The assumption of equal marginal utility of income: how much does it matter?

Hege Medin a,1, Karine Nyborg a,*, Ian Bateman b

a Division for Environmental and Resource Economics, Research Department, Statistics Norway, P.O. Box 8131 Dep., N-0033 Oslo, Norway
b Centre for Social and Economic Research on the Global Environment (CSERGE), School of Environmental Sciences, University of East Anglia, Norwich NR4 7TJ, UK

Received 7 February 2000; received in revised form 9 August 2000; accepted 9 August 2000

Abstract

In most applied cost–benefit analyses, individual willingness to pay (WTP) is aggregated without using explicit welfare weights. This can be justified by postulating a utilitarian social welfare function along with the assumption of equal marginal utility of income for all individuals. However, since marginal utility is a cardinal concept, there is no generally accepted way to verify the plausibility of this latter assumption, nor its empirical importance. In this paper, we use data from seven contingent valuation studies to illustrate that if one instead assumes equal marginal utility of the public good for all individuals, aggregate monetary benefit estimates change dramatically. © 2001 Elsevier Science B.V. All rights reserved.

Keywords: Utility comparisons; Cost–benefit analysis; Choice of numeraire; Environmental goods

JEL classification: D61; D62; D63; H41; Q2

1. Introduction

Undergraduate students of economics usually spend considerable time and energy grappling with the concept of ordinal utility. This is not surprising given that ordinal utility is not comparable between persons and does not tell us anything about intensities, being simply a tool for describing an individual’s binary choices. As such, it is a concept very much deprived of normative content (Sen, 1977), although it is extremely useful for purely descriptive analyses.

When faced with the task of conducting an applied cost–benefit analysis after graduation,
however, many economists need to go through the reverse process, since applied cost–benefit analysis requires a cardinal, interpersonally comparable utility concept (Arrow, 1951). After spending so much time getting accustomed to ordinal utility, the economist now has to grasp the implicit consequences of assuming, instead, that utility does indeed tell us something about intensities, and that one can compare benefits between persons.

There is currently no generally accepted way to measure cardinal and interpersonally comparable utility. In applied cost–benefit analysis, the most common way to proceed is to use the unweighted sum of individuals’ net willingness to pay as a measure of social benefits. This is consistent with using a utilitarian social welfare function along with an assumption of equal marginal utility of income for all individuals. However, since the latter assumption does not rest on any empirical evidence, it introduces a certain kind of arbitrariness into the analysis. Unlike the choice of social welfare function, which can be discussed on ethical grounds, assumptions about the cardinality and comparability aspects of utility functions may be regarded as positive rather than normative; but the problem is that, in the absence of measurement methods, these assumptions cannot be empirically verified. One simply does not know whether the assumption of equal marginal utility of income is reasonable.

However, although we cannot directly test the plausibility of this assumption, we may still perform sensitivity analyses regarding the robustness of results to alternative ways of comparing cardinal utility between persons. This paper is an attempt to do precisely that. Based on data from seven contingent valuation (CV) studies of environmental changes, we calculate aggregate monetary benefits using an alternative operationalisation of cardinal and interpersonally comparable utility, namely, that individuals have an equal marginal utility of the environmental good in question.

This latter assumption corresponds to using units of the environmental good as the numeraire when aggregating individual benefits, instead of the usual approach of using money for this purpose. Until recently, it was a common belief that the choice of numeraire does not matter in cost–benefit analysis. However, Brekke (1997) demonstrated that the unit of measurement did indeed matter (see also Drèze, 1998; Johansson, 1998). Brekke pointed out that, when it comes to public goods, different individuals generally have different marginal rates of substitution, since the amount of the public good is necessarily equal for all; implying that individuals have different marginal conversion rates between those goods that may alternatively be chosen as the numeraire. Consequently, when individual benefit estimates are aggregated, the interests of different individuals are given a different emphasis depending on which measurement unit is used.

Brekke’s result may be dismissed as irrelevant by some, arguing that it is not practicable to use environmental units. For example, one cannot always in practice pay compensations in environmental units, and survey questions using environmental units may be very difficult for respondents to understand. However, we believe that the importance of Brekke’s result lies elsewhere. As already mentioned, using environmental units as the numeraire corresponds to an alternative operationalisation of cardinal, interpersonally comparable utility. As such, it provides a means to check the empirical importance of the seemingly innocuous, but admittedly arbitrary, assumption of equal marginal utility of income used in most applied cost–benefit analyses. Brekke presents one empirical example in his paper. Using data from a survey by Strand (1985), he found that the maximum per person cost that would make the project’s net benefits positive were 22 times higher if money were used as a numeraire than if one used environmental units. In other words, exchanging the assumption of equal marginal utility of income for an assumption of equal marginal utility of the public good changed the result dramatically in this particular case. It is hard to see that the former assumption is more plausible than the latter from a theoretical point of view: Arguing in favour of one or the other requires reasoning about cardinality and interpersonal

2 However, see Unsworth and Bishop (1994) for a discussion of in-kind compensation.
comparisons of utility, which is rarely found in economic theory. However, apart from Brekke’s own example, we know of no attempts in the literature to analyse empirically the sensitivity of aggregate benefit estimates with respect to the choice of assumptions concerning marginal utilities.

The widespread use of aggregate unweighted WTP as an estimate of social benefits has been criticised by many scholars (for example, Kelman, 1981; Bromley, 1990; Hammond, 1990; Vatn and Bromley, 1994). If alternative and seemingly equally a priori plausible ways to operationalise interpersonally comparable cardinal utility yielded dramatically different results than the standard procedure, this should be of concern for all practitioners and users of cost–benefit analysis. In such a case, the traditional assumption cannot be defended by convenience alone, and one needs to take the issue of interpersonal comparisons of utility more seriously. We believe, therefore, that it is important to investigate whether Brekkes’ finding of dramatically different aggregate benefit estimates holds more generally.

