TØI report
491/2000

PB2002-105072

Valuation of nonmarket goods for use in cost-benefit analyses

Methodological issues

Dissertation for the Degree of Doctor Scientiarum
Department of Economics and Social Sciences
Agricultural University of Norway
1999

Kjartan Sælensminde

ISSN 0802-0175
ISBN 82-480-0166-0

Oslo, september 2000

Title: Valuation of nonmarket goods for use in cost-benefit analyses: Methodological issues

Author(s): *Kjartan Sælensminde*

TØI report 491/2000
Oslo, September 2000
121 pages
ISBN 82-480-0166-0
ISSN 0802-0175

Financed by:
Norwegian Research Council and Institute of Transport Economics

Project: 2270 Valuation of nonmarket goods for use in cost-benefit analyses

Project manager: Kjartan Sælensminde

Quality manager: Marika Kolbenstvedt

Key words:
Cost-benefit analyses; Stated Preference valuation methods; Inconsistent choices; Lexicographic choices; Interaction effects

Summary:
The subject of this thesis is how to value nonmarket goods when the valuations are to be used in cost-benefit analysis. Two different Stated Preference valuation methods are used in the five papers that constitute the thesis, Stated Choice and Contingent Valuation. A main result in the thesis is that Stated Choice is a valuation method that may collect so much data noise that the valuation results are affected. Therefore, the elicitation methods should be made sufficiently simple that people are able to state their preferences in response to the presented choice set. Another main result is that a simultaneous valuation procedure, accounting for interactions between the nonmarket goods included in cost-benefit analysis for road investments, significantly reduces the valuations compared to a separate valuation procedure. In the methodological discussion of how to value nonmarket goods for use in cost-benefit analysis, the choice of valuation context seems, therefore, to be an important issue.

Tittel: Valuation of nonmarket goods for use in cost-benefit analyses: Methodological issues

Forfatter(e) *Kjartan Sælensminde*

TØI rapport 491/2000
Oslo: September 2000
121 sider
ISBN 82-480-0166-0
ISSN 0802-0175

Finansieringskilde:
Norges Forskningsråd og Transportøkonomisk institutt

Prosjekt: 2270 Verdsetting av ikke-markedsgoder for bruk i nytte-kostnadsanalyser

Prosjektleder: Kjartan Sælensminde

Kvalitetsansvarlig: Marika Kolbenstvedt

Emneord:
Nytte-kostnadsanalyser; hypotetiske verdsettingsmetoder; inkonsistente valg; leksikografiske valg; interaksjonseffekter

Sammendrag:
Temaet for avhandlingen er hvordan man kan verdsette goder som ikke omsettes i markeder dersom disse verdsettingene skal brukes i nytte-kostnadsanalyser. To ulike hypotetiske verdsettingsmetoder brukes i de fem artiklene som utgjør avhandlingen; samvalganalyse (Stated Choice) og "betinget verdsetting" (Contingent Valuation). En hovedkonklusjon fra avhandlingen er at samvalganalyse er en så komplisert verdsettingsmetode at respondentene kan få problemer med å velge konsistent. Verdsettingsmetoder bør derfor gjøres så enkle at respondentene blir i stand til å tilkjennegi sine preferanser i de valgene som presenteres. En annen hovedkonklusjon er at en simultan verdsettingsstudie, som tar hensyn til interaksjonseffekter mellom de ikke-markedsgodene som inkluderes i NKA av veginvesteringer, gir signifikant lavere verdsetting enn en får ved separate verdsettingsstudier. For å få riktig relativ verdsetting av reisetid, trafikksikkerhet, støy og luftforurensning i NKA av veginvesteringsprosjekter må disse ikke-markedsgodene derfor verdsettes simultant i samme studie.

Language of report: English

The report can be ordered from:
Institute of Transport Economics, the library
P O Box 6110 Etterstad N-0602 Oslo, Norway
Telephone +47 22 57 38 00 Telefax +47 22 57 02 90
Price NOK 200

Rapporten kan bestilles fra:
Transportøkonomisk institutt, biblioteket
Postboks 6110 Etterstad, 0602 Oslo
Telefon 22 57 38 00 - Telefaks 22 57 02 90
Pris 200

Preface by the Institute of Transport Economics

This report contains a dissertation for the degree of Doctor Scientiarum at the Agricultural University of Norway. It is based on five papers; one published in the scientific journal *Transportation Research D*, the other submitted for publication.

The Institute of Transport Economics would like to thank the Norwegian Research Council, the Ministry of Transport and Communication, the Public Roads Administration of Norway and the Norwegian Railways for sponsoring the research that made this dissertation possible.

This is the first dissertation written at the Institute of Transport Economics that conducts methodological studies related to valuation of nonmarket goods for use in cost-benefit analyses. In this dissertation two Stated Preference methods, Stated Choice (Conjoint Analysis) and Contingent Valuation, has been applied. The cases and the data are related to valuation of travel time, environmental impacts and traffic safety included in cost-benefit analyses of road investments, but the results and the conclusions are relevant to valuation of nonmarket goods and considerations about cost-benefit analyses in general.

The permission of Elsevier Science Ltd, publisher of Transportation Research Part D to reprint the first paper of this dissertation is gratefully acknowledged.

The Institute of Transport Economics would like to thank the Agrivultural University of Norway for evaluating this dissertation. This kind of external evaluation provides a check on the quality of the research done at the Institute, which is important both from the Institute's point of view and in relation to our sponsoring partners.

Oslo, September 2000
INSTITUTE OF TRANSPORT ECONOMICS

Knut Østmoe
Managing Director

Marika Kolbenstvedt
Head of Department

Authors Preface

Several people have contributed in one way or another to the work presented in this dissertation. I would like to thank my supervisor, Ståle Navrud, for encouragement and professional guidance. In addition I have had valuable help from more experienced colleagues at Institute of Transport Economics. I would in particular mention Ronny Klæboe, Rune Elvik and Kreg Anders Lindberg who have given valuable comments to all five papers in the dissertation and Marika Kolbenstvedt and Knut Østmoe who encouraged me and gave me the opportunity to start this doctoral study.

Among other people that have contributed I would like to thank Staffan Algers for help with the modelling in paper 4, Olvar Bergland for comments to paper 2 and 5 and moral support during my stay at University of Sydney, Frode Hammer whose collaboration was essential to the work reported in paper 1, David Hensher for his hospitality and comments to papers 2, 3 and 5, Jan-Erik Lindjord for programming and discussions related to paper 2 and 3, Jordan Louviere for a wider perspective on the topics of paper 2 and 5, and Farideh Ramjerdi for exciting discussions and inspiration related to all five papers.

Finally, I would like to thank the Norwegian Research Council and Institute of Transport Economics for financing the work on my dissertation. In addition, I would like to thank the Ministry of Transport and Communication, the Public Roads Administration and the Norwegian Railways, who financed the extra data collection needed to investigate possible interaction effects connected to the Norwegian value of time study.

Oslo, September 2000

Kjartan Sælensminde

Contents

TØI report 491/2000
Author: Kjartan Sælensminde
Oslo 2000, 121 pages English language

Summary:

Valuation of nonmarket goods for use in cost-benefit analyses: Methodological issues

The subject of this thesis is how to value nonmarket goods when the valuations are to be used in cost-benefit analysis (CBA). To be more specific, the aim of the thesis is to investigate i) how the Stated Choice method can be used in valuation of environmental goods, ii) how methodological problems related to Stated Choice influence the valuation results, and iii) how the choice context influences the valuations of nonmarket goods. The cases and the data are related to valuation of travel time, environmental impacts and traffic safety included in CBA of road investments, but the results and the conclusions are relevant to valuation of nonmarket goods and considerations about CBA in general.

Background and limitation of the focus on Stated Preference methods

In Norway, CBA are carried out for all road investment projects, and a benefit/cost ratio is calculated for each project. An optimal decision rule would be to rank the projects according to a decreasing benefit/cost ratio and then carry out the projects in that order, until the budget is depleted. However, several recent Scandinavian studies show very weak – if any – association between the priority ranking assigned to a given road investment project and the project's benefit/cost ratio. If doubt amongst decision makers about the accuracy of the valuation of the goods included in CBA partly can explain this weak association, improving the valuation methods could be one way to strengthen it.

If a good is to be included in a CBA, the good has to have a monetary value associated with it. However, some of the goods decision makers want to include in CBA are not directly bought and sold in actual markets. Such nonmarket goods thus have to be valued either in Stated Preference (SP) studies or in Revealed Preference (RP) studies. SP data are collected in surveys where people in a constructed market are asked hypothetical questions and/or are presented hypothetical choice tasks. RP data, on the other hand, are based on people's actual choices in real (surrogate) markets. For example, the labour or housing market can be used to indirectly value safety or environmental goods (i.e. hedonic pricing techniques).

In this dissertation the focus is on SP methods. The main reason for this focus is that SP methods in many cases are the only means to construct the relevant

valuation contexts and, therefore, the only way to measure people's preferences for nonmarket goods.

Methodological issues

Two different SP valuation methods are used in the five papers that constitute this thesis, Stated Choice and Contingent Valuation. Stated Choice is, briefly described, a method in which nonmarket goods are assessed relative to each other and not in absolute amounts as with the more common Contingent Valuation method.

Data from two different national Norwegian surveys consisting of in-person interviews are used in the papers. The first study, conducted in 1993-94, is an environmental study valuing air pollution and noise related to road traffic. The second study, conducted in 1995-96, is a so-called "value of time study" measuring travellers' willingness to pay for a reduction in travel time.

In paper 1 environmental impacts of urban traffic are valued by Stated Choice. The results from this study have been used by the Norwegian Public Roads Administration in their CBA since 1995. Paper 1 points to the fact that considerable uncertainty remains with regard to:

- the impact on the valuations of interaction effects and other general methodological problems with Stated Preference methods because the respondents in this study only had the possibility to state their preferences for environmental goods and not other goods included in the CBA; and
- problems related to the complex choice situation of Stated Choice.

Paper 2 shows that inconsistent choices (i.e. violations of the transitivity axiom) commonly occur in several Stated Choice tasks and have a significant impact on the valuation of reduced travel time. It is shown that different abilities (i.e. level of education) cause inconsistent choices. The occurrence of inconsistent choices is shown to be largest in the beginning of the choice sequence and is reduced for subsequent choices. As a conclusion to the results in paper 2 it is suggested that respondents may need more training and help to choose consistently in Stated Choice studies.

Paper 3 investigates the causes and consequences of lexicographic choices (i.e. violation of the continuity axiom) in Stated Choice studies. By lexicographic choices we mean a set of choices in which the respondent consistently chooses the alternative that is best with respect to one particular attribute. The analyses in paper 3 shows that lexicographic choices commonly occur in Stated Choice tasks and that lexicographic choices are not the same as lexicographic preferences because lexicographic choices are partly a result of:

- Study designs with too large differences between the presented alternatives. Such choices give less information about preferences, but this is normally not a serious modelling problem;
- Simplification of the choice task. Such lexicographic choices contribute to the large variance in Stated Choice data and might therefore have a significant

impact on the valuation of nonmarket goods if it is not corrected in the analysis.

Paper 4 makes use of the logit scaling approach to handle variance increases caused by inconsistent choices in Stated Choice data. The scaling approach is a statistical estimation method that allows for differences in the amount of unexplained variance in different types of data, which can then be used together in the analysis. The amount of unexplained variance is shown to increase as the number of inconsistent choices increases. The main conclusion from the analyses in paper 4 is that scaling due to inconsistencies significantly improves the models and reduces the valuations of travel time. In addition, the scaling approach makes the valuations of travel time from the Stated Choice data more consistent with the valuations from Contingent Valuation data included in the same study. Another important conclusion is that scaling due to education (cf. paper 2) gives no improvement of the model.

The subject of paper 5 is possible interaction effects in valuation of nonmarket goods included in CBA. CBA for road investments includes nonmarket goods such as travel time savings, traffic safety, noise and air pollution. Traditionally, these nonmarket goods are valued through separate willingness-to-pay studies without any attention to interactions between them. Paper 5 shows that a simultaneous valuation procedure, accounting for interactions between the nonmarket goods included in CBA for road investments, significantly reduces the valuations of travel time savings, noise and air-pollution compared to a separate valuation procedure. The results presented in paper 5 show that the choice of valuation context is an important issue in the overall methodological discussion of how to value nonmarket goods for use in CBA.

Conclusions, questions and further research needs

The results in paper 5 show that when one wishes to include nonmarket goods such as travel time, traffic safety, noise and air-pollution in CBA for road investment projects, these nonmarket goods should be valued simultaneously in the same study. However, there are reasons to question whether the relevant shadow prices for use in CBA should be based on short-term valuations (as in today's valuation practice) or more long-term valuations. A more long-term valuation procedure can ensure that:

- the valuation context is in better accordance with a desired future context (i.e. the society people want) instead of today's context which may be undesired and too limited to present in a questionnaire as an acceptable choice set where people are to state their preferences; and
- the budget allocation process is completed, as is assumed in standard economic theory.

In paper 2 and 3 it is shown that Stated Choice is a valuation method that may collect so much data noise that the valuation results are affected. Therefore, it should be a goal for practitioners to make the elicitation methods sufficiently simple that people are able to state their preferences in response to the choice set they are presented.

Paper 4 shows that the scaling approach may be a way to handle variance differences due to inconsistent choices and that accounting for such heteroscedasticity may lower the estimated value of travel time. Similar results are obtained in other studies accounting for heteroscedasticity due to taste/preference variations and heteroscedasticity due to variations connected with different elicitation methods. Accounting for variance differences in SP data seems to be a crucial issue for future research, and analyses that simultaneously correct for heteroscedasticity due to inconsistencies, taste differences and other sources are needed.

Sammendrag:

Verdsetting av ikke-markedsgoder for bruk i nytte-kostnadsanalyser: Metodeproblemer

Temaet for denne avhandlingen er hvordan man kan verdsette goder som ikke omsettes i markeder dersom disse verdsettingene skal brukes i nytte-kostnadsanalyser (NKA). Eller for å være mer konkret, målsettingen er å undersøke: i) hvordan metoden "Stated Choice", som er basert på hypotetiske valg, kan brukes ved verdsetting av miljøgoder, ii) hvordan metodiske problemer relatert til "Stated Choice" påvirker verdsettingsresultatene og iii) hvordan verdsettingskonteksten påvirker verdsettingen av ikke-markedsgoder. For å belyse disse spørsmålene brukes data der reisetid, miljøkonsekvenser og trafikksikkerhet er verdsatt for å inkluderes i NKA av vegprosjekter. Man skal være forsiktig med generaliseringer, men resultatene og konklusjonene fra avhandlingen kan også ha gyldighet som generelle betraktninger om verdsetting av ikke-markedsgoder og NKA.

Bakgrunn og begrensning av fokus til "Stated Preference" metoder

I Norge gjør vegmyndighetene NKA av alle planlagte veginvesteringsprosjekter og beregner en nytte/kostnadsbrøk for hvert prosjekt. Basert på NKA ville en optimal beslutningsregel være å rangere vegprosjektene etter en avtagende nytte/kostnadsbrøk og deretter iverksette prosjektene i denne rekkefølgen helt til vegbudsjettet var brukt opp. Mange studier viser likevel liten eller ingen sammenheng mellom rekkefølgen vegprosjekter utføres i og prosjektenes nytte/kostnadsbrøk. Dersom dette delvis skyldes at beslutningstakere tviler på riktigheten av verdsettingen av ikke-markedsgodene, kan forbedring av verdsettingsmetodene styrke samsvaret mellom prosjektenes nytte/kostnadsbrøk og utføringsrekkefølgen.

Dersom et gode skal inkluderes i en NKA må godet ha en verdi som kan måles i penger fordi dette er måleenheten i NKA. Noen av godene som beslutningstakere ønsker å inkludere i NKA blir imidlertid ikke omsatt i faktiske markeder. Slike ikke-markedsgoder må derfor verdsettes enten i hypotetiske valgstudier ("Stated Preference" eller SP studier) eller i faktiske valgstudier ("Revealed Preference" eller RP studier). SP-data innsamles gjennom spørreundersøkelser der respondentene i et konstruert marked blir stilt hypotetiske spørsmål eller blir forelagt hypotetiske valgoppgaver. RP-data, som kan oppfattes som motstykket til SP-data, er data som beskriver folks faktiske valg i virkelige (surrogat) markeder.

Rapporten kan bestilles fra:
Transportøkonomisk institutt, Postboks 6110 Etterstad, 0602 Oslo
Telefon: 22 57 38 00 Telefax: 22 57 02 90

Et eksempel er bruk av arbeids- eller boligmarkedet til en indirekte verdsetting av sikkerhet og miljøgoder.

I denne avhandlingen er det fokusert kun på SP-metoder. Den viktigste grunnen til dette valget er at bruk av SP-metoder i mange tilfeller er eneste mulighet for å få frem den riktige konteksten for verdsetting og derfor er den eneste måten folks preferanser for mange ikke-markedsgoder kan måles på.

Metodiske forhold

To ulike SP-metoder, "Stated Choice" og "Contingent Valuation", er brukt i de fem artiklene som utgjør avhandlingen. "Contingent Valuation" er av mange kalt "betinget verdsetting" på norsk, men alle SP-metoder er egentlig betinget verdsetting, så i fortsettelsen brukes de engelske metodebetegnelsene for å unngå forvirring mht. terminologi. Stated Choice er, kort beskrevet, en metode der ikke-markedsgoder blir verdsatt relativt til hverandre og ikke i absolutte termer slik som i den mer brukte metoden Contingent Valuation.

I de fem artiklene brukes data fra to ulike norske spørreundersøkelser som begge brukte PC og personlige intervju ved datainnsamlingen. Den første er en miljøverdsettingsstudie fra 1993-94 der luftforurensning og støy relatert til vegtrafikk ble verdsatt. Den andre er en tidsverdistudie fra 1995-96 der folks betalingsvillighet for redusert reisetid på ulike transportmidler ble målt.

Artikkel 1 verdsetter miljøproblemer av vegtrafikk ved bruk av Stated Choice. Resultatene fra denne studien har vært brukt av norske vegmyndigheter i deres NKA siden 1995. I artikkel 1 pekes det på at betydelig usikkerhet gjenstår mht:

- Hvordan verdsettingen påvirkes av generelle metodiske problemer med SP-metoder. Særlig er det pekt på den såkalte interaksjons- eller fokuseringseffekten som opptrer dersom respondentene i SP-studier bare får mulighet til å tilkjennegi sine preferanser for eksempel for miljøgoder og ikke andre goder som senere også inkluderes i de samme NKA.
- Om den komplekse valgsituasjonen respondentene stilles overfor i Stated Choice studier påvirker verdsettingsresultatene.

Artikkel 2 viser at inkonsistente valg (definert som brudd på konsument-teoriens aksiom om transitivitet) er vanlig i Stated Choice studier og at dette har en signifikant innvirkning på verdsettingen av redusert reisetid. Det er vist at ulike evner (her indikert ved utdanningsnivå) signifikant påvirker antall inkonsistente valg. Forekomsten av inkonsistente valg viser seg å være størst først i valgsekvensen og reduseres for etterfølgende valg. Som en konklusjon på disse resultatene foreslås det i artikkel 2 at respondentene bør få hjelp slik at de i større grad kan velge konsistent i Stated Choice studier.

Artikkel 3 undersøker årsaker til leksikografiske valg (definert som brudd på konsumentteoriens aksiom om kontinuitet) og hvilke konsekvenser slike valg har i Stated Choice studier. Med leksikografiske valg menes her et sett av valg der respondenten i alle valgene har valgt det alternativet som er best mht nivået for bare en av attributtene som inngår, f eks bare valgt alternativet med lavest pris. Analysene i artikkel 3 viser at leksikografiske valg er vanlig i Stated Choice studier. Dessuten vises det at leksikografiske valg ikke er det samme som leksikografiske preferanser fordi leksikografiske valg er et resultat av:

- En studiedesign med for store forskjeller mellom de presenterte alternativene. Slike valg gir mindre informasjon om respondentenes preferanser, men er vanligvis ikke noe alvorlig modelleringsproblem.
- Forenkling av valgoppgaven. Slike leksikografiske bidrar til større varians i Stated Choice data og kan påvirke verdsettingen av ikke-markedsgodene dersom dette ikke korrigeres for i analysen.

I artikkel 4 brukes en skaleringsmetode ("logit scaling approach") i analysemodellene for å korrigere for den økning i varians som forårsakes av inkonsistente valg i Stated Choice data. Denne skaleringsmetoden er en statistisk estimeringsmetode som korrigerer for ulik mengde uforklart varians i ulike typer av data slik at data fra ulike kilder kan analyseres sammen. Det vises at mengden av uforklart varians øker med økende antall inkonsistente valg. Artikkel 4 konkluderer med at skalering som korrigerer for ulik mengde inkonsistente valg gir signifikant forbedring av modellene og reduserer verdsettingen av reisetid. Det viser seg også at denne skaleringen gjør at verdsettingen av reisetid fra Stated Choice dataene samsvarer bedre med verdsettingen fra Contingent Valuation dataene som er inkludert i den samme studien. En annen viktig konklusjon er at skalering som korrigerer for forskjeller i utdanning (jfr artikkel 2) ikke forbedrer modellene.

Artikkel 5 undersøker om interaksjons- eller fokuseringseffekter påvirker verdsettingen av ikke-markedsgoder som inkluderes i NKA. NKA av veginvesteringsprosjekter i Norge inkluderer ikke-markedsgoder som reisetid, trafikksikkerhet, støy og luftforurensning. Tradisjonelt har disse ikke-markedsgodene blitt verdsatt i separate betalingsvillighetsstudier uten at det har blitt tatt hensyn til eventuelle interaksjonseffekter mellom dem. Artikkel 5 viser at en simultan verdsettingsstudie, som tar hensyn til interaksjonseffekter mellom ikke-markedsgodene som inkluderes i NKA av veginvesteringer, gir signifikant lavere verdsetting av reisetid, støy og luftforurensning enn en får ved separate verdsettingsstudier. Resultatene i artikkel 5 viser at valg av kontekst for verdsettingen er et viktig tema i den generelle metodiske diskusjonen om hvordan ikke-markedsgoder skal verdsettes når resultatene skal brukes i NKA.

Konklusjoner, spørsmål og videre forskning

Artikkel 5 viser at dersom en ønsker å inkludere ikke-markedsgoder som reisetid, trafikksikkerhet, støy og luftforurensning i NKA av veginvesterings-prosjekter, bør disse ikke-markedsgodene verdsettes simultant i samme studie. Men det er et uavklart spørsmål om skyggepriser for bruk i NKA bør baseres på kortsiktige verdsettinger (som i dagens verdsettingspraksis) eller mer langsiktige verdsettinger. En mer langsiktig verdsettingsprosedyre kan sikre at:

- konteksten for verdsettingen blir i bedre samsvar med en ønsket fremtidig kontekst (f eks det samfunn folk ønsker) i stedet for dagens kontekst som både kan være uønsket og for begrenset til at den kan fungere som en ramme der folk kan tilkjennegi sine virkelige preferanser.

- allokeringen av ressurser mellom ulike kortsiktige og langsiktige budsjetter blir fullstendig, slik som økonomisk teori forutsetter.

Artikkel 2 og 3 viser at Stated Choice data kan inneholde så mye "støy" at verdsettingsresultatene kan påvirkes. Det bør derfor være en målsetting å gjøre datainnsamlingsmetodene så enkle at folk blir i stand til å tilkjennegi sine preferanser gjennom de valg som presenteres.

Artikkel 4 viser at en skaleringsmetode er mulig å bruke for å håndtere variansforskjeller som skyldes inkonsistente valg og at korrigering av denne heteroskedastisiteten reduserer verdsettingen av reisetid. Tilsvarende resultater er fremkommet i andre studier som har korrigert for heteroskedastisitet som skyldes smaks-/preferanseforskjeller og metodeforskjeller. Håndtering av variansforskjeller i SP-data ser altså ut til å være et viktig tema for fremtidig forskning, og analyser der en simultant korrigerer for heteroskedastisitet som skyldes både inkonsistente valg, smaksforskjeller og andre forhold er nødvendig.

An overview of the thesis

December 13, 1999

1. Aim and motivation

In Norway, cost-benefit analyses (CBAs) are carried out for all road investment projects, and a benefit/cost ratio is calculated for each project. An optimal decision rule would be to rank the projects according to a decreasing benefit/cost ratio and then carry out the projects in that order, until the budget is depleted. However, several recent Scandinavian studies show very weak – if any – association between the priority ranking assign to a given road investment project and the project's benefit/cost ratio (see e.g. Odeck 1996 and Fridstrøm and Elvik 1997). If doubt among decision makers about the correctness of the valuation of the goods included in CBAs partly can explain this weak association, improving the valuation methods could be one way to strengthen it.

CBAs are usually part of a more comprehensive investigation of consequences in which changes in the quantity and/or quality of goods resulting from the project are identified. If a good is to be included in a CBA, it has to have a monetary value associated with it. According to economic theory, CBAs are premised on the notion that the monetary values assigned to changes in the included goods, favourable or unfavourable, should be those of the affected individuals, not the values held by economists, moral philosophers, or others (Arrow *et al.* 1996).

However, some of the goods decision makers want to include in CBAs are not directly bought and sold in actual markets. This is the case for traffic safety, environmental impacts and travel time included in CBAs for road investments. Such nonmarket goods thus have to be valued either in Stated Preference (SP) studies or in Revealed Preference (RP) studies. SP data are collected in surveys where people are asked hypothetical questions and/or are presented hypothetical choice tasks in a constructed market. RP data are based on people's actual choices in real (surrogate) markets. For example, the labour or housing market can be used to indirectly value safety or environmental goods (i.e. hedonic pricing techniques).

An important question in connection to the practice and use of CBAs is whether the priority or weighting of goods included in CBAs gives an accurate picture of people's preferences. Therefore, it may be questioned whether CBAs gives decision makers information about which projects are profitable to society. Grudemo (1994) gives an example that illustrates problems current practice in CBAs may cause. He describes an actual case where a profitable (according to CBA) road project in Linköping, Sweden, was judged as non-profitable by most people in the affected area. They therefore voted "no" in a referendum where they were asked if they wanted this project to be implemented. The main reason why people in Linköping did not want this project was that the new road would affect an area used for recreation. Because the value of this recreation area was not included in the CBA, the CBA did not accurately measure whether the road project was profitable to the society in the theoretical meaning of such a concept. This example shows that the concept "profitability to the society" has to be used with care if goods that are important to affected individuals are excluded from the CBA. Similar problems may occur if nonmarket goods are valued in a context different than the context later used in CBAs (i.e. benefit transfer).

Since valuation of nonmarket goods is difficult, uncertain and often involves ethical questions it has been discussed whether or not nonmarket goods should be valued at all, and if they should, which nonmarket goods can be valued with sufficient precision for inclusion in CBAs. Despite problems like the one described above, it is assumed in this dissertation that it *is* meaningful to include nonmarket goods in CBAs. This assumption is mainly based on the view that a CBA may give valuable systematic information to the decision making process and that CBAs ensure that there is consistency between valuation of nonmarket goods in different decision processes.

The aim of the thesis is to investigate i) how the Stated Choice method can be used in valuation of environmental goods, ii) how methodological problems related to Stated Choice influence the valuation results and iii) how the choice context influence the valuations of nonmarket goods. The cases and the data are related to valuation of travel time, environmental problems and traffic safety included in CBAs of road investments, but the results and the conclusions are relevant to valuation of nonmarket goods and considerations about CBA in general. The different papers are written with a critical, but (hopefully) constructive view of today's practice of valuation of nonmarket goods and inclusion of such valuations in CBAs.

2. Why focus on Stated Preference methods only?

This chapter first presents and discusses some central concepts related to valuation of nonmarket goods for use in CBAs and then explains why the focus of this dissertation is on SP methods only, and not on RP methods. One reason for this limitation is that valuation of nonmarket goods is an extensive field and one has to limit the focus in order to contribute. However, a more important reason is that SP methods based on constructed markets often is the only way to construct a relevant valuation context and, therefore, the only way to measure people's preferences for certain nonmarket goods.[1] This view may be controversial, but nevertheless easy to argue for. To show that this is neither new knowledge nor inaccessible knowledge, I have cited some "old" Dictionaries of Economics in this argumentation (e.g. Pearce 1986).

[1] This strength of SP methods is also its greatest fault: it is always possible to create contexts that bias the answers systematically.

2.1 Cost-benefit analysis: A definition

Pearce (1986) defines CBA as

"a conceptual framework for the evaluation of investment projects in the government sector, although it can be extended to any private sector project. It differs from any straightforward financial appraisal in that it considers all gains (benefits) and losses (costs) regardless of to whom they accrue (although usually confined to the inhabitants of one nation.) A benefit is then any gain in utility[2] *and a cost is any loss of utility as measured by the opportunity cost of the project in question. In practice, many benefits (which may be positive or negative) will not be capable of quantification in money terms (e.g. loss of wildlife, destruction of natural beauty, destruction of community ties etc.) while the costs will be measured in terms of the actual money costs of the project. Where money measured are secured, however, they should be corrected for any divergence between the shadow price and the market price if possible. Strictly pursued, CBA would value all outputs and inputs at their shadow prices.*

A project is potentially worthwhile if the discounted[3] *value of the benefits exceeds the costs. It does not enable us to recommend outright acceptance since capital rationing may be present. In this case it is necessary to adopt some other rule for ranking projects. This is often done by the size of the benefit/cost ratio (i.e. those with the highest ratios being accepted and the list of projects being adopted until the budget is exhausted).*

This is in fact not a foolproof ranking procedure, but it is a first rough guide to desirability. Note that the requirement that benefits exceed costs can be expressed more formally in terms of the compensation principle, namely that beneficiaries should be able hypothetically to compensate the bearers of the costs."

2.2 Limitations of cost-benefit analysis

Starret (1988) uses the term "project analysis" as something different than standard "first best" CBA. Starret shows that it is almost certainly not optimal to follow the first best rules of CBA in deciding on project levels. Such a rule would involve equating marginal benefit to marginal cost and would only be optimal if the marginal project has no equity effects and did not affect the production of "monopolised goods" or consumption of taxed goods. Simplifications are, however, often done in practical CBAs; for example, distributional effects are not included or accounted for separately. In the following we concentrate on the issue of how to assign a correct shadow price on nonmarket goods in a CBA. This is important irrespective of whether the shadow price is to be used as input in

---

[2] Utility is widely constructed in economics to be synonymous with "welfare", economic welfare, satisfaction, and, occasionally, happiness. More strictly, however, to say that someone derives utility from a good or event is to say that they prefer the good to exist rather than not to exist. To say that they derive more utility from good X than good Y is simply to say that X is preferred to Y. (Pearce 1986)

[3] Discussion of how to discount the values included in CBA is not included in this paper.

standard first-best CBA or in a more comprehensive project analysis. It must nevertheless be stressed that the shadow prices discussed in this paper are appropriate only for the evaluation of small projects. For large changes, techniques based exclusively on first-order terms are no longer adequate.

2.3 Shadow pricing

Shadow pricing is defined by Starret (1988) as the study and use of first-order welfare impacts associated with changes in the levels of particular goods or groups of goods. He argues that this definition accords with the common view of what a shadow price stands for. Suppose we are talking about some input to a government project. Then Starret states that the shadow price ought to be the opportunity cost per unit of increasing this input, and the correct way to measure this opportunity cost is by the marginal welfare forgone. Ambiguity in the definition arises over the *ceteris paribus* assumption. Pearce (1986) defines shadow price as

"an imputed valuation of a commodity or service which has no market price. Shadow prices are used in CBA and in the application of mathematical programming to a planned economy. They represent the opportunity cost of producing or consuming a commodity which is generally not traded in the economy. Even in a market economy certain outputs such as health, education and environmental quality do not attract a market price. A set of market prices representing consumers' marginal rates of substitution[4] or producers' marginal rates of transformation between such commodities may be calculated reflecting the marginal costs of production or the marginal value of their use as inputs. To the extent that market prices do not reflect opportunity costs, CBA may substitute shadow prices."

Dréze and Stern (1994) state that

"when the social opportunity cost or shadow price of a good is defined in terms of the marginal effect on social welfare of the availability of an extra unit, it leads directly to a "cost-benefit test", i.e., projects which make positive profits at shadow prices should be accepted because they increase welfare. Indeed, it should be clear that no other definition of a shadow price can have this property... The shadow prices themselves will depend on the social welfare function and on how the economy, including the government, functions. But there is nevertheless essentially only one definition of the concept which allows their systematic link to cost-benefit tests."

2.4 Opportunity cost and the ceteris paribus assumption

Opportunity cost is a central concept in both Starret's and Pearce's explanations of shadow price. Pearce (1986) describes opportunity cost as

---

[4] In consumer demand theory the marginal rate of substitution refers to the amount of one good, say Y, that is required to compensate the consumer for giving up an amount of another good, say X, such that the consumer has the same level of welfare (utility) as before (Pearce 1986).

"perhaps the most fundamental concept in economics. The opportunity cost of an action is the value of the forgone alternative action. Opportunity cost can only arise in a world where the resources available to meet wants are limited so that all wants cannot be satisfied. If resources were limitless no action would be at the expense of any other – all could be undertaken – and the opportunity cost of any single action, the value of the "next best" alternative, would be zero. Clearly, in a real world of scarcity opportunity cost is positive."

