them]

The following terms were defined and explained (translated from Norwegian here):

- **Biological diversity (or biodiversity):** The variation in Earth's life forms (plants, animals and micro organisms, their genetic material and the complicated interplay they are part of. A common measure of biodiversity is number of species. (Source: Norwegian Directorate for Nature Management)
- **Species:** Groups of naturally occurring populations which can breed with each other, and which is isolated in this sense from other such groups. (Source: Norwegian Red List 2006)
- **Vulnerable and threatened species:** Collective term for species classified as critically threatened, strongly threatened or vulnerable in the Norwegian Red List, that is the list of species which are classified in the different categories depending on their risk of extinction the next 100 years. (Source: Norwegian Red List 2006)
- **Insects:** A group of animals which constitute 8 out of 10 known species on the Earth and which can live in most environments, except the sea (Source: Wikipedia)
- **Fungi:** A group of organisms that is part of its own overall kingdom, the fungi kingdom, on par with plants, animals, bacteria. Can live almost anywhere and has no chlorophyll. Feeds on dead or living organic material. (Source: CAPLEX)
- **Lichens:** A group of organisms where each individual has both fungi cells and algae cells. The intimate way fungi and plants live together is advantageous for both (symbiosis), and the lichens are very hardy organisms that can live where other plants must give up, for example in polar regions and in high mountain areas (Source: CAPLEX)
- **Mosses:** Low, green plants which grow through spores. Grows mostly in moist environments. Ca 25 000 known mosses are distributed all over the world. (Source: CAPLEX)
- **Virgin-like or old-growth forest:** Forest that has been little influenced by harvest or other human impacts for a long time (more than 100 years)
- **Nature reserve:** The strictest form of area protection which exists in Norway and which is established under the law of nature protection. Nature reserves consist of untouched or near-untouched nature, or they constitute a special nature type of scientific or predagogical importance. A nature reserve is either totally protected, or protected for a special purpose. (Source: Wikipedia)
- **Productive forest area:** Forest that produces, i.e. has a growth of at lest 0.125 m3 wood per mål and year. Norway has a productive forest area of around 66 000 km2, around 17.7 percent of total land area (Source: CAPLEX)
- **National park:** A national park is an area which a government protects to conserve important nature values. In Norway a national park is regarded as medium strict protection. A weaker protection status is landscape protected area ("landskapsvernområde"), while stricter protection is called a nature reserve. Further, national parks are typically much larger than these two. According to Norwegian law a national park should primarily contain publicly owned land. (Source: Wikipedia).
- **Eastern Norway:** Within this context the counties of Hedmark, Oppland, Buskerud, Akershus, Oslo, Telemark, Vestfold, Østfold, Aust- og Vest-Agder.

- **Mål:** A mål is a measure of an area which is equal to 1000 square metres (that is the area of a square which sides are ca 32 metres). 1 mål = 1 dekar(daa), 1 km^2= 1000 mål/dekar (Source: CAPLEX)
- **Square kilometre (km^2):** The area of a square where each side is one kilometre long. 1 km^2= 1000 mål/dekar (Source: CAPLEX).

[**Variables in webpanel database:** In addition to the CV data collected through the questionnaire above, around 50 background variables were recorded from respondents in the webpanel earlier the same autumn of the survey. The most relevant of these variables include gender, age, number of household members, position in the household, number of children, age of children, civil status, type of job and position, type of house, own or rent, internet use frequency, general attitude questions, geographical location on the municipality and regional levels, individual and household income (which also was collected in the CV survey), etc.]

Henrik Lindhjem

Henrik Lindhjem was born in Larvik, Norway in 1973. He holds a Master of Science degree in Economics (Cand. Oecon.) from the University of Oslo (1998) and a Master of Philosophy degree in Environment and Development from University of Cambridge, UK, (1999).

This thesis consists of four papers on methodological issues in meta-analysis (MA), benefit transfer (BT) and environmental valuation. The first paper presents a MA of stated preference studies valuing non-timber benefits in Norway, Sweden and Finland over the last 20 years. It investigates using different meta-regression models to what extent willingness to pay (WTP) estimates conform with standard expectations, tests a number of novel hypotheses and identifies gaps in the literature. Papers 2-4 then each pick up an important research theme following from the first paper. The second paper utilises the same data to investigate the precision in using MA for international BT, as compared with simpler and more common BT techniques. The third paper investigates, using a different and more extensive dataset of biodiversity and nature conservation values from Asia and Oceania, how the heterogeneity or scope of the MA data influences the results of different meta-regression models and their precision when used for BT. The fourth paper tests using a primary contingent valuation data set of WTP for forest protection in Norway, whether people state different WTP if asked as individuals or on behalf of their household. This paper tests in a more controlled way the question also investigated in the first paper. Results from the four papers are encouraging in contributing to our understanding of people's preferences for complex environmental goods. However, more research is required to determine the conditions under which MA may be reliably used for BT.

Associate Professor Ståle Navrud was Henrik's supervisor

Henrik Lindhjem is currently senior researcher/consultant at Econ Pöyry, Oslo.

E-mail: henrik.lindhjem@poyry.com
Home page: www.lindhjem.info

Department of Economics and Resource Management
Norwegian University of Life Sciences
PO Box 5003
N-1432 Ås,
Norway

Telephone: +47 6496 5700
Telefax: +47 6496 5701
e-mail: *ior@umb.no*
http:/www.umb.no/ior

ISSN: 1503-1667
ISBN: 978-82-575-0819-7