We should stress at the outset, however, that our aim is neither to argue that using environmental units is necessarily more relevant than using money as a numeraire, nor to identify the ‘best’ way to aggregate individual welfare. We recognise that money, being a much more generally exchangeable numeraire than environmental units, is the most convenient measurement unit in many contexts. Under certain conditions, it may also be argued that assuming equal marginal utility of money is more reasonable than assuming equal marginal utility of the public good. For example, if there are respondents with negative WTPs, the latter assumption would imply that some people have a negative marginal utility of money, which seems implausible. As discussed below, however, there are also conditions under which equal marginal utility of the public good is the most reasonable assumption of the two. Rejection of one of these assumptions does not imply that the other is correct. Maybe both are wrong. Given that we do not know how to measure interpersonally comparable cardinal utility, we are simply examining the empirical implications of replacing current practice with an alternative approach that is, in theory, equally valid. Our result is, in brief, that the two methods yield very different results.

2. Some central concepts

Below, we present a simple model explaining the main concepts which will be used in our calculations. The reader is referred to Brekke (1993, 1997), Medin (1999) for further details.

2.1. Aggregating net willingness to pay

Assume that there are \( n \) heterogeneous consumers with utility functions

\[
U_i = u_i(Y, E)
\]

for all \( i = (1, \ldots, n) \), where \( Y_i \) is individual \( i \)'s income, and \( E \) is a pure public good, which we will think of as being provided by the environment. \( E \) will be measured in physical units, for example, the estimated number of fish in a lake, or km\(^2\) of wilderness. Utility is assumed to be increasing in income and the public good.

Consider a project where the environmental good is increased by \( dE > 0 \), at a total cost \( \Sigma C_i \), where \( C_i \) is the amount of money person \( i \) has to pay if the project is implemented. To avoid complicating matters unnecessarily, we will assume that \( C_i = -dY_i = C \) for every \( i = (1, \ldots, n) \), i.e. every individual faces the same cost. A money measure \( dU_i^\gamma \) of the project’s net effect on \( i \)'s utility can be derived by differentiating Eq. (1) and dividing by the individual’s marginal utility of income, \( u_{iY} \):

\[
dU_i^\gamma = \frac{u_{iE}}{u_{iY}} dE - C = \text{WTP}_i dE - C
\]

where \( u_{iE} \) is the individual’s marginal utility of the public good. \( \text{WTP}_i \) is the individual’s marginal gross willingness to pay for the environmental improvement, and \( dU_i^\gamma \) is her net willingness to pay for the entire project (i.e. her maximum willingness to pay to ensure the environmental change
minus the costs she has to pay, \( C \).\(^3\) We will take as a starting point that \( C \) and \( dE \) are known. Individuals’ marginal rates of substitution \( u_{ik}/u_{ij} \) can in principle be observed by asking their willingness to pay for a one unit change in the environmental good; below, we will abstract from all practical problems of actually eliciting true willingness to pay, and simply assume that it can be observed. Assume further that social welfare \( W \) can be written as a differentiable function of individual utilities:

\[ W = V(U_i(Y_i, E), ..., U_n(Y_n, E)) \]  

(3)

Assume that the project is marginal for all individuals, in the sense that any changes in individuals’ marginal rates of substitution due to the project’s implementation are small enough to be disregarded. Then, the project’s effect on social welfare is

\[ dW = \sum_{i=1}^{n} \left( V_i(u_{iy} dY_i + u_{ik} dE) \right) = \sum_{i=1}^{n} \left( V_i u_{iy} dU_i \right) \]  

(4)

where \( V_i \) denotes the partial derivative of the social welfare function with respect to person \( i \)’s utility.

In applied cost–benefit analysis, explicit welfare weights \( V_i u_{iy} \) are rarely used (Little and Mirrlees, 1994). The most common procedure is simply to estimate social benefits as the unweighted sum of individuals’ net willingness to pay, which is consistent with Eq. (4) only if \( V_i u_{iy} = V_j u_{iy} \) for all \( i, j \).\(^4\) If the social welfare function is such that \( V_i \) should always be inversely proportional to \( i \)’s marginal utility of income, equality of welfare weights \( V_i u_{iy} \) will, of course, always hold.\(^5\) Generally, however, these welfare weights depend on the marginal utility of income. For example, if we use a utilitarian social welfare function, the net willingness to pay of individuals with a high marginal utility of income should get a larger weight in the cost–benefit analysis than others. However, as there is no generally accepted method to measure the marginal utility of income in an interpersonally comparable way (although some measurement attempts have certainly been made; see Frisch, 1932; van Praag, 1991), it is virtually impossible to verify the assumptions one makes on this parameter.

Some economists have argued that efficiency and distribution are two separate issues which could and should dealt with separately (e.g. Hicks, 1939), and defend unweighted aggregation of net willingness to pay on that ground. However, such reasoning usually relies on the assumption that costless lump-sum redistribution of income is available, which is rarely the case in large economies with private information (Hammond, 1979).\(^6\) Unweighted aggregation of net willingness to pay may further be justified by assuming that the status quo income distribution is socially optimal. Note, however, that not only would a decision-maker need an extreme degree of power to actually implement the income distribution of her choice, she would also encounter exactly the same problem of non-observability as the cost–benefit analyst. In order to identify the social optimum, the decision-maker too must know every individual’s interpersonally comparable marginal utility of income \( u_{iy} \). It is not clear how such information, which is unobservable for the analyst, could be available to decision-makers.\(^7\)

---

\(^3\) Note the difference between gross and net willingness to pay. In the latter, costs are deducted. Here, \( dU_i \) is \( i \)’s compensating variation for the entire project (costs taken into account). For a project in which \( dE < 0 \), \( dU_i \) would correspond to the equivalent rather than compensating variation. However, the difference between compensating variation and equivalent variation does not matter for marginal projects under standard neoclassical assumptions. For a critical discussion of the latter with respect to such welfare measures, see Bateman et al. (1997a).