Another central concept in Starret's explanation of shadow pricing is *the ceteris paribus* assumption. *Ceteris paribus* is a Latin expression meaning "other things being equal". Economic analysis frequently proceeds by considering the effect of varying one or a few *independent variables*[5] while other things remain unchanged. To indicate that this is being done, the term *ceteris paribus* is employed.

In a discussion of the ceteris paribus assumption Starret (1988) continues with the following question: When an input level is changed, what other things are allowed to change with it? His point on this issue is that it is useful to think differently about the ceteris paribus assumption in different contexts and that it is important to be precise about the assumptions in each context. Starret ends this discussion with the following conclusions:

"We can achieve precision by specifying which variables are to be treated as independent in each instance. We may choose these so that there are feasibility constraints linking the independent variables, in which case the associated shadow prices will be hypothetical. Having chosen the independent variables, the model must be specified in such a way that equilibrium determines all remaining (dependent) variables as functions of the independent ones." [6]

2.5 Partial versus general equilibrium models

Starret's discussion clearly points in the direction of scepticism to the use of a partial equilibrium model i.e. the study of a market for a commodity in isolation. Unfortunately, such "in isolation studies" have been common when valuing nonmarket goods. With respect to partial equilibrium Pearce (1986) states:

"Given the prices of all other commodities the conditions for equilibrium in a single market are examined. This technique ignores the effects of changes in the price of a commodity on all other related market prices including the prices of factors of production. These changes may have feedback effects on the original market which can only be analysed in a system of general equilibrium."

---

[5] Pearce (1986) gives the following explanation of an independent variable. A variable appearing on the right-hand side of the equality sign in an equation, so called because its value is determined "independently" of, or outside, the equation.

[6] In econometric terms this would be described as reduced form equations.

This view is in agreement with Bannock *et al.* (1984) who state that

"the assumption underlying partial equilibrium analysis is that the interaction between the market under study and the rest of the economy can be ignored, since it has little effect on the final results. In effect, we make the "other things being equal" assumption that the rest of the economy remains the same throughout the analysis and that there are no "feedback effects" on the single market under study. Many economists would regard partial equilibrium analysis as being valid only for expositional purposes or for studying very short-run effects. In general, interdependence among markets in the economy is strong enough to warrant a general equilibrium analysis."

According to the discussion above, the partial equilibrium approach appears inappropriate for finding shadow prices of nonmarket goods if these are to be used in CBAs that include more than one nonmarket good. This conclusion seems also to be in accordance with the view of Randall and Hoehn (1996) about the embedding problem. The term embedding is used to describe a situation when willingness to pay for a particular good may vary "over a wide range depending on whether the good is assessed on its own or embedded as part of a more inclusive package" (Kahneman and Knetsch 1992). Randall and Hoehn (1996) conclude that "embedding effects are standard economic phenomena induced by substitution relationships and constrained endowments, and these effects may be of substantial magnitude."

2.6 Choice of valuation context determines the choice of possible valuation methods

Another conclusion that may be drawn from the discussion above is that context seems to be a key issue in a judgement of how to find the relevant shadow prices of nonmarket goods for use in CBA.

Starret's (1988) conclusion was that it is useful to think differently about the ceteris paribus assumption in different contexts and that we can achieve accuracy by specifying which variables are to be treated as independent in each instance. But which variables can clearly be said to be independent?

One important question is: Given the context for use of a nonmarket good's shadow price, e.g. a CBA together with other nonmarket goods and market goods, what is theoretically the correct context for finding the shadow price?

If more than one (nonmarket) good is to be included in the CBA, it should be clear that at least all those nonmarket goods included should be treated as dependent in the valuation procedure. This conclusion is supported by Randall and Hoehn's (1996) conclusions above about embedding effects as standard economic phenomena. Only in such a simultaneous valuation procedure is it possible to capture the correct relative shadow prices for the relevant nonmarket goods.

Another important question is: Given the context for use, and therefore the correct context for elicitation of the shadow prices, which valuation methods are capable of measuring the correct shadow price? Will valuation methods based on revealed preference data in surrogate markets capture the relevant context for valuation? Or are stated preference methods based on constructed markets the only way to construct the relevant valuation context? (See, for example, Braden and Kolstad 1991 for an overview of valuation methods.) The answers to these questions should be clear from the discussion above. It will be very difficult to find surrogate markets that represents the relevant context in a CBA where more than one nonmarket good is included.

The next sequence of questions that arises is related to the general criticism against methods based on constructed markets (e.g. Hausman 1993, Diamond and Hausman 1994, Haneman 1994, McFadden 1994 and Portney 1994). For example, will people state their true willingness to pay? Will they respond strategically? Will they understand the constructed market that is presented to them and respond as if it was a real market? How can we trust valuations from methods based on constructed markets when such methods give different valuations? Many such questions have been raised, and there are probably not yet good answers to all such questions. But as an example of the on-going research on these issues, Cameron *et al.* (1999) show that handling variance differences from different elicitation methods gives a constructive and promising answer to the important question about why valuations from methods based on constructed markets apparently seem to give different valuations of the same good.

Given the conclusion that the relevant context for valuation in many cases only can be supported by constructed markets, the next question is: Can we be sure that shadow prices found in constructed markets would be in accordance with shadow prices found in actual markets or surrogate markets if similar contexts were possible? Fortunately, this question may be answered in part because there have been studies that compare the valuation, in similar contexts, of nonmarket goods using both SP and RP approaches. Carson *et al.* (1996) have conducted a meta-analysis of studies that compare Contingent Valuation and Revealed Preference valuations of essentially the same quasi-public goods. They conclude that the meta-analysis provides support for convergent validity of the two different valuation approaches. Wardman (1998) has compared the valuation of travel time from revealed and stated preference methods from several different studies. His conclusion is that there is an encouraging level of correspondence between RP and SP methods.

2.7 Valuation of existence values: Another argument for the use of Stated Preference methods

Under the heading "Estimating shadow prices", Dasgupta and Mäler (1994) stress that "the production function approach" allows one to capture only the current use value of a nonmarket good. Their conclusion is that the shadow price of a nonmarket good is the sum of its use-value and its non-use value, i.e. the shadow price is equal the total value. Randall and Stoll (1982) and Bergland (1993) give an overview of the total value approach and a presentation of the different subcomponents included in the concepts total use value and non-use-value. With the knowledge that the shadow price of a nonmarket good should equal the total value it is easy to conclude that SP methods based on constructed markets are the only way to find shadow prices of nonmarket goods that have a non-use value component greater than zero. Another obvious conclusion is that values other than current use values are also difficult to capture by methods other than SP methods, i.e. future use values can not be captured by other valuation methods.

3. Overview of the papers

This chapter describes the relationship between the subjects presented in the different papers and gives a short presentation of the conclusions. The dissertation consists of five papers. In order of appearance, these papers are:

1. Stated Choice valuation of urban traffic air pollution and noise (Sælensminde 1999a)
2. The impact of choice inconsistencies in Stated Choice studies (Sælensminde 1999b)
3. Causes and consequences of lexicographic choices in Stated Choice studies (Sælensminde 1999c)
4. Inconsistent choices in Stated Choice data: use of the logit scaling approach to handle resulting variance increases (Sælensminde 1999e)
5. Interaction effects in valuation of nonmarket goods (Sælensminde 1999d)

Two different valuation methods are used in the five papers, Stated Choice and Contingent Valuation. (See for example Mitchell and Carson 1989 for an extensive description of the Contingent Valuation method, probably the most used valuation method based on constructed markets, and Hensher 1994 for an overview of Stated Choice and its use in transport research.) Stated Choice is, briefly described, a Stated Preference method in which nonmarket goods are assessed relative to each other and not in absolute amounts as with the more common Contingent Valuation method. Paper 1 presents results from a Stated Choice study only. Paper 2, 3 and 4 presents methodological problems related to the complex choice situation in Stated Choice studies, while valuations from the Contingent Valuation method are used for comparisons. In paper 5 the Contingent Valuation method is used to investigate possible interaction effects, while the valuations from the Stated Choice study presented in paper 1 is used for comparisons.

Data from two different national Norwegian surveys consisting of in-person interviews are used in the papers. The first study, conducted in 1993-94, is an environmental study valuing air pollution and noise related to road traffic (Sælensminde and Hammer 1994). The second study, conducted in 1995-96, is a so-called "value of time study" measuring travellers' willingness to pay for a reduction in travel time (Ramjerdi *et al.* 1997). These two studies, respectively, provide the Norwegian road authorities with valuation of environmental benefits and travel time savings for use in CBA.

Paper 1 only uses data from the environmental survey in the analysis. Data from the environmental survey is also used in papers 2, 3 and 4, while the value of time study is used as the main data source in these papers. Paper 5 only uses data from the value of time survey in the analysis, but the results from the environmental survey and studies that value traffic safety are used for comparisons. The fact that data sources to a large extent are common for the five papers is one reason for overlap. Another reason for overlap is the fact that the subjects of the different papers are strongly related. Because the papers are written as independent units for publication in different journals, and because each paper provides an overview of related literature, overlap is difficult to avoid.

In paper 1 (Sælensminde 1999a), environmental problems related to urban traffic are valued by Stated Choice. This paper presents one of the first studies that used Stated Choice to value environmental goods (Sælensminde and Hammer 1994).

The use of Stated Choice to estimate people's willingness to pay to reduce environmental problems caused by road traffic can be seen as an extension of the established use of Stated Choice in transport research. The results from this study have been used by the Norwegian Public Roads Administration in their CBAs since 1995. Paper 1 points to the fact that considerable uncertainty remains with regard to:

- the area of validity of the results because the respondents have only stated their preferences for a specific range of environmental changes;
- the impact on the valuations of interaction effects and other general methodological problems with Stated Preference methods because the respondents in this study only had the opportunity to state their preferences for environmental goods and not other goods included in CBAs; and
- problems related to the complex choice situation of Stated Choice.

In this way paper 1 recommends studies that investigate the impact of choice context and choice complexity, which are the subjects for the other papers in this dissertation.

The subject of *paper 2* (Sælensminde 1999b) is the occurrence of inconsistent choices (i.e. violations of the transitivity axiom) in Stated Choice studies and the impact such choices may have on the valuations of the nonmarket goods.

Paper 2 presents a test procedure based on a so-called "ray-diagram". This procedure is applied to the Stated Choice data and allows determination of whether choices by each respondent are jointly consistent. The analysis shows that inconsistent choices commonly occur in several Stated Choice tasks, and they have a significant impact on the valuation of reduced travel time. In a regression model it is shown that different abilities (i.e. level of education) cause inconsistent choices. The occurrence of inconsistent choices is shown to be largest early in the choice sequence and is reduced in later choices. As a conclusion, it is suggested that respondents may need more training and help to choose consistently in Stated Choice studies.

Paper 3 (Sælensminde 1999c) investigates the causes and consequences of lexicographic choices (i.e. violation of the continuity axiom) in Stated Choice studies. By lexicographic choices we mean a set of choices in which the respondent consistently chooses the alternative that is best with respect to one particular attribute, e.g. lowest price, neglecting all other attributes.

The analyses in paper 3 show that lexicographic choices commonly occur in several Stated Choice tasks and that lexicographic choices are not the same as lexicographic preferences because lexicographic choices are partly a result of:

- study designs with too large differences between the alternatives; and
- simplification of the choice task.

Study design as a reason for lexicographic choices gives less information about preferences, but this is normally not a serious modelling problem. Simplification as a reason for lexicographic choices is shown to be a result of respondents' different abilities to choose. Such lexicographic choices contribute to the larger variance in SC data compared to less cognitive demanding valuation methods and might therefore have a significant impact on the valuation of nonmarket goods if it is not corrected in the analysis.

Paper 4 (Sælensminde 1999e) makes use of the logit scaling approach to handle variance increases caused by inconsistent choices in Stated Choice data. The scaling approach is a statistical estimation method that allows for differences in the amount of unexplained variance in different types of data, which can then be combined in the analysis. This approach has been mostly used in the context of combining Stated Preference and Revealed Preference data, but has also been used as a method for identifying systematic differences in the variance of choices within a single Stated Preference data set, e.g. for investigation of learning and fatigue effects.

The amount of unexplained variance is shown to increase as the number of inconsistent choices increases. A scaling variable in different models in paper 4 is used for:

- each respondent's number of inconsistent choices (based on the test of violations of the transitivity axiom described in paper 2); and
- each respondent's education.

The analyses are done with and without inclusion of lexicographic choices. The influence of lexicographic choices seems to be small, but this is not fully analysed in paper 4.

The main conclusion from the analyses in paper 4 is that scaling due to inconsistencies significantly improves the models and reduces the valuations of travel time. In addition, the scaling approach makes the valuations of travel time from the Stated Choice data more consistent with the valuations from Contingent Valuation data included in the same study. Another important conclusion is that scaling due to education gives no improvement of the model. This result occurs despite the fact that education is the only significant explanatory variable for the number of inconsistent choices (cf. paper 2). These conclusions indicate that the scaling approach is not an easy solution to the problem of inconsistent choices due to the complex choice situation in Stated Choice studies.

In *paper 5* (Sælensminde 1999d) possible interaction effects in valuation of nonmarket goods included in CBAs is the subject. CBAs undertaken for the road sector in Norway is chosen as a case for investigating interaction effects. CBAs for road investments include nonmarket goods such as travel time savings, traffic safety, noise and air pollution. Traditionally, these nonmarket goods are valued through separate willingness-to-pay studies without attention to interactions between them.

Paper 5 shows that a simultaneous valuation procedure, accounting for interactions between the nonmarket goods included in CBAs for road investments, significantly reduces the valuations of travel time savings, noise and air pollution compared to a separate valuation procedure. Another result is that a reminder of substitutes and budget constraints seems less necessary if the valuation procedure is more holistic. The results presented in paper 5 show that the choice of valuation context is an important issue in the overall methodological discussion of how to value nonmarket goods for use in CBAs.

4. Questions and further research needs

4.1 Context and budget allocation

There is now an increasing amount of evidence suggesting that the conventional cost-benefit procedure, where each element of a multiple impact policy is evaluated independently and the total valuation is obtained by summing across the independent component valuations, in most cases (i.e. when the elements are substitutes) leads to overestimates of total valuation. (See e.g. Hoehn and Randall 1989 and Paper 5). However, if the elements are complements, this summing-procedure may underestimate total valuation. The relevant context for estimating shadow prices of nonmarket goods may be that which is similar to the context in which these shadow prices subsequently will be used. If the case is CBA for road projects and one wishes to include nonmarket goods such as travel time, traffic safety, noise and air-pollution, these nonmarket goods should be valued simultaneously. Only in that way can the respondents pay sufficient attention to interaction effects.

Should the relevant shadow prices for use in CBA be based on short-term valuations (as in today's valuation practice) or more long-term valuations? A more long term valuation procedure can ensure that i) the valuation context is in better accordance with a desired future context (i.e. the society people want) instead of today's context which may be undesired and too limited to present in a questionnaire as an acceptable choice set where people are to state their preferences, and ii) the budget allocation process is completed, as is assumed in standard economic theory.

4.1.1 What is a too narrow context?

An example that illustrates how the context may be too narrow came during respondent reaction to the Norwegian value of time study (Ramjerdi *et al.* 1997). The choice context was a specific journey the respondent had undertaken, and the choice involved respondent willingness to pay to reduce travel time on that particular car journey. However, the respondent protested that she could not state her preferences with respect to reduced travel time connected to that car-journey. What she really wanted was to use her bicycle on that journey, but she was forced to use her car because it was too dangerous and polluted to go by bicycle. That is, she preferred a broader context, an alternative choice task.

4.1.2 The budget allocation process

Randall and Hoehn (1996) illustrate possible problems with the budget allocation process with a term they call "multi-stage budgeting" and explain the connections to the embedding problem mentioned above. Multi-stage budgeting is explained as the formal development of the idea that all expenditures are variable in the long run but many kinds of expenditures are fixed in the short run. As an example of multi-stage budgeting they say that Stage 1 may be a stage where rebudgeting to determine WTP for a proposed environmental policy takes place within a short-run discretionary account that includes only environmental goods. In Stage 2 rebudgeting is confined to short-run discretionary expenditures but may occur across budget categories, e.g., environment, recreation and vacations, food, clothing etc. In the final stage rebudgeting occurs across all short-term and long-term accounts. Randall and Hoehn (1996) conclude that if the budget allocation process is incomplete, discretionary income is reduced and, theory predicts, the effects of embedding are exacerbated.

4.1.3 Questions for discussion

Will a long-term valuation procedure that takes into account a less limited future context and a more complete budget allocation process be considered as too hypothetical and therefore less acceptable as foundation for decisions? Will such long-term valuation procedures result in respondents acting more as citizens than as consumers, and thus not in accordance with economic theory? (See e.g. Blamey *et al.* 1995.)

The answers to such questions are interesting, and a discussion is needed of which components are sufficient to incorporate in shadow prices and what context is most relevant for valuation of nonmarket goods that are to be included in CBAs.

4.2 Noise or signals?

4.2.1 Collecting less noise - Issues of study design

In papers 2, 3 and 4 (Sælensminde 1999b, c, e) it is shown that Stated Choice is a valuation method that may collect so much noise that the valuation results are affected. This is because a cognitively demanding valuation method may cause respondents to use simplifying decision procedures (heuristics) or to choose in an inconsistent manner. It should be a goal for practitioners to make the elicitation methods sufficiently simple that people are able to state their preferences in response to the choice set they are presented. If peoples' abilities to choose determine whether their preferences shall count in valuation studies we may have a problem concerning the representativeness of such studies.

Paper 4 suggests that a method based on Frisch (1972), where the difference between the presented choice alternatives are larger than in today's practice, might be less cognitive demanding. Whether this is a promising approach remains to be seen, but it may be a step forward collecting more signals and less noise. Such a data collection method is needed.

4.2.2 Separating signals from noise - Issues of analysis

In paper 4 (Sælensminde 1999e) methodological problems related to variation in the statistical models' error terms (heteroscedasticity) is the issue. Heteroscedasticity, when not accounted for in a logit model, may result in biased parameter estimates and therefore may also bias the valuation of the attributes included in the analysis (see e.g. Maher *et al.* 1999). *Within a single data set* heteroscedasticity could be caused by:

a) inconsistencies due to variations in the competence of individuals to perform SP tasks; and

b) taste/preference variations.

Between data sets (e.g. from different elicitation methods) heteroscedasticity could be a result of:

c) different complexity levels (e.g. number of attributes);

d) differences in how to state the preferences (e.g. handling of preference uncertainty); and

e) different choice contexts.

An additional source of systematic differences in valuations obtained by different elicitation methods is use of different functional forms for the utility functions in the analysis for different methods (e.g. Munizaga *et al.* 1997 and Halvorsen and Sælensminde 1998). Paper 4 shows that the scaling approach may be a way to handle heteroscedasticity due to inconsistent choices and that accounting for such heteroscedasticity may lower the estimated value of travel time. Similar results are obtained in recent papers by Hultkrantz and Mortazavi (1998) and Algers *et al.* (1999), which, using different approaches, handle heteroscedasticity due to taste/preference variations. In contrast, Kim (1998) concludes that such handling of taste differences results in higher valuations than the standard logit model. To make this overview more complete it should also be mentioned that recent studies by Adamowicz *et al.* (1998), Halvorsen and Sælensminde (1998) and Cameron *et al.* (1999) conclude that the valuations from different elicitation methods seem to be more similar when one corrects for heteroscedasticity.

Accounting for heteroscedasticity seems to be a crucial issue for future research (Hensher *et al.* 1999 and Swait and Bernardino 2000), and analyses that simultaneously correct for heteroscedasticity due to inconsistencies, taste differences and other sources are needed.

References

Adamowicz, W. L., P. Boxall, M. Williams and J. Louviere (1998) Stated Preference approaches to measuring passive use values: Choice Experiments and Contingent Valuation. *American Journal of Agricultural Economics,* 80(1): 64-75.

Algers, S., P. Bergström, M. Dahlberg and J. L. Dillén (1999) Mixed logit estimation of the value of travel time. In: Dillén, J. L. and S. Algers (1999) The value of time – Further research on data collected in 1994. KFB-Rapport 1999:11, Stockholm.

Arrow, K.J., M.L. Cropper, G. C. Eads, R. W. Hahn, L. B. Lave, R. G. Noll, P. R. Portney, M. Russell, R. Schmalensee, V. K. Smith and R. N. Stavins (1996) Benefit-cost analysis in environmental, health and safety regulation. A statement of principles, American Enterprise Institute for Public Policy Research, Washington 1996.

Bannock, G., R. E. Baxter and R. Rees (1986*) The Penguin Dictionary of Economics*. Third edition, Penguin Books 1984.

Bergland, O. (1993) Externalities in economic theory and literature. Paper presented at the seminar on External Effects in the Utilisation of Renewable Energy, 16 September 1993, the Technical University of Denmark, Lyngby.

Blamey, R., M. Common and J. Quiggin (1995) Respondents to Contingent Valuation surveys: Consumers or citizens. *Australian Journal of Agricultural Economics*, 39(3): 263-288.

Braden, J. B. and Kolstad C. D. (eds.) (1991) *Measuring the demand for environmental quality.* Contributions to economic analysis, no. 198, North-Holland, 1991.

Cameron, T. A., G. L. Poe, R. G. Ethier and W. D. Schulze (1999) Alternative nonmarket valauation-elicitation methods: Are the underlying preferences the same? Paper presented at the W-133 Annual Meeting in Tucson, February 1999.

Carson, R. T., N. E. Flores, K. M. Martin and J. L. Wright (1996) Contingent valuation and reveal preference methodologies: Comparing the estimates for quasi-public goods, *Land Economics* 72(1): 80-90.

Dasgupta, P. and K.-G. Mäler (1994) The environment: The environment and emerging development issues. In: Layard, R. and S. Glaister (1994) Cost-Benefit Analysis. Second edition, Cambridge University Press.

Diamond, P. A. and J. A. Hausman (1994) Contingent Valuation: Is some number better than no number? *Journal of Economic Perspectives*, 8(4): 45-64.

Drèze, J. and N. Stern (1994) Shadow prices and markets: Policy reform, shadow prices and market prices. In: Layard, R. and S. Glaister (1994) Cost-Benefit Analysis. Second edition, Cambridge University Press.

Fridstrøm, L. and R. Elvik (1997) The barely revealed preference behind road investment priorities. *Public Choice* 92: 145-168.

Frisch, R. (1972) Cooperation between politicians and econometricians on the formalization of political preferences. Reprint series no. 90. Institute of Economics, University of Oslo, Oslo 1972.

Grudemo (1994) Nya vägars intrångskostnader. En sammanställning av resultat av CVM-undersökningar ock "för eller emot"-studier. VTI-rapport nr 744, 1994, Väg- och transportforskningsinstitutet, Linköping.

Halvorsen, B. and K. Sælensminde (1998) Differences between willingness-to-pay estimates from open-ended and discrete choice contingent valuation method: The effect of heteroscedasticity, *Land Economics,* 74(2): 262-282.

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

Hausman, J. A. (ed.) (1993) *Contingent Valuation. A Critical Assessment.* Contributions to economic analysis, no 220, North-Holland, 1993.

Hensher, D. A. (1994) Stated preference analysis of travel choices: the state of practice. *Transportation*, 21(2): 107-133.

Hensher, D. A., J. Louviere and J. Swait (1999) Combining sources of preference data. *Journal of Econometrics*, 89:197-221.

Hoehn, J. P. and A. Randall (1989) Too many proposals pass the Benefit Cost test. *American Economic Review* 79(3): 544-551.

Hultkrantz, L. And R. Mortazavi (1998) The value of travel time changes in a random nonlinear utility model. Unpublished paper. Dalarna University, Sweden.

Kahneman, D. and J. L. Knetsch (1992) Valuing Public Goods: The Purchase of Moral Satisfaction, *Journal of Environmental Economics and Management*, 22, 57-70.

Kim, K.-S. (1998) Application of random coefficient logit model. Paper presented at the 26th European Transport Conference 1998, PTRC, England.

Maher, M. A. Rosa and J. Kontolemakis (1999) Bias in discrete choice models estimates. Paper presented at the 27th European Transport Conference 1999, PTRC, Cambridge.

McFadden, D. (1994) Contingent valuation and social choice. *American Journal of Agricultural Economics*, 76: 689-708.

Mitchell, R. C. and R. T. Carson (1989) *Using surveys to value public goods: The contingent valuation method.* Resources for the Future, The John Hopkins University Press, 1989.

Munizaga, M. A., B. G. Heydecker and J. de D. Ortúzar (1987) On the error structure of discrete choice models. *Traffic Engineering and Control* 38(11): 593-597.

Odeck, J. (1996) Ranking of regional road investment in Norway: Does socioeconomic analysis matter? *Transportation* 23: 123-140.

Pearce, D. W. (1986) *The MIT dictionary of modern economics.* Third edition, The Macmillan Press Ltd 1986.

Portney, P. R. (1994) The contingent valuation debate: Why economists should care. *Journal of Economic Perspectives*, 8(4): 3-18.

Ramjerdi, F., L. Rand, I. A. F. Sætermo and K. Sælensminde (1997) The Norwegian value of time study, Report 379/1997, Institute of Transport Economics, Oslo.

Randall, A. and J. R. Stoll (1982) Existence value in a total valuation framework. In Rowe, R. D. and L. G. Chestnut (eds.) (1982) *Managing air quality and scenic resources at national parks and wilderness areas.* Boulder, Colorado, Westview Press.

Randall, A. and J. P. Hoehn (1996) Embedding in Market Demand Systems, *Journal of Environmental Economics and Management,* 30: 369-380.

Starret, D. A. (1988) *Foundations of public economics.* Cambridge University Press 1988.

Sælensminde, K. (1999a) Stated Choice valuation of urban traffic pollution. *Transportation Research Part D* 4: 13-27.

Sælensminde, K. (1999b) The impact of choice inconsistencies in Stated Choice studies. Working paper, Institute of Transport Economics, Oslo.

Sælensminde, K. (1999c) Causes and consequences of lexicographic choices in Stated Choice studies. Working paper, Institute of Transport Economics, Oslo.

Sælensminde, K. (1999d) Interaction effects in valuation of nonmarket goods. Working paper, Institute of Transport Economics, Oslo.

Sælensminde, K. (1999e) Inconsistent choices in Stated Choice data: Use of the logit scaling approach to handle resulting variance increases. Working paper, Institute of Transport Economics, Oslo.

Sælensminde, K. and F. Hammer (1994) Verdsetting av miljøgoder ved bruk av samvalganalyse. (Assessing environmental benefits by means of conjoint analysis.) TØI-rapport no. 251/1994, Institute of transport economics, Oslo. (In Norwegian, with English summary.)

Swait, J. and A. Bernardino (2000) Distinguishing taste variation from error structure in discrete choice data. *Transportation Research Part B*, 34: 1-15.

Wardman, M. (1998) The Value of Travel Time. A Review of British Evidence, *Journal of Transport Economics and Policy*, 32(3): 285-316.

Paper 1:
Stated Choice valuation of urban traffic air pollution and noise

Kjartan Sælensminde
December 13, 1999

1. Introduction

Abstract

In this study environmental problems related to urban traffic are valued by Stated Choice. Stated Choice is a Stated Preference method in which non-market goods are assessed relatively to each other and not in absolute amounts as with the more common Contingent Valuation Method. The use of Stated Choice to estimate people's willingness to pay to reduce environmental problems caused by road traffic can be seen as an extension of the established use of Stated Choice in transport research. Despite the fact that the results from this study have been used by the Norwegian Public Roads Administration in their cost benefit analyses since 1995, this paper points to the fact that considerable uncertainty remains with regard to (1) the area of validity of the results, (2) the impact on the valuations of interaction effects and other general methodological problems with Stated Preference methods, and (3) problems related to the complex choice situation of Stated Choice. © 1998 Elsevier Science Ltd. All rights reserved.

Keywords: Stated choice; Air pollution; Noise; Norway

1. Introduction

The determination of Willingness to Pay (WTP) for environmental benefits by means of Stated Preference (SP) data has traditionally been associated with the Contingent Valuation (CV) method. Briefly, the CV method consists of asking people what they are willing to pay for a benefit, for instance an improvement in air quality, or how much they are Willing to Accept (WTA) in compensation for a deterioration.[1]

Subsequent to the Exxon Valdez disaster in Alaska in 1989, international literature in this area is roughly divided into two camps. One consists of those who believe in the CV method, as

* Author for correspondence. Fax: +47-22-57-02-90; e-mail: ks@toi.no

[1] Chiefly owing to the state of the law where 'road traffic and the environment' is concerned, but also in order to induce respondents to take their budgets into account, we decided to use questions and choice situations of the WTP type in this study. Both compensating and equivalent variation are estimated as measures of welfare gain or loss.

PII: S1361-9209(98)00020-0

represented for instance by contributions in Navrud (1992), Braden and Kolstad (1991) and Mitchell and Carson (1989). In the other camp are many economists who are highly critical of CV, some of whom contributed to Hausman (1993). The contributions of McFadden and Leonard (1993) and of Kemp and Maxwell (1993) both point to the Stated Choice (SC) method as an alternative to CV. Useful light on this discussion has been shed by Arrow et al. (1993), who called attention to the strict standards CV studies have to satisfy to ensure satisfactory results. In their recommendations for future research, Arrow et al. (1993) also propose the use of SC as a method for assessing environmental benefits.

SC is an SP method in which environmental benefits are assessed relatively to each other and not in absolute amounts as with CV. A review of the historical development and the various types of valuation methods (ranking, rating and choice) in transport research can be found in Hensher (1994). Hensher (1994) emphasises that: "Stated *choice* experiments are now the most popular form of SP method in transportation and are growing in popularity in other areas such as marketing, geography, regional science and tourism. The papers by Louviere and Hensher (1982) and Louviere and Woodworth (1983) have become the historical reference sources for stated choice modelling in transportation."

The use of stated choice experiments in the valuation of environmental benefits related to road traffic in this study can probably be seen as an extension of the use of SC that has gradually become established in transport research. Lately, SC-experiments have also been applied to environmental management problems such as recreational use of rivers (Adamowicz et al., 1994), recreational moose hunting (Adamowicz et al., 1996; Boxall et al., 1996), protection of old-growth forests (Adamowicz et al., 1998), and landscape and wildlife protection in Scotland (Hanley et al., 1997).

2. Study design

The final design of the main SC study, presented here, is based on experience with a pilot survey carried out in 1992.[2] To make the interview situation as realistic as possible, we based it on a particular journey which the respondent had made. The journey presented in the choice exercises was chosen by lot among all local journeys the last day that the respondents had undertaken one or more journeys by car or public transport. All travel objectives are therefore covered in this study. The respondents were asked to choose between two alternative journeys, each of which was described in terms of some of its characteristics (e.g. in vehicle time, cost and various noise and air pollution levels). Table 1 shows the factors (attributes) included and the numbers of respondents who took part in the various choice exercises. The choice of a maximum of four attributes in the exercises was made on the assumption that the respondents may have problems in handling more than four attributes at the same time. Later studies of lexicographic and inconsistent choices in stated choice studies have shown that such problems occur and influence

[2] In the pilot study 179 respondents from Oslo took part (Sælensminde and Hammer, 1993). The specific objective of the piolet project was to test whether SC seemed to be a good method for measuring the willingness of road users to pay to reduce the environmental problems (in this connection, noise and air pollution) caused by road traffic in urban areas. As in the main project, we used SC applied to a particular journey. The data were analysed by means of logit models.

2.1 Choice exercise 1

Table 1
Choice exercises and factors/attributes included in the Environmental Assessment Survey in Oslo/Akershus, 1993

Factors/attributes	Choice exercise				
	1	2	3	4	5
Cost	X	X	X	X	X
In-vehicle travel time	X	X	X	X	X
Seat availability (public transport)	X				
Walking time (car users)	X				
Noise		X		X	
Local air pollution		X		X	
Dust and dirt from road wear			X		X
CO_2[a]			X		X
No. of respondents (car/public transport)	897/580	596/373	289/196	1179	683

[a] CO_2 is included to give a complete an assessment as possible of environmental problems due to road traffic and so as to prevent confusion of local and global environmental problems.

the valuation even if the experiment contains only three attributes (Sælensminde, 1998a). Maybe people are able to handle four or more attributes, but it was considered that the design was 'safer' when the number of attributes in this study was restricted to four. The inclusion of a maximum of 4 attributes in each exercise, and therefore different attributes included in the different exercises, may cause scale effects not accounted for in the analysis.[3] Considerable care is therefore required when comparing the results from the different exercises/models.

All those who had made a journey by car or by public transport first took part in choice exercise 1. They were then randomly divided into two groups. One group took part in choice exercises 2 and 4 and the other in choice exercises 3 and 5. Except for choice exercise 1, which always came first, the other choice exercises were presented in random order. We wanted the journey in choice exercise 1, 2 and 3 to be fresh in the respondents' memory, the respondents who had made no journey during the last 2 weeks therefore only took part in choice exercises 4 and 5. Choice exercise 1 produces four observations (four paired choices) per respondent, and the other choice exercises produce six observations per respondent.

2.1. Choice exercise 1

Choice exercise 1 was a relatively easy introductory choice exercise in which the respondents had to choose between two journeys in which known factors as cost, journey time, and seating varied. The respondent is supposed to imagine that the choice exercise relates to the actual journey, the same means of transport, the same purpose, the same time of day, etc., and that only the factors included in the choice exercise change.