\(^4\) Strictly speaking, it is consistent with Eq. (4) only if \( V_i u_{iy} = 1 \) for all \( i \). However, if we have instead \( V_i u_{iy} = K \) for all \( i \), where \( K > 0 \), the choice of the constant \( K \) would not matter for the ranking of projects.

\(^5\) This implies that if the marginal utility of income decreases in income, an individual’s interests receives more weight in the social welfare function the richer he is.

\(^6\) Dreze and Stern (1987) discuss several arguments for the view that distributional concerns can be disregarded in a cost–benefit analysis, and reject all of them. Their conclusion is that explicit welfare weights should be used in cost–benefit analysis, for example, by estimating the welfare weights implied by earlier policy decisions.

\(^7\) Further, a decision-maker who is omnipotent and omniscient enough to identify and implement an optimal income distribution would hardly need cost–benefit analysis.
2.2. Aggregating public good requirements

Above, individual utility changes were measured in monetary units. Quite equivalently, one may measure utility changes using some other numeraire than money. In our simple two-good model, thus, individual utility changes may alternatively be measured using units of the environmental good. Differentiation of the utility function (Eq. (1)) and dividing by $u_{iE}$ yields a measure of individual net benefits in environmental units, $dU^E_i$:

$$dU^E_i = -\frac{u_{iY}}{u_{iE}} C + dE = -\frac{1}{\text{WTP}_i} C + dE$$

(1/WTP,)C tells us how large an increase in the environmental good individual $i$ demands in order to be willing to pay $C$. We will call this measure $i$’s public good requirement. By differentiating Eq. (3) it can easily be seen that net social benefits may be expressed as a function of individuals’ net benefits in environmental units (Brekke, 1997):

$$dW = \sum_{i=1}^{n} (V_i u_{iE} dU^E_i)$$

Corresponding to the case of monetary units, simple unweighted aggregation of individuals’ net benefits in environmental units requires that $V_i u_{iE} = V_j u_{jE}$ for all $i, j$. Again, these weights depend on unobservable aspects of individual utility, and cannot generally be determined through knowledge of the social welfare function alone.

2.3. Marginal utility: two alternative assumptions

In the following, we will replace Eq. (3) by a utilitarian social welfare function:

$$W = \sum_{i=1}^{n} U_i$$

This simplifies the calculations and makes the analysis more transparent, allowing us to focus on utility measurement and comparability rather than the social objective function itself.\(^8\) Note, however, that although our empirical results would look different with another choice of social welfare function, the problems posed by unobservability of cardinal and interpersonally comparable utility would not disappear.

We have chosen to study two particularly simple assumptions concerning cardinal utility.

- **Alternative I**
  The marginal utility of income is equal for all. This would, for example, hold if a utilitarian and omniscient decision-maker had already redistributed income to implement a socially optimal income distribution. It would also hold if the level of income were equal for every consumer, and utility functions were of the following additively separable form:

$$U_i = f(Y_i) + g(E)$$

- **Alternative II**
  The marginal utility of the public good is equal for all. This would, for example, always hold, regardless of individual income levels, if utility functions were of the following additively separable form:

$$U_i = f_i(Y_i) + g(E)$$

Under Alternative I, a money measure of social benefits $dW^Y$ can be derived by unweighted aggregation of net willingness to pay (alternatively, unweighted aggregation of gross willingness to pay minus aggregate costs), such that the project is welfare-improving if $dW^Y > 0$.

$$dW^Y = \sum_{i=1}^{n} dU^Y_i = \left( \sum_{i=1}^{n} \frac{u_{iE}}{u_{iY}} \right) dE - nC$$

This corresponds to the net benefit estimate from a standard unweighted cost–benefit analysis.

Under Alternative II, unweighted aggregation of individual net benefits in environmental units yields a social benefit estimate $dW^E$, measured in environmental units. If this indicator is positive, the increase in the environmental good is large enough to justify the costs $nC$.

$$dW^E = \sum_{i=1}^{n} dU^E_i = n \left( \sum_{i=1}^{n} \frac{u_{iY}}{u_{iE}} \right) C$$

While the individual net benefit estimators $dU^Y_i$ and $dU^E_i$ will always have the same sign as the
individual’s utility change, regardless of the chosen numeraire, the aggregate benefit estimators \(dW^Y\) and \(dW^E\) may have different signs (Brekke, 1997). This is caused by the different assumptions about interpersonal comparability of cardinal utility underlying these indicators. Changing the way one operationalises interpersonal comparisons of utility is equivalent to a change of normative weights \(V_i\).

2.4. Maximum acceptable costs

Apart from the possibility of different signs, it is difficult to use \(dW^Y\) and \(dW^E\) directly to judge the empirical importance of a particular choice of aggregation method. They are measured in different units, and since each person may have a different ‘exchange rate’ between units (i.e. different marginal rates of substitution), it is not obvious which conversion rate one should use in an attempt to make them directly comparable. However, an interesting comparison can be made by looking at the per person costs which would leave the project with exactly zero net benefits using the two methods.

If we denote by \(C^*\) the per person cost that implies \(dW^Y = 0\), i.e. the maximum acceptable per person cost when equal marginal utility of income is assumed, we have from Eq. (10) that

\[
C^* = \frac{1}{n} \left( \sum_{i=1}^{n} \frac{u_i^E}{u_i^Y} \right) dE \tag{12}
\]

Similarly, we can denote by \(C^{**}\) the maximum allowable per person cost when equal marginal utility of the environmental good is assumed. \(C^{**}\) can be defined by

\[
C^{**} = \frac{n}{\sum_{i=1}^{n} (u_i^Y/u_i^E)} \ dE \tag{13}
\]

which is the per person cost implying exactly \(dW^E = 0\). \(C^*\) and \(C^{**}\) can both be regarded as monetary measures of aggregate benefits from increasing the public good supply.\(^9\)

An interesting indicator for the empirical importance of the choice of numeraire is \(C^*/C^{**}\). This will be the central indicator in our empirical results. We will denote this the maximum acceptable cost (MAC) ratio. Mathematically, the MAC ratio is given by

\[
MAC = \frac{C^*}{C^{**}} = \frac{1}{n} \left( \sum_{i=1}^{n} \frac{u_i^E}{u_i^Y} \right) \left( \frac{1}{\sum_{i=1}^{n} \frac{u_i^Y}{u_i^E}} \right) \tag{14}
\]

It can easily be seen from Eq. (14) that if all individuals have the same marginal rate of substitution between income and the environmental good, the MAC ratio = 1. Hence, if we are concerned only with ordinary market goods in a perfectly competitive market (assuming no corner solutions), the maximum acceptable per person costs will be the same using both measurement methods. However, whenever marginal rates of substitution differ between individuals, the numeraire problem will arise. This will be the case in a number of circumstances, e.g. when some goods are rationed (see Drèze, 1998); but for simplicity, we will concentrate on the case where the good is a public (environmental) good.

For later reference, we note that Eq. (14) could alternatively be expressed as (see also Eqs. (2) and (5)):

\[
MAC = \frac{1}{n} \left( \sum_{i=1}^{n} \frac{1}{WTP_i} \right) \left( \frac{1}{\sum_{i=1}^{n} WTP_i} \right) \tag{15}
\]

where

\[
WTP = \frac{1}{n} \sum_{i=1}^{n} WTP_i
\]

is the average of all respondents’ marginal willingness to pay, and

\[
WTP^{-1} = \frac{1}{n} \sum_{i=1}^{n} \frac{1}{WTP_i}
\]

is the average of all respondents’ inverse willingness to pay. Thus, even if we do not have direct information on people’s public good requirements, MAC ratios can be calculated using individual willingness to pay data only, assuming that the project is marginal.\(^10\) The MAC ratios reported in the next section were calculated using Eq. (15).

\(^9\) Note the similarity to the uniform variation measures proposed by Hammond (1994). The uniform compensating variation is defined as the total amount that society is willing to pay, in the form of a uniform poll tax on all individuals, in order to be allowed to move from the status quo to an alternative social state.

\(^10\) For a discussion of non-marginal projects, see Medin et al. (1998), Appendix 1.
Generally, a given numeraire will favour the interest of a person if the numeraire is of relatively low value to that person (Brekke, 1997). If, for example, a person only gets a relatively small utility increment from receiving an extra dollar, this person’s net benefits of a project expressed in money terms must become a large number. However, if the same person gets a relatively high utility increment of an extra unit of the public good, her net benefits expressed in environmental units must be a small number. Consequently, using money as the numeraire will favour those with a relatively high valuation of the environment compared with using environmental units as the numeraire.

The MAC ratio presupposes that costs are shared equally between individuals. Under this assumption, unless the project is a Pareto improvement, those who have the lowest valuation of the environmental good will always be the project’s opponents because the cost they have to pay always exceeds their willingness to pay. Thus, the benefit measure \( C^* \) will systematically give less weight to the interests of the project’s opponents than \( C^{**} \), implying that the MAC ratio \( \geq 1 \) will always hold. In other words, given the assumptions employed here, assuming that everybody has the same marginal utility of income will always favour the project, compared with the alternative assumption of equal marginal utility of the public good.

The indicators presented above presuppose that the project can be regarded as marginal. If the project is non-marginal in the sense that individuals’ marginal rates of substitution change significantly due to the project’s implementation, the above indicators provide only approximations. However, regarding the MAC ratio, errors caused by changes being non-marginal will generally go in both directions because the public good requirement is overestimated for those who have positive net benefit from the project and underestimated for those who have negative net benefit from the project. We thus cannot know a priori whether the MAC ratio is over- or underestimated in the case of a non-marginal public good change. In the special case of quasi-linear utility, however, the MAC ratio will be correct even if the public good change is non-marginal.\(^\text{12}\)

### 3. Empirical results

The disturbing part of the theoretical results discussed above is that one way of operationalising cardinal and interpersonally comparable utility systematically favours certain interest groups, compared with another, equally simple method. However, if this bias was of a small empirical magnitude, it might still not be of much practical importance. To examine the empirical significance for applied cost–benefit analysis of the choice of assumption regarding cardinal utilities, we have calculated the MAC ratio from seven contingent valuation studies using individual willingness to pay data. Unfortunately, our results indicate that the choice of numeraire (corresponding to a certain choice of assumption on cardinal utilities) may be extremely important.

#### 3.1. Data

The studies we have used are those of Loomis (1987), Navrud (1993), Bateman and Langford (1997), Bateman et al. (1995, 1997b), Magnussen et al. (1997), Strand and Wahl (1997). All the studies examine willingness to pay to avoid reductions in certain specified recreational services, ex-

---

\(^{11}\) It is possible to calculate similar indicators with other assumptions concerning the distribution of costs. A more generally applicable indicator may be denoted the TMAC ratio, the ratio of maximum allowable total costs, where \( \text{TMAC} = \frac{\text{MAC}}{C_i} \) if \( C_i = C \) for every \( i \). One will generally have a TMAC ratio \( > 1 \) if costs are shared ‘under-proportionally’ with individual marginal willingness to pay for the environmental good, i.e. \( C_i = K(u_i/u_i^E)^x \), where \( K \) and \( x \) are positive constants, and \( x < 1 \). Equal distribution of the costs is a special case of under-proportionally cost distribution. If costs are ‘over-proportionally’ distributed (\( x > 1 \)), we get a TMAC ratio \( \leq 1 \). However, overproportional cost sharing seems to be a fairly peculiar sharing rule, leading to severe incentive compatibility problems. Proportional cost distribution will give a TMAC ratio \( = 1 \) (see Brekke, 1993; Medin, 1999 for detail on this issue).