Before the choice exercises and environmental factors were presented, the following were carried out to prepare the respondents: (a) To help them to clarify their own preferences, the

[3] In Sælensminde (1998a) the sufficiency of the scaling approach is discussed with regard to handling problems connected to noisy choices because of simplifications and inconsistencies.

respondents were asked a number of questions on their attitudes and priorities with regard to the health services, education, defence, employment and various environmental problems; (b) We gave definitions of local air pollution, noise, dust/dirt and CO_2 emissions, which were the four environmental factors we intended to focus on; (c) The respondents were given background information on environmental problems caused by road traffic, i.e. the damage and nuisances that *can* arise; (d) They were reminded (both in text and in figures) that any money for environmental improvements would have to come out of their own private budget;[4] (e) They were told which public bodies had commissioned the project, and that their answers could influence both the work of the authorities on the environment and the rates set for road user charges, petrol taxes, and public transport charges as well as the tax level in general; (f) Finally, we used traffic surveys for different days and periods on roads where the respondents were familiar with the volume of traffic so as to give a picture of what a given percentage reduction of pollution could mean in practice.

Before choice exercises 2 and 3, it was also pointed out that the following conditions applied to the choices to be made: (a) Respondents were to imagine that the choice exercises related to the actual journey (i.e. the same means of transport, the same purpose, the same time of day, etc.); (b) Present fuels were to be *replaced* by a new type of fuel. The respondents were to choose which fuel they wanted to use in future; (c) The new fuel would be available at all petrol stations, and the performance of the vehicles (e.g. acceleration, comfort, and range) would be the same as with present fuels; (d) The measure outlined (new fuel) will *only* affect the factors included in the choice situations.

2.2. Choice exercises 2 and 3

Choice exercises 2 and 3 offered choices between environmental consequences (new levels for environmental factors) of using new types of fuel. The choice exercises related to two specific journeys in which the purpose and length of the journey and the means of transport were the same as in the actual journey. Cost, journey time and the degree of environmental improvement or deterioration varied. The choice situation in choice exercise 2 is shown in Fig. 1.

The levels of the factors in choice exercises 2 and 3 are shown in Table 2. The respondents were divided randomly into groups A, B and C. Thirty percent of the respondents were placed in group A, 60% in group B and 10% in group C. A minor share of the respondents were allocated to group C because it was uncertain if the respondents would distinguish between the attribute levels designed to group C. The environmental attribute levels vary around three different 'average levels'. These have been set at 20, 40 and 60% reductions, labelled A2, B2 and C2, respectively, in Table 2. That means that the reductions for group A were −20% (i.e. a 20% deterioration), 20% (the average level), and 60%. The basic level for environmental factors is the present level; in other words 'the situation as experienced on Norwegian roads today'. The basic levels for the cost

[4] The inclusion of a budget reminder is one of many proposals from the NOAA panel (Arrow et al., 1993) to improve the validity of CV studies. The effect of a budget reminder is uncertain and the proposal from the NOAA panel has been investigated and discussed in many studies. See Sælensminde (1998b) for an overview. In connection to the discussion about a budget reminder a useful distinction may be between an experiment where the respondent assesses the goods in terms of their private utility vs an assessment as a citizen. Daniels (1997) investigates such a distinction.

and time factors are the actual travel cost and the actual in-vehicle travel time. The difference in the calculation of the cost factor levels for car journeys and journeys by public transport makes up for the fact that car travellers have a lower basic level (actual journey cost) than travellers by public transport. The use of the respondent's WTP per year for environmental improvements from a previous CV question divided by the respondent's total Travel Costs (TC) per year (WTP/TC in Table 2) is a way to customise the levels of the price attribute in the SC experiment to different respondents.[5]

Car journey A	**Car journey B**
Noise from road traffic reduced by 20% Local air pollution caused by road traffic reduced by 20% In-vehicle time = 30 min. Fuel cost for the journey = NOK 25	Noise from road traffic reduced by 20% Local air pollution caused by road traffic reduced by 60% In-vehicle time = 40 min. Fuel cost for the journey = NOK 30

Fig. 1. Example of one choice from excercise 2. Interviewees are asked to choose between journey A and journey B

Table 2
Levels of factors/attributes in choice exercises 2 and 3

Factors specified	Levels for car journey		Levels for journey by public transport
Cost[a]	1. Basic = actual travel cost 2. Basic×(1 + WTP/TC) 3. (Basic×(1 + WTP/TC))×1.5		1. Basic = actual travel cost 2. Basic×(1 + WTP/(2×TC)) 3. Basic×(1 + WTP/TC)
Travel time (min on this journey)		1. −25% 2. Basic = actual travel time 3. +25%	
Environmental attributes (% reduction from present level)	A1. −20[b] A2. 20 A3. 60	B1. −10[b] B2. 40 B3. 90	C1. 30 C2. 60 C3. 90

[a] WTP is willingness to Pay per year according to a previous CV question in the questionnaire. TC is travel costs per year by car or public transport according to previous questions.
[b] Negative reduction = increase.

[5] By use of the CV data from this study, such a customisation (of the bid-vector) has proven to be a possible way to reduce, or even eliminate, the difference between WTP from open-ended and closed-ended CV questions (Halvorsen and Sælensminde, 1998).

3. Modelling

4. Data collection

3. Modelling

Based on the assumptions that all choices are independent from irrelevant alternatives (IIA)[6] and the errors (the random and unobserved part of the indirect utility) are Gumbel-distributed, the choice data were analysed by means of logit models [Eq. (1)][7]; see McFadden (1973).

$$Pr(y_n = 1) = \frac{e^{V_{nA}}}{e^{V_{nA}} + e^{V_{nB}}} = \frac{1}{1 + e^{-(V_{nA} - V_{nB})}} \tag{1}$$

In Eq. (1), y_n is individual n's choice between the two alternatives $i, i = A, B$. (A is set equal to 1 and B is set equal to 0.) The systematic (observable) part of the (indirect) utility, V, of the alternative chosen, i, can be stated as

$$V_{ni} = \beta_0 + \beta_1 \ln X_{ni1} + \ldots + \beta_K X_{niK}, \tag{2}$$

where $\beta_k, k = 0, \ldots, K$, are the $K + 1$ unknown parameters to be estimated and $X_k, k = 1, \ldots, K$, are the K explanatory variables in the model. In Eq. (2) variable 1 is the price variable. The choice of nonlinear functions (log of the price variable) in the models is a result of tests with Box–Cox models that consider higher order effects.

The fact that respondents answer multiple choice tasks may give rise to specific interdependency problems in the error structure (autocorrelation) which are ignored in the standard logit or probit models usually applied in this context. Ouwersloot and Rietveld (1996) show how the solution from the panel data field to such autocorrelation problems can be applied in the context of stated choice experiments. Despite the results of their analysis, that the effect of autocorrelation is very modest and statistically insignificant, this may be an important topic for further research, but it was not addressed in this study.

How the different factors included in the choice exercises are valued is expressed as the *relation* between two estimators (the marginal rate of substitution between two attributes). The uncertainty of the calculated valuations will thus depend on the standard deviation of the parameters in addition to the correlation between them.

4. Data collection

In this study, undertaken in the autumn of 1993, five different SC exercises were used. Each choice exercise involved choices between two alternatives. An example of such a choice is given in Fig. 1. The 1680 respondents, a representative sample of residents in Oslo and the county of Akershus, took part in a maximum of three choice exercises each. The respondents were interviewed face-to-face in their homes and the interviewers used portable computers.

[6] The IIA property states that "..for any individual, the ratio of choice probabilities of any two alternatives is entirely unaffected by the systematic utilities of any other alternatives" (Ben-Akiva and Lerman, 1985). Data sets of actual choices may be tested for the IIA property; this was not done in this study.

[7] Eq. (1) implies constant error variance, which is usually assumed in logit models. This restriction may be relaxed if one multiplies the indirect utility by a scale parameter. See for example Bhat (1995) and Bradley and Daly (1994) for use of such heteroscedastic models.

5. Main results of the analysis

5.1 Choice exercise 1 – assessment of journey time, seat availability and walking time

The Norwegian Gallup Institute undertook the data collection. The sample was drawn at random from the telephone register. In the recruitment of the respondents the interview was described to last between 30 min and 1 h, but 'environmental issues' was not mentioned as a subject. The completion rate for questionnaires was about 40%. This is about the same rates as for other Norwegian studies with telephone recruitment.

Our sample was compared to the whole population in Oslo and the county of Akershus and we found very little differences with respect to sex and age. With respect to education we found some differences, but this may be a result of how the question about education is stated. Due to missing statistics at county level, it is not possible to compare income in our sample with income in the population of Oslo and Akershus.

5. Main results of the analysis

In the following we present the main results of choice exercises 1, 2 and 3 included in the stated choice.[8] Both the assessments of the different factors, calculated as the relation between the estimated parameters, and the estimated coefficients from the logit models are given.[9]

For choice exercises 2 and 3 we estimated a common model for travellers by car and public transport. This was done because the results suggest that the mode of transport is of little significance in the relative assessment of the factors in the choice exercise, and because we thereby obtained an average model for travellers by car and public transport.[10]

5.1. Choice exercise 1—assessment of journey time, seat availability and walking time

Tables 3 and 4 show the estimated coefficients and the calculated valuations for choice exercise 1. As mentioned above, it is necessary to show caution when comparing the results from the different exercises/models. Nevertheless, it may be interesting to compare the results from exercise I with the results from exercises 2 and 3. Introducing the environmental attributes may influence the role of the traditional attributes and for example the valuation of travel time. In terms of *valuation*, Tables 3, 5 and 7 show that travel time preserves its dominating role. But in terms of *t-values of the parameters* in the models and *lexicographic choices*[11] the environmental factors seem to be more important for the respondents than the traditional attributes of cost and travel time.

[8] Results of choice exercises 1, 4 and 5, and results from different segmentations are presented in Sælensminde and Hammer (1994).

[9] The data were analysed using the ALOGIT programme package from Hague Consulting Group.

[10] A better way to analyse the data may be by use of covariates or scale parameters in a common model. In such a model current mode effects could be tested.

[11] Lexicographic choices are defined as choices where the respondent in all his/her choices has chosen the best level for one of the attributes. More respondents chose lexicographically according to the environmental attributes than price and travel time in exercises 2 and 3. The share of lexicographic choices in exercise 2 are 7.3, 5.9, 43.7 and 6.5 for the attributes price, travel time, dust/dirt and CO_2, respectively. In exercise 1 more respondents have chosen lexicographically according to the price attribute. The lexicographic shares in exercise 1 are 41.2, 14.2 and 17.0 for the attributes price, travel time and walking time/seat, respectively. [See Sælensminde and Hammer (1994) and Sælensminde (1998a) for an overview of the impacts of lexicographic choices in stated choice studies.]

Table 3
Estimated coefficients from the logit model. Results from choice exercise 1 (*t*-ratio in parentheses)

Variable name	Car drivers (N = 3577)	Public transport (N = 2318)
Constant	0.0248 (−0.7)	0.08641 (2.0)
Travel cost	−2.448 (−22.3)	−1.633 (−9.3)
Travel time	−0.06084 (−10.8)	−0.02348 (−4.2)
Walking time	−0.05727 (−2.0)	
Seat availability		0.5406 (7.1)
ρ^2 w.r.t. zero	0.1573	0.0313
Log-likelihood (final value)	−2089	−1557

Table 4
Assessments of travel time, walking time and seat availability. Results from choice exercise 1. Denomination: NOK per h for travel time and walking time and NOK per journey for a seat on public transport (*t*-ratio in parentheses) (1 NOK≈0.15 $)

Variable name	Car drivers (N=3577)	Public transport (N=2318)
Travel time	20.70 (47.7)	13.11 (20.0)
Walking time	19.48 (7.7)	
Seat availability		5.03 (37.7)

Table 5
Estimated coefficients from the logit model. Groups A, B and C with 20, 40 and 60% average reductions of environmental factors and for the whole sample. Results from choice exercise 2

Variable name	Estimated coefficients (*t*-ratio in parentheses)			
	Group A (N=1757)	Group B (N=3497)	Group C (N=558)	Whole sample (N=5812)
Constant	−0.09471 (−1.7)	0.005699 (0.1)	0.02128 (0.2)	−0.02106 (−0.7)
Travel cost	−1.561 (−7.5)	−1.519 (−10.7)	−2.245 (−6.1)	−1.525 (−13.9)
Travel time	−0.06312 (−6.8)	−0.06077 (−9.4)	−0.06730 (−4.7)	−0.06070 (−12.3)
Local air Pollution	−0.03533 (−17.6)	−0.02722 (−25.2)	−0.03371 (−8.6)	−0.02927 (−31.8)
Noise	−0.01775 (−9.9)	−0.009355 (−9.7)	−0.01442 (−4.0)	−0.01137 (−13.9)
ρ^2 w.r.t. zero	0.2154	0.2210	0.1410	0.2057
Log-likelihood (final value)	−955	−1888	−332	−3199

Table 6
Assessments of travel time, local air pollution and noise for groups A, B and C with 20, 40 and 60% average reductions of environmental factors, and for the whole sample. Results from choice exercise 2. Denomination: NOK per h for journey time and NOK per percentage point per journey for the environmental factors (*t*-ratio in parentheses) (1 NOK≈0.15 $).

	Group A (N=1757)	Group B (N=3497)	Group C (N=558)	Whole sample (N=5812)
Journey time	42.65 (32.6)	43.11 (33.4)	32.99 (23.5)	42.70 (45.9)
Local air pollution	0.40 (40.7)	0.32 (57.1)	0.27 (34.6)	0.34 (74.4)
Noise	0.20 (44.6)	0.11 (53.8)	0.12 (21.9)	0.12 (73.2)

5.2 Choice exercise 2 – assessment of journey time, local air pollution and noise

5.3 Choice exercise 3 – assessment of journey time, dust/dirt and CO_2

5.2. Choice exercise 2—assessment of journey time, local air pollution and noise

Tables 5 and 6 show the results using a model for the whole sample and the results when a selection was made according to average levels of change in the environmental factors (cf. Table 2). The results show that the estimated parameters have the expected sign and are significantly different from zero. The results of choice exercise 2 show that reduction of local air pollution and noise is assessed more highly per percentage point when the reduction[12] is small, as for group A, than when it is large, as for group C (i.e. declining marginal utility).

5.3. Choice exercise 3—assessment of journey time, dust/dirt and CO_2

The results from choice exercise 3 are reported in the same way as for choice exercise 2. Tables 7 and 8 show that the assessments of all the factors have the correct signs and are significantly different from zero.

In contrast to choice exercise 2, the results of choice exercise 3 show that reductions in dust/dirt and CO_2 are valued about as highly per percentage point whether they are small or large.

Table 7
Estimated coefficients from the logit model. Groups A, B and C with 20, 40 and 60% average reductions of environmental factors and for the whole sample. Results from choice exercise 3

Variable name	Estimated coefficients (*t*-ratio in parentheses)			
	Group A ($N=924$)	Group B ($N=1685$)	Group C ($N=294$)	Whole sample ($N=2903$)
Constant	0.09756 (1.3)	−0.1385 (−2.6)	−0.1033 (−0.8)	−0.06353 (−1.6)
Travel cost	−1.732 (−5.6)	−1.167 (−5.8)	−1.487 (−3.2)	−1.287 (−8.3)
Travel time	−0.04099 (−3.7)	−0.03610 (−4.8)	−0.07315 (−3.4)	−0.03889 (−6.8)
Dust and dirt	−0.02566 (−9.7)	−0.01769 (−12.9)	−0.02294 (−4.4)	−0.01966 (−16.7)
CO_2	−0.03068 (−11.4)	−0.01874 (−13.6)	−0.02512 (−4.7)	−0.02169 (−18.3)
ρ^2 w.r.t. zero	0.1804	0.1392	0.0854	0.1394
Log-likelihood (final value)	−525	−1005	−186	−1731

Table 8
Assessment of travel time, dust/dirt and CO_2 for groups A, B and C with 20, 40 and 60%, respectively, average reductions of environmental factors and for the whole sample. Results from choice exercise 3. Denomination: NOK per h for journey time and NOK per percentage point per journey for the environmental factors. (*t*-ratio in parentheses.)(1 NOK ≈ 0.15$).

	Group A ($N=924$)	Group B ($N=1685$)	Group C ($N=294$)	Whole sample ($N=2903$)
Journey time	25.79 (15.7)	34.00 (13.4)	53.10 (12.9)	33.92 (25.6)
Dust/dirt	0.27 (34.4)	0.28 (31.2)	0.28 (18.3)	0.28 (46.2)
CO_2	0.32 (33.2)	0.29 (31.1)	0.30 (18.1)	0.31 (45.5)

[12] The fact that groups A and B also experienced increased local air pollution and increased noise levels in the choice exercises is probably important for this result too.

5.4 "Focusing effects"

5.5 Assessment per annoyed person

5.4. 'Focusing effects'

'Focusing effects', 'interaction effects' or 'package effects' (as we can call them in SC) can result in higher than 'correct' assessments and often occur if only one or a few of the factors are assessed which 'naturally belong in a package'. This problem affects SP methods generally and may cause doubt if one can aggregate WTP estimates of different public goods elicited from different studies or contexts. Hoehn and Randall (1989) and Randall (1991) discuss aggregation bias that occurs when a project is valued in isolation when it is a part of a more general political agenda. These studies show that the sum of the partially valued projects on the agenda exceeds the simultaneous valuation as the number of projects on the agenda increases. See Sælensminde (1998b) for an overview of the subject of interaction effects and results from an empirical study of the valuation of reduced travel time, road safety and environmental factors included in cost/benefit analyses for the road sector in Norway.

In the current SC study we have assessed two of the environmental factors in each choice exercise and not the whole 'package' of four factors (cf. Table 1). In an attempt to take this into account, we have assumed that this has resulted *at the most* in an assessment of the environmental factors in the choice exercises that is twice as high as the 'correct' one.[13] This view is also in line with the recommendations from National Oceanic and Atmospheric Administration (NOAA, 1994).[14] By halving the assessment of the environmental factors in the choice exercises, we thus arrive at an interval that includes the 'methodological uncertainty' attached to the focusing effect.

Intervals in valuation of environmental factors are shown in Table 9. The lower limit of each interval takes account of the focusing effect. The upper limit results from the choice exercises without regard to the focusing effect. Choice exercises 2 and 3 give the assessment of environmental factors per journey. To convert this into assessments per year as in Table 9, a total of 750 journeys per year were assumed.[15]

5.5. Assessment per annoyed person

In order to utilise the results in cost/benefit analyses, WTP was calculated per person for whom road traffic is a nuisance.[16] As an example, WTP per household for a 50% reduction in the environmental factors was calculated by multiplying the figures in Table 9 by 50. By multiplying this WTP per household by the total number of households (410,000), the 'WTP of the total population of

[13] This *ad hoc* correction is based on the fact the valuation of the four factors are not 'too different'.

[14] According to Navrud and Pruckner (1997) this recommended ad hoc reduction of the estimated WTP by 50% from NOAA is an oddity. The rationale behind this was the lack of external validation and a set of studies that indicate that hypothetical WTP turns out to overstate actual WTP. However, in the presence of NOAA's efforts to eliminate or at least greatly reduce potential biases in CV surveys and thus guarantee conservative results, the requirement to discount WTP by 50% seems counter-intuitive.

[15] Source: The Norwegian National Travel Survey 1990.

[16] The Norwegian Public Roads Administration (NPRA) have models that, based on dose response relations, compute the number of persons annoyed by noise and air pollution for different roads and traffic conditions (NPRA, 1995a). In their cost/benefit analysis NPRA use the results from exercises 4 and 5, which are slightly lower than the results from Exercises 2 and 3. NPRA also use higher valuations for detoriations of the environment than for improvements of the environment. This, together with the fact that only highly annoyed persons are included in their models, makes it difficult to recognise the valuations in Tables 8, 9 and Table 10 in the description of the cost/benefit analysis given in NPRA (1995a,b)

Oslo/Akershus' was calculated. This is shown in Table 10. 'WTP per annoyed person' was calculated by dividing the population's total WTP by the total number of persons annoyed by smells, noise and dust/dirt, respectively. The number of persons annoyed by various road traffic problems was calculated on the basis of questions in the survey about the experience of various forms of nuisance. The results from this SC study are in the same magnitude as results from other Scandinavian WTP studies where the CV method is used. Studies that have used hedonic pricing, calculated effort costs or damage cost by use of dose response relations show somewhat different results. Some numerical examples of other valuation studies are cited in Table 11 to show both similarities

Table 9
Intervals in valuation of environmental factors in the stated choice study in Oslo/Akershus 1993. Denomination: NOK per percentage point of change per year per household (1 NOK ≈ 0.15 $).

Factor	Assessments from choice exercises 2 and 3
Local air pollution	127–255
Noise	45–90
Dust/dirt	105–205
CO_2	116–233

Table 10
Calculated interval for valuation of a 50% reduction in the level of four environmental factors. Valuation based on exercises 2 and 3 in the stated choice study in Oslo/Akershus 1993

Factor	WTP of the total population of Oslo/Akershus[a] (denomination: Mill NOK per year)	WTP per annoyed person[b] (denomination: NOK per year, round figures)
Local air pollution	2614–5227	13,700–27,400
Noise	923–1845	3550–7100
Dust/dirt	2153–4305	6900–13,800
CO_2	2388–4777	

[a] 410,000 households.
[b] About 191,000 people were annoyed by the smell, about 260,000 by noise and 313,000 by dust/dit in Oslo and Akershus in 1993. The same weighting was given to 'slightly' and 'seriously' annoyed people.

Table 11
Costs due to noise and air pollution exposure from road traffic. Calculated in a Norwegian context with application of results from previous studies

Source	Valued	Method	Costs per person (1993; NOK/year)
Larsen (1985)	Noise	HP	1600
Nielsen and Solberg (1985)	Noise	EC	5700
Wenstøp et al. (1993)	Noise	DP	6100
Leksell (1987)	Air pollution	DC	19,200
Strand (1985)	Air pollution	EC	10,800
Strand (1985)	Air pollution	CV	13,200–28,200

HP = Hedonic price method; EC = effort costs; DP = decision panels, willingness to pay; DC = damage cost, CV = contingent valuation.
Source: Sælensminde (1992) and Sælensminde and Hammer (1994).

and differences depending on valuation method. See Sælensminde (1992) and Sælensminde and Hammer (1994) for a more comprehensive presentation of these and other valuation studies.

6. Discussion

The results of this SC study have been used by the Norwegian Public Roads Administration in their cost/benefit analyses since 1995. This study was a first attempt to value environmental goods using SC in Norway and the design of such a study might certainly be improved in many ways. A possible weakness is the presentation of environmental change in percentage of the pollution situation today which is more difficult for the respondents to imagine than end-effects based on dose–response relations. One improvement is therefore to base new studies on the knowledge about dose–response relations that has been presented since 1993. Another improvement may be to include a 'non-choice' alternative enabling the respondents to express that none of the presented alternatives are acceptable and a 'don't know' and/or an 'equal' alternative enabling the respondents to express that they are uncertain about which alternative to choose. If the respondents are able to distinguish between such alternatives the 'non-choice' alternative may move the valuations from a relative context towards an absolute context and the 'don't know' and 'equal' alternatives may be used in models that take explicit account of preference uncertainty.

One advantage of SC compared to CV is that in SC the environmental benefits can be assessed relative to each other within a familiar framework such as a concrete journey. This makes the choice situation less hypothetical to the respondent. A second advantage is related to the fact that in SC studies the respondent is usually given more than one choice in each choice exercise. This gives the researcher a possibility to investigate how each respondent has completed the presented choice exercise(s), for example, whether his/her choices are consistent or if he/she has behaved in a lexicographic fashion. [Such investigations have been undertaken by Foster and Mourato (1997) and Sælensminde (1998a).] By lexicographic choices is meant that the respondent always chooses that alternative where one of the factors (e.g. price) is at its best level. Lexicographic choices may be a result of simplification if the respondent finds it difficult to choose between the presented alternatives or a result of actual preferences if too different levels of the factors are included in the choice experiment. In consequence, lexicographic choices in a choice experiment do not necessarily mean that the respondent has lexicographic preferences. A third advantage is that SC may avoid 'yea-saying' which is a problem in discrete choice CV studies. 'Yea-saying' occurs if some respondents agree with an interviewer's request regardless of their true views (Mitchell and Carson, 1989).

At this point, we cannot conclude that SC is an answer to the important question: 'How, by means of hypothetical questions, can one be certain that the benefits one wishes to assess have been correctly quantified?' This view is based on the fact that general problems with SP methods (both CV and SC), such as focusing, ordering, embedding and scope effects, will influence the results. Ordering effects occur when the sequence in which the valuation questions are asked affects the finale valuation of different goods, when a package of goods is valued in a sequence. Embedding effects occur when the valuation is insensitive to the inclusiveness of the package of goods for which valuation is sought.[17]

[17] Adamowicz (1995) and Adamowicz et al. (1998) have speculated that SC experiments might be a good way around the embedding problem encountered in CV studies. Halvorsen et al. (1996) does not share this view.

References

A particular version of the embedding problem illustrated by Desvousges et al. (1993) is when the valuation of a good is insensitive to the amount in which it is provided. This effect is what Carson and Mitchell (1993) denote as "problems with scope". Much recent criticism of CV put forth in Desvousges et al. (1993), Kahneman and Knetsch (1992), Hausman (1993) and Diamond and Hausman (1994) is related to these problems. See Sælensminde (1998b) for an overview.

In addition to the general SP problems, the complicated choice situation for the respondents and a more complicated design for the researcher are the two main disadvantages of SC. It is still an unanswered question what effect simplified decision-making procedures (e.g. lexicographic answers) have on SC results, and how lexicographic answers due to simplification can be distinguished from answers due to dominant factors/real preferences. There is also much more to be done with the possibilities in SC studies when exploring the consistencies of individual respondent's choices (Sælensminde, 1998a). Halvorsen et al. (1996) compared advantages and disadvantages with CV, SC and Decision Panels and conclude that none of these valuation methods are preferable to others in essential details.

Despite the fact that Norwegian authorities make use of the results from the current SC study in their cost/benefit analyses, considerable uncertainty remains with regard to (1) the area of validity of the results (especially their limited validity where deterioration of air pollution and noise levels is concerned), (2) the impact on the results of focusing/interaction effects and other general SP problems, and (3) problems related to the complex choice situation in SC studies. Further surveys are therefore required to correct for such sources of uncertainty.

Acknowledgements

The Ministry of Transport and Communications in Norway and The Norwegian Public Roads Administration financed the research. The helpful comments of Ståle Navrud, Lasse Fridstrøm, Rune Elvik and two anonymous referees are greatly appreciated, as is the excellent research assistance and comments of Frode Hammer. Of course, the responsibility for any remaining errors is mine.

References

Adamowicz, W. 1995. Alternative valuation techniques: A comparison and movement toward a synthesis. In Environmental Valuation: New Perspectives, eds, K. Willis and J. Corkindale. CAB International, Oxford.

Adamowicz, W., Louviere, J., Williams, M., 1994. Combining revealed and stated preference methods for valuing environmental amenities. Journal of Environmental Economics and Management 26(3), 271–292.

Adamowicz, W., Swait, J., Boxall, P., Louviere, J., Williams, M. 1996. Perceptions versus objective measures of environmental quality in combined revealed and stated preference models of environmental valuation. Mimeo, Department of Rural Economy, University of Alberta, Canada.

Adamowicz, W., Boxall, P., Williams, M., Louviere, J., 1998. Stated preference approaches to measuring passive use values: choice experiments versus contingent valuation. Mimeo, Department of Rural Economy, University of Alberta, Canada. American Journal of Agricultural Economics 80, 64–75.

Arrow, K., Solow, R., Portney, P.R., Leamer, E.E., Radner, R., Schuman, H. 1993. Report of the NOAA Panel on Contingent Valuation. Federal Register, 15 January, 58(10), 4601–4614.

Ben-Akiva, M., Lerman, S.R. 1985. Discrete Choice Analysis. The MIT Press, Cambridge, MA.

Bhat, C.R., 1995. A heteroseedastic extreme value model of intercity travel mode choice. Transportation Research 29B(6), 471–483.

Boxall, P.C., Adamowicz, W.L., Swait, J., Williams, M., Louviere, J., 1996. A comparison of stated preference methods for environmental valuation. Ecological Economics 18(3), 243–253.

Braden, J.B., Kolstad, C.D. (eds) 1991. Measuring the Demand for Environmental Quality. Contributions to economic analysis, no. 198, North Holland, Amsterdam.

Bradley, M., Daly, A., 1994. Use of the logit scaling approach to test for rank-order and fatigue effects in stated preference data. Transportation 21(2), 167–184.

Carson, R.T., Mitchell, R.C., 1993. The issue of scope in contingent valuation studies. American Journal of Agricultural Economics 75, 1263–1267.

Daniels, R. 1997. A stated choice framework for the valuation of environmental externalities of road projects. Doctor of Philosophy thesis, Institute of Transport Studies, Graduate School of Business, The University of Sydney, 1997.

Desvousges, W.H., Reed Johnson, F., Dunford, R.W., Hudson, S.P., Wilson, K.N., Boyle, K.J. 1993. Measuring natural resource damages with contingent valuation: tests of validity and reliability, chapter 3. In Contingent Valuation. A Critical Assessment, ed. J.A. Hausman. Contributions to economic analysis, no. 220. North-Holland, Amsterdam.

Diamond, P.A., Hausman, J.A., 1994. Contingent valuation: is some number better than no number?Journal of Economic Perspectives 8(4), 45–64.

Foster, V., Mourato, S. 1997. Are consumers rational? Evidence from a contingent ranking experiment. Paper presented at the 8th Annual Conference of the European Association of Environmental and Resource Economists, 26–28 June 1997, Tilburg, The Netherlands.

Halvorsen, B., Strand, J., Sælensminde, K., Wenstøp 1996. Comparing contingent valuation, conjoint analysis and decision panels: an application to the valuation of reduced damages from air pollution in Norway. Memorandum no. 37 1996 from Department of Economics, University of Oslo.

Halvorsen, B., Sælensminde, K. (in press) Differences between willingness-to-pay estimates from open-ended and discrete choice contingent valuation method: the effect of heteroscedasticity. Land Economics.

Hanley, N., MacMillan, D., Wright, R., Bullock, C., Simpson, I., Parsission, D., Crabtree, R. 1997. Contingent valuation versus choice experiments: estimating the benefits of environmentally sensitive areas in Scotland. Paper presented at the 8th Annual Conference of the European Association of Environmental and Resource Economists, 26–28 June 1997, Tilburg, The Netherlands.

Hausman, J.A. (ed.) 1993. Contingent Valuation. A Critical Assessment. Contributions to economic analysis, no 220, North-Holland, Amsterdam.

Hensher, D.A., 1994. Stated preference analysis of travel choices: the state of practice. Transportation 21(2), 107–133.

Hoehn, J.P., Randall, A., 1989. Too many proposals pass the benefit cost test. American Economic Review 79, 544–551.

Kahneman, D., Knetsch, J.L. 1992. Contingent valuation and the value of public goods: reply. Journal of Environmental Economics and Management 22 90–94.

Kemp, M.A., Maxwell, C. 1993. Exploring budget context for contingent valuation estimates. In Contingent Valuation. A Critical Assessment, ed. J.A. Hausman. Contributions to economic analysis, no. 220, North-Holland, Amsterdam.

Larsen, O.I. 1985. Vegtrafikk og Boligpriser. TØI-rapport (ISBN 82-7133-4980), Transportøkonimisk institutt, Oslo (in Norwegian).

Leksell, I. 1987. Samhäallsekonomisk vädrdering av bilavgasutsläpp. DsK 1987:6, Kommunikationsdepartementet, Stockholm (in Swedish).

Louviere, J.J., Hensher, D.A., 1982. On the design and analysis of simulated or allocation experiments in travel choice modelling. Transportation Research Record 890, 11–17.

Louviere, J.J., Woodworth, G.G., 1983. Design and analysis of simulated choice or allocation experiments: an approach based on aggregated data. Journal of Marketing Research 20, 350–367.

McFadden, D. 1973. Conditional logit analysis of qualitative choice behavior. In Frontiers in Econometrics, ed. P. Zaremmbka. Academic Press, New York.

McFadden, D., Leonard, G.K. 1993. Issues in the contingent valuation of environmental goods: methodologies for data collection and analysis. In Contingent Valuation. A Critical Assessment, ed. J.A. Hausman. Contributions to economic analysis, no. 220, North-Holland, Amsterdam.

Mitchell, R.C., Carson, R.T. 1989. Using surveys to value public goods: the contingent valuation method. Resources for the Future. The Johns Hopkins University Press, Baltimore, MD.

Navrud, S. (ed.) 1992. Pricing the European Environment. Scandinavian University Press, Oslo.

Navrud, S., Pruckner, G.J., 1997. Environmental valuation—to use or not to use. Environmental and Resource Economics 10, 1–26.

Nielsen, G., Solberg, S. 1985. Kostnader og effekter av tiltak mot vegtrafikkstøy. Grunntag for et nasjonalt handlingsprogram. TØI-rapport (ISBN 82-7133-497-2), Transporto/konomisk institutt, Oslo (in Norwegian).

NOAA 1994. Natural Resource Damage Assessments, Notice of Proposed Rulemaking. National Oceanic and Atmospheric Administration, 59 Federal Register, 1062, 7 January.

NPRA (1995a) Håndbok 140, Konsekvensanalyser (del I-III). Statens vegvesen (in Norwegian).

NPRA (1995b) Brukerveiledning og dokumentasjon, MIKO. Statens vegvesen (in Norwegian).

Ouwersloot, H., Rietveld, P., 1996. Stated choice experiments with repeated observations. Journal of Transport Economics and Policy 30(2), 203–212.

Randall, A. 1991. Total and nonuse values. Chapter 10 In Measuring the Demand for Environmental quality, ed. J.B. Braden, C.D. Kolstad. North-Holland, Amsterdam.

Strand, J. 1985. Verdsetting av reduserte luftforurensninger fra biler i Norge. (Valuation of reduced air pollution from road traffic in Norway.) Memorandum nr. 1/1985, Sosialøkonomisk institutt, Universitetet i Oslo, Oslo (In Norwegian).

Sælensminde, K. 1992. Miljøkostnader av vegtrafikk i byområder—resultater fra tidiigere studier. (Environmental costs caused by road traffic in urban areas—results from previous studies.) TØI-rapport nr. 115/1992, Transportøkonomisk institutt, Oslo (in Norwegian, with English summary).

Sælensminde, K. (1998a) Inconsistent choice's impact on the valuation of travel time in stated choice studies. Working paper, Institute of Transport Economics, Oslo.

Sælensminde, K. (1998b) Interaction effects between nonmarket goods included in benefit-cost analysis for road investments: valuation of travel time savings, traffic safety and environmental goods. Working paper, Institute of Transport Economics, Oslo.

Sælensminde, K., Hammer, F. 1993. Samvalganalyse som metode for verdsetting av miljøgoder—pilotundersøkelse. (Conjoint analysis as a method of assessing enviromental benefits—Pilot study.) TØI-rapport no. 184/1993, Transportøkonomisk institutt, Oslo (in Norwegian, with English summary).

Sælensminde, K., Hammer, F. 1994. Verdsetting av miljøgoder ved bruk av samvalganalyse—Hovedundersoøkelse. (Assessing environmental benefits by means of conjoint analysis—Main study.) TØI-rapport no. 251/1994, Transportøkonomisk institutt, Oslo (in Norwegian, with English summary).

Wenstøp, F., Carlsen, A.J., Bergland, O., Magnus, P. 1994. Valuation of environmental goods with expert panels. Research report 1994/1, BI Research Center, Oslo.

Paper 2: The impact of choice inconsistencies in Stated Choice studies

Kjartan Sælensminde

December 13, 1999

Abstract

A new test procedure revealing choices that are jointly inconsistent has been developed and applied to Stated Choice data. Our analysis shows that inconsistent choices commonly occur in several Stated Choice tasks and have a significant impact on the valuation of reduced travel time. Different abilities to choose (i.e. less education) cause inconsistent choices. The occurrence of inconsistent choices is shown to be most frequent early in the choice sequence and reduced for later choices. These results suggest that respondents may need more training and assistance to choose consistently in Stated Choice studies.

1. Introduction

Stated Choice (SC), and other methods based on choice experiments, have been used for many years in transport economics and marketing research (see e.g. Hensher 1994 for an overview). They are now becoming increasingly popular for the valuation of environmental goods (Widlert *et al.* 1993, Sælensminde and Hammer 1994, Boxall *et al.* 1996, Roe *et al.* 1996, Adamowicz *et al.* 1998, Hanley *et al.* 1998 and Sælensminde 1999).

Traditionally, however, environmental goods have been valued in studies using the Contingent Valuation (CV) method. CV has been used in numerous studies during the last 30 years and has been subject to much critique and debate (see for example Diamond and Hausman 1994, Hanemann 1994 and Portney 1994 for an overview). One important issue in the CV debate is how questions should be formulated to make them more familiar to respondents and in that way eases their choice task. The NOAA panel (Arrow *et al.* 1993) suggested that the discrete choice (DC) CV format was preferable to the open-ended (OE) CV format. However, issues concerning statistical uncertainty, stability and bias were not addressed by the panel, and the debate concerning the preferability of DC-CV to OE-CV continues (Neill 1995, Boyle *et al.* 1996, Dubourg *et al.* 1997, Frykblom 1997, Randall 1997, Boyle *et al.* 1998, Green *et al.* 1998 and Halvorsen and Sælensminde 1998). A move away from valuation of environmental goods by OE or DC-CV to use of SC, with two or more attributes in each choice, will again increase the complexity of the respondent's choice task. Task complexity is in this case determined by the number of alternatives, the number of attributes that vary in each alternative and the differences between the presented alternatives, which in turn is determined by the differences between the levels of each attribute. A move from CV to SC in valuation of environmental goods should, therefore, include consideration of the impact a more complex choice situation would have on respondent ability to state their preferences correctly. It is also important to investigate how possible "erroneous choices", due to a too complex choice situation, influence the analysis and therefore the valuation of the goods.

The aim of this study is to investigate whether complexity in the choice situation in SC studies results in inconsistent choices, and whether such choices have a significant impact on the valuation of the goods in the study.

Inconsistent choices refers to choices that are not consistent with previous or subsequent choices, i.e. choices that violate the transitivity axiom of consumer theory. Inconsistent choices are revealed because each choice gives an interval of the valuation of each attribute included in the choice task. If the valuation intervals for two choices do not overlap, we find these two choices to be mutually inconsistent. The occurrence of choice inconsistencies in SC studies has, apparently, not previously been investigated in this manner.

The current empirical study of inconsistent choices uses data from two Norwegian surveys. These were respectively designed to revise the values for travel time (Ramjerdi *et al.* 1997) and environmental goods (Sælensminde and Hammer 1994) used in cost-benefit analyses (CBAs). Two experimental designs were implemented in these surveys: a SC and an OE-CV study. The SC tasks used in the current study include only three attributes, and the context is related to a real journey. Occurrence of inconsistent choices in such simple choice situations may therefore concern practitioners of cognitively demanding data collection methods in general.

The paper commences with an overview of the existing literature on inconsistent choices in choice studies. Section three describes how tests of inconsistency are carried out. In section four, we present the random utility model used for the empirical analysis. Section five describes the surveys applied here. The sixth section presents the results from the analysis. Finally, section seven provides concluding remarks and the implications of our results for the design and analysis of future SC studies.

2. Previous literature on "problematic choices"

In previous literature both simple inspections and parametric approaches have been used in the investigation of problems occurring in choice studies because of choice task complexity. One approach is to inspect each respondent's choices. In such an approach one can detect the occurrence of lexicographic choices which may constitute a violation of the continuity axiom (see e.g. Widlert 1994), the occurrences of choices that contravene the IIA property and choices that violate the axioms of non-satiation and transitivity.[1] The design of a study by Foster and Mourato (1997) enable them to use this inspection approach to examine all such violations of consumer theory axioms. Another approach, parametric, is to use a statistical estimation method that allows for different error variances within a single model. Scaling, one such parametric approach (Ben-Akiva and Lerman 1985), has mainly been used in studies that utilize a mixture of Revealed Preference (RP) and Stated Preference (SP) data[2]. In the context of this paper, a more interesting use of scaling is as a tool in the investigation of the sources of variance (i.e. "noisy" choices) *within* a single SP data set (Widlert 1994, Bradley and Daly 1994, Mazzotta and Opaluch 1995, Swait and Adamowicz 1996 and Hensher *et al.* 1997). In another parametric approach applied by Johnson and Desvousges (1997), each respondent's multiple observations are used to estimate a model for each individual. A third parametric approach, also applied by Johnson and Desvousges (op.cit.), is to estimate a panel model with respondent-specific scale parameters for the latent randomutility distribution.[3] A fourth parametric approach is to include decision strategy selection as an explicit factor in the choice model (Swait and Adamowicz 1999). Both inspection of the data and parametric approaches provide the opportunity to investigate how inconsistent choices influence the results from a statistical model. One important difference is between approaches that can identify each respondent's problematic choices and approaches that only are able to detect that (some) respondents have made problematic choices. To see what is really going on, it may be important to identify both each problematic choice and each respondent's number of problematic choices. This view is in line with Foster and Mourato (1997), who conclude that "parametric analysis is highly sensitive to the presence of irrational responses in the dataset. It is therefore important to design contingent ranking surveys in such a way as to permit non-parametric tests of the fundamental axioms of consumer theory." They continue by noting that "the fact that this has not been standard in the literature to date raises questions about the accuracy of valuation estimates extracted from existing studies."

Bates (1994) points out that the danger of presenting respondents with tasks that are too complicated is well known to practitioners. What is less clear, however, is where the boundaries lie, and, in particular, how they might vary

---

[1] The continuity axiom of consumer theory states that two commodity bundles that are similar to each other will be ranked close together in the consumer's preference ordering. This axiom precludes lexicographic ordering of bundles by requiring respondents to trade-off gains in one commodity against losses in another. The non-satiation axiom states that a consumer must prefer a bundle that is in all respect superior to another. The transitivity axiom requires that if a consumer prefers option A over option B and option B over option C, then he must necessarily prefer option A over option C. The IIA (Independence from Irrelevant Alternatives) property, which is central to the logit framework of econometric analysis, requires that the ranking between two bundles in a choice set is not affected by the content of the remaining bundles in the set.

[2] See e.g. Ben-Akiva and Morikawa (1990), Bradley and Daly (1992), Hensher and Bradley (1993), Adamowicz *et al.* (1994) and Adamowicz *et al.* (1997).

[3] This is actually the extreme option of the scaling approach treating each person in the sample as a separate data set with its own scale factor.

between respondents of different intellectual ability. According to Bates, there is a need for much more research into how respondents actually carry out the tasks, using de-briefing techniques and alternative decision rules.

Inspired by Bates (1994), the current study investigates the effects of inconsistent choices and how such choice problems vary between respondents. (See Sælensminde 1998 for an analysis of effects of lexicographic choices.) The distribution of inconsistent choices in the choice sequence is also used to describe possible learning and fatigue effects.

3. Tests of consistency

The tests undertaken in the current study include all the choices that each respondent has made. This is a strong and comprehensive test procedure that will detect all inconsistencies that are present. Test procedures that only make use of a few of the respondents' choices may detect some of the violations against the axioms of consumer theory, but will not detect inconsistencies between all the choices made by each respondent.

This section introduces and explains the tests used to investigate whether the respondents choose consistently in the SC task. The results of these tests are then used to group the respondents according to their number of inconsistent choices as an indication of "task problems", but the modelling is done by use of a random utility model as is standard for SC data.

3.1 Test of consistency

The consistency test is based on the assumption that respondents have a given preference structure and that their choices should satisfy the axiom of transitivity according to consumer theory. The tests are applied to studies of people's choices between two different journeys with the same mode. In the example shown in Figures 1 and 2 the three attributes, price, travel time and headway (time between each departure), describe the journey with public transport. (See Section 4 for an overview of the attributes for the different modes.) The preference structure that characterises rational choices is based on the assumption that the respondents prefer to pay less, use less travel time and get a shorter headway.[4] Bates (1994) shows how a "ray diagram" can be used in the design of choice studies. In the current study the idea behind such a ray diagram is used to test if the choices made by each respondent are mutually consistent.

Figure 1 shows four different discrete choices. Each choice is between two long distance train journeys described by the three attributes, price, travel time and headway. The data used in the current study include nine such choices. The respondent's task is to choose, in each choice, if he/she prefers the journey

[4] In the choice between two longer car journeys the third attribute was automatic speed control; i.e. how many such speed control units that were to be used. No preference structure was assumed in the design of the study for that attribute.

described on the left-hand side (LHS) or the journey described on the right hand side (RHS). Each of the four choices in Figure 1 is illustrated graphically in the "ray diagram" in Figure 2.[5] Ray number 1 in Figure 2 is a graphical representation of choice number 1 in Figure 1 and represents the following: if the respondent's WTP for reduced headway is zero, he/she will choose LHS or RHS if his/her WTP for reduced travel time is more or less, respectively, than 100 NOK/h. 100 NOK/h is therefore a point on the axis denoted "valuation of travel time". Similarly, if the respondent's WTP for reduced travel time is zero, we achieve the point 50 NOK/h on the axis denoted "valuation of headway". The line between the two points represents linear combinations of the valuation of travel time and headway and, according to the continuity axiom, it is assumed that the respondents trade-off gains in travel time against losses in headway, and vice versa.

	Choice no. 1		Choice no. 2		Choice no. 3		Choice no. 4	
	LHS	RHS	LHS	RHS	LHS	RHS	LHS	RHS
Price (NOK)	300	200	300	200	300	225	300	225
Travel time (hour)	4	5	4	6	4	5	4	4,5
Headway (hour)	4	6	4	5	4	5	4	5

Figure 1. Four different choices between journeys described by the three attributes, price, travel time and headway. In each choice the respondent chooses whether he/she prefers the journey presented on the left hand side (LHS) or the journey presented on the right hand side (RHS).

If the respondent in choice 1 chooses RHS, then he/she is not willing to pay 100 NOK more for a journey where both travel time and headway are improved by the relevant amounts. Such a choice indicates that his/her valuation of reduced travel time is less than 100 NOK/h and that his/her valuation of reduced headway is less than 50 NOK/hour. In this case, the valuation of travel time and headway therefore lies in the area restricted by the two axes and ray 1. This conclusion is based on the *a priori* assumptions about preference structure and that an improvement in both travel-time and headway is not valued higher than the aggregate of an isolated valuation of these attributes.[6]

[5] All the choices in Figure 1 are configured such that by choosing the most expensive journey the respondent gets an improvement in both travel time and headway. This is done to simplify the graphical illustration.

[6] This is an assumption that may be wrong for some respondents. Sælensminde (1995) shows that a valuation of travel time and headway in a package may give a higher valuation than the sum of the valuation of travel time and headway valued separately. For most of the respondents these package effects are negative. *A priori*, it is difficult to determine the sign of such package effects for the individual respondent.

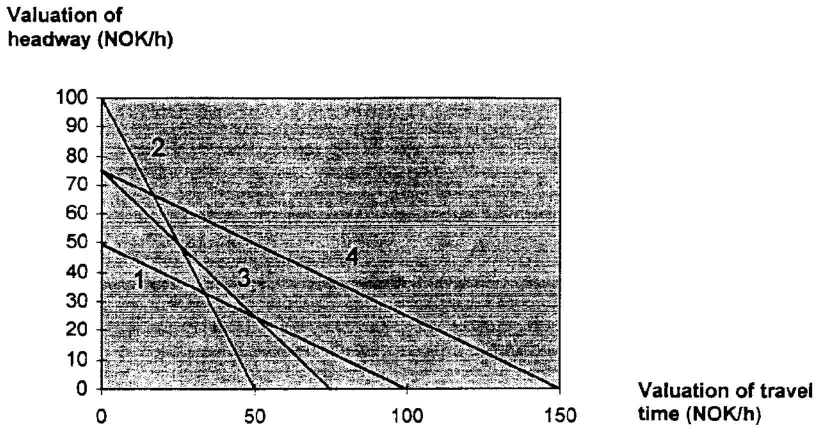

Figure 2: "Ray diagram" based on the four choices in Figure 1.

In a similar manner as for choice 1, it is possible to determine which side of the rays 2, 3 and 4 the respondent's valuation is positioned given the choices he/she makes in choice 2, 3 and 4. If the valuation area determined by one choice is part (not part) of the area determined by another choice that means that these two choices are mutually consistent (inconsistent).

3.2 Two degrees of consistency

The least restrictive test of consistency is when each choice made by one respondent is only tested against each of the respondent's previous choices separately.[7] In the examples presented in Figures 1 and 2 this means that choice 2 is tested against choice 1. According to Figure 2, choices 1 and 2 are mutually consistent irrespective of the choice made in these tasks. In similar manner, we see that choices 1, 2 and 3 are all mutually consistent when tested against each other separately. A failure of this least restrictive test can be illustrated by the choice of RHS in choice 1 and LHS in choice 4.

The most restrictive test of consistency is when each choice made by one respondent is tested against the "aggregate" of the previous choices. This means that choice 2 is first tested against choice 1. If he/she, for example, chooses RHS in both these choices, his/her valuation of travel time and headway lies in the area restricted by the two axes and the rays numbered 1 and 2 in Figure 2. When the respondent makes the third choice, this choice will not be consistent with the "sum" of the two previous choices unless he/she chooses the RHS alternative. To use the aggregate of the previous choices is therefore a more restrictive consistency test than if the choices are tested against each other separately.

[7] If one choice (e.g. no. 3) is inconsistent with a previous choice (e.g. no. 2) then the previous choice (no. 2) will be inconsistent with the subsequent choice (no. 3). Therefore it is sufficient to test each choice for consistency against previous choices to know "what's going on".

The most restrictive consistency test is probably in greatest agreement with the idea of rationality in economic theory. Despite this, the tests of consistency in this study are based on the least restrictive test. The reason for this lies in the fact that the respondents are engaged in a cognitive process during the course of the choice experiment, and the first choices may therefore be more troublesome for the respondents than later choices in the choice sequence. This is also why Bradley and Daly (1994) and Johnson and Desvousges (1997) advise against use of an adaptive design in choice studies.[8]

3.3 Lexicographic choices are consistent choices

By lexicographic choices we mean a set of choices in which the respondent consistently chooses that alternative which is best with respect to one particular attribute, e.g. lowest price, neglecting all other attributes. Lexicographic choices do not necessarily imply that the respondent has lexicographic preferences, e.g. that he/she has a willingness to preserve the environment at any cost.[9] Lexicographic choices may be a result of simplification because the respondent finds the choice task too difficult or a result of actual preferences because the presented choice alternatives are too different.

By use of the least restrictive consistency test all the respondents who make lexicographic choices according to the price attribute will be classified as consistent respondents. The reason for this is that the use of the price attribute as *numeraire* in the consistency test implies that choices with equal levels of the price attribute are inconclusive and therefore excluded from the test. A respondent who always chooses the alternative with the lowest price, will (in Figure 2) always lie to the left of the rays, i.e. all his/her choices are mutually consistent.

This unambiguous relation between lexicographic choices and consistency is not present for the attributes not chosen as *numeraire*. The reason for this is that choices where the levels for these factors are equal are included in the consistency test. Equal levels will appear either as vertical or horizontal rays in the ray diagram. Nevertheless, this may be a problem because the respondents who choose lexicographically may choose randomly if the levels of his/her sorting attribute are equal. Such choices therefore may, or may not, be detected as inconsistent in the consistency test.

---

[8] Inconsistent choices may also influence tests of the effects of repeated observations from the same respondent in Stated Choice studies. See, for example, Ouwersloot and Rietveld (1996) who, without considering inconsistencies, conclude that repeated observations do not cause significant autocorrelation problems, and Kim (1998) who, by use of the scaling approach, conclude that heteroscedasticity of observations should be considered in the repeated measurement problems along with individual heterogeneity.

[9] According to Foster and Mourato (1997) earlier analyses of lexicographic preferences in the context of environmental valuation studies (e.g. Stevens *et al.* 1991 and Spash and Hanley 1995) have been based on responses to attitudinal questions indicating a willingness to preserve the environment at any cost.

To avoid this source of confusion, the respondents who have chosen lexicographically, thus violating the axiom of continuity, were removed before the consistency test. (The occurrence of lexicographic choices and their effect on valuations is described in Sælensminde 1998.) This approach was used to ensure that respondents included in the consistency test have actually considered the levels of more than one attribute.[10] The share of respondents in this group that have chosen inconsistently may give an indication of how cognitively demanding the choice task has been. There were no tests of violation of the non-satiation axiom in the data used in the current study; all inconsistencies detected are therefore due to violations of the axiom of transitivity.

4. Modelling

Based on the assumptions that all choices are independent from irrelevant alternatives (IIA)[11] and the errors (the random and unobserved part of the indirect utility) are Gumbel-distributed, the SC data were analysed by means of logit models (equation 1)[12]; see McFadden (1973).

$$\Pr(y_n = 1) = \frac{e^{V_{nA}}}{e^{V_{nA}} + e^{V_{nB}}} = \frac{1}{1 + e^{-(V_{nA} - V_{nB})}} \tag{1}$$

In equation (1), y_n is individual n's choice between the two alternatives i, i = A, B. (A is set equal to 1 and B is set equal to 0.) The systematic (observable) part of the (indirect) utility, V, of the alternative chosen, i, can be stated as

$$V_{ni} = \beta_0 + \beta_1 X_{ni1} + \ldots + \beta_K X_{niK} \ , \tag{2}$$

where β_k, k = 0, ..., K, are the K+1 unknown parameters to be estimated and X_k, k = 1, ..., K, are the K explanatory variables in the model.

The fact that respondents answer multiple choice tasks may give rise to specific interdependency problems in the error structure (autocorrelation) which are

---

[10] Unfortunately, we can not be absolutely certain that respondents who have not chosen lexicographically have considered the levels of more than one attribute. The reason for this is that some respondents may have simplified the decision by choosing for example the journey on the left hand side on all choices or by choosing alternately "left, right, left,...". By the use of such choice strategies the respondents are not judging the levels of the attributes. Tests for these types of simplification strategies have not been carried out in this study.

[11] The IIA property states that "...for any individual, the ratio of choice probabilities of any two alternatives is entirely unaffected by the systematic utilities of any other alternatives." (Ben-Akiva and Lerman 1995). Data sets of actual choices may be tested for the IIA property; this was not done in this study.

[12] Equation 1 implies constant error variance, which is usually assumed in logit models. This restriction may be relaxed if one multiplies the indirect utility by a scale parameter. (See for example Bhat (1995) and Bradley and Daly (1994) for use of such heteroscedastic models.)

ignored in the standard logit or probit models usually applied in this context. Ouwersloot and Rietveld (1996) show how the solution from the panel data field to such autocorrelation problems can be applied in the context of SC experiments. The results of their analysis, that the effect of autocorrelation is very modest and statistically insignificant, could be influenced by inconsistencies in the choice sequence. A possible relationship between autocorrelation and choice inconsistencies in SC data may be a topic for further research, but it was not addressed in the current study. (See e.g. Ortúzar *et al.* 1997 for an overview of the autocorrelation problem.)

How the different factors included in the choice exercises are valued is expressed as the *relation* between two estimators (the marginal rate of substitution between two attributes). The uncertainty of the calculated valuations will thus depend on the standard deviation of the parameters in addition to the correlation between them (equation 3).

$$S_{\left(\frac{\hat{\beta}_1}{\hat{\beta}_2}\right)} = \sqrt{\operatorname{var}(\hat{\beta}_1/\hat{\beta}_2)} = \frac{1}{\hat{\beta}_2}\sqrt{\left\{\operatorname{var}(\hat{\beta}_1) - 2\cdot\frac{\hat{\beta}_1}{\hat{\beta}_2}\operatorname{cov}(\hat{\beta}_1,\hat{\beta}_2) + \frac{\hat{\beta}_1^{\,2}}{\hat{\beta}_2^{\,2}}\operatorname{var}(\hat{\beta}_2)\right\}} \quad (3)$$

5. Description of the surveys

Data from two different surveys are used in this paper to study inconsistency in SC studies. The data from the Norwegian value of time study (Ramjerdi *et al.* 1997) are used as the main data source. Data from a study providing the Norwegian road authorities with valuation of environmental benefits for use in cost-benefit analysis is used as a supplementary data scource (Sælensminde and Hammer 1994). This section mainly describes how the choice tasks and the OE-CV questions were formulated in the value of time survey and only briefly presents the first choice task in the environmental study.[13]

5.1 The Norwegian value of time survey

The Norwegian value of time survey consisted of personal interviews undertaken in 1994-1996 of car drivers and travellers by public transport. The respondents were randomly drawn from the population of the largest cities in Norway. Of the 4556 interviews made in the value of time survey, 2568 are used in the current study. Respondents who recently had undertaken a private[14] journey first answered questions in the SC experiment and were then asked more directly

[13] Only the first choice task in the environmental study is used in the current study. This task is similar (the same attributes and the same context) to the choice tasks in the value of time study. The complete questionnaires from both studies are available from the author on request.

[14] The value of time survey also includes data for business travellers and travellers with air and ferry. These data are not included in the current study, but valuations from these travellers are reported in Ramjerdi *et al.* (1997).

about their willingness to pay for a reduction of 25 % in travel time on that particular journey [15]. Four typical choices from the SC sequence are shown in Figure 1. The direct questions used to determine respondent willingness to pay for reduced travel time on the particular car journey were formulated as so-called Transfer Price questions. What have been called Transfer Price questions in the context of value of time estimation are very similar to OE-CV questions (see, for example, Layard and Glaister 1994 p.258 for an outline). Respondents should be quite familiar with their preferences for reduced travel time on a particular journey when they were presented the OE-CV question, since it followed the choice experiment.

In the SC task the respondents were to choose between two journeys using the travel mode they had used on the particular journey. The journeys were described by three attributes: price, in-vehicle travel time and a third attribute. The third attribute was "automatic speed control" or so-called "photo-boxes" for long (more than 50 km) car journeys and "chance of delay" for short car journeys. In the choice tasks presented to travellers by public transport the third attribute was "headway" or "chance of delay". In addition to the choice task between two journeys with actual transport mode, respondents were also given a choice task between two journeys with their most preferred alternative transport mode. In this task all the travellers with car had to choose between two journeys with public transport.

The survey was administered using the MINT computer-assisted personal interview (CAPI) software. Three design variables were used, each with three or four levels. SPEED software was used to present an orthogonal fractional factorial design of sixteen alternatives randomly selected for each respondent from the full factorial design of all possible combinations. Respondents were presented with a series of nine pairwise choices from among these sixteen options. Pairs of alternatives were presented in random order. Binary logit models were estimated by ALOGIT using the pairwise choice data and linear functions of the three design variables.[16]

5.2 The environmental survey

The environmental survey consisted of 1680 personal interviews undertaken in 1993 of car drivers and travellers by public transport randomly drawn from the population in Oslo and the neighbouring county of Akershus. Respondents who recently had undertaken a private local journey first answered simple SC questions and then they were given more complicated choice tasks that included environmental attributes. Only the first choice task, completed by 1473 respondents, was used in the current study.

[15] They were also asked about their willingness to pay for a reduction of 10 % in travel time on that particular journey and about what compensation was needed if travel time were increased with 25 %. The results from these questions are reported in Ramjerdi *et al.* (1997).

[16] MINT, SPEED and ALOGIT are all software from Hague Consulting Group.

In the first choice task of the environmental survey respondents chose between two journeys with the travel mode they had used on the particular journey. The journeys were described by three attributes: price, in-vehicle travel time and "walking time to parking place" for car travellers, and "chance of seating" for travellers by public transport. As in the value of time survey, the environmental survey was administered using MINT. A similar design procedure was used and similar binary logit models were used for estimation. One important difference between the two surveys is that respondents were presented with a series of only four pairwise choices in the first choice task of the environmental survey.

6. The results

This section first gives an overview of the occurrence of inconsistent choices and the impact of such choices of the valuation of reduced travel time based on the discrete choice data in a logit model. The last part of the section investigates whether the OE-CV valuations show systematic differences between respondents with consistent versus inconsistent choices in the SC data, and whether the occurrence of inconsistent choices differs during the choice sequence.

6.1 The occurrence of inconsistent choices and their effect on valuations

As explained in Section 3, all respondents that made lexicographic choices were removed before the consistency tests. A respondent has either chosen such that all his choices are mutually consistent (i.e. zero inconsistent choices) or such that one or more choices are inconsistent with other choices. For example, if choices 7 and 9 are not consistent with choices 1, 2, 3, 4, 5 and 6, the test will show that two choices are inconsistent with previous choices (i.e. IPC=2). But the test will also show that six choices are inconsistent with subsequent choices (i.e. ISC=6). In Table 1 the number of inconsistent choices for each respondent is defined as the minimum of IPC and ISC.

Table 1. The percentage of respondents with each number of inconsistent choices, and the average number of inconsistent choices for different modes.

	Long journeys (more than 50 km)			Short journeys (less than 50 km)	
Number of inconsistent choices	Car (N=490)	Coach (N=330)	Train (N=330)	Car (N=393)	Public transport (N=257)
0	33.1	25.2	24.5	33.1	39.7
1	39.0	31.8	33.0	31.0	31.1
2	16.9	24.2	22.7	19.8	16.7
3	7.6	11.8	12.7	9.7	9.3
4	2.4	5.9	5.2	4.6	2.7
5	1.0	0.9	1.5	1.0	0.4
6		0.3	0.3	0.5	
7				0.3	
Average number of inconsistent choices	1.10	1.45	1.47	1.28	1.05

For long journeys, Table 1 shows that car travellers chose more consistently than did travellers by public transport. The reason for this may be that longer car journeys are undertaken more frequently than similar journeys by public transport and that car travellers therefore have more knowledge about their preferences than do travellers by public transport.

For short journeys it is shown that car travellers chose less consistently than did travellers by public transport. The reason for this is probably that the attributes in the choice task (e.g. "price per journey") are common when people use public transport, but less commonly associated with local car journeys.

Table 1 shows a large difference in consistency between long and short journeys by public transport. The main explanation for this result is probably differences in travel frequency between short and long journeys and therefore a much clearer preference structure (i.e. knowledge about the goods in the choice task) for short journeys than for long journeys.

Table 2. Valuation of travel time from the SC data for non-lexicographic respondents with each number of inconsistent choices. Unit: NOK/h. (1 NOK ≈ 0.125 \$.) Standard errors presented in parentheses.

Sample	Long journeys (more than 50 km)			Short journeys (less than 50 km)		
	Car	Coach	Train	Car	Public transport [a]	Public transport [b]
All non-lex. resp.	76 (2.9)	44 (4.2)	59 (4.0)	37 (2.5)	17 (3.3)	28 (3.8)
0 incons. choices	71 (3.4)	37 (4.4)	51 (5.3)	16 (2.7)	8 (3.0)	17 (4.6)
1 incons. choices	69 (3.9)	16 (9.4)	37 (7.1)	44 (4.6)	23 (5.1)	23 (5.1)
2 incons. choices	89 (9.0)	43 (13.5)	58 (6.6)	61 (7.0)	14 (12.9)	30 (16.5)
3+ incons. choices	95 (21.4)	101 (14.8)	134 (22.6)	86 (19.0)	96 (25.5)	96 (25.5)

a All data are included in the analysis. b Respondents with i) headway as the third attribute in the exercise and two inconsistent choices and ii) delay as the third attribute in the exercise and zero inconsistent choices are excluded from the analysis. (Inclusion of these respondents causes problems with wrong signs and nonsignificant parameters.) The result is a "pattern" more similar to the other modes.

The valuations of travel time savings from the SC data in Table 2 show significant differences between respondents that have chosen consistently (denoted: zero incons. choices) and respondents that have one or more inconsistent choices. This is a general result for short journeys, but longer journeys by public transport also show significant differences between the group with three or more inconsistent choices and those groups with no inconsistencies. For longer car journeys the valuations show the same pattern, but the differences between the groups are not as large and are not significant. The Likelihood Ratio test is used to test whether the different models are statistically different. The test statistics shown in Table 3 indicate that the null hypothesis (coefficients from the estimated logit models are similar across segments with different numbers of inconsistent choices) can be rejected for all modes in table 2.[17]

Table 3. Results of likelihood ratio tests of the null hypothesis that the coefficients from the estimated logit-models are similar across the segments with different numbers of inconsistent choices.

Sample	Long journeys (more than 50 km)			Short journeys (less than 50 km)		
	Car	Coach	Train	Car	Public transport [a]	Public transport [b]
LR test statistics	114.6	86.6	70.4	168.2	92.0	119.0
Degrees of freedom	12	12	12	16	20	20
$\chi^2_{0.05}$	21.4	21.4	21.4	26.3	31.4	31.4

The differences in valuations between the groups in Table 2 may be explained by differences in real preferences or by differences in the cognitive ability to express real preferences in the SC task presented to respondents. The valuations from OE-CV in Table 4 show no systematic differences between respondents with consistent and inconsistent choices in the SC data. This result indicates that there is no difference in real preferences between these groups, and that differences in the cognitive ability to express the real preferences in the discrete choice task seem to be the most reasonable explanation for the differences shown in Table 2.[18] These results may also indicate that it is more burdensome for the respondents to state their preferences in a SC task than in a CV task.

[17] The estimation results for the 30 different logit models are available from the author on request.

[18] Table 2 shows that inconsistent choices disturb model results and increase the valuation of travel time from the Stated Choice data. Together with the result that the occurrence of inconsistent choices is largest first in the choice sequence (Table 6), one possible solution to the problem with inconsistent choices might be to exclude the first (one or two) choices from the analysis. After analysis based on each choice separately it was decided to exclude the two first choices from the analysis because these choices cause nonsignificant parameters. The valuation of travel time (NOK/hour) for "all non-lex. respondents" when the two first choices are excluded are 81, 47, 58, 37, 19 and 34 for the modes from left to right in Table 2, respectively. These results are still above the valuations from the group with 0 inconsistent choices (and the valuation from the OE-CV), and the reason is probably that there are enough inconsistencies left in the remaining choices to affect the valuations. These analyses suggest that the exclusion of the first choices in the choice sequence does not seem to solve the problem caused by inconsistent choices in Stated Choice data.