\(^{12}\) See Medin et al. (1998), Appendix 1, for a formalisation of this argument.
cept the Magnussen et al. (1997), Bateman and Langford (1997) surveys, which measure willingness to pay for an increase in recreational services. Several of the studies varied the survey design between subsamples; for those studies, results are reported for each subsample. The table below reports results for a total of 18 subsamples. A brief summary of each study is provided in Medin et al. (1998), (Appendix 2). All calculations are based on open-ended WTP data.

### 3.2. Outliers and zero bids

A well-known problem in contingent valuation research is that average WTP estimates, and thus $C^*$, can be sensitive to extremely high reported individual values. A common approach to such ‘outliers’ is simply to omit them from the data assuming that they are caused by errors, misunderstandings, strategic responses, or protest reactions on respondents’ part. On the other hand, if these ‘extreme’ observations do reflect respondents’ valuations, one may obviously understate average (and aggregate) willingness to pay by omitting them.

When environmental units are used as numeraire in the aggregation of individual welfare effects, one encounters a similar problem regarding extremely low observations of individual willingness to pay (low $u_{E}/u_{Y}$, or correspondingly, high $u_{Y}/u_{E}$). Taken literally, a zero willingness to pay for a public good implies that an infinite amount of the public good is required to compensate for the cost this particular person has to pay. Correspondingly, the social welfare loss measured in environmental units caused by forcing such persons to pay a positive cost is also counted as infinite, and the project will not be socially desirable, regardless of how much other persons are willing to pay.

Just as contingent valuation practitioners have to think carefully about how to treat extremely high and ‘infinite’ willingness to pay-bids, we have to consider how to treat zero willingness to pay-bids when using environmental units in the aggregation of individual values. These zero bids may be given different interpretations. One possible interpretation is that zero bidders have a positive, but very small willingness to pay. The most extreme assumption to make in our context, however, is to take the zero bids literally, since this implies that some respondents have infinite public good requirements. Correspondingly, the presence of a zero bid will always imply that neither $C^{**}$ nor the MAC-ratio is well defined.

In all studies except those by Loomis (1987) and Strand and Wahl (1997), before the willingness to pay question, respondents were faced with a ‘payment principle’ question concerning whether they were, in principle, willing to pay anything at all for the environmental change. Those who responded negatively to this question were not asked to state their WTP. It seems reasonable to assume that some of these ‘no-bidders’ did so to protest against the very idea of valuing the environment in monetary terms, or against accepting a personal responsibility for the problem at hand. The environmental good may still be important to such respondents’ welfare. Others might have responded ‘no’ because their marginal valuation was indeed zero. Finally, some respondents may

---

13 If changes are non-marginal, WTP to avoid loss measures the individual’s equivalent variation, while WTP for a gain measures the compensating variation.

14 In this elicitation procedure, respondents are asked questions about how much they are willing to pay, and are free to state whatever amount they want. An alternative procedure is that of dichotomous choice, in which respondents are asked yes/no-questions such as ‘would you be willing to pay X?’, where X is varied across the sample. See Medin et al. (1998), Appendix 2, for a description of the method used in each survey.

15 This discussion illustrates that the numeraire problem arises only if at least one person is worse off after the project’s implementation. If losers were in fact compensated, so that no one was worse off, the choice of numeraire would not matter for the conclusion regarding the social desirability of the project.

16 The problem on how to interpret extremely high WTP bids is, of course, in principle the same here as in an ordinary CV study. However, while the results from a CV study might be very sensitive to the inclusion or omission of high bids, our results are much more sensitive to the interpretation of small bids (see Section 3.4 for a discussion).
reply ‘no’ to the payment principle question because they actually have a negative WTP. However, it is very difficult to judge which respondents belong to which group, and what level of ‘true’ WTP, if any, such ‘no’ responses might correspond to.

Since the correct treatment of ‘zero’-bids is far from obvious, and the same is true for ‘no’-bids, we have calculated two different versions of MAC ratios. The first version is based on the assumption that all ‘zero’- and ‘no’-bids reflect very small, but positive WTPs. How small is determined, somewhat arbitrarily, as 5% of the lowest strictly positive bid reported in that survey. This assumption probably implies an underestimation of some respondents’ public good requirements because a WTP of zero, taken literally, would imply an infinite public good requirement and thus an infinite MAC ratio. Since our assumption here does not allow anyone to have a lower WTP than 5% of the lowest strictly positive observation, MAC ratios will not go to infinity. However, for ‘no’-bidders whose ‘true’, but unobserved valuations are significantly higher than zero, our assumption may imply that MAC ratios are overestimated: Very small observations tend to yield large MAC ratios, while medium-sized observations have much less dramatic impact on MAC ratios.

The second version of the MAC ratio is calculated after omitting all ‘no’- and ‘zero’-bids from the datasets. This amounts to an assumption that the ‘true’ valuations underlying such observations are distributed in exactly the same way as the set of strictly positive observations. This yields conservative estimates of MAC ratios compared with the previous version since values close to zero tend to imply high MAC ratios.

It turns out that the assumptions one makes on ‘no’- and ‘zero’-bids are quite essential for the magnitude of the MAC ratios. Ideally, then, we should have more information about these observations, but since the surveys were designed with elicitation of monetary values in mind, such information is not available. Below, we will focus on the second, most conservative version.