Table 4. Valuation of travel time from the OE-CV data for non-lexicographic respondents with each number of inconsistent choices from the SCs. Unit: NOK/h. (1 NOK ≈ 0.125 $.) Standard errors presented in parentheses.*

Sample	Long journeys (more than 50 km)			Short journeys (less than 50 km)	
	Car	Coach	Train	Car	Public transport
All non-lex. resp.	57 (2.3)	39 (2.0)	42 (2.2)	23 (1.7)	16 (1.2)
0 incons. choices	62 (4.3)	35 (2.6)	40 (4.1)	23 (3.6)	15 (1.8)
1 incons. choices	54 (3.4)	38 (4.4)	41 (3.3)	19 (2,4)	13 (1.8)
2 incons. choices	53 (5.0)	42 (4.4)	40 (4.0)	28 (3.8)	21 (4.0)
3+ incons. choices	59 (7.6)	43 (4.1)	50 (6.5)	23 (3.7)	17 (2.5)

* The valuation of travel time is an average of WTP to obtain a 25% reduction in travel time and WTP to avoid a 25% increase in travel time. This makes the results from the OE-CV data comparable to the results from analysis of the SC data.

6.2 Relationships between observable respondent characteristics and the tendency to choose inconsistently

If inconsistent choices in general cause significant effects on the valuation from SC data, it is important to discuss whether such choices should be deleted from the analysis. It might be objected that the removal of "problematic" responses from the survey sample potentially creates a different kind of problem – one of self-selection bias – if test failures are systematically related to observable respondent characteristics. If this were so, it would be a case of trading-off the bias created by the inclusion of "noisy" responses against the bias created by the loss of population representation in the "cleaned" sample. Unfortunately, the data from the value of time study do not include education, which is an important variable for testing whether a relationship exists between observable respondent characteristics and tendency to choose inconsistently. However, by use of the first choice task from the environmental study described above, which is similar to the choice tasks in the value of time study, this relationship can be investigated. The results from logistic regressions to explain inconsistent choices by socio-economic variables in the first SC task from the environmental study are reported in Table 5.

Table 5. Logistic regression of inconsistent choices as a function of socio-economic variables in a SC task. T-statistics presented in parentheses. N=408.

Variable	Parameter estimate
Gender (0=female, 1=male)	-0.2008 (-0.93)
Age	-0.0107 (-1.19)
Income (in 1000 NOK)	-0.0290 (-0.15)
Education (no. of years)	-0.0831 (-2.35)
Pensioner (0=no, 1=yes)	-0.0680 (-0.16)
Difficult to choose?(0=no, 1=yes) [a]	0.0344 (0.12)
Difficult to concentrate? (0=no, 1=yes) [a]	0.0409 (0.15)
Constant	1.1243 (1.46)

a Answers to control questions asked after the respondents had completed the choice task.

Table 5 shows that the education variable was the only significant explanatory variable in the model. This result is in contrast to the results from Foster and

Mourato (1997) and Johnson and Desvousges (1997), and is probably a result of more comprehensive tests for "noisy" choices in the current study. Therefore, in the context of the present sample, the potential problems created by removal of "noisy" responses from the survey may be an issue for further research.

6.3 The occurrence of inconsistent choices is more frequent early in the choice sequence

If the choice task is difficult to understand at first glance one can imagine that respondents need education or training in the choice task such that they are able to express their real preferences in the framework presented to them. If this is the case, one would expect that the occurrence of inconsistencies is greater among the first choices in the sequence than among later choices. If the choice task is cognitively demanding and/or the choice sequence is long, one might expect that fatigue effects could result. In the presence of fatigue effects one would expect that the occurrence of inconsistent choices becomes greater later in the choice sequence. One might see a combination of both learning and fatigue effects in SC data.

In the current study consistency tests are made both for Choice Task 1, nine choices between two journeys with the respondents' actual modes, and Choice Task 2, nine choices between two journeys with alternative modes. The results of these tests are presented in Table 6. (As in Table 1, the share of inconsistent choices in Table 6 is defined as the minimum of IPC and ISC.)

Table 6. The percentage of inconsistent choices for each choice number in the choice sequence. Choice Task 1 (=actual mode) / Choice Task 2 (=alternative mode).

	Long journeys (more than 50 km)			Short journeys (less than 50 km)	
Choice number	Car (N=490/492)	Coach (N=330/328)	Train (N=330/341)	Car (N=393/351)	Public transport (N=257/121)
1	25 / 37	35 / 29	38 / 32	35 / 25	23 / 24
2	18 / 17	23 / 17	22 / 18	18 / 14	17 / 17
3	14 / 17	16 / 13	17 / 16	18 / 9	11 / 10
4	9 / 14	16 / 13	15 / 11	11 / 9	14 / 15
5	10 / 16	14 / 13	10 / 10	12 / 10	12 / 12
6	10 / 11	9 / 12	14 / 13	10 / 11	7 / 7
7	9 / 11	12 / 9	11 / 10	12 / 10	7 / 8
8	8 / 9	10 / 6	11 / 8	7 / 6	7 / 5
9	8 / 6	11 / 8	9 / 8	6 / 7	8 / 6

For public transport (both long and short journeys) and for short car journeys Table 6 shows that the share of inconsistent choices is lower for Choice Task 2 (alternative mode) than for Choice Task 1 (actual mode). Such results suggest that it is the choice task itself that is problematic for the respondents and that it is not particularly more difficult to make such choices for modes used more seldomly. For long car journeys there are more inconsistent choices in Task 2 than Task 1. A possible explanation is that respondents who have used a car on a long journey very seldom, or never, use public transport for such journeys. The unfamiliar context in Choice Task 2 presented to car users on long journeys seems therefore to increase the choice problems more than the learning process reduces them.

Table 6 also shows a declining share of inconsistency for both Choice Tasks 1 and 2. This indicates that training is necessary in order to achieve more consistent choices and that fatigue is not a large problem in this study.[19] The relatively short duration of the interview (15-20 minutes), combined with a relatively easy choice task possibly, may explain why one observes little sign of fatigue effects in this study.

7. Discussion and conclusion

In this paper a test procedure that provides the possibility of identifying each respondent's mutually inconsistent choices is presented. Together with a simple test for lexicographic choices (see Sælensminde 1998), this provides the opportunity to examine "problematic choices" more thoroughly than by other approaches. The empirical analysis in this paper illustrates that inconsistent choices are common in several SC tasks and that such choices have a significant impact on the valuation of reduced travel time. Other important findings are that differing abilities to choose are an important explanation of inconsistent choices. The occurrence of inconsistent choices is shown to be greatest at the beginning of the choice sequence, and is lower for later choices. These results suggest that respondents may need more training and help to choose consistently in SC studies.

A discussion of implications resulting from this and similar studies require consideration of possible explanations for "noisy choices" in SC studies. Unfortunately, we can not assume that genuinely irrational individuals and people unwilling to complete the task in a proper manner are excluded in SC surveys. Such respondents may use the choice task to give protest answers, but their choices will probably be either lexicographic or inconsistent. They can therefore be detected in the test procedure described in the current study. If this was the only reason for the occurrence of "problematic choices", these choices could easily be deleted from the sample. Unfortunately, other possible explanations for lexicographic and inconsistent choices include that respondents a) have difficulties with the task, b) have unstable or ill defined preferences, c) learn about their actual preferences during the task, d) are indifferent to the presented options, e) become fatigued during the task, or f) have preferences not covered by the presented alternatives. Fortunately, the occurrence of "problematic choices"

[19] The share of inconsistent choices in Table 6 is defined as the minimum of IPC and ISC. This is the same way as the number of inconsistent choices for each respondent is defined in Table 1. If the sum of IPC and ISC are tabled we see a relatively large share of inconsistent choices in the first choice and a uniform distribution of inconsistent choices for the eight subsequent choices in the choice sequence. This is indicating that there are more inconsistencies in later choices than the numbers presented in Table 6 seems to tell us.

caused by the latter reasons may be reduced by improvements in study design[20] and by giving more help to the respondents during the task. In addition, the influence of "problematic choices" may be reduced by use of better statistical models.

The results of this, and other studies, show that early choices may be problematic. It is therefore a good idea to give respondents a few learning choices to ensure that they understand the task (Bradley and Daly 1994). Another approach that may ease the task for respondents is to undertake tests of consistency during the choice task similar to that undertaken in Multi-Attribute Utility Theory (MAUT) (see for example Gregory *et al.* 1993, Schkade and Payne 1994 and Baron and Greene 1996). This will probably help the respondents to answer consistently according to their real preferences, but may result in a more time consuming data collection procedure. Letting the respondents become more aware of their preferences before the choice task begins may reduce problems caused by unstable or ill-defined preferences and learning about own preferences during the task. This may be done by use of simple questions about how they prefer the actual goods compared to other goods, if they are willing to pay for improvements in the actual good's quality at all and possible by use of OE-CV questions like those used in the current study.

Indifference problems in the choice task may be reduced by giving the respondent the opportunity to choose that he/she is "indifferent" to the presented alternatives and/or to state that he/she "does not know" which alternative he/she prefers. By giving such choice alternatives one can probably reduce the number of inconsistent choices, but if the respondents use these alternatives as an easy way to complete the task one will probably also reduce the number of informative choices. If inconsistent choices generally cause such large problems as in the current study, the inclusion of "indifferent" and/or "don't know" as choice alternatives will probably be an improvement of SC studies. Whether respondents choosing "indifferent" really are indifferent or use this choice alternative as a simplification of the task can be detected in the test procedure used in the current study.

Preference uncertainty may also be explicitly modelled in the analysis of the data. By use of a follow-up certainty question or by letting respondents directly indicate how certain they are of their choice, one can incorporate respondent uncertainty in the analysis. Li and Mattsson (1995) and Ready *et al.* (1995) use such an approach in DC-CV studies. Both studies conclude that ignorance of preference uncertainty lead to an upwardly biased estimate of WTP. The fact that DC-CV and SC are related methods makes these results interesting. If the

[20] Toner *et al.* (1998) contend that the standard errors of coefficient estimates derived from a logit model are not necessarily minimised when using an orthogonal design in Stated Choice studies. Following a different approach, they have obtained expressions for minimising the variance of the estimated parameters of a logit model which indicate that the two choices offered to respondents should have probabilities of being chosen of 0.917 and 0.083, respectively. Such large differences in the presented alternatives will probably reduce problems with inconsistent choices, but may cause more lexicographic choices. Still, it seems like a promising approach for design improvements.

respondents that make inconsistent choices state that they are more uncertain about their choices in the choice task than do respondents that choose consistently, then this approach will result in less weight on inconsistent choices (i.e. respondents with task problems) in the analysis of the data. Preference uncertainty may therefore be a possible explanation of why respondents that chose inconsistently had a higher valuation of travel time in the current SC study. Champ *et al.* (1997) state that use of a follow-up certainty question is a promising approach as a means of providing a lower bound to the theoretical construct. They conclude that respondents who answer consistently are more certain of their real preferences.

As illustrated in the current study, "noise" in a SC study is more than unexplained variation in a statistical model. If one investigates a data sample as a whole, it will typically appear both non-lexicographic and mostly consistent. The fact that one person's lexicographic or inconsistent choices, when combined with other persons' choices, seems "reasonable" results in very few of the lexicographic and inconsistent choices being detected as model "outliers", i.e. choices with low probability of the observed choice. Therefore, it is probably not enough to design statistical models without a thorough investigation of each respondent's choices.

In the data collection we should use designs that collect more signals and less noise, and in the data analysis we should use models that can separate signals from noise. (In order to deal with these issues Swait and Adamowicz 1996 present a promising approach.) But there is obviously a limit regarding how much noise can be separated from the signals in a model and therefore how much noise such models can stand and yet produce a useful valuation of the goods of interest. Further investigation of the limits of a direct test procedure that can investigate whether each respondent has completed the task in a proper manner, like the one presented in the current study, is valuable. It can be used to investigate how different choice complexity levels influence the share of inconsistent choices and therefore produce input to the discussion of "optimal complexity levels". And it is valuable because it can be used to investigate how many of the inconsistent choices are detected, and if they are handled satisfactorily by statistical models that claim to "separate signals from noise".

Almost all empirical choice modelling work in the literature assumes that individuals behave in a compensatory (i.e. non-lexicographic and consistent) fashion. Further, the specification of choice models tends to assume a utility maximising, full information, indefatigable decision-maker who is able to assign values to alternatives, and choose the alternatives with the highest value, independent of context. If the occurrence of non-compensatory choices is generally large in SC studies, it is questionable whether the data meet the assumptions upon which the analysis relies. It is therefore important to detect whether limits in peoples' cognitive abilities make some data collection methods, or the complexity levels of such methods, unsuitable as a framework for people to state their preferences. It may, for example, be a problem if only highly educated people are able to choose in a consistent manner in SC studies and that their preferences therefore will count more than those of others.

References

Adamowicz, W. L., P. Boxall, M. Williams and J. Louviere (1998) Stated Preference approaches to measuring passive use values: Choice Experiments and Contingent Valuation. *American Journal of Agricultural Economics,* 80(1): 64-75.

Adamowicz, W., J. Louviere and M. Williams (1994) Combining Revealed and Stated Preference Methods for Valuing Environmental Amenities. *Journal of Environmental Economics and Management*, 26: 271-292.

Adamowicz, W., J. Swait, P. Boxall, J. Louviere and M. Williams (1997) Perceptions versus objective measures of environmental quality in combined revealed and stated preference models of environmental valuation, *Journal of Environmental Economics and Management* 32, 65-84.

Arrow, K., R. Solow, P. R. Portney, E. E. Leamer, R. Radner and H. Schuman (1993) Report of the NOAA Panel on Contingent Valuation, Federal Register, January 15, vol. 58, no. 10, pp. 4601-4614.

Baron, J. and J. Greene (1996) Determinants of Insensitivity to Quantity in Valuation of Public Goods: Contribution, Warm Glow, Budget Constraints, Availability, and Prominence, *Journal of Experimental Psychology: Applied,* 2(2): 107-125.

Bates, J. J. (1994) Reflections on stated preferences: theory and practice. Paper presented on Seventh International Conference on Travel Behaviour, 13-16 June 1994, Santiago, Chile.

Ben-Akiva, M. and S. R. Lerman (1985) *Discrete Choice Analysis*, The MIT Press, 1985.

Ben-Akiva, M. and T. Morikawa (1990) Estimation of travel demand models from multiple data sources. Paper presented at the 11th International Symposium on Transportation and Traffic Theory, Yokohama.

Bhat, C. R. (1995) A heteroscedastic extreme value model of intercity travel mode choice. *Transportation Research B*, 29(6): 471-483.

Boxall, P. C., W. L. Adamowicz, J. Swait, M. Williams and J. Louviere (1996) A comparison of stated preference methods for environmental valuation, *Ecological Economics*, 18 (3): 243-253.

Boyle, K. J., F. R. Johnson, D. W. McCollum, W. H. Desvousges, R. W. Dunford, S. P. Hudson (1996) Valuing Public Goods: Discrete versus Continuous Contingent-Valuation Responses, *Land Economics*, 72 (3), pp 381-396

Boyle, K. J., H. F. MacDonald, H. Cheng and D. W. McCollum (1998) Bid design and yea saying in single-bounded, dichotomous-choice questions. *Land Economics* 74(1): 49-64.

Bradley, M. and A. Daly (1992) Estimation of logit choice models using mixed stated preference and revealed preference information. Paper presented at the 6th international Conference on Travel Behaviour, Quebec.

Bradley, M. and A. Daly (1994) Use of the logit scaling approach to test for rank-order and fatigue effects in stated preference data, *Transportation* 21(2): 167-184.

Champ, P. A., R. C. Bishop, T. C. Brown and D. W. McCollum (1997) Using Donation Mechanisms to Value Nonuse Benefits from Public Goods, *Journal of Environmental Economics and Management* 33, 151-162.

Diamond, P. A. and J. A. Hausman (1994) Contingent Valuation: Is Some Number Better Than No Number? *Journal of Economic Perspectives*, 8(4): 45-64.

Dubourg, W. R., M. W. Jones-Lee and G. Loomes (1997) Imprecise Preferences and Survey Design in Contingent Valuation. *Economica* 64: 681-702.

Foster, V. and S. Mourato (1997) Are Consumers Rational? Evidence from a Contingent Ranking Experiment, Paper presented at the 8th Annual Conference of the European Association of Environmental and Resource Economists, Tilburg, The Netherlands, 26th-28th June 1997.

Frykblom, P. (1997) Hypothetical question modes and real willingness to pay, *Journal of Environmental Economics and Management*, 34(3): 275-287.

Green, D., K. E. Jacowitz, D. Kahneman and D. McFadden (1998) Referendum contingent valuation, anchoring, and willingness to pay for public goods. *Resource and Energy Economics* 20(2): 85-116.

Gregory, R., S. Lichtenstein and P. Slovic (1993) Valuing Environmental Resources: A Constructive Approach, *Journal of Risk and Uncertainty*, 7:177-197, 1993.

Halvorsen, B. and K. Sælensminde (1998) Differences between Willingness-to-pay estimates from Open-Ended and Diskrete Choice Contingent Valuation Method: The effect of heteroscedasticity, *Land Economics*, 74(2): 262-282.

Hanemann, W. M. (1994) Valuing the Environment Through Contingent Valuation. *Journal of Economic Perspectives*, vol 8, no 4: 19-43, 1994.

Hanley, N., R. E. Wright and V. Adamowicz (1998) Using Choice Experiments to value the environment. *Environmental end Resource Economics*, 11(3-4): 413-428.

Hensher, D. A. (1994) Stated preference analysis of travel choices: the state of practice. *Transportation*, 21(2), 1994.

Hensher, D. A. and M. Bradley (1993) Using stated choice data to enrich revealed preference discrete choice models. *Marketing Letters* 4(2): 139-152.

Hensher, D. A., J. Louviere and J. Swait (1997) Combining sources of Preference data: The case of the lurking λ's, Working paper 97-8, Institute of Transport Studies, Sydney.

Johnson, F. R. and W. H. Desvousges (1997) Estimating Stated Preferences with Rated-Pair Data: Environmental, Health, and Employment Effects of Energy Programs, *Journal of Environmental Economics and Management*, 34(1): 79-99.

Kim, K-S. (1998) Analysing Repeated Measurement Problems in SP Data Modelling. Paper presented at the 8th World Conference on Transport Research (WCTR), Antwerp, July 12-17, 1998.

Layard, R. and Glaister, S. (1994) *Cost-Benefit Analysis*, Cambridge University Press.

Li, C.-Z. and L. Mattsson (1995) Discrete Choice under Preference Uncertainty: An Improved Structural Model for Contingent Valuation, *Journal of Environmental Economics and Management*, 28: 256-269.

McFadden, D. (1973) Conditional Logit Analysis of Qualitative Choice Behavior. In Zaremmbka, P. (ed.) (1973): *Frontiers in Econometrics*. Academic Press, New York.

Mazzotta, M. J. and J. J. Opaluch (1995): Decision making when choices are complex: A test of Heiner's hypothesis, *Land Economics* 71(4): 500-15.

Neill, H. R. (1995) The Context for Substitutes in CVM Studies: Some Emperical Observations, *Journal of Environmental Economics and Management,* 29: 393-397.

Ortúzar, J. de D., D. A. Roncagliolo and U. C. Velarde (1997) Interactions and independence in Stated Preference modelling. Proceedings from the 25th European Transport Forum, vol. P415, pp. 143-157.

Ouwersloot, H. and P. Rietveld (1996) Stated Choice Experiments with Repeated Observations, *Journal of Transport Economics and Policy*, May 1996: 203-212.

Portney, P. R. (1994) The Contingent Valuation Debate: Why Economists Should Care, *Journal of Economic Perspectives*, vol. 8, no. 4, pp. 3-18.

Ramjerdi, F. L. Rand, I. A. F. Sætermo and K. Sælensminde (1997) The Norwegian value of time study, Report no. 379/1997, Institute of Transport Economics, Oslo.

Randall, A. (1997) The NOAA Panel Report: A Beginning or the End of an Era? *American Journal of Agricultural Economics* 79(5): 1489-1494.

Ready, R., J. C. Whitehead and G. C. Blomquist (1995) Contingent Valuation When Respondents Are Ambivalent. *Journal of Environmental Economics and Management* 29: 181-196.

Roe, B., K. J. Boyle and M. F. Teisl (1996) Using Conjoint Analysis to Derive Estimates of Compensating Variation, *Journal of Environmental Economics and Management* 31(2): 145-159.

Schkade, D. A. and J. W. Payne (1994) How People Respond to Contingent Valuation Questions: A Verbal Protocol Analysis of Willingness to Pay for an Environmental Regulation. *Journal of Environmental Economics and Management* 26: 88-109.

Spash, C. L. and Hanley, N. (1995) Preferences, information and biodiversity preservation, *Ecological Economics*, 12: 191-208.

Stevens, T. H., J. Echevarria, R. J. Glass, T. Hager and T. A. More (1991) Measuring the existence value of wildlife: what do CVM estimates really show? *Land Economics*, 67(4): 390-400.

Swait, J. and W. Adamowicz (1996) The effect of choice environment and task demands on consumer behavior: Discriminating between contribution and confusion, Working Paper, Dept. of Rural Economy, University of Alberta, June 1996.

Swait, J. and W. Adamowicz (1999) Choice task complexity and decision strategy selection. Working paper, Dept. of Marketing, College of Business Administration, University of Florida, Gainesville, September 1999.

Sælensminde, K. (1995) Pakkeeffekter og budsjettpåminnelse - Hvordan påvirkes verdsettingen av redusert reisetid i den norske tidsverdistudien? (Effects of packaging and a budget reminder on the valuation av reduced travel time.) Working paper TP/0934/1995, Institute of Transport Economics, Oslo. (In Norwegian.)

Sælensminde, K. (1998) The impact of choice inconsistencies on the valuation of travel time in Stated Choice studies. Paper presented at the First World Congress of Environmental and Resource Economists, Venice June 25-27, 1998.

Sælensminde, K. (1999) Stated Choice valuation of urban traffic pollution. *Transportation Research Part D,* 4: 13-27.

Sælensminde, K. og F. Hammer (1994) Verdsetting av miljøgoder ved bruk av samvalganalyse, (Assessing environmental benefits by means of conjoint analysis) TØI-rapport no. 251/1994, Institute of transport economics, Oslo. (In Norwegian, with English summary.)

Toner, J. P., S. D. Clark, S. M. Grant-Muller and A. S. Fowkes (1998) Anything you can do, we can do better: a provocative introduction to a new approach to Stated Preference design. Paper presented at the 8th World Conference on Transport Research (WCTR), Antwerp, July 12-17, 1998.

Widlert, S. (1994) Stated preference studies - The design affects the results. Paper presented on Seventh International Conference on Travel Behaviour, 13-16 june 1994, Santiago, Chile.

Widlert, S., G. Lind, E. Lindqvist, J. Lindqvist and U. Lindstedt (1993) Värdering av miljöfaktorer, Transek, Stockholm.

Paper 3:
Causes and consequences of lexicographic choices in Stated Choice studies

Kjartan Sælensminde

December 13, 1999

Abstract

Stated Choice (SC) methods are now becoming increasingly popular for the valuation of environmental goods. This paper shows that lexicographic choices (LCs) commonly occur in a variety of SC tasks. LCs are partly a result of (i) study designs with too-large differences between the alternatives and (ii) simplification of the choice task. Study designs that cause LCs provide less information about preferences, but this normally is not a serious modeling problem. Simplification is argued to be a consequence of respondents' differing abilities to choose. Such LCs contribute to the larger variances of SC data relative to less cognitively demanding valuation methods and might therefore have a significant impact on the implied valuation of non-market goods if it is not corrected in the analysis.

1. Introduction

Conjoint Analyses and Stated Choice (SC) methods are now becoming increasingly popular for the valuation of environmental goods (Boxall *et al.* 1996, Roe *et al.* 1996, Adamowicz *et al.* 1998a, Hanley *et al.* 1998 and Sælensminde 1999). These methods have been used for many years in transport economics and marketing research (see e.g. Hensher 1994 for an overview).

Traditionally, however, environmental goods have been valued in studies using the Contingent Valuation (CV) method. CV has been used in many studies over the last 30 years and has been subject to much criticism and debate (see for example Diamond and Hausman 1994, Hanemann 1994 and Portney 1994 for an overview). One important issue in the CV debate is how questions should be formulated to make them more familiar to the respondent and thereby ease his/her choice task. In response, the NOAA panel (Arrow *et al.* 1993) suggested that the discrete choice (DC) CV format was preferable to the open-ended (OE) CV format. However, issues concerning statistical uncertainty, stability and bias were not addressed by the panel, and the debate concerning the preferability of DC-CV to OE-CV is not closed (Neill 1995, Boyle *et al.* 1996, Dubourg *et al.* 1997, Frykblom 1997, Randall 1997, Boyle *et al.* 1998, Green *et al.* 1998 and Halvorsen and Sælensminde 1998). Two recent contributions to this debate, Adamowicz *et al.* (1998a) and Cameron *et al.* (1999), conclude that different elicitation methods are compatible with the same underlying preferences, providing differences in error variances across methods are permitted. Based on this knowledge, Cameron *et al.* suggest that further research should be directed to explain why error variances differ across methods.

The aim of the current study is to investigate whether the complexity of the choice scenario in SC studies is related to the frequency of lexicographic choices[1], whether lexicographic choices could be a source of increased error variance, and whether such choices therefore may have a significant impact on the implied values of the goods in the study. By lexicographic choices we mean a set of choices in which the respondent consistently chooses the alternative that is best with respect to one particular attribute, e.g. lowest price, neglecting all other attributes. Lexicographic choices do not imply that the respondent has lexicographic preferences, e.g. that he/she has a willingness to preserve the environment at any cost.[2] Lexicographic choices may be a result of i) simplification because the respondent finds the choice task too difficult or ii) actual preferences and a study design that includes widely differing choice alternatives. It is difficult to determine the reason for lexicographic choices on the basis of SC data alone. However, inclusion of OE-CV questions in the current study provides an opportunity for further investigation of this issue. The SC tasks used in the current study include only three attributes, and the context is related to an actual journey. Occurrence of "problematic choices" in such simple choice situations may therefore represent a source of concern for practitioners using cognitively demanding data collection methods in general.

The paper begins with an overview of the existing literature on lexicographic choices, and other types of problematic choices, in choice studies. Section 3 describes how tests for the presence of lexicographic choices are carried out. In Section 4, we present a simple random utility model used for our empirical analysis. Section 5 describes the surveys applied here. The empirical results are presented in Section 6. Finally, Section 7 provides some concluding remarks and the implications of our results for the design and analysis of future SC studies.

---

[1] The complex choice situation in SC studies may also result in mutually inconsistent choices, e.g. choices that violate the transitivity axiom. The impact of such choices is presented in Sælensminde (1998a).

[2] According to Foster and Mourato (1997) earlier analyses of lexicographic preferences in the context of environmental valuation studies (e.g. Stevens *et al.* 1991 and Spash and Hanley 1995) have been based on responses to attitudinal questions indicating a willingness to preserve the environment at any cost.

2. Previous literature on "problematic choices"

Both simple inspections and parametric approaches have been used in the investigation of problems occurring in choice studies as a result of choice task complexity. One approach is to inspect each respondent's choices. In such an approach, one can detect the occurrence of lexicographic choices that may constitute a violation of the continuity axiom (see e.g. Widlert 1994), the occurrences of choices that violate the IIA property and choices that violate the axioms of non-satiation and transitivity.[3] The design of a study by Foster and Mourato (1997) enables them to use this inspection approach to identify all such violations of consumer theory axioms. Another approach, parametric, is to use a statistical estimation method that allows for different error variances within a single model. Scaling, one such parametric approach (Ben-Akiva and Lerman 1985), has been employed mainly in studies that use a mixture of Revealed Preference (RP) and Stated Preference (SP) data[4]. For the topic of this paper, a more interesting use of scaling is as a tool in the investigation of the sources of variance (i.e. "noisy" choices) *within* a single SP data set (Widlert 1994, Bradley and Daly 1994, Mazzotta and Opaluch 1995, Swait and Adamowicz 1996 and Hensher *et al.* 1997). In another parametric approach applied by Johnson and Desvousges (1997), each respondent's multiple observations are used to estimate a model for each individual. A third parametric approach, also applied by Johnson and Desvousges (op.cit.), is to estimate a panel model with respondent-specific scale parameters for the latent random utility distribution.[5] A fourth parametric approach is to include decision strategy selection as an explicit factor in the choice model (Swait and Adamowicz 1999). Both inspection of the data and parametric approaches provide the opportunity to investigate how "problematic choices" influence the results from a statistical model. One important difference is between approaches that can identify each respondent's problematic choices and approaches that only are able to detect that (some) respondents have had more or fewer problems with the choice task. To see what is really going on, it may be important to identify both.

Bates (1994) points out that the danger of presenting respondents with tasks that are too complicated is well known to practitioners, and, at a minimum, choice

[3] The continuity axiom of consumer theory states that two commodity bundles that are similar to each other will be ranked close together in the consumer's preference ordering. This axiom precludes lexicographic ordering of bundles by requiring respondents to tradeoff gains in one commodity against losses in another. The non-satiation axiom states that a consumer must prefer a bundle which is in all respect superior to another. The transitivity axiom requires that if a consumer prefers option A over option B and option B over option C, then he must necessarily prefer option A over option C. The IIA (Independence from Irrelevant Alternatives) property, which is central to the logit framework of econometric analysis, requires that the ranking between two bundles in a choice set is not affected by the characteristics of the remaining bundles in the set.

[4] See e.g. Ben-Akiva and Morikawa (1990), Bradley and Daly (1992), Hensher and Bradley (1993), Adamowicz *et al.* (1994) and Adamowicz *et al.* (1997).

[5] This is actually the extreme option of the scaling approach, treating each person in the sample as a separate data set with its own scale factor.

data should be checked for possible lexicographic effects. Inspired by Bates (1994), the current study investigates the impact of lexicographic choices and how such problems vary between respondents. The current study also investigates *the causes of* lexicographic answers through supplementary OE-CV questions and socioeconomic variables in statistical models.

3. Test for lexicographic choices

The test for lexicographic choices undertaken in the current study includes all the choices made by each respondent. Test procedures that make use of only a few of each respondent's choices may only detect violations of some of the axioms of consumer theory.

3.1 Testing for lexicographic choices in the stated choice study

Lexicographic choices in choice studies occur as a result of (i) simplification if the respondent finds the choice task too difficult to handle or (ii) too-large differences in attribute levels; i.e. as a result of actual preferences. It is difficult to determine whether a respondent has chosen lexicographically because he/she wanted to simplify the choice task or because the differences in attribute levels were too large.

In the test of lexicographic choices performed in the current study, only whether the respondent has consistently chosen the alternative with the best level for *one* of the attributes included in the task is considered. For example, if he/she has chosen the alternative with lowest price in all the choices, one would say that he/she has chosen "lexicographically regarding the price attribute". How lexicographic respondents made choices when the level was equal on the attribute they used for sorting is not investigated here (see for example Foster and Mourato 1997 for the results of such a strong test). This simple test is used because a stronger test will probably not detect all respondents that use a lexicographic choice heuristic to simplify their choice task.

3.2 More choices in the task means fewer "apparent" lexicographic choices

In tests of lexicographic choices it is important to be aware of the share of "natural" or "apparent" lexicography. Analytically, the share of "apparent" lexicographic choices is determined by the probability that a random choice[6] is judged as lexicographic according to the procedure described above. For each discrete random choice, this probability[7] is 3/4 if the attribute has two levels and

---

[6] By random choice we mean a choice made without considering the levels of the attributes, or just by chance selection (or lot).

[7] The probability that the level of the attribute, of the randomly chosen alternative, is better or equal to the level of that same attribute of the alternative not chosen. This is the same algorithm that is used to investigate the share of lexicographic choices in the current study.

2/3 if the attribute has three levels.[8] Table 1 shows how the share of "apparent" lexicographic choices is reduced when the respondents are given more choices in a sequence.

Table 1. The percentage of "apparent" lexicographic choices for an attribute in a choice task is reduced with increased number of choices and increased number of levels of the attribute.

	Attribute, 2 levels	Attribute, 3 levels	Attribute, 4 levels
1	75.0 %	66.7 %	62.5 %
2	56.3 %	44.4 %	39.1 %
3	42.2 %	29.6 %	24.4 %
4	31.6 %	19.7 %	15.3 %
5	23.7 %	13.2 %	9.5 %
6	17.8 %	8.8 %	6.0 %
7	13.3 %	5.9 %	3.7 %
8	10.0 %	3.9 %	2.2 %
9	7.5 %	2.6 %	1.5 %

The share of "apparent" lexicographic choices in a real choice task will be reduced if dominant choices (i.e. choices where one of the alternatives is better or equal for all the attributes) are not part of the task. Dominant choices may be used as an introduction to the task (Bradley and Daly 1994) and/or for test purposes (Foster and Mourato 1997).