### 3.3. MAC ratios: results

Table 1 reports our empirical results. Column 1 shows the number of observations in each study, while column 2 reports the percentage no- and zero-bids. No- and zero-bids were lumped together in the table because we could not separate them in all surveys. For those studies where more detailed information was available, the proportion of the total sample who refused the payment principle ranged from 6 to almost 14% in the use value studies (Bateman et al., 1995, 1997b) and was over 46% in the study of non-user values (Bateman and Langford, 1997). In each case, respondents were asked to state the reasons underpinning this response. Analysis of this data showed that (even if we classify those who failed to give such a reason as being ‘protest’ voters) as a proportion of their respective total samples, only between 0.6 and 1.75% of respondents in the use value surveys could be classed as protestors rising to 7% in the non-user survey. These low rates suggest that the surveys were, at least in this respect, successful in that they were considered plausible and credible by respondents.

---

17 Using, for example, 1% of the strictly positive bid yields dramatically higher MAC ratios (see Medin, 1999).

18 The assumption that utility is increasing in both income and the public good implies that no respondent has a negative WTP. If in fact some of the no or zero bids reflect negative ‘true’ WTPs, some individuals must have negative marginal utility of either money or the public good. The former is inconsistent with the assumption of equal marginal utility of money, and thus employment of $C^*$ as a welfare estimate, while the latter is inconsistent with equal marginal utility of the public good, and thus implies that $C^{**}$ is not a correct welfare estimate.

19 If zero- and no-bids in fact reflect protesters with positive ‘true’ WTPs, using the ‘true’ WTPs in the calculations could have produced either higher or lower MAC ratios than those obtained by omitting zero bids. Roughly, if zero bidders’ ‘true’ WTP values were mainly in the middle range, omission of zero bids would produce too high MAC ratios; if ‘true’ values were very high and/or very low, omission of zero bids would yield too low MAC ratios.

20 This is not unexpected for, as Carson (1999) discusses, passive use surveys by their nature encompass a large proportion of respondents who do not value the good in question.
Table 1. *MAC* ratios (ratio of maximum acceptable costs) under different assumptions on zero- and no-bids and the single highest or lowest bid. \(^{21,22}\) (Note that for *MAC* ratios < 10, values are reported with one decimal)\(^a\)

<table>
<thead>
<tr>
<th>Survey</th>
<th>MAC-ratio</th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td></td>
<td><em>N</em></td>
<td>Zero- and no-bids, per cent of <em>N</em></td>
<td>Each zero- and no-bid = 5% of smallest positive bid</td>
<td>All zero- and no-bids removed</td>
<td>Zero- and no-bids and smallest bid removed</td>
</tr>
<tr>
<td>Bateman et al., 1995</td>
<td>Subsample 1</td>
<td>846 15</td>
<td>20,202</td>
<td>38</td>
<td>27</td>
</tr>
<tr>
<td></td>
<td>Subsample 2</td>
<td>2051 15</td>
<td>22,434</td>
<td>11</td>
<td>5.9</td>
</tr>
<tr>
<td>Bateman and Langford, 1997</td>
<td>Subsample 1</td>
<td>93 37</td>
<td>8,647</td>
<td>70</td>
<td>40</td>
</tr>
<tr>
<td></td>
<td>Subsample 2</td>
<td>90 63</td>
<td>378</td>
<td>6.7</td>
<td>4.5</td>
</tr>
<tr>
<td></td>
<td>Subsample 3</td>
<td>88 7</td>
<td>93</td>
<td>5.2</td>
<td>4.5</td>
</tr>
<tr>
<td></td>
<td>Subsample 4</td>
<td>80 16</td>
<td>5,894</td>
<td>83</td>
<td>53</td>
</tr>
<tr>
<td>Bateman et al., 1997a</td>
<td>Subsample 1</td>
<td>143 18</td>
<td>11,598</td>
<td>169</td>
<td>138</td>
</tr>
<tr>
<td></td>
<td>Subsample 2</td>
<td>126 10</td>
<td>18,003</td>
<td>307</td>
<td>226</td>
</tr>
<tr>
<td>Loomis, 1987</td>
<td>78 17</td>
<td>82</td>
<td>3.2</td>
<td>2.8</td>
<td>3.0</td>
</tr>
<tr>
<td>Magnussen et al., 1997</td>
<td>Subsample M1</td>
<td>143 60</td>
<td>101</td>
<td>3.1</td>
<td>2.8</td>
</tr>
<tr>
<td></td>
<td>Subsample M2</td>
<td>139 59</td>
<td>34</td>
<td>2.1</td>
<td>2.1</td>
</tr>
<tr>
<td></td>
<td>Subsample S1</td>
<td>139 47</td>
<td>97</td>
<td>2.9</td>
<td>2.7</td>
</tr>
<tr>
<td></td>
<td>Subsample S2</td>
<td>132 49</td>
<td>87</td>
<td>2.3</td>
<td>2.1</td>
</tr>
<tr>
<td>Navrud, 1993</td>
<td>161 32</td>
<td>806</td>
<td>4.2</td>
<td>2.6</td>
<td>4.1</td>
</tr>
<tr>
<td>Strand and Wahl, 1997</td>
<td>Subsample 1</td>
<td>140 14</td>
<td>23</td>
<td>1.8</td>
<td>1.8</td>
</tr>
<tr>
<td></td>
<td>Subsample 2</td>
<td>140 13</td>
<td>30</td>
<td>1.8</td>
<td>1.8</td>
</tr>
<tr>
<td></td>
<td>Subsample 3</td>
<td>138 28</td>
<td>60</td>
<td>1.8</td>
<td>1.7</td>
</tr>
<tr>
<td></td>
<td>Subsample 4</td>
<td>145 21</td>
<td>69</td>
<td>2.3</td>
<td>2.2</td>
</tr>
</tbody>
</table>

\(^a\) *N*, number of respondents; zero-bids, = respondents reporting a zero WTP; no-bids, = respondents responding ‘no’ to the payment principle question.

\(^{21}\) If choice of aggregation unit does not matter, the *MAC* ratio = 1.