4. Modelling

Based on the assumptions that all choice probabilities are independent from the presence of irrelevant alternatives (IIA) and the errors (the random and unobserved part of the indirect utility) are Gumbel-distributed, the SC data are analyzed by means of random utility logit models (McFadden 1973 and Ben-Akiva and Lerman 1985).

$$\Pr(y_n = 1) = \frac{e^{V_{nA}}}{e^{V_{nA}} + e^{V_{nB}}} = \frac{1}{1 + e^{-(V_{nA} - V_{nB})}} \qquad (1)$$

In equation (1), y_n is individual n's choice between the two alternatives i, i = A, B. (A is set equal to 1 and B is set equal to 0.) Equation 1 implies constant error

[8] An attribute with two levels (high and low) gives four different combinations in the two choice alternatives; 1) low/low, 2) low/high, 3) high/low and 4) high/high. If we assume that these four combinations have equal probability of being presented as a choice in the choice task, a random choice will be classified as lexicographic in 3 of 4 cases, i.e. better or equal. Correspondingly, an attribute with three levels gives nine different combinations in the two choice alternatives, and a random choice will therefore be classified as lexicographic in 6 of 9 cases.

variance, which is usually assumed in logit models. The systematic (observable) part of the (indirect) utility, V, of the alternative chosen, i, can be stated as

$$V_{ni} = \beta_0 + \beta_1 X_{ni1} + ... + \beta_K X_{niK} \ , \qquad (2)$$

where β_k, k = 0, ..., K, are the K+1 unknown parameters to be estimated and X_k, k = 1, ..., K, are the K explanatory variables in the model.

The fact that respondents answer multiple choice tasks may give rise to specific interdependency problems in the error structure (autocorrelation) that are ignored in the standard logit or probit models usually applied in this context. Ouwersloot and Rietveld (1996) show how techniques borrowed from panel data analysis to accomodate such autocorrelation problems can be applied in the context of SC experiments. Despite the results of their analysis, that the extent of autocorrelation is very modest and statistically insignificant, this may be an important topic for further research. However, it will not be addressed in the current study.

The attributes included in the choice exercises are valued by examining the *relation* between two parameter estimates (the marginal rate of substitution between two attributes). The uncertainty of the calculated valuations will thus depend on the standard deviation of the parameters in addition to the correlation between them.

5. Description of the surveys

Data from two different surveys are used in this paper to study inconsistency in SC studies. The data from the Norwegian value of time study (Ramjerdi *et al.* 1997) are used as the main data source. Data from a study providing the Norwegian road authorities with valuation of environmental benefits for use in cost-benefit analysis is used as a supplementary data scource (Sælensminde and Hammer 1994). This section mainly describes how the choice tasks and the OE-CV questions were formulated in the value of time survey and only briefly presents the first choice task in the environmental study.[9]

5.1 The Norwegian value of time survey

The Norwegian value of time survey consisted of personal interviews undertaken in 1994-1996 of car drivers and public transport passengers. The respondents were randomly drawn from the populations of the largest cities in Norway. 2560 of the interviews from the value of time survey are used in the current study. Respondents who had recently undertaken a private trip first participated in a SC experiment and were then asked CV[10] questions. The

---

[9] Only the first choice task in the environmental study is used in the current study. This task is similar (the same attributes and the same context) to the choice tasks in the value of time study. The complete questionnaires from both studies are available from the author on request.

[10] The CV questions used to determine respondent willingness to pay for reduced travel time on the particular car journey were formulated as so-called Transfer Price questions. What has been

respondents should be quite familiar with their preferences for changes in travel time on a particular journey when they were presented with the OE-CV question, since it followed the choice experiment. Both the SC task and the CV questions were designed to elicit the respondents' willingness to pay for a reduction of 25 % in travel time and what compensation that was needed if the travel time were increased by 25 % on that particular journey. A typical willingness to pay choice from the SC sequence is shown in Figure 1 together with the corresponding CV question.

One choice from the SC task	
Train journey A	**Train journey B**
Travel costs = 250 NOK. Travel time = 3 hours. Headway = 2 hours	Travel costs = 200 NOK. Travel time = 4 hours. Headway = 2 hours
The corresponding CV question	
Your actual train journey took 4 hours and the price was 200 NOK. Imagine that this journey by train could be completed in 3 hours. What is the maximum price you would be willing to pay for the journey with reduced travel time? I.e. at what price are the following two journeys of equal value for you? Journey 1. Travel time=4 hours and price=200 NOK. Journey 2. Travel time=3 hours and price=? NOK.	

Figure 1. Example of one choice from the SC task and the corresponding CV question exercise presented to the travelers by train in the value of time study. Respondents were asked to choose between train journey A and train journey B in the SC task and to state their willingness to pay in the CV question.

In the SC task, the respondents were to choose between two trips using the travel mode they had used on the particular journey. The trips were described by three attributes: price (four levels), in-vehicle travel time (four levels) and a third attribute (three levels). The third attribute for long (more than 50 km) car trips was "automatic speed control" or so-called "photo-boxes" and for short car journeys, the third attribute was "delay". In the choice tasks presented to travelers using public transport the third attribute was "headway" (i.e. time between each departure) or "delay".

The survey was administered using the MINT[11] computer-assisted personal interview (CAPI) software. Three design variables were used, each with three or four levels. By use of SPEED, an orthogonal[12] fractional factorial design of

called Transfer Price questions in the context of value of time estimation are very similar to OE-CV questions (see, for example, Layard and Glaister 1994 p.258 for an outline).

[11] MINT, SPEED and ALOGIT are all software from Hague Consulting Group frequently used in transport research.

[12] Toner *et al.* (1998) contend that the standard errors of coefficient estimates derived from logit models are not necessarily minimized when using an orthogonal design. They show that larger

sixteen alternatives was randomly selected for each respondent from the full factorial design of all possible combinations. Respondents were presented with a series of nine pairwise choices from among these sixteen options. Pairs of alternatives were presented in random order. Binary logit models were estimated by ALOGIT using the pairwise choice data and linear functions of the three design variables.

5.2 The environmental survey

The environmental survey consisted of 1680 personal interviews undertaken in 1993 of car drivers and public transport passengers randomly drawn from the population in Oslo and the neighboring county of Akershus. Respondents who had recently undertaken a local privat trip first answered simple SC questions and then they were given more complicated choice tasks that included environmental attributes. Only the first choice task, completed by 1473 respondents, is used in the current study.

In the first choice task of the environmental survey respondents were asked to choose between two journeys with the same travel mode they had used on their recent journey. The journeys were described by three attributes: price, in-vehicle travel time and "walking time to parking place" for car travelers, and "chance of seating" for public transport passengers. As in the value of time survey, the environmental survey was administered using MINT. A similar design procedure was used and similar binary logit models were used for estimation. One important difference between the two surveys is that respondents were presented with a series of only four pairwise choices in the first choice task of the environmental survey.

6. The results

This section first gives an overview of the frequency of lexicographic choices and the impact of such choices on the valuation of reduced travel time based on the discrete choice data in a logit model. Thereafter, the reason for such choices is discussed by use of the valuation of travel time based on OE-CV data from the same respondents.

6.1 The occurrence of lexicographic choices and their effect on valuations

Table 2 shows that there is a larger share of lexicographic choices for long journeys than for short journeys. Price is the most common "sorting attribute" for all modes except car (long), for which travel time is most common, and train, for which price and travel time are essentially equally common.

differences between the presented alternatives will reduce the variance of the estimated parameters. Such design issues may be a direction for future research on the impact of "problematic choices" in SC data.

Table 2. The percentage of lexicographic choices for the different attributes (used as a "sorting attribute" by the respondents) and modes.

	Long journeys (more than 50 km)			Short journeys (less than 50 km)	
Attribute	Car (N=685)	Coach (N=440)	Train (N=593)	Car (N=509)	Public transport (N=333)
Price	7.3	11.4	19.8	21.0	19.8
Travel time	19.8	5.7	20.9	2.6	1.8
Third. attribute[a]	3.8	8.0	3.2	4.5	5.4
Total share	30.9	25.1	43.9	28.1	27.0

a) For the car mode of travel the third attribute is "automatic speed control" for long journeys and "delay" for short journeys. For public transport (including coach and train) the third attribute is "headway" for long journeys and "headway" or "delay" for short journeys.

As illustrated in Table 3, the lexicographic choices in the SC data increased the apparent valuation of travel time for long car journeys. This is a natural consequence of the fact that travel time is the main "sorting attribute" for long car journeys. For long journeys with public transport and for short journeys the use of lexicographic choices decreases the valuation of travel time. This result is explained by the fact that price is the main "sorting attribute" for these journeys. For long journeys by coach and for short journeys, it is not possible to estimate a valuation of travel time for the respondents who chose lexicographically. The reason for this is that relatively few of these respondents have used travel time as the "sorting attribute", and consequently too few of the respondents in these groups show willingness to pay for reduced travel time in the SC data.

Table 3. Valuation of travel time from the SC data. Unit: NOK/h. (1 NOK ≈ 0.125 $.) Standard errors in parentheses.

	Long journeys (more than 50 km)			Short journeys (less than 50 km)	
Sample	Car	Coach	Train	Car	Public transport
All respondents	85 (3.1)	35 (4.1)	55 (4.0)	23 (1.6)	11 (2.9)
Respondents that chose lex.	120 (12.4)	Wrong sign	48 (6.8)	Wrong sign	Wrong sign
Respondents that did not choose lex.	76 (2.9)	44 (4.2)	59 (4.0)	37 (2.5)	17 (3.3)

A Likelihood Ratio test is used to test whether the alternative models are statistically different. The test statistics in Table 4 show that the null hypothesis, that the coefficients from the estimated logit models are similar across the segment with lexicographic choices and the segment with non-lexicographic choices, can be rejected for all modes in Table 3.[13]

[13] The estimation results for the 15 different logit models are available from the author on request.

Table 4. Results of likelihood ratio tests of the null hypothesis that the coefficients from the estimated logit-models are similar across the two segments of respondents that i) chose lexicographically, and ii) did not choose lexicographically.

	Long journeys (more than 50 km)			Short journeys (less than 50 km)	
Sample	Car	Coach	Train	Car	Public transport
LR test statistics	52.2	34.0	48.0	273.2	53.8
Degrees of freedom	4	4	4	4	5
$\chi^2_{0.05}$	9.49	9.49	9.49	9.49	11.07

6.2 Lexicographic choices are partly a result of real preferences

As described, lexicographic choices may be a result of simplification of the choice task or a result of real preferences. Respondents who use the price attribute as a "sorting attribute" (i.e. always choose the alternative with lowest price) because they want to simplify the task will, in the analysis of the SC data, generate a lower valuation (if it is possible to calculate a valuation) than correct valuation. Respondents who use the price attribute as a "sorting attribute" because they are not willing to pay the amount that is the difference between the two alternatives will, in the analysis of the SC data, get a higher valuation than their correct valuation. To decide if a specific respondent has chosen lexicographically because of a) simplification or b) real preferences, one must know his/her "correct" valuation.

If one assumes that CV is a less cognitively demanding valuation method than is SC, and that CV therefore will produce preference information with a higher quality (i.e. less variance) as shown by Cameron *et al.* (1999)[14], the OE-CV data may be used as an indication of the basis for lexicographic choices in the SC data. This assumption is based on the fact that SC is shown to produce a significant share of inconsistent choices depending on choice set complexity (Swait and Adamowicz 1996, Sælensminde 1998a and DeShazo and Fermo 1999) and that choice complexity seems to be larger in SC than in CV. Still, it may be debatable whether SC is a more burdensome elicitation method for the respondents than is CV (Adamowicz *et al.* 1998b). It is important to note that the following simple indication test uses only the valuations from the OE-CV data. This is not a comparison of the valuations from SC with the valuations from CV. Such comparisons may be invalid if there is heteroscedasticity across methods.

If the respondents who have chosen lexicographically with price as the "sorting attribute" have done so based on their real preferences, one can expect that the valuation from the OE-CV data is lower for this group than for the group that has not chosen lexicographically. If real preferences have resulted in some other attribute being the "sorting attribute", one can expect that the valuation from the OE-CV data is higher for this group than for the group that has not chosen

[14] Actually, in Cameron *et al.*'s sample the value estimates from OE-CV are more variable than the value estimates from DC-CV. They state that this larger variance in the OE-CV data may be accounted for by the existence of a handful of large outliers and therefore that the conventional wisdom among "CV researchers" that OE-CV estimates should be less variable might still hold.

lexicographically. In contrast, if the lexicographic choices are caused by simplification, one can expect that the valuation from the OE-CV data will be about the same for the lexicographic group and the non-lexicographic group. These expectations rely on the assumption that other factors are not influencing the valuation.

As can be seen from Table 5, respondents who have used price as the "sorting attribute" (denoted: lex. price) in the choice task have a significantly lower valuation of travel time derived from their OE-CV responses than do respondents who made non-lexicographic choices (denoted: non-lex.). (Short journeys by public transport are an exception.) It can also be seen that respondents who have used travel time as the "sorting attribute" (denoted: lex. travel time) have a significantly higher valuation of travel time, for most of the modes, than do respondents who made non-lexicographic choices. A conclusion from this investigation is that lexicographic choices in this SC study are at least partly a result of the respondents' real preferences. Again, this shows that real preferences that cause a lexicographic choice procedure in a choice task are not the same as lexicographic preferences.

Table 5. Valuation of travel time, from the OE-CV data. Unit: NOK/h. (1 NOK ≈ 0.125 $.) Standard errors presented in parenthesis.*

	Long journeys (more than 50 km)			Short journeys (less than 50 km)	
Sample	Car	Coach	Train	Car	Public transport
All	62 (2.3)	38 (1.7)	41 (1.9)	21 (1.4)	15 (1.0)
Lex. total	75 (5.4)	33 (3.1)	39 (3.2)	16 (2.7)	14 (2.3)
Lex. price	25 (4.3)	21 (2.9)	27 (3.6)	11 (4.6)	17 (3.4)
Lex. travel time	98 (6.9)	47 (8.3)	54 (5.7)	39 (7.2)	42 (8.3)
Non-lex.	57 (2.3)	39 (2.0)	42 (2.2)	23 (1.7)	16 (1.2)

* The valuation of travel time is an average of WTP to obtain a 25% reduction in travel time and WTP to avoid a 25% increase in travel time. This makes the results from the OE-CV data comparable to the results from analysis of the SC data. Respondents who used "headway" as a sorting attribute have a higher valuation of headway than respondents who did not make lexicographic choices. This is not shown in the table.

6.3 Relationships between observable respondent characteristics and the tendency to choose lexicographically

If lexicographic choices in general cause significant effects on apparent valuations from SC data, it is important to discuss whether such choices should be deleted from the analysis. It might be objected that the removal of "noisy" responses from the survey sample potentially creates a different kind of problem – one of endogeneous-selection bias – if test failures are systematically related to observable respondent characteristics. If this were so, it would be a case of trading-off the bias created by the inclusion of "noisy" responses against the bias created by the loss of segment representation in the "cleaned" sample. Unfortunately, the data from the value of time study do not include education, which is an important variable for testing whether a relationship exists between observable respondent characteristics and tendency to choose lexicographically or inconsistently. However, by switching to the use of the first choice task from the environmental study described above (similar to the choice tasks in the value of time study), this potential relationship can be investigated. The results from logistic regressions to explain lexicographic choices by socioeconomic variables in the first SC task from the environmental study are reported in Table 6.[15]

Table 6. Logistic regression of lexicographic choices in a SC task as a function of socioeconomic variables. The dependent varible = 1 if respondent chose lexicigraphically. T-values in parentheses. (N=147)

Variable	Parameter estimate
Gender (0=female, 1=male)	-0.1539 (-1.29)
Age	-0.0053 (-1.04)
Income (in 1000 NOK)	-0.0860 (-0.73)
Education (no. of years)	-0.0534 (-2.77)
Pensioner (0=no, 1=yes)	0.2097 (0.87)
Difficult to choose? (0=no, 1=yes) *[a]	-0.1427 (0.91)
Difficult to concentrate? (0=no, 1=yes) *	0.3399 (2.11)
Constant	1.7823 (4.15)

* Answers to control questions asked after the respondents had completed the choice task.

Table 6 shows statistically significant relationships between the variables "education" and "difficult to concentrate" and the tendency to choose lexicographically in this survey. These results occur despite the fact that the environmental survey data contain only four choices, which causes more "apparent" lexicographic choices than if more choices were given. This strengthens the hypotheses that simplification also is an important reason for lexicographic choices.

6.4 Simplification is a major reason for choosing lexicographically

The two previous sections suggest that lexicographic choices are caused partly by actual preferences and partly by simplification of the choice task. Unfortunately, nothing can be said about the relative shares of these causes of lexicographic choices from those investigations. In this section a simple consistency test of each respondent's valuation of travel time from SC and OE-CV is used to investigate whether lexicographic choices are made because of actual preferences or simplification.

The SC task includes only three attributes (price, travel time and a third attribute), and for 95 percent of the respondents the level of the third attribute was equal in one or more of the nine choices that were presented. This fact allows us to

[15] Results from OE-CV questions may also be used as explanatory variables in the logit model for lexicographic choices, but OE-CV questions, valuing reduced travel time, were not included in the environmental survey. However, by use of the data from the value of time survey, logistic regression models including answers to the OE-CV questions are used to explain lexicographic choices. These models confirm the above results that actual preferences are also an important factor affecting lexicographic choices in this particular case.

find an upper limit of the valuation of travel time ($VotSC_{Max}$) from each respondent who used price as the sorting attribute and a lower limit of the valuation of travel time ($VotSC_{Min}$) from each respondent who used travel time as the sorting attribute. If these SC valuations are consistent with the valuations of travel time from the CVquestions (VotCV), i.e. if $VotSC_{Max} \geq VotCV$, or if $VotSC_{Min} \leq VotCV$, it is assumed that actual preferences are the most likely reason for lexicographic choices. On the other hand, if $VotSC_{Max} < VotCV$ or $VotSC_{Min} > VotCV$ it is assumed that simplification is the most likely reason for lexicographic choices.

Of course, inconsistency between a single respondent's SC and CV valuations can only give an indication of the reasons for lexicographic choices. Such inconsistencies may also be a result of, for example, unstable or ill-defined preferences or the posibility that respondents learn about their preferences during the task. The consistency test used here is a test of transitivity of valuations from SC and CV. Such a test may be judged as valuable for at least two reasons. First, respondents who made nine internally consistent (transitive and non-lexicographic) choices in this SC task (i.e. small variance in the SC data) are shown to have similar valuations from SC and OE-CV (Sælensminde 1998a). Second, lexicographic choices due to simplification will contribute more to the unexplained variance in SC data than will lexicographic choices due to actual preferences. In other words, if the lexicographic choices are due to actual preferences, one would expect consistency between SC and CV valuations. The result of this simple test of consistency between SC and OE-CV is presented in Table 7.

Table 7. The percentage of lexicographic choices caused by i) actual preferences (P), and ii) simplification (S) divided across price and travel time attributes, and different modes of transport.

Attribute	Long journeys (more than 50 km)						Short journeys (less than 50 km)			
	Car		Coach		Train		Car		Public transport	
	P	S	P	S	P	S	P	S	P	S
Price	72	28	42	58	59	41	63	37	67	33
Travel time	18	82	30	70	15	85	19	81	17	83

Table 7 indicates that a majority (61 percent of the total sample) of the respondents who chose lexicographically based on the price attribute did so in agreement with their actual preferences. An interesting result is that the consistency test indicates that a minority (only 16 percent of the total sample) of the respondents that have chosen lexicographically based on the time attribute did so in agreement with their actual preferences. The attribute levels in the study design could explain these results if the design contains relatively larger differences in the levels of the price attribute than in the levels of the time attribute. Such attribute level differences should also result in larger shares of lexicographic choices based on the price attribute than based on the time attribute. Table 1 shows that this is the case for short journeys, but not for long journeys.

Together with the fact that the results of the consistency test in table 7 are very similar for long and short journeys, one should probably look for other possible explanatory factors beside study design to explain the pattern in Table 7. One such explanation may be that SC tasks capture respondents' relative valuation rather than their absolute valuation (Roe *et al.* 1996 and Sælensminde 1998b.) If the respondents choose more in accordance with their budget constraints in CV than in SC it should be expected that the valuations from OE-CV will be less than the valuations from SC. If this is the case, the consistency test between SC and CV described above will result in greater consistency between SC and CV for the respondents who have used "price" as a sorting attribute than for the respondents who have used "time" as the sorting attribute.

It is important to stress that even if these data show that the lexicographic choices are partly a result of real preferences, this result can hardly be generalized. The reason for this is that the share of lexicographic choices that is caused by simplification probably will increase with the number of attributes in the task and if respondents have less *a priori* knowledge of the attributes. In the current study each alternative involves only three attributes, and those three are well known to the respondents. It is therefore expected that these choice tasks are considered relatively easy by most of the respondents and that they do not find it necessary to simplify the choice task by choosing lexicographically. This view is supported by findings in Mazzotta and Opaluch (1995).

7. Discussion and conclusion

In this paper, a simple test for lexicographic choices is used to examine certain types of "problematic choices". The empirical analysis illustrates that lexicographic choices are common in several SC tasks and that such choices have a significant impact on the implied value of reduced travel time. Other important findings are that both different abilities to choose and actual preferences can lead to lexicographic choices.

This result is in contrast to the results from Foster and Mourato (1997) and Johnson and Desvousges (1997), and is probably attributable to the more comprehensive tests for "noisy" choices in the current study. Therefore, in the context of the present sample, the potential problems created by removal of "noisy" responses from the survey may be an issue for further research.

As illustrated in the current study, "noise" in a SC study is more than just unexplained variation in a statistical model. If one investigates a data set as a whole, it will in most cases seem both non-lexicographic and mostly consistent. . The fact that one person's lexicographic or inconsistent choices, when combined with other persons' choices, seems "reasonable" results in very few of the lexicographic and inconsistent choices being detected as model "outliers", i.e. choices with low probability of the observed choice. Therefore, it is probably not enough to design statistical models without a thorough investigation of each respondent's choices if one want to study the impact of "problematic choices" in SC-data.

In data collection, we should use survey designs that collect more signals and less noise, and in data analysis we should use models that can separate signals from noise. (In order to deal with these issues, Swait and Adamowicz 1996 present a promising approach that chararacterize task demands and incorporates them into the analysis.) But there is obviously a limit to how much noise can be separated from signals in a model and therefore how much noise such models can stand and still produce useful valuation estimates for the goods of interest. For further investigation of such limits, a direct test procedure that can investigate whether each respondent has completed the task in a proper manner, like the one presented in the current study, is valuable. It can be used to investigate how different choice complexity levels influence the share of "problematic choices" and therefore produce input into the discussion of "optimal complexity levels". And it is valuable because it can be used to investigate what proportion of the "problematic choices" are detected, and if they are handled satisfactorily by statistical models that claim to "separate signals from noise".

Cameron *et al.* (1999) show that choices under different elicitation methods are entirely compatible with the same underlying set of homogeneous preferences, providing heteroscedastic errors across methods are permitted, but they do not explain why error variances differ across methods. In the current study it is argued that lexicographic choices due to simplification of the choice task is one reason for larger error variances in SC data, and Sælensminde (1998a) shows that internal inconsistencies in the SC-data is another. Both lexicographic and inconsistent choices in SC-data are probably directly influenced by the complexity of the choice task.

Almost all empirical choice modeling work in the literature assumes that individuals behave in a compensatory (i.e. non-lexicographic and consistent) fashion. Further, the specification of choice models tends to assume a utility maximizing, full information, indefatigable decision-maker who is able to assign utility levels to alternatives, and choose the alternatives with the highest utility, independent of context. If the occurrence of "problematic choices" is generally large in SC studies, it is questionable whether these data meet the assumptions upon which the analysis relies. It is therefore important to detect whether limits in peoples'cognitive abilities make some data collection methods, or complexity levels of such methods, unsuitable as a framework for people to state their preferences. It may, for example, be a problem if only highly educated people are able to choose in a compensatory manner in SC studies and that their preferences therefore will count more than those of others.

References

Adamowicz, W., J. Louviere and M. Williams (1994) Combining Revealed and Stated Preference Methods for Valuing Environmental Amenities. *Journal of Environmental Economics and Management*, 26: 271-292.

Adamowicz, W., J. Swait, P. Boxall, J. Louviere and M. Williams (1997) Perceptions versus objective measures of environmental quality in combined revealed and stated preference models of environmental valuation, *Journal of Environmental Economics and Management* 32, 65-84.

Adamowicz, W., P. Boxall, M. Williams and J. Louviere (1998a) Stated Preference approaches to measuring passive use values: Choice Experiments and Contingent Valuation. *American Journal of Agricultural Economics,* 80(1): 64-75.

Adamowicz, W., J. Louviere and J. Swait (1998b) Introduction to attribute-based Stated Choice methods. Report submitted to Resource Valuation Branch Damage Assessment Center, NOAA- National Oceanic and Atmospheric Administration, US Department of Commerce, January 1998.

Arrow, K., R. Solow, P. R. Portney, E. E. Leamer, R. Radner and H. Schuman (1993) Report of the NOAA Panel on Contingent Valuation, Federal Register, January 15, vol. 58, no. 10, pp. 4601-4614.

Bates, J. J. (1994) Reflections on stated preferences: theory and practice. Paper presented on Seventh International Conference on Travel Behaviour, 13-16 June 1994, Santiago, Chile.

Ben-Akiva, M. and S. R. Lerman (1985) *Discrete Choice Analysis*, The MIT Press, 1985.

Ben-Akiva, M. and T. Morikawa (1990) Estimation of travel demand models from multiple data sources. Paper presented at the 11th International Symposium on Transportation and Traffic Theory, Yokohama.

Boxall, P. C., W. L. Adamowicz, J. Swait, M. Williams and J. Louviere (1996) A comparison of stated preference methods for environmental valuation, *Ecological Economics*, 18 (3): 243-253.

Boyle, K. J., F. R. Johnson, D. W. McCollum, W. H. Desvousges, R. W. Dunford, S. P. Hudson (1996) Valuing Public Goods: Discrete versus Continuous Contingent-Valuation Responses, *Land Economics*, 72 (3), pp 381-396

Boyle, K. J., H. F. MacDonald, H. Cheng and D. W. McCollum (1998) Bid design and yea saying in single-bounded, dichotomous-choice questions. *Land Economics* 74(1) 49-64.

Bradley, M. and A. Daly (1992) Estimation of logit choice models using mixed stated preference and revealed preference information. Paper presented at the 6th international Conference on Travel Behaviour, Quebec.

Bradley, M. and A. Daly (1994) Use of the logit scaling approach to test for rank-order and fatigue effects in stated preference data, *Transportation* 21(2): 167-184.

Cameron, T. A., G. L. Poe, R. G. Ethier and W. D. Schulze (1999) Alternative nonmarket valuation-elicitation methods: Are the underlying preferences the same? Paper presented at the W-133 Annual Meeting in Tucson, February 1999.

DeShazo, J. R. and G. Fermo (1999) Designing choice sets for Stated Preference methods: The effects of complexity on choice consistency. Paper presented at the W-133 Annual Meeting in Tucson, February 1999.

Diamond, P. A. and J. A. Hausman (1994) Contingent Valuation: Is Some Number Better Than No Number? *Journal of Economic Perspectives*, 8(4): 45-64.

Dubourg, W. R., M. W. Jones-Lee and G. Loomes (1997) Imprecise Preferences and Survey Design in Contingent Valuation. *Economica* 64: 681-702.

Foster, V. and S. Mourato (1997) Are Consumers Rational? Evidence from a Contingent Ranking Experiment, Paper presented at the 8th Annual Conference of the European Association of Environmental and Resource Economists, Tilburg, The Netherlands, 26th-28th June 1997.

Frykblom, P. (1997) Hypothetical question modes and real willingness to pay, *Journal of Environmental Economics and Management,* 34(3): 275-287.

Green, D., K. E. Jacowitz, D. Kahneman and D. McFadden (1998) Referendum contingent valuation, anchoring, and willingness to pay for public goods. *Resource and Energy Economics* 20(2): 85-116.

Halvorsen, B. and K. Sælensminde (1998) Differences between Willingness-to-pay estimates from Open-Ended and Diskrete Choice Contingent Valuation Method: The effect of heteroscedasticity, *Land Economics,* 74(2): 262-282.

Hanemann, W. M. (1994) Valuing the Environment Through Contingent Valuation. *Journal of Economic Perspectives,* vol 8, no 4: 19-43, 1994.

Hanley, N., R. E. Wright and V. Adamowicz (1998) Using Choice Experiments to value the environment. *Environmental end Resource Economics*, 11(3-4): 413-428.

Hensher, D. A. (1994) Stated preference analysis of travel choices: the state of practice. *Transportation*, 21(2), 1994.

Hensher, D. A. and M. Bradley (1993) Using SC data to enrich revealed preference discrete choice models. *Marketing Letters* 4(2): 139-152.

Hensher, D. A., J. Louviere and J. Swait (1997) Combining sources of Preference data: The case of the lurking λ's, Working paper 97-8, Institute of Transport Studies, Sydney.

Johnson, F. R. and W. H. Desvousges (1997) Estimating Stated Preferences with Rated-Pair Data: Environmental, Health, and Employment Effects of Energy Programs, *Journal of Environmental Economics and Management*, 34(1): 79-99.

Layard, R. and Glaister, S. (1994) *Cost-Benefit Analysis*, Cambridge University Press.

McFadden, D. (1973) Conditional Logit Analysis of Qualitative Choice Behavior. In Zaremmbka, P. (ed.) (1973) *Frontiers in Econometrics*. Academic Press, New York.

Mazzotta, M. J. and J. J. Opaluch (1995) Decision making when choices are complex: A test of Heiner's hypothesis, *Land Economics* 71(4): 500-15.

Neill, H. R. (1995) The Context for Substitutes in CVM Studies: Some Empirical Observations, *Journal of Environmental Economics and Management,* 29: 393-397.

Ouwersloot, H. and P. Rietveld (1996) SC Experiments with Repeated Observations, *Journal of Transport Economics and Policy*, May 1996: 203-212.

Portney, P. R. (1994) The Contingent Valuation Debate: Why Economists Should Care, *Journal of Economic Perspectives*, vol. 8, no. 4, pp. 3-18.

Ramjerdi, F., L. Rand, I. A. F. Sætermo and K. Sælensminde (1997) The Norwegian value of time study, Report no. 379/1997, Institute of Transport Economics, Oslo.

Randall, A. (1997) The NOAA Panel Report: A Beginning or the End of an Era? *American Journal of Agricultural Economics,* 79(5): 1489-1494.

Roe, B., K. J. Boyle and M. F. Teisl (1996) Using Conjoint Analysis to Derive Estimates of Compensating Variation, *Journal of Environmental Economics and Management* 31(2): 145-159.

Spash, C. L. and Hanley, N. (1995) Preferences, information and biodiversity preservation, *Ecological Economics*, 12: 191-208.

Stevens, T. H., J. Echevarria, R. J. Glass, T. Hager and T. A. More (1991) Measuring the existence value of wildlife: what do CVM estimates really show? *Land Economics*, 67(4): 390-400.

Swait, J. and W. Adamowicz (1996) The effect of choice environment and task demands on consumer behavior: Discriminating between contribution and confusion, Working Paper, Dept. of Rural Economy, University of Alberta, June 1996.

Swait, J. and W. Adamowicz (1999) Choice task complexity and decision strategy selection. Working paper, Dept. of Marketing, College of Business Administration, University of Florida, Gainesville, September 1999.

Sælensminde, K. (1998a) The impact of choice inconsistencies on the valuation of travel time in Stated Choice studies. Paper presented at the First World Congress of Environmental and Resource Economists, Venice June 25-27, 1998.

Sælensminde, K. (1998b) Interaction effects between nonmarket goods included in benefit–cost analysis for road investments: valuation of travel time savings, traffic safety and environmental goods, Working Paper, Institute of Transport Economics, Oslo.

Sælensminde, K. (1999) Stated Choice valuation of urban traffic air pollution and noise. *Transportation Research D,* 4:13-27.

Sælensminde, K. and F. Hammer (1994) Verdsetting av miljøgoder ved bruk av samvalganalyse, (Assessing environmental benefits by means of conjoint analysis) TØI-rapport no. 251/1994, Institute of transport economics, Oslo. (In Norwegian, with English summary.)

Toner, J. P., S. D. Clark, S. M. Grant-Muller and A. S. Fowkes (1998) Anything you can do, we can do better: a provocative introduction to a new approach to Stated Preference design. Paper presented at the 8th World Conference on Transport Research (WCTR), Antwerp, July 12-17, 1998.

Widlert, S. (1994) Stated preference studies - The design affects the results. Paper presented on Seventh International Conference on Travel Behaviour, 13-16 June 1994, Santiago, Chile.

Paper 4: Inconsistent choices in Stated Choice data: Use of the logit scaling approach to handle resulting variance increases

Kjartan Sælensminde

December 13, 1999

Abstract

The scaling approach is a statistical estimation method that allows for differences in the amount of unexplained variation in different types of data, which can then be used together in the analysis. This approach has been mostly used in context of combining Stated Preference and Revealed Preference data, but has also been used as a method of identifying systematic differences in the variance of choices within a single Stated Preference data set, e.g. for investigation of learning and fatigue effects. This paper investigates whether a scaling approach is suitable for handling inconsistencies in Stated Choice data. Both the number of inconsistent choices, based on a test of violations of the transitivity axiom, and education are used as scaling variables. Scaling effects appear to exist due to inconsistent choices, and the amount of unexplained variance is shown to increase as the number of inconsistent choice increase. Scaling due to inconsistencies significantly improves the models and reduces the valuations of travel time. In addition, the scaling approach makes the valuations of travel time from the Stated Choice data more consistent with the valuations from Contingent Valuation data included in the same study. In spite of the fact that education is the only significant explanatory variable for the number of inconsistent choices, scaling due to education gives no improvement in the model.