\(^{22}\) Most surveys also include a number of respondents who say they don’t know, or who do not answer the WTP question. These respondents have been omitted from the data in all studies, except from the Magnussen et al. (1997) survey, where we could not distinguish such respondents from zero bidders.
Column 3 reports the first version of the MAC ratios assuming that no- and zero-bids reflect a very small, but positive WTP (5% of the lowest bid in that survey). This yields extremely high MAC ratios, ranging from 23 (subsample 1, Strand and Wahl, 1997) to 22 434 (!) (subsample 2, Bateman et al., 1995). Thus, if one accepts the assumptions underlying the first version of MAC, the estimated aggregate monetary benefit indicator is reduced by a factor of up to about 22 000 by replacing the conventional assumption of equal marginal utility of income by an assumption of equal marginal utility of the environmental good.

Column 4 reports the more conservative version of MAC ratios, i.e. after all zero- and no-bids are omitted from the dataset. This approach yields considerably less extreme MAC ratios varying between approximately 2 (the four subsamples in Strand and Wahl, 1997 and two subsamples from Magnussen et al., 1997) and 307 (subsample 2 in Bateman and Langford, 1997). However, even the smallest MAC ratios of approximately 2 are, in one sense, large, since they imply that using environmental units as numeraire instead of money almost halves the maximum acceptable per person cost which leaves the project socially desirable. In the study with highest MAC ratio, the maximum acceptable per person cost varies with a factor of up to 307, depending on whether one employs an assumption of, respectively, equal marginal utility of income or of the environmental good.

Table 1 shows that the MAC ratios emerging from the data of Bateman et al. (1995, 1997b) and subsample 1 and 4 in Bateman and Langford (1997) are considerably higher than those of the other surveys. One reason for this might be that these studies contain several very small WTP bids in the sense that the ratio between the smallest and the highest bid reported is large. In subsample 2 from Bateman and Langford (1997), for example, the highest bid reported was £1000, while the lowest strictly positive bid was £0.005. Thus, in this survey, using money as the numeraire implies that the net benefits of the person with the highest WTP is weighted 200 000 times more than the net benefits of the person with the lowest WTP, as compared with the procedure of using environmental units as the numeraire (see Brekke, 1993).

3.4. Sensitivity to extreme observations

Column 5 reports MAC ratios when the single smallest WTP bid is removed from the data, in addition to removing the zero- and no-bidders. It turns out that the MAC ratio can be surprisingly sensitive to such removal of one single observation. Particularly interesting is subsample 2 from Bateman et al. (1995) where despite the unusually large sample size of about 1800 observations (after the removal of all zero- and no-bids), the MAC-ratio is almost halved by removing the single smallest strictly positive WTP bid from the data.

To understand this phenomenon, recall the expression for the MAC ratio used in Eq. (15), \( \text{MAC} = \frac{(WTP)}{(WTP)} \). Somewhat imprecisely, one might say that the effect of omitting one observation from the dataset depends on whether this observation’s relative impact on these two averages is very different, or rather, asymmetric. Removing a very small bid may have a large impact on \( \overline{WTP} \), while \( \overline{WTP} \) may be quite unaffected. \(^{24}\) Removing a high bid, on the other hand, is likely to affect \( \overline{WTP} \) much more than \( \overline{WTP} \). Column 6 reports MAC ratios when the single highest WTP bid is removed from the data, in addition to removing the zero- and no-bidders. We see that in the surveys examined here, removing the single highest bid did not affect MAC ratios nearly as much as removing the smallest strictly positive bid.

The above illustrates the importance of obtaining accurate information on low WTP bids. To calculate correct MAC ratios, the ability to distin-

\(^{23}\) To understand this seemingly bizarre result, recall that if zero bids are taken literally, the MAC ratio goes to infinity, since this implies infinite public good requirements.

\(^{24}\) For example, imagine a survey where \( N = 10 \). Say that nine respondents report a WTP of $10, while one reports $0.05. Omitting the latter observation would change WTP from 9.005 to 10, while WTP \(^{-1}\) would change from 2.09 to 0.1. Correspondingly, the MAC ratio would change from 18.8 to 1.
guish between small WTP numbers is essential. It is not obvious that respondents are capable of such fine distinction. Further, existing CV surveys have, of course, not been designed with this problem in mind. One way to alleviate the problem may be to ask respondents in CV surveys to report their public good requirements in addition to the usual WTP question.\textsuperscript{25}

3.5. Social net benefits

Some of the studies also estimated the costs of the project. For example, in Magnussen et al. (1997) (subsamples S1 and S2), the annual total costs were estimated at between $0.38 and 0.51 million.\textsuperscript{26} These costs were to be divided between approximately 8800 households, implying annual costs per household of $43–65. The average monetary benefit per household (assuming equal marginal utility of income) was estimated at between $111 and 132 per annum. Thus, benefits appeared to substantially exceed costs, and the project was deemed to be socially desirable. Would the policy recommendation of this study be changed if one had, instead, assumed equal marginal utility of the environmental good? If all zero- and no-bids are removed from the dataset, corresponding to the ‘conservative’ MAC ratios reported in column 4, maximum acceptable per household costs ($C^{**}$) will be between $85 and 97, which is still above the estimated costs; and the project still yields positive social benefits.\textsuperscript{27} Thus, in this particular case, the conclusion of the analysis seems robust.

Note, however, that since all zero-bids are omitted, many respondents who would most likely get a negative net benefit have been excluded from the analysis. Thus, both $C^*$ and $C^{**}$ may overestimate the projects’ net benefits. Further, if all no- and zero bids are included, and interpreted as 5\% of the lowest strictly positive bid, the $C^{**}$ estimate is reduced to $1.3, which is far less than the estimated annual per household costs, and the conclusion of the cost-benefit analysis is changed.