1. Introduction

Many of the contributions in a special issue of *Journal of Transport Economics and Policy* from 1988, which was focused on Stated Preference (SP) methods, mentioned methodological problems related to variation in the statistical models' error terms (heteroscedasticity) as an important topic for further research. *Within a single data set* heteroscedasticity could be caused by a) inconsistencies due to variations in the competence of individuals to perform SP tasks (e.g. Bates 1988, Bradley 1988 and Hensher *et al.* 1988) and b) taste/preference variations (e.g. Bates 1988, Fowkes and Wardman 1988, and Kroes and Sheldon 1988). *Between data sets* (e.g. from different elicitation methods) heteroscedasticity could be a result of c) different complexity levels (e.g. number of attributes, Hensher *et al.* 1988), d) differences in how to state the preferences (e.g. handling of preference uncertainty, Bradley 1988 and Wardman 1988) and e) different choice contexts (Bradley 1988, Hensher *et al.* 1988 and Wardman 1988). Heteroscedasticity, when not accounted for in a logit model, may result in biased parameter estimates and therefore may also bias the valuation of the attributes included in the analysis (see e.g. Maher *et al.* 1999). An additional source of systematic differences in valuations obtained by different elicitation methods is the use of different functional forms for the utility functions in the analysis by different methods (e.g. Munizaga *et al.* 1997 and Halvorsen and Sælensminde 1998).

The current study of inconsistent choices is directly related to recent studies by Adamowicz *et al.* (1998), Halvorsen and Sælensminde (1998) and Cameron *et al.* (1999). These studies conclude that if one corrects for differences in error variances (heteroscedasticity) between different non-market valuation methods, the valuations from these different methods seem to be more similar. For this conclusion to be valid one must, of course, value the non-market goods in the same context (cf. discussion in the next section). Cameron *et al.* (1999) also conclude that this is not the end of the story and that we are now compelled to explain why error variances differ across methods. These studies, in connection with the knowledge that inconsistent choices cause larger error variance in Stated Choice (SC) studies (Sælensminde 1999b) and that such inconsistencies probably are a result of the complex choice situation in SC studies, makes the answer to the following question very interesting. Is a scaling approach a way to overcome or reduce the problem of inconsistent choices due to the more complex choice situation often presented to respondents in SC studies? Inclusion of both SC data and Contingent Valuation (CV) data in the current study makes it possible to compare the valuation of travel time from the CV data with the valuations from the SC data with and without use of a scaling approach.

The rest of the paper is organised as follows. The next section presents suggested solutions to the heteroscedasticity problem and other relevant literature. Section three explains the logit scaling approach and presents the models used. A description of the data sources used in the analysis and the data from the consistency test are given in section four. Section five presents the results of the analysis. Finally, conclusions are drawn and the results are discussed in sections six and seven.

2. Solutions to the heteroscedasticity problem – relevant literature

Possible solutions to the heteroscedasticity problems related to SP data were presented as early as in 1988. The solutions suggested were either to improve the study design such that the number of inconsistent choices was reduced in the data collection process and/or to improve the modelling/analysis process such that the influence of inconsistent choices on the valuations could be reduced. For example Fowkes and Wardman (1988) suggested that making the differences between choice alternatives larger could reduce errors. This would simplify the choices and therefore reduce the number of inconsistent choices. This suggestion runs against the more common view that in order to get more precise information about the preferences the differences between the two alternatives should be small. Fowkes and Wardman (1988) also suggest excluding inconsistent choices from the analyses. Bradley (1988) and Hensher *et al.* (1988) suggested including tests of internal consistency in the choice task. In addition, Hensher *et al.* (1988) suggested limiting the number of attributes to three or less. Weighting schemes or scaling procedures, where observations with larger variance can be given less weight in the statistical modelling, is suggested most frequently (e.g. Bates 1988, Kroes and Weldon 1988 and Louviere 1988). However, it can be difficult to find indicators of the systematic differences in precision that might be utilised in such a weighting scheme. Bates (1988) states that this is an unresearched area, but mentioned an example of how one might proceed. If it could be demonstrated that, say, education played an important part in explaining variations in the competence of individuals to perform SP tasks, this variable could be incorporated explicitly into the analysis.

This paper investigates whether a scaling approach is suitable for handling increased variance due to inconsistencies in SC data. As scaling variables, the number of inconsistent choices based on additional data from a direct test of violations of the transitivity axiom is used, as well as education as suggested by Bates (1988). An additional motivation for using education is that it has been shown to be a significant explanatory variable – in fact the only one - in an econometric model where the number of inconsistent choices is explained by socio-economic variables (Sælensminde 1999b). Other uses of the scaling approach in the investigation of the sources of variance within a single SP data set can be seen in the papers of Widlert (1994), Bradley and Daly (1994), Mazzotta and Opaluch (1995), Swait and Adamowicz (1996), Hensher *et al.* (1997) and Cho and Kim (1999).

The following studies provide additional discussion of methodical problems related to inconsistencies in SP data. In Sælensminde (1999b) it is shown that inconsistent choices (i.e. violation of the transitivity axiom) are common in several SC studies even when only three attributes were used in the choice alternatives. Foster and Mourato (1997) present similar results. Sælensminde (1999b) also show that inconsistent choices both result in increased variance of the estimated coefficients and affect the relative magnitude of the coefficients. Inconsistent choices therefore influence the valuation of the goods in these SC studies. Such inconsistencies are probably a result of a combination of the complex choice situation in SC studies and a limited respondent cognitive ability (Mazzotta and Opaluch 1995, Swait and Adamowicz 1999 and DeShazo and Fermo 1999). Complexity in a SC task is said to increase with i) the number of alternatives in the choice set, i) the number of attributes or goods included, iii) the number of levels used for each attribute, and iv) the number of attributes allowed to have different levels in each choice. Inconsistent choices may also result from respondent fatigue (e.g. Widlert 1994 and Bradley and Daly 1994).

Problems caused by increased complexity due to more attributes or goods included in the choice task has to be weighed against the gain in realism of the choice context and the possibility of valuation of more attributes of a composite good. The additional number of attributes one can value is a clear strength of the SC method compared to, for example, the Contingent Valuation (CV) method, where often one (environmental) program or one composite good is valued as a whole. On the other hand, the gain in realism due to a more holistic choice context is a potential strength of the CV method compared to the SC method, where often the valuation of attributes related to a composite good is the focus. As both these methods are SP methods and therefore based on hypothetical or constructed markets the importance of choosing a realistic choice context and a choice context that is consistent with later use of the valuation results should be stressed. This is not the topic of the current paper, but it is an important issue in the overall methodological discussion of how to value non-market goods for use in cost benefit analysis. See e.g. Hoehn and Randall (1989), and Randall and Hoehn (1996), Carson *et al.* (1998) and Sælensminde (1999d) for a discussion of choice context and possible interaction effects between nonmarked goods.

3. Modelling – the logit scaling approach

This section is based on Bradley and Daly (1994), who illustrate the use of the logit scaling approach for investigation of sources of variance within a single SP data set. A brief overview of such scaled modelling is given here; readers may see for example Daly (1987), Ben-Akiva and Morikawa (1990) and Bradley and Daly (1992, 1994) for more details of the particular procedure that it used.

Based on the assumptions that all choices are independent from irrelevant alternatives (IIA) (see e.g. Ben-Akiva and Lerman 1985) and the error terms are Gumbel-distributed, the Stated Choice data are commonly analysed by means of the multinomial logit model (MNL). This model has the form (McFadden 1973):

$$\Pr(y_n = 1) = \frac{e^{\mu V_{nA}}}{e^{\mu V_{nA}} + e^{\mu V_{nB}}} = \frac{1}{1 + e^{-(\mu V_{nA} - \mu V_{nB})}} \qquad (1)$$

where μ is a scale parameter that is related to the variance σ^2 of the error term ($\mu = \pi/(6\sigma)^{1/2}$). This term is unidentifiable, and can be fixed to any value without affecting the model beyond complementary scaling of the systematic component of the utility; it is usually set equal to unity. In equation (1), y_n is individual n's choice between the two alternatives i, i = A, B. (A is set equal to 1 and B is set equal to 0.) With only one type of data (assuming constant variance across observations), the systematic (observable) part of the (indirect) utility, V, of the alternative chosen, i, can be stated as

$$V_{ni} = \beta_0 + \beta_1 X_{ni1} + \ldots + \beta_K X_{niK} \ , \qquad (2)$$

where β_k, k = 0, ..., K, are the K+1 unknown parameters to be estimated and X_k, k = 1, ..., K, are the K explanatory variables in the model.

Assuming two sets of data, with observations from one set associated with larger variance than those from the other, the utility functions (U_I, U_{II}) to be estimated may appear as:

$$U_I = \boldsymbol{\beta}\mathbf{X}_I + \boldsymbol{\alpha}\mathbf{Y} + \varepsilon_I \ , \qquad (3)$$

$$U_{II} = \boldsymbol{\beta}\mathbf{X}_{II} + \boldsymbol{\tau}\mathbf{Z} + \varepsilon_{II} \ , \qquad (4)$$

where $\boldsymbol{\beta}$ is a vector of parameters to be estimated, assumed to have the same values in both data sets, while $\mathbf{X}_I$ and $\mathbf{X}_{II}$ are vectors of observed values of variables common to the two data sets; $\mathbf{Y}$ and $\mathbf{Z}$ are vectors of observed variables which may be specific to one data set or the other; $\boldsymbol{\alpha}$ and $\boldsymbol{\tau}$ are vectors of parameters to be estimated for the data-specific independent variables; ε_I and ε_{II} represents the amount of residual, unexplained variance in the choices in the two data sets.

The scaling approach addresses the problem of ε_I and ε_{II} not having the same distribution by allowing different types of data to have different error variances within a single model. Suppose that:

$$\mu^2 = \frac{\mathrm{var}(\varepsilon_I)}{\mathrm{var}(\varepsilon_{II})} \ , \qquad (5)$$

In scaling the utility of data set II such that:

$$U_{II}' = \mu U_{II} \ , \qquad (6)$$

we thus allow U_{II}' to be estimated in a single model with U_I in an efficient and unbiased manner. With the multiplicative scale parameter μ, however, the utility function U_{II}' is no longer linear in the parameters, and standard model estimation methods are no longer appropriate. Fortunately, Ben-Akiva and Morikawa (1990) developed a procedure to estimate the scale factor(s) (μ) at the same time as the other unknown parameters ($\boldsymbol{\beta}$, $\boldsymbol{\alpha}$ and $\boldsymbol{\tau}$). In order to make the estimation procedure of Ben-Akiva and Morikawa more applicable, Bradley and Daly (1992) incorporated it into one that can be carried out with any logit estimation software capable of estimating models with nested "tree" structures (Daly 1987).

In the current paper the logit scaling approach is implemented using an "artificial tree" structure to take advantage of the existing software capabilities in the ALOGIT package (Hague Consulting Group 1992).

4. Data used in the analysis

Data from two different surveys are used in this paper to study inconsistency in SC studies. The data from the Norwegian value of time study (Ramjerdi *et al.* 1997) are used as the main data source. Data from a study providing the Norwegian road authorities with valuation of environmental benefits for use in cost-benefit analysis is used as a supplementary data source (Sælensminde and Hammer 1994). This section mainly describes how the choice tasks and the OE-CV questions were formulated in the value of time survey and only briefly presents the first choice task in the environmental study. (The questionnaires from both studies are available from the author on request.)

4.1 The Norwegian value of time survey

The Norwegian value of time survey consists of 4556 personal interviews of car drivers and travellers by public transport undertaken in 1994-1996. The respondents were drawn randomly from the population of the largest cities in Norway. 509 of the interviews made in the value of time survey are used in the current study. Respondents who recently had undertaken a private car journey of less than 50 km first answered questions in the SC experiment and were thereafter asked more directly about their willingness to pay. In the SC task the respondents were to choose between two journeys using the same travel mode (in this case: car) they had used on the particular journey. The short distance car journeys were described by three attributes: price, in-vehicle travel time and "chance of a delay".

The direct questions used to determine respondent willingness to pay for reduced travel time on the particular car journey were formulated as so-called Transfer Price questions. What have been called Transfer Price questions in the context of value of time estimation are very similar to CV questions (see, for example, Layard and Glaister 1994 p.258 for an outline). Respondents should be quite familiar with their preferences for reduced travel time on a particular journey when they were presented the CV question, since it followed the choice experiment.

Both the SC task and the CV questions were designed to elicit respondents willingness to pay for a reduction of 25 % in travel time and the compensation required if travel time were increased by 25 % on that particular journey. A typical willingness-to-pay choice from the SC sequence is presented in Figure 1, together with the corresponding CV question.

One choice from the SC task	
Car journey A	**Car journey B**
Travel costs = 25 NOK. Travel time = 30 minutes. 1 in 10 car journeys is delayed by 5 minutes.	Travel costs = 20 NOK. Travel time = 40 minutes. 1 in 10 car journeys is delayed by 5 minutes.
The corresponding CV question	
Your actual car journey took 40 minutes and the price was 20 NOK. Imagine that this journey by car could be completed in 30 minutes. What is the maximum price you would be willing to pay for the journey with reduced travel time? I.e. at what price are the following two journeys of equal value for you? Journey 1. Travel time=40 minutes and price=20 NOK. Journey 2. Travel time=30 minutes and price=? NOK.	

Figure 1: Example of one choice from the SC task and the corresponding CV question presented to the travellers with car in the value of time study. Respondents were asked to choose between car journey A and car journey B in the SC task and to state their willingness to pay in the CV question.

The survey was administered using the MINT computer-assisted personal interview (CAPI) software. Three design variables were used, each with three or four levels. An orthogonal fractional factorial design of sixteen alternatives was randomly selected for each respondent from the full factorial design of all possible combinations using the program SPEED. From these sixteen options respondents were presented with a series of nine pairwise choices presented in random order. Binary logit models were estimated by ALOGIT on the pairwise choice data, using linear functions of the three design variables. (MINT, SPEED and ALOGIT are all software from Hague Consulting Group.)

4.2 The environmental survey

The environmental survey consisted of 1680 personal interviews undertaken in 1993 of car drivers and travellers by public transport randomly drawn from the population in Oslo and the neighbouring county of Akershus. Respondents who recently had undertaken a private local journey first answered simple SC questions and then they were given more complicated choice tasks also included environmental attributes. Only the first choice task, completed by 1473 respondents, was used in the current study.

In the first choice task of the environmental survey respondents chose between two journeys with the travel mode they had used on the particular journey. The journeys were described by three attributes: price, in-vehicle travel time and "walking time to parking place" for car travellers, and "chance of seating" for travellers by public transport. As in the value of time survey, the environmental survey was administered using MINT. A similar design procedure was used and similar binary logit models were used for estimation. One important difference between the two surveys is that respondents were presented with a series of only four pairwise choices in the first choice task of the environmental survey.

4.3 Data from consistency tests

Sælensminde (1999b) describes in detail how the tests of inconsistent choices are performed. Briefly explained, the consistency test is based on the assumption that the respondents have a given preference structure and that their choices should satisfy the axiom of transitivity from consumer theory. The tests are applied to studies of people's choices between two different journeys with the same mode. The preference structure characterising consistent choices is based on the assumption that respondents prefer to pay less, use less travel time and have less chance of delays. There were no tests of violation of the non-satiation axiom in the data used in the current study; all inconsistencies detected are therefore due to violations of the axiom of transitivity. The number of inconsistent choices for car travellers in the two data sets is presented in Table 1. (Statistical tests of the distributions of the observed number of inconsistent choices with the Poisson distribution show that there is significant difference between these two distributions. This indicates that there is some type of dependency between the observed number of inconsistent choices in Table 1.)

Table 1. The percentage of respondents with a different number of inconsistent choices and the average number of inconsistent choices for car travellers in two different surveys.

Number of inconsistent choices	The value of time survey. 9 stated choices. (N=393)	The environmental survey. 4 stated choices. (N=240)
0	33.1	63.3
1	31.0	35.4
2	19.8	1.3
3	9.7	
4	4.6	
5	1.0	
6	0.5	
7	0.3	
Average number of inconsistent choices	1.28	0.38

Lexicographic choices satisfy the axiom of transitivity, but to avoid confusion and possibly add precision, the respondents who have chosen lexicographically, thus violating the axiom of continuity, are partly removed and partly included in the analysis. By lexicographic choices we mean a set of choices in which the respondent consistently chooses the alternative that is best with respect to one particular attribute, e.g. lowest price, neglecting all other attributes. Lexicographic choices may be a way of simplifying the choice task if the respondent finds it too difficult or a result of levels for one of the attributes in the presented choice alternatives that are too different. In the latter case the respondent chooses according to his/her actual preferences, but because the chosen levels for one of the attributes cause this attribute to dominate the choice task, it is not possible for the respondent to state his/her preferences for the other attributes included in the task. In the value of time survey and the environmental survey, respectively, 143 and 654 (28.1 and 73.2 percent) of the car travellers chose lexicographically in the first choice task. A more comprehensive analysis of the occurrence of lexicographic choices, their effect on valuations and an investigation of possible reasons (simplification or actual preferences) for lexicographic choices is presented in Sælensminde (1999c).

5. Modelling results

In this section the logit scaling approach is used to answer three main questions.

1. Are learning and/or fatigue effects present in the SC data and can such phenomena explain the observed inconsistencies and therefore the variance differences? In addition, can scaling based on the consecutive order of the choices improve the models and reduce the effect of variance differences?
2. Will a scaling procedure directly based on the number of inconsistent choices improve the models, reduce the effect of the variance differences and therefore affect the valuations?
3. Is it possible to use education as an indicator of inconsistent choices, and will a scaling procedure based on education improve the models and reduce the effect of the variance differences in similar ways as a scaling based on inconsistent choices?

5.1 Learning and fatigue

In an attempt to capture possible learning and fatigue effects, scaling based on the consecutive order of the choices was carried out. The results are presented in Table 2, and similar results are achieved regardless of whether lexicographic choices are included.

Table 2. Scaling procedure for detecting and adjusting variance differences due to possible learning and fatigue effects. Data source: The value of time study.

Model	Lexicographic choices included (4572 observations)		Lexicographic choices excluded (3537 observations)	
	Base	Scaled	Base	Scaled
Log-likelihood	-2876.6	-2864.3	-2264.9	-2255.8
Coefficients (t-statistics w.r.t. 0)				
Travel Cost (NOK)	-0.1527 (-17.1)	-0.1785 (-8.9)	-0.1410 (-14.8)	-0.1718 (-7.6)
Travel time (min.)	-0.0574 (-7.7)	-0.0640 (-6.7)	-0.0878 (-10.4)	-0.1023 (-7.0)
Chance of delay (min.)	-0.0949 (-2.4)	-0.1258 (-2.9)	-0.2225 (-5.0)	-0.2700 (-4.6)
Constant	-0.0503 (-1.6)	-0.0533 (-1.5)	-0.0620 (-1.8)	-0.0701 (-1.6)
Scale factors (t-statistics w.r.t. 1)				
Choice 1 (base)		1.000		1.000
Choice 2		0.781 (-1.5)		0.630 (-2.4)
Choice 3		0.725 (-1.9)		0.657 (-1.8)
Choice 4		0.590 (-3.5)		0.591 (-2.9)
Choice 5		0.703 (-2.1)		0.662 (-2.0)
Choice 6		1.152 (0.7)		1.221 (0.9)
Choice 7		1.313 (1.4)		1.169 (0.7)
Choice 8		1.006 (0.0)		0.941 (-0.3)
Choice 9		0.813 (-1.0)		0.802 (-1.0)
Valuations (t-statistics w.r.t. 0)				
Travel time (NOK/h)	22.5 (10.0)	21.5 (9.5)	37.4 (14.8)	35.7 (5.4)
Chance of delay (NOK/h)	37.2 (2.2)	42.3 (3.3)	94.7 (6.1)	94.2 (6.8)

According to the Likelihood Ratio (LR) test this scaling procedure gives no improvement in the overall fit of the models (only 12 and 9 units in the log-likelihood for the addition of 8 parameters). The LR test was chosen because it simultaneously tests for improvements in all parameters in contrast to tests for improvements in each parameter separately. The test statistics for the null hypothesis that the restrictions are true is asymptotically distributed as chi squared with degrees of freedom equal to the number of restrictions being tested (see e.g. Ben-Akiva and Lerman 1985). The models in Table 2 show no clear learning or fatigue effects, and the scaling has little effect on the relative magnitude of the coefficients compared to the unscaled base models.

These results are similar to those presented by Bradley and Daly (1994), but the studies differ in an important manner. In Bradley and Daly's study, the respondents were presented 10 to 16 choices (in contrast to the 9 choices in the current study), and they seem to find a fatigue effect from choice number 10. This fatigue effect may also be the reason why Bradley and Daly conclude that the overall fit of their model substantially improves due to the addition of the scale parameters. The conclusion in the current study is supported by the results in Adamowicz *et al.*'s (1998) SC study indicating no learning and fatigue effects in the case of 8 subsequent choices. As suggested by Swait and Adamowicz (1996), the degree to which learning and fatigue effects are present in SC studies seems to be dependent on the context and complexity of the decision environment.

5.2 Inconsistent choices

Table 3 presents the result of a scaling procedure by the number of inconsistent choices showed in Table 1. These results also seem quite similar regardless of whether lexicographic choices are included. The LR test shows that scaling based on inconsistent choices significantly improves both models (improvement in the log-likelihood of 143 and 30 units for the addition of 3 parameters), and the results show reduced valuations of travel time and chance of delay.

In the case of lexicographic choices excluded, this reduction in the valuation of travel time leads to greater consistency with the valuations of travel time from the CV data (23 NOK/h) included in the same study. However, in the case of lexicographic choices included, this reduction in the valuation of travel time from the SC data leads to less consistency with the valuations of travel time from the CV data (21 NOK/h).

This result occurs despite the fact that the model is significantly improved by the scaling approach, and, unfortunately, there seem to be no obvious explanations. A possible explanation is that lexicographic choices may distort the models so much that it also affects the parameter estimates and the valuations. Such effects are presented in Sælensminde (1999c), and it is concluded that lexicographic choices have a significant impact on the valuation of reduced travel time. The results from this scaling approach may therefore be seen as an additional argument for the view that lexicographic choices should be avoided in SC data.

Table 3. Scaling procedure for detecting and adjusting variance differences based on inconsistent choices (i.e. violations of the transitivity axiom). Data source: The value of time study.

Model	Lexicographic choices included (4572 observations)		Lexicographic choices excluded (3537 observations)	
	Base	Scaled	Base	Scaled
Log-likelihood	-2876.6	-2733.6	-2264.9	-2234.5
Coefficients (t-statistics w.r.t. 0)				
Travel Cost (NOK)	-0.1527 (-17.1)	-0.3379 (-18.0)	-0.1410 (-14.8)	-0.2614 (-12.1)
Travel time (min.)	-0.0574 (-7.7)	-0.0569 (-4.8)	-0.0878 (-10.4)	-0.1172 (-9.0)
Chance of delay (min.)	-0.0949 (-2.4)	-0.1010 (-1.6)	-0.2225 (-5.0)	-0.3358 (-4.9)
Constant	-0.0503 (-1.6)	0.0354 (0.7)	-0.0620 (-1.8)	-0.0039 (-0.1)
Scale factors (t-statistics w.r.t. 1)				
0-inconsistent choices (base)		1.000		1.000
1-inconsistent choice		0.296 (-16.3)		0.556 (-5.4)
2-inconsistent choices		0.161 (-23.5)		0.360 (-9.5)
3+-inconsistent choices		0.094 (-25.7)		0.215 (-13.1)
Valuations (t-statistics w.r.t. 0)				
Travel time (NOK/h)	22.5 (10.0)	10.1 (5.4)	37.4 (14.8)	26.9 (11.2)
Chance of delay (NOK/h)	37.2 (2.2)	17.9 (1.7)	94.7 (6.1)	77.1 (5.9)

5.3 Education used as an indicator for inconsistent choices

The data from the value of time survey do not include education as a socio-economic variable. Therefore, we have to use data from the environmental survey to investigate whether it is possible to use education as an indicator of inconsistent choices, and if a scaling procedure based on education will improve the models.

Even though 72 % of the respondents chose lexicographically in the first choice task of the environmental survey, exclusion of respondents with lexicographic choices gave model results similar to those presented in Table 3. With each respondent only making four choices in this choice task, many respondents scored as giving lexicographic answers are probably misclassified. This misclassification is due the scoring procedure's inability to distinguish between valid choice combinations giving by chance the same results, as would a lexicographic choice process. (According to Sælensminde (1999c) the theoretical share of "apparent" lexicographic choices in a choice task with four choices is almost 20 percent for an attribute with three levels.) For these reasons, model results are presented with "lexicographic" choices included. An additional comment is that non-linear functions (log of the price variable)) are used in these models and that the base model, together with models for valuation of environmental attributes, are previously documented in Sælensminde (1999a). The choice of non-linear functions is a result of tests with Box-Cox models that consider higher order effects. The scaling procedure, applied to linear functions gave similar results. (See e.g. Gaudry *et al.* 1989 for computation of the valuation of the attributes included in Box-Cox models.)

The education variable used in the scaling procedure was measured on a 2-point scale where "less education" means 10 years or less at school (19 % of the sample) and "more education" means 11 years or more at school (81 % of the sample). Analyses using 3-point and 4-point education scales did not result in the education parameters being significantly different from zero.

Table 4 shows that scaling based on inconsistent choices applied to the data from the environmental survey significantly improve the model (a 37 unit improvement in the log-likelihood for the addition of one parameter). In this case no significant changes in the valuations of travel time and walking time are registered. This may be as a result of the fact that the percentage of inconsistent choices is much lower in this data set compared to that of the value of time survey (cf. Table 1).

Table 4. Scaling procedure for detecting and adjusting variance differences based on A) inconsistent choices (i.e. violations of the transitivity axiom) and B) education. Data source: The environmental study.

	Lexicographic choices included (3576 observations)		
Model	Base	Scaled-A	Scaled-B
Log-likelihood	-2089.4	-2051.9	-2087.5
Coefficients (t-statistics w.r.t. 0)			
Travel Cost (NOK)	-2.4479 (-22.3)	-2.7190 (-22.5)	-2.5390 (-21.0)
Travel time (min.)	-0.0608 (-10.8)	-0.0586 (-9.8)	-0.0663 (-10.2)
Walking time (min.)	-0.05727 (-2.0)	-0.0608 (-2.0)	-0.0628 (-2.1)
Constant	0.02480 (0.7)	0.0573 (1.3)	0.0245 (0.6)
Scale factors (t-statistics w.r.t. 1)			
0-inconsistent choices (base)		1.000	
1 or 2 inconsistent choices		0.273 (-10.8)	
More education (base)			1.000
Less education			0.786 (-2.1)
Valuations (t-statistics w.r.t. 0)			
Travel time (NOK/h)	20.7 (47.7)	17.9 (41.8)	21.7 (47.1)
Walking time (NOK/h)	19.5 (7.7)	18.6 (7.8)	20.6 (8.2)

Scaling based on education gives no improvement of the model (only a two unit reduction in the log-likelihood for the addition of one parameter). This result is achieved despite the fact that education is the only significant explanatory variable for the number of inconsistent choices (Sælensminde 1999b). Barring poor quality of the data, the models in Table 4 show that scaling based on education does not improve the models the same way as scaling based on the number of inconsistent choices. It does not, therefore, appear possible to use education as an indicator of inconsistent choices in the scaling procedure.

6. Conclusions

The analyses show that:

1. Scaling based on the consecutive order of the choices gives no improvement to the overall fit of the models. Possible learning and fatigue effects seem to be dominated by more severe inconsistencies.
2. Scaling based on inconsistent choices significantly improves the models and affects the relative magnitude of the coefficient when compared to the unscaled (base) models. In the case of lexicographic choices excluded, the scaling procedure makes the valuation of travel time more consistent with the valuations of travel time from the CV data.
3. Scaling based on education does not improve the model. Therefore, it does not appear possible to use education as an indicator of inconsistent choices in the scaling procedure.

These results indicate that the scaling approach is not an easy solution to the problem of inconsistent choices due to the complex choice situation in CS studies. This conclusion is based on the fact that a consistency test may be both resource consuming and of limited use if the SC task is so complex that the data consists of more "noise" (e.g. inconsistent choices and lexicographic choices as a result of simplification) than "signals". In addition, it is thought provoking if only highly educated people are able to choose in a consistent manner in SC studies and that their preferences therefore will count more than that of others (Sælensminde 1999b). In any case, a data collection procedure that collects a large proportion of "noise" should be amenable to improvement. A short discussion of some of the suggestions in the special issue of *Journal of Transport Economics and Policy* from 1988, in light of new results, could therefore be fruitful for the direction of future research.

7. Discussion

Bradley (1988) suggested that use of different elicitation methods should be able to handle error differences. This suggestion implies that more resources are needed both in the data collection process and for modelling, but in light of the results from Adamowicz *et al.* (1998), Halvorsen and Sælensminde (1998) and Cameron *et al.* (1999) it is worth further attention. In a phase where different elicitation methods are judged with respect to their advantages and disadvantages, collection of data in the same choice context, but by different methods, is crucial.

As mentioned by Bradley (1988) and Wardman (1988) heteroscedasticity between methods could be a result of differences in how preferences are stated. Use of methods that give respondents the opportunity to state if they are uncertain of their preferences could therefore reduce inconsistencies, and such information could later be used as input for scaling of the statistical models. Champ *et al.* (1997) state that use of a follow-up certainty question is a promising approach and conclude that respondents who answer consistently are more certain of their real preferences. Studies by Li and Mattsson (1995), Ready *et al.* (1995) and Blamey *et al.* (1999), which use different kinds of certainty questions in their elicitation formats, conclude that ignorance of preference uncertainty and so-called "yea-saying" lead to an upwardly biased estimate of willingness to pay in discrete choice CV studies. The fact that discrete choice CV and SC are related methods and that the current study generates similar conclusions makes these results interesting.

If the respondents that make inconsistent choices state that they are more uncertain about their choices in the choice task than are respondents that choose consistently, then this approach will result in less weight on inconsistent choices (i.e. respondents with task problems) in the analysis of the data. The use of a preference uncertainty approach may therefore result in less weight on choices from less educated people in the same way as the scaling procedure based on results from a consistency test.

A better way to reduce the problem with inconsistent and other "noisy" choices in SP data is probably through methods that guide the respondents or in other ways ensure that their choices are in accordance with the transitivity axiom. In this connection, the suggestion by Fowkes and Wardman (1988), making differences between choice alternatives larger in order to simplify the choices, may be of special interest. Following Fowkes and Wardman's suggestion, Toner *et al.* (1998) have obtained expressions for minimising the variance of the estimated parameters of a logit model, which indicate that the two choices offered to respondents should have probabilities of being chosen of 0.917 and 0.083, respectively. Such large differences in the presented alternatives will probably reduce problems with inconsistent choices, but may cause more lexicographic choices. Still, it seems like a promising approach for design improvements (Toner *et al.* 1999). This view is supported by the findings of Peterson and Brown (1998), who use the psychometric method of paired comparison to investigate the reliability and transitivity of binary choices. They conclude that the primary causes of inconsistency appear to be indifference and simple mistakes. Their conclusion is based on the findings that the likelihood of inconsistency declines rapidly with increasing value contrast between the two items in each choice, and that inconsistent choices are much more likely to be switched on retrial than are consistent choices.

As a comment to this "new" insight into the problem of how to minimise the share of inconsistent choices in SP data it is interesting to see that one of the first winners of the Nobel Prizes in Economics (1969), Ragnar Frisch, presented an elicitation method based on these principles in his Nobel lecture (see e.g. Frisch 1972). The basic principle of Frisch's suggestion is to start with such large differences between the two choice items that the choice is viewed as obvious and simple by the respondent. By use of such large differences there should be less reason to advise against use of an adaptive design in choice studies (Bradley and Daly 1994 and Johnson and Desvousges 1997). In order to receive more information about the respondent's preferences Frisch made the differences between the two choice items subsequently smaller until the respondent stated that he/she could no longer say that one of the items were better than the other. Admittedly, Frisch's method is presented as a method to elicit the preference functions of politicians, but there is nothing to prevent the use of this method to elicit the preferences of ordinary people in the same way as other SP methods.

Given the large number of valuation studies that have taken place in the last 20-30 years, surprisingly little progress has been made in the attempt to solve methodological problems like the one discussed in this paper. It is therefore easy to agree with Deacon *et al.* (1998) when they conclude that valuation research has been dominated by applications rather than testing and refining the basic theory. Deacon *et al.* are therefore calling for research that identifies specific gaps or inconsistencies in the current state of the art, and develops empirical or theoretical strategies to close them. The current study contributes on the issue of heteroscedasticity caused by inconsistent choices in SP data and answers some questions about how to handle such problems in the future, but a lot more research on these topics is certainly needed.

References

Adamowicz, W. L., P. Boxall, M. Williams and J. Louviere (1998) Stated Preference approaches to measuring passive use values: Choice Experiments and Contingent Valuation. *American Journal of Agricultural Economics,* 80(1): 64-75.

Bates, J. (1988) Econometric issues in Stated Preference analysis. *Journal of Transport Economics and Policy* 22 (1): 59-69.

Ben-Akiva, M. and S. R. Lerman (1985) *Discrete Choice Analysis*, The MIT Press, 1985.

Ben-Akiva, M. and T. Morikawa (1990) Estimation of travel demand models from multiple data sources. Paper presented at the 11th International Symposium on Transportation and Traffic Theory, Yokohama.