We have also calculated the net social benefits of the projects studied in Loomis (1987), Navrud (1993), Strand and Wahl (1997).\textsuperscript{28} Estimated net benefits are positive for all these projects when equal marginal utility of income is assumed. If one assumes, instead, equal marginal utility of the public good, net benefits remain positive for the projects in Loomis (1987), Navrud (1993), Strand and Wahl (1997) subsample 1 and 2, provided that no- and zero-bids are removed from the datasets. For Strand and Wahl (1997) subsample 3 and 4, net benefits are negative if a 7\% discount rate is assumed, but positive if the discount rate is set to 3.5\%. Considering the high MAC ratios, the robustness of the conclusions may be somewhat surprising. Note, however, that when no- and zero-bids are included and counted as 5\% of the lowest strictly positive bid, net benefits are negative for all projects.

3.6. Can the results be generalised?

The MAC ratios reported in Table 1 indicate that the way one compares utility between persons may be extremely important in applied cost–benefit analysis. However, the generality of our results depends, of course, on the extent to which the studies we have used are representative of the ‘typical’ response pattern in CVM studies. Also, some of the simplifications employed in our theoretical model may not hold in practice. One objection is concerned with the fact that our theoretical model assumes only marginal changes in the public good supply, in the sense that any changes in

\begin{footnotesize}
\begin{enumerate}
\item A public good requirement question could, for example, be formulated as follows, if a proposed program to increase the salmon population in river X were implemented, you would have to pay $10 per year in increased income taxes. What is the lowest percentage increase in the salmon population which would make you willing to pay this cost?
\item We assume the exchange rate between NKR and USD to be 7.828 (31 August, 1998).
\item ‘Person’ is here used interchangeably with ‘households’, thus, we disregard intra-household conflicts of interest in this example.
\end{enumerate}
\end{footnotesize}
individuals’ marginal rates of substitution between the public good and income due to the project can be disregarded. In practice, for many environmental projects, this will not hold for at least some individuals. Regarding the studies mentioned in Table 1, this seems particularly questionable for subsample 1 in Bateman and Langford (1997), and for both subsamples in Bateman et al. (1995). However, as mentioned above, one cannot know a priori whether the calculated MAC ratios will be too high or too low if the public good change is in fact non-marginal, as the errors will generally go in both directions; and in the special case of quasi-linear utility functions, Eq. (15) can be used to calculate MAC ratios correctly, even if willingness to pay data does not represent marginal changes.

It is also somewhat difficult intuitively to understand what measuring in ‘environmental units’ really means. Some may dismiss our results on the grounds that the environmental unit has not been well-defined enough in some or all of the studies we have used. All the surveys do consider measurement problems and problems related to giving a precise definition of the public good, and all authors appear to have given serious consideration to this problem in their survey design. It is certainly often difficult to specify the environmental good in a precise enough way, and this is a problem which all contingent valuation studies have to grapple with. Since the surveys we have used were designed with the income numeraire in mind, questionnaires cannot be expected to have focused on aspects which are considerably more important in our context than in the traditional context. However, since MAC ratios can be expressed using only monetary valuations (see Eq. (15)), respondents do not in practice need to express their valuations using environmental units (i.e. their public good requirements). The issue of defining and understanding what ‘an environmental unit’ means only represents a problem in our context to the extent that misunderstandings regarding this prevented respondents from actually reporting their true WTPs (in monetary terms). If

the overall pattern of responses in the surveys we have used are typical for CVM surveys, our results on the MAC ratios will also be typical.

4. Conclusions

The results reported above indicate that aggregate social benefit estimates may be extremely sensitive to alternative ways of comparing different individuals’ utility changes. Making non-verifiable assumptions on cardinal and interpersonally comparable aspects of individuals’ utility functions introduces a non-negligible element of arbitrariness into cost–benefit analysis.

The numerical results are very sensitive to changes in the treatment of extreme WTP observations, in particular no- and zero bids. However, even our most conservative estimates indicate that if one assumes equal marginal utility of the public good for everyone, instead of the usual assumption of equal marginal utility of income, aggregate monetary benefit estimates are reduced by a factor of between 2 and 307.

We wish to stress that our aim has not been to argue in favour of one or the other method of making utility interpersonally comparable. Rather, the lesson from our study is that operationalisation of interpersonal utility comparisons is extremely important for empirical cost–benefit analysis. In the light of this, we believe that cost–benefit practitioners should take on a much more active attitude towards this issue.

One alternative is to take considerably more care when interpreting aggregate social benefit estimates, and emphasise that what one reports in a cost–benefit analysis is not aggregate utility, but aggregate willingness to pay. At the present state of art, we simply do not know whether the two coincide reasonably well.

As a practical approach, Johansson (1998) suggests to report costs and benefits in monetary units for subgroups of the population. However, this will only reduce the aggregation problem if one has reasons to believe that the marginal utility of income is similar within each group (Nyborg, 2000). Since marginal utilities are unobservable, such judgements must necessarily be subjective.

---

29 See Medin et al. (1998), Appendix 2, for details.
Another alternative is to report the number of individuals with positive and negative net benefits, respectively, which would in principle yield the same result as a referendum. This alternative is based on ordinal information only, and does not require particular assumptions on marginal utilities. Nor does it require WTP data; a ‘yes’ or ‘no’ would be sufficient. However, the result of such a voting procedure does not necessarily reflect the sign of the change in social welfare. To actually estimate aggregate social benefits, welfare economists must face the question of utility comparisons explicitly, and address the issue of which methods are actually defensible.

Acknowledgements

We wish to thank John Loomis, Kristin Magnussen and Olver Bergland who have kindly provided us with disaggregated data from their own research, and to Kjell Arne Brekke and Rolf Aaberge for comments.

References


Nyborg, K., 2000. Project analysis as input to public debate: environmental valuation versus physical unit indicators. Ecol. Econ. 34 (3), 393–408.