Blamey, R. K., J. W. Bennett and M. D. Morrison (1999) Yea.saying in Contingent Valuation surveys. Land Economics, 75 (1): 126-141.

Bradley, M. (1988) Realism and adaptation in designing hypothetical travel choice concepts. *Journal of Transport Economics and Policy* 22 (1): 121-137.

Bradley, M. and A. Daly (1992) Estimation of logit choice models using mixed stated preference and revealed preference information. Paper presented at the 6th international Conference on Travel Behaviour, Quebec.

Bradley, M. and A. Daly (1994) Use of the logit scaling approach to test for rank-order and fatigue effects in stated preference data, *Transportation* 21(2): 167-184.

Cameron, T. A., G. L. Poe, R. G. Ethier and W. D. Schulze (1999) Alternative nonmarket valuation-elicitation methods: Are the underlying preferences the same? Paper presented at the W-133 Annual Meeting in Tucson, February 1999.

Carson, R. T., N. E. Flores and W. M. Hanemann (1998) Sequencing and Valuing Public Goods, *Journal of Environmental Economics and Management*, 36: 314-323.

Champ, P. A., R. C. Bishop, T. C. Brown and D. W. McCollum (1997) Using Donation Mechanisms to Value Nonuse Benefits from Public Goods, *Journal of Environmental Economics and Management* 33, 151-162.

Cho, H.-J. and K.-S. Kim (1999) Combined analysis of the heteroscedasticity and correlation of repeated observations in SP data. Paper presented at European Transport Conference, PTRC, September 1999, Cambridge.

Daly, A. J. (1987) Estimation of "Tree" Logit Models. *Transportation Research* 21B(4): 251-267.

Deacon, R. T., D. S. Brookshire, A. S. Fisher, A. V. Kneese, C. D. Kolstad, D. Scrogin, V. K. Smith, M. Ward and J. Wilen (1998) Research trends and opportunities in Environmental and Natural Resource Economics. *Environmental and Resource Economics*, 11(3-4): 383-397.

DeShazo, J. R. and G. Fermo (1999) Designing choice sets for Stated Preference methods: The effects of complexity on choice consistency. Paper presented at the W-133 Annual Meeting in Tucson, February 1999.

Foster, V. and S. Mourato (1997) Are Consumers Rational? Evidence from a Contingent Ranking Experiment, Paper presented at the 8th Annual Conference of the European Association of Environmental and Resource Economists, Tilburg, The Netherlands, 26th-28th June 1997.

Fowkes, T. and M. Wardman (1988) The design of Stated Preference Travel choice experiments. With special reference to interpersonal taste variations. *Journal of Transport Economics and Policy* 22 (1): 27-44.

Frisch, R. (1972) Cooperation between politicians and econometricians on the formalization of political preferences. Reprint series no. 90. Institute of Economics, University of Oslo, Oslo 1972.

Gaudry, M., S. R. Jara-Diaz and J. D. Ortúzar (1989) Value of time sensitivity to model specification. *Transportation Research* 23B(2): 151-158.

Hague Consulting Group (1992) ALOGIT 3.2. Users Guide. Hague Consulting Group, The Hague, 1992.

Halvorsen, B. and K. Sælensminde (1998) Differences between Willingness-to-pay estimates from Open-Ended and Discrete Choice Contingent Valuation Method: The effect of heteroscedasticity, *Land Economics*, 74(2): 262-282.

Hensher, D. A., P. O. Barnard and P. T. Truong (1988) The role of Stated Preference methods in studies of travel choice. *Journal of Transport Economics and Policy* 22 (1): 45-58.

Hensher, D. A., J. Louviere and J. Swait (1997) Combining sources of Preference data: The case of the lurking λ's, Working paper 97-8, Institute of Transport Studies, Sydney.

Hoehn, J. P. and A. Randall (1989) Too Many Proposals Pass the Benefit Cost Test. *American Economic Review* 79, 1989, 544-551.

Johnson, F. R. and W. H. Desvousges (1997) Estimating Stated Preferences with Rated-Pair Data: Environmental, Health, and Employment Effects of Energy Programs, *Journal of Environmental Economics and Management*, 34(1): 79-99.

Kroes, E. P. and R. J. Sheldon (1988) Stated Preference methods. *Journal of Transport Economics and Policy* 22 (1): 11-25.

Layard, R. and Glaister, S. (1994) *Cost-Benefit Analysis*, Cambridge University Press.

Li, C.-Z. and L. Mattsson (1995) Discrete Choice under Preference Uncertainty: An Improved Structural Model for Contingent Valuation, *Journal of Environmental Economics and Management*, 28: 256-269.

Louviere, J. J. (1988) Conjoint analysis modelling of stated preferences. *Journal of Transport Economics and Policy* 22 (1): 93-119.

Maher, M. A. Rosa and J. Kontolemakis (1999) Bias in discrete choice models estimates. Paper presented at the 27th European Transport Conference 1999, PTRC, Cambridge.

Mazzotta, M. J. and J. J. Opaluch (1995) Decision making when choices are complex: A test of Heiner's hypothesis, *Land Economics* 71(4): 500-15.

McFadden, D. (1973) Conditional Logit Analysis of Qualitative Choice Behavior. In Zaremmbka, P. (ed.) (1973) *Frontiers in Econometrics*. Academic Press, New York.

Munizaga, M. A., B. G. Heydecker and J. de D. Ortúzar (1997) On the error structure of discrete choice models. *Traffic Engeneering and Control* 38(11): 593-597.

Peterson, G. L., and T. C. Brown (1998) Economic valuation by the method of Paired Comparison, with emphasis on evaluation of the transitivity axiom. *Land Economics* 72 (2): 240-261.

Ramjerdi, F. L. Rand, I. A. F. Sætermo and K. Sælensminde (1997) The Norwegian value of time study, Report no. 379/1997, Institute of Transport Economics, Oslo.

Randall, A. and J. P. Hoehn (1996) Embedding in Market Demand Systems, *Journal of Environmental Economics and Management,* 30: 369-380.

Ready, R., J. C. Whitehead and G. C. Blomquist (1995) Contingent Valuation When Respondents Are Ambivalent. *Journal of Environmental Economics and Management* 29: 181-196.

Swait, J. and W. Adamowicz (1996) The effect of choice environment and task demands on consumer behavior: Discriminating between contribution and confusion, Working Paper, Dept. of Rural Economy, University of Alberta, June 1996.

Swait, J. and W. Adamowicz (1999) Choice task complexity and decision strategy selection. Working paper, Dept. of Marketing, College of Business Administration, University of Florida, Gainesville, September 1999.

Sælensminde, K. (1999a) Stated Choice valuation of urban traffic pollution. *Transportation Research Part D* 4: 13-27.

Sælensminde, K. (1999b) The Impact of Choice Inconsistencies in Stated Choice Studies. Working paper, Institute of Transport Economics, Oslo.

Sælensminde, K. (1999c) The Cause and Impact of Lexicographic Choices in Stated Choice Studies. Working paper, Institute of Transport Economics, Oslo.

Sælensminde, K. (1999d) Interaction Effects in Valuation of Non-Market Goods. Working paper, Institute of Transport Economics, Oslo.

Sælensminde, K. and F. Hammer (1994) Verdsetting av miljøgoder ved bruk av samvalganalyse. (Assessing environmental benefits by means of conjoint analysis.) TØI-rapport no. 251/1994, Institute of transport economics, Oslo. (In Norwegian, with English summary.)

Toner, J. P., S. D. Clark, S. M. Grant-Muller and A. S. Fowkes (1998) Anything you can do, we can do better: a provocative introduction to a new approach to Stated Preference design. Paper presented at the 8th World Conference on Transport Research (WCTR), Antwerp, July 12-17, 1998.

Toner, J. P., M. Wardman and G. Whelan (1999) Testing recent advances in Stated Preference Design. Paper presented at European Transport Conference, PTRC, September 1999, Cambridge.

Wardman, M. (1988) A comparison of Revealed Preference and Stated Preference models of travel behaviour. *Journal of Transport Economics and Policy* 22 (1): 71-91.

Widlert, S. (1994) Stated preference studies - The design affects the results. Paper presented on Seventh International Conference on Travel Behaviour, 13-16 June 1994, Santiago, Chile.

Paper 5: Interaction effects in valuation of nonmarket goods

Kjartan Sælensminde

December 13, 1999

Abstract

Cost-benefit analyses undertaken for the road sector in Norway includes non-market goods such as travel time savings, traffic safety, noise and air pollution. Traditionally, these non-market goods are valued through separate willingness to pay studies without attention to interactions between them. This paper shows that a simultaneous valuation procedure, accounting for interactions between the non-market goods included in cost-benefit analyses for road investments, significantly reduces the valuations of travel time savings, noise and air pollution compared to a separate valuation procedure. Another result is that a reminder of substitutes and budget constraints seems less necessary if the valuation procedure is more holistic.

1. Introduction

A range of non-market goods are included in cost-benefit analyses (CBAs) undertaken to evaluate road projects in Norway, and probably in many other countries. The Norwegian Roads Administration currently includes the value of travel time, traffic accidents and environmental impacts such as air pollution and noise in their CBAs. These non-market goods have been valued in separate willingness to pay (WTP) studies without taking account of interaction effects (e.g. substitution or complementary effects) between them. A total valuation is obtained by summing across the independent component valuations.

Carson *et al.* (1996b) suggest that CBAs tends to proceed on such a piecemeal basis due to the time and expense associated with determining the interactions between the different goods and the belief or hope that such effects are small. However, whether such interaction effects are small or large is an empirical question, and many recent studies have concluded that more research is needed to answer this question (e.g. Hoehn and Loomis 1993, Randall and Hoehn 1996, Macdonald and McKenney 1996 and Brown *et al.* 1995).

The current study uses data collected in a Norwegian study primarily designed to revise the value of travel time savings used in economic evaluation, but also designed to study interaction effects between valuations of travel time savings, accidents, noise and air pollution. Two experimental designs were implemented: Conjoint Analysis (or more precisely the Stated Choice method) and open-ended Contingent Valuation (CV). The current study of interaction effects uses the CV responses.[1] In the CV question sequence, respondents first answered separate questions about their WTP for reduced travel time. They then answered WTP questions where reduced travel time was included in a package together with reduced number of traffic accidents and environmental impacts. The context of the package was the same as that used in CBAs. Although the current empirical study investigates possible interaction effects by use of a study primarily designed to value travel time savings, the same methodological issues apply to using CV methods to value environmental and other goods. Perhaps even more so, because CV is more prominent in those valuations than in value of time studies.

The reminder of the current section gives a brief overview of theoretical and conflicting empirical evidence related to interaction effects, and hypotheses to be tested. The next two sections describe the survey, present the results from the analysis and compare these results with the values that are used in CBAs for road projects in Norway. Section four discusses directions for future research. Finally, in section five, some concluding remarks are presented concerning the implications of the results for the choice between a separate valuation procedure and a simultaneous valuation procedure that takes account of multiple changes.

1.1 Previous theoretical contributions

Interaction effects are closely connected to the terms part-whole bias and embedding. Part-whole bias and embedding have been used to describe how WTP for a particular good may vary "over a wide range depending on whether the good is assessed on its own or embedded as part of a more inclusive package" (Kahneman and Knetsch 1992).

Randall *et al.* (1981) is the first study to report systematic embedding effects in a CV study. They conjectured that these observed effects could be explained by standard economic interactions involving substitution and budget constraints. Hoehn and Randall (1989) identify a valid welfare change measure and show how it deviates from the invalid "in isolation" measure. Two sources of deviation are mentioned: (1) substitution and complementarity relationships and (2) resource scarcity. Mitchell and Carson (1989) note that difficulty in generalising from parts to the whole (and vice versa) is by no means a unique property of the CV method, but is characteristic of other benefit measurement methods.

Hoehn and Randall (1989), Hoehn (1991) and Madden (1991) have articulated a general theory of the relationship between the total value of a complex package of goods and the value of its various components and Randall and Hoehn (1996) summarise the important theoretical findings in five points. The fifth of these

[1] The results from the Stated Choice study are reported in Ramjerdi et al. (1997). Methodological issues such as tests of consistency and lexicographic choices in the Stated Choice data are reported in Sælensminde (1998).

points concerns multi-stage-budgeting, which means that all expenditures are variable in the long run but many kinds of expenditures are fixed in the short run. Standard theory assumes that the budget allocation process is complete, i.e. rebudgeting proceeds through the final stage. However, if rebudgeting is incomplete, discretionary income is reduced and, as predicted by the theory, the effects of embedding are exacerbated. Furthermore, the effect of incomplete rebudgeting is more pronounced the smaller the initial budget allocation to an expenditure category. For environmental goods the respondent's initial direct expenditures are likely to be small and the impact of incomplete budgeting is likely to be large.

Hanemann (1991) demonstrated that substitution effects have a more pronounced influence when quantity-rationed public goods are being evaluated than for market goods where welfare evaluation typically involves price changes and not changes in quantity and/or quality. This implies that embedding effects might be larger for nonmarket goods than for market goods. However, economists have little guidance, whether from theory or empirical evidence, regarding the magnitude of embedding effects that are to be expected.

1.2 Conflicting empirical evidence

A question related to interaction and embedding effects is how to present information to the respondent about substitutes (and complements) to the good being valued. The recommendation from the NOAA panel (Arrow *et al.* 1993) that information regarding substitutes and reminders of budget constraints should be included in CV studies has been the subject of empirical testing in many studies.

1.2.1 The effect of information about substitutes and budget constraints

The studies that value non-market goods, and have only tested whether information about substitutes and budget constraints influences the valuation of non-market goods, conclude differently. Whitehead and Blomquist (1991, 1995, 1999) and Hoevenagel and van der Linden (1993) found that information about substitutes and/or a budget reminder did lower the valuation. However, Loomis *et al.* (1994, 1995) and Kotchen and Reiling (1999) found that such information did not influence the valuation.

1.2.2 Separate valuation versus valuation in a package

Despite valuation of quite different non-market goods, the studies that tested whether the valuation of a good is different if it is valued "in-isolation" or as part of a larger package, arrive at the same conclusion. Valuation of a good in a package lowers the value of the good compared to separate valuation (Samples and Hollyer 1990, Hoehn 1991, Magnussen 1992, Hoehn and Loomis 1993, Loomis *et al.* 1993, Cummings *et al.* 1994, Boxall *et al.* 1996, Jones 1997, Nelson and Towriss 1999 and Penin and Riera 1999). In these cases the goods in the package were not complementary goods.

1.2.3 The effect of information about substitutes and separate valuation versus valuation in a package

Different conclusions are reached in the studies that have tested whether the valuation of a non-market good is different if it is valued "in-isolation" or as part of a larger package, and in the same study tested the effect of additional information about substitutes and budget constraints. Three studies, Brown *et al.* (1995), Neill (1995) and Barro *et al.* (1996), conclude that valuation of a good in a package lowers the value of the good compared to separate valuation, and that the amount of information had little impact on the good's value. Macdonald and McKenney (1996), however, conclude the opposite.

1.2.4 Interaction effects and valuation of market goods [2]

The issue of interaction effects and substitutes is not unique to non-market goods (e.g. Adamowicz *et al.* 1993, Loomis *et al.* 1993, 1995 and 1996, Neill *et al.* 1994 and Bateman *et al.* 1997). Randall and Hoehn (1996) examined the effects of embedding by use of a market demand system. The conclusion from this study is that embedding is a routine economic phenomenon that may be observed with market as well as non-market valuation. They also demonstrated with market goods that incomplete rebudgeting tends to reduce WTP and increase WTA from their fully optimal values, thus exacerbating the embedding effects.

1.3 Hypothesis to be tested

In line with the literature above three essential hypotheses are tested in the current paper. These are the null hypotheses that i) separate valuation equals valuation in a package, ii) valuation without a budget reminder equals valuation with a budget reminder, and iii) valuation without a budget reminder equals valuation with a budget reminder independent of whether the good is valued separately or as part of a larger package.

2. The Survey

Data collected in connection with a value of time survey (Ramjerdi *et al.* 1997) are used to study the interaction effects between the non-market goods included in CBAs for the road sector in Norway. This section describes how the WTP

[2] Many studies of valuation of market goods that address these issues use small samples. They are therefore subject to the critique by Carson and Mitchell (1995) that if studies are to make a contribution, the sample should be large enough to reject reasonable differences. In addition, the respondents should be in the market for the actual good (Loomis et al. 1996 and Roe et al. 1996). Loomis et al. (1996) conclude that experiments using market goods may not provide an unambiguous test of criterion validity of CV for non-market goods.

questions regarding interaction effects were formulated. An English translation of an extract from the questionnaire is given in Appendix 1.[3]

The survey consisted of 728 personal interviews, undertaken in 1996, of car drivers randomly drawn from the population of the five largest cities in Norway. Respondents who had recently undertaken a local car journey were asked about i) their WTP for a reduction of 25 % in travel time on that particular car journey, ii) if their WTP for reduced travel time has general validity or if it is connected to the particular car journey only, and iii) their WTP for a reduction in travel time when travel time is part of a package also including increased traffic safety and/or a reduction in environmental problems caused by road traffic. The respondents were assumed to be quite familiar with their preferences for reduced travel time on a particular journey when they were presented the CV questions since it followed the Stated Choice questions.

Before valuation of the package, the respondents were divided into three groups, each receiving a package with a different content. Group 1 received a package of reduced travel time and increased traffic safety, Group 2 had a package of reduced travel time and a reduction in environmental problems, and Group 3 had reduced travel time, increased traffic safety and a reduction in environmental problems in their package. To test the relationship between interaction effects and such information half of the respondents were given a reminder of substitutes and their budget restrictions. In the following sections this reminder is called the "budget-reminder". (See the questionnaire in Appendix 1 for an overview of the information given to the different subgroups and the wording of the reminder.)

Differences in demographic variables between samples with and without a budget reminder may influence later test results. One-way ANOVAs for the total sample, with and without a budget reminder, were performed for age ($F = 3.4$, $p = 0.07$), gender ($F = 1.1$, $p = 0.30$), personal income ($F = 0.4$, $p = 0.51$), household income ($F = 0.2$, $p = 0.66$) and driving distance per year ($F = 2.1$, $p = 0.16$). This showed that the total samples are not statistically different with respect to these demographic variables at the 0.05 significance level. However, age is different at the 0.10 significance level for the total sample, driving distance is different at the 0.10 significance level for Group 2, personal income is different at the 0.05 significance level for Group 3 and gender is different at the 0.10 significance level for Group 3.

One objection to this study may be, as concluded by the Arrow *et al.* (1993), that open-ended CV questions are less familiar to the respondents and therefore less accurate than discrete-choice CV questions. According to the NOAA panel discrete-choice CV should be the preferred CV question format. However, issues concerning statistical uncertainty, stability and bias are not addressed by the panel, and the debate concerning the preferability of discrete-choice CV to open-ended CV is certainly not closed (Neill 1995, Boyle *et al.* 1996, Dubourg *et al.* 1997, Frykblom 1997, Randall 1997, Boyle *et al.* 1998, Green *et al.* 1998, Halvorsen and Sælensminde 1998, and Welsh and Poe 1998).

[3] The complete questionnaire is available from the author on request.

Strictly speaking, the questions used to elicit respondent WTP for reduced travel time on the particular car journey were formulated as so-called "Transfer Price" questions. These are very similar to open-ended CV questions (see, for example, Layard and Glaister 1994 p.258 for an outline). The term "CV" is henceforth used as shorthand for open-ended CV (or these "Transfer Price" questions) when talking about the current study.

3. Valuation results and testing of hypotheses

3.1 Valuation results

3.1.1 The share of respondents with WTP equal to zero

As the package of goods valued becomes more inclusive, it is reasonable to assume that more respondents will benefit from having it. It is therefore assumed that the percentage of the respondents with WTP equal to zero will decline as the package of good(s) valued grows from reduced travel time on one particular local car journey via reduced travel time on all car journeys, to a package including reduced travel time, increased traffic safety and a reduction in environmental impacts. Table 1 confirms this assumption with a percentage of WTP equal to zero as high as 81.7 for a reduction in travel time of 10 percent on one particular car journey to only 15.3 for the most complete package including travel time, traffic safety and the environment.

Table 1. Percentage of the respondents with WTP=0 for the different goods valued.

The good(s) valued	Percentage of the respondents with WTP=0
Reduced travel time (10 %) on a specific car journey	81.7
Reduced travel time (25 %) on a specific car journey	60.3
Reduced travel time (25 %) on all local car journeys	50.6
Reduced travel time (25 %) on all car journeys	40.5
Reduced travel time (25 %) on all car journeys and 50 % reduction in the risk for traffic accidents	17.3
Reduced travel time (25 %) on all car journeys and 50 % reduction in environmental problems	17.8
Reduced travel time (25 %) on all car journeys, 50 % reduction in the risk for traffic accidents and 50 % reduction in environmental problems	15.3

3.1.2 Relatively good conformity between WTP for reduced travel time on a specific car journey and all car journeys

In Table 2 the WTP for reduced travel time on one particular car journey is compared to the WTP for reduced travel time on all car journeys. The WTP is stated for local and longer car journeys. The respondents are divided into two groups, one which was given a budget reminder and another without such a reminder. The results show larger conformity between WTP for reduced travel time on a specific car journey and all car journeys if the journey is local, i.e. is undertaken more frequently, than if the journey is a longer journey undertaken at less frequent intervals. This result can be explained by i) two different data sources being used to compare the WTP for longer car journeys, ii) the WTP question about reduced travel time on all longer car journeys being presented after the WTP question about reduced travel time on all local car journeys (i.e. a sequencing effect), and/or iii) that the respondents have a clearer view about the value of a good (and the attributes of the good) and more thoroughly consider their budget when a good is purchased more frequently.

Table 2. WTP for reduced travel time on one (local and long) specific car journey versus all car journeys, and the effect of a budget reminder. Standard error in parenthesis. (1 NOK ≈ 0.125 $.)

	Mean WTP for reduced travel time with no budget reminder (NOK/h)	Mean WTP for reduced travel time with a budget reminder (NOK/h)	Percentage reduction in mean WTP due to a budget reminder
One specific local car journey (N=725)	28.86 (2.79)	24.90 (2.58)	14.2
All local car journeys (N=725)	27.85 (2.29)	23.09 (2.17)	17.2
One specific long car journey[1] (N=351)	49.63 (5.05)	40.40 (3.60)	20.0
All longer car journeys (N=725)	44.31 (5.18)	27.14 (3.28)	38.9**

[1] These results are from the value of time study for longer journeys (Sælensminde 1995). ** = Significantly different at 0.01 level.

If the results in Table 2 are due to point iii) above, one can assume that the budget reminder has more significant influence on the WTP of reduced travel time on longer car journeys than on local car journeys. Such a hypothesis seems to be confirmed by noting that a budget reminder reduces the WTP of reduced travel time more for longer journeys than for local journeys, but it is only for all longer car journeys that the budget reminder has a significant influence at the 0.05 level.

Table 2 also seems to confirm that if one asks for WTP and the payment is for a short period (per journey) a larger WTP per unit is obtained than if the payment is for a longer period (per year). However, the only significant difference (0.05 level) is between the WTP for reduced travel time on one particular long car journey (40.40 NOK/hour) and the WTP for reduced travel time on all longer car journeys (27.14 NOK/hour) for the group that received a budget reminder. There is good correlation between WTP for reduced travel time on one particular local car journey and WTP for reduced travel time on all local car journeys.

3.1.3 Separate valuation versus valuation in a package

In the analysis of how the WTP for reduced travel time is affected when it is valued separately versus in a package where traffic safety and/or the environment were improved, only respondents with a WTP larger than zero for the total package are included. This is done because only respondents with WTP larger than zero for the total package had the possibility of dividing their WTP on the different goods in the package (see the questionnaire in appendix 1). The percentage of the respondents with zero WTP for the total package of improvements is only 16.5 and the sample with WTP larger than zero is 606 people.

Table 3. Mean and median WTP for a) a package including travel time, traffic safety and/or environmental impacts, b) the specific goods in the package and c) travel time valued separately. Total samples (with and without a budget reminder) for each of the three groups.[4] (1 NOK ≈ 0.125 $.)

	Mean/Median WTP (Std err). Unit: NOK/year.	Percentage reduction in WTP due to the budget reminder
GROUP 1 (N=204)		(N, reminder=94)
Package with travel time and traffic safety	5010 / 3000 (302)	18 / 38
Traffic safety in the package	2675 / 1500 (229)	4 / 5
Travel time in the package	2335 / 787 (262)	33 / 59
Travel time valued separately	2909 / 1200 (302)	19 / 38
GROUP 2 (N=180)		(N, reminder=93)
Package with travel time and environment	4141 / 2075 (318)	45 / 46
Environment in the package	2160 / 1000 (205)	34 / 50
Travel time in the package	1981 / 1000 (200)	55 / 75
Travel time valued separately	2470 / 1000 (263)	54 / 65
GROUP 3 (N=221)		(N, reminder=115)
Package with travel time, traffic safety and environment	6144 / 3650 (521)	3 / 18
Traffic safety in the package	2235 / 1000 (264)	3 / 22
Environment in the package	1661 / 1000 (164)	9 / 0
Travel time in the package	2248 / 1000 (372)	-2 / 20
Travel time valued separately	3310 / 1250 (419)	1 / 0

Interaction effects may be different depending on which and how many goods are valued together. Therefore, the results from the three groups (c.f. section 2) are presented separately. Calculated percentage increase in mean WTP for reduced travel time valued separately versus in the packages presented in Table 3 are 25, 25 and 47 percent for the two two-goods packages and the three-goods package, respectively. Because of similarities with the current study it is interesting to mention that Hoehn and Loomis (1993) show that independent valuation and summation of environmental programs overstates the benefits of two and three program policies by an average of 24 and 54 percent, respectively.

[4] The results from the subgroups with and without a budget reminder, respectively, are available from the author on request.

3.1.4 Comparisons with the values that are used in CBAs for road projects in Norway

Table 4 shows a comparison of the simultaneous valuations done in "CBA context" (Group 3 in Table 3), separate valuations from WTP-studies (Elvik 1993 and Sælensminde and Hammer 1994) and the valuations used by Public Roads Administration (PRA) in CBAs for road projects in Norway (PRA 1995). The valuation of reduced travel time used by PRA is based on several Revealed Preference (RP) studies of people's actual choices of road, speed and transport mode (Johansen 1994). A comparison of the valuations of reduced travel time used by PRA with the valuation done in "CBA context" in the current study, is therefore, not only a comparison of valuations in different contexts, but also a comparison of valuation from RP versus SP data. This may look like a poor basis for comparisons, but comparisons of valuations from RP and SP data done in similar contexts show very similar valuation results (Carson *et al.* 1996a, Wardman 1988, 1998, Randall 1998 and Yoo and Ashford 1998). An interpretation of valuation differences as a result of context differences may therefore not be appropriate.

The valuation of WTP for traffic safety used by PRA is from a meta-analysis of international studies that have valued safety and health (Elvik 1993). The valuation of noise and air pollution used by PRA is from a Stated Choice (SC) study (Sælensminde and Hammer 1994). In that SC study a "bottom-up" approach, i.e. a separate valuation of noise and air pollution, was used. The "bottom-up" approach may produce higher valuations than a "top-down" approach, i.e. a simultaneous valuation of environmental problems related to road traffic, which is used in the "CBA context" valuation in the current study. In addition, the SC-study probably suffers from the influence of inconsistent choices which may bias the valuation results upwards (Sælensminde 1998). To make this comparison of the influence of context differences more meaningful the valuations from the "top-down" CV study (Sælensminde and Hammer 1994), conducted on the same respondents as the SC study, is included in Table 4.

Table 4 shows that the valuations of travel time and environmental problems that are used in by PRA in their CBAs are much larger than the valuations from the current study done in "CBA context". It also shows that the relative valuations of travel time, traffic safety and environmental problems actually used in CBAs are very different from the relative valuations of these goods found in "CBA context". The values used in CBAs seem therefore to favour projects that mainly make improvements in travel time and/or the environment at the expense of projects that mainly improve traffic safety. A comment to these conclusions is that the valuation of travel time in "CBA context" probably will be lower in a study where travel time, traffic safety and environmental issues are given a more equal framing and where travel time is not always valued first in the sequence (Mitchell and Carson 1989). Likewise, the valuation of noise and air pollution, which was always placed last in the sequence, is probably a low estimate.

Table 4. Comparisons of the simultaneous valuations done in "CBA context", separate valuations and the valuations actually used in CBAs for road projects in Norway. Unit: WTP per household in 1996 NOK/year.

Context/Source	Travel time (25 % reduction)	Traffic accidents (50 % reduction)	Noise and air pollution (50 % reduction)
"CBA context"/ The current study	2248	2235	1661
Separate/ The current study	3310		
Separate/ (Elvik 1993)		2835	
Separate/ (Sælensminde and Hammer 1994)			2395
Values used in CBAs/ (PRA 1995)	7282	2835	12,563
Percentage increase: "CBA context" to Separate	47	27	44
Percentage increase: "CBA context" to Values used in CBAs	224	27	656

3.2 Testing of hypotheses

Both parametric and non-parametric tests were considered for testing the null hypotheses that i) separate valuation equals valuation in a package and ii) valuation without a budget reminder equals valuation with a budget reminder. The choice of non-parametric tests was motivated by the fact that there are differences between mean and median WTP in Table 3 and by inspection of the skewness and kurtosis statistics which range from 2.7 to 6.0 and from 10.1 to 68.0, respectively. This is far above the skewness and the kurtosis values of the normal distribution which are 0 and 3 respectively. Paired sample tests are used to account for possible correlation that may complicate relevant comparisons between expected benefits across scenarios. Such complications similarly arise in testing for within-subject embedding effects in which values placed on a comprehensive good are compared with values for a subset of the comprehensive good. (See Carson and Mitchell (1995) for an outline and Poe *et al.* (1997) for a solution to this problem related to discrete-choice CV data.)

3.2.1 Separate valuation versus valuation in a package

The non-parametric Wilcoxon signed-rank test (see e.g. Noether 1967), a test of whether two samples come from the same distribution, was chosen to test the null hypothesis that separate valuation equals valuation in a package. The Wilcoxon signed-rank test considers information about both the sign of the differences and the magnitude of the differences between pairs and is therefore more powerful than the more common Sign test.

Tests were made for the total sample and for the three different groups with a budget reminder and without a budget reminder. All tests were two-sided. The

Wilcoxon signed rank test of equal valuation is rejected at 0.01 level in all cases except for the subsample of Group 2 that received a budget reminder.

3.2.2 WTP with and without a reminder of substitutes and budget constraints

The non-parametric Kolmogorov-Smirnov test (see e.g. Noether 1967) was chosen to test the null hypothesis that valuation without a budget reminder equals valuation with a budget reminder. The Kolmogorov-Smirnov test detects differences in both the locations and the shapes of the distributions of independent samples and is therefore a more powerful test than e.g. the Mann-Whitney test.

The Kolmogorov-Smirnov tests show significant (at 0.05 level) reductions in all goods due to the budget reminder for Group 2. For Group 1 only travel time valued in the package is significantly reduced due to the budget reminder. Group 3 shows small differences due to the budget reminder (c.f. Table 3), and none of these differences is significant. The average reduction due to the budget reminder for all three groups together in a) the WTP for the total package, b) reduced travel time in the package, c) reduced travel time valued separate, d) traffic safety in the package and e) environment in the package are 18, 29, 22, 3 and 24 percent, respectively. The first three of these reductions are significant at the 0.05 level; the two latter are not.

Group differences in terms of demographics may have influenced the tests of the effect of a budget reminder. To isolate the effect of the budget reminder, some simple regression models were constructed.

3.2.3 Regression models and the isolated effect of the budget reminder

To isolate the effect of the budget reminder five different OLS regression models and five different logit models are considered. In both the OLS models and the logit models the dependent variables are 1) WTP for the total package, 2) WTP for travel time valued in the package, 3) WTP for travel time valued separately, 4) WTP for traffic safety valued in the package and 5) WTP for the environment valued in the package. In each of the OLS models of package valuation, the budget reminder was at first included as a single dummy variable. Regardless of the number of goods in the package, the dummy variable coefficients for the budget reminder were not significantly different from zero in these models.

To investigate whether a budget reminder and valuation in a package may be part of "the same thing", two dummy variables for the budget reminder were included: one dummy variable if the package consisted of two goods (BR-2), and another dummy variable if the package consisted of three goods (BR-3). If the budget reminder and valuation in a package are part of "the same thing", one would expect that the parameter for BR-2 will be larger and more significant than the parameter for BR-3. The hypothesis is therefore that respondents are more capable of accounting for their budget if they are given a more complete package to value.

Only respondents with WTP larger than zero are included in the OLS models. Due to the fact that the WTP distributions seem to be log-normal and the residuals in OLS models are assumed to be normally distributed, the OLS models are logarithmic in the continuous variables. A Tobit model was considered, but as stated by Greene (1993) "non-normality is an especially difficult problem in this setting". Instead of presenting Tobit models without non-normality, logit regressions are presented to explain whether WTP are equal to or larger than zero.[5] The same explanatory variables as in the OLS models were included in the logit regressions.

Table 5. Logit models to predict whether WTP > 0. T-statistics in parentheses.

	1) total package N=717	2) travel time in the package N=717	3) travel time valued separately N=717	4) traffic safety in the package N=425	5) environment in the package N=396
BR2, budget reminder and two goods in the package, dummy (0=no, 1=yes)	-0.371 (-1.6)	-0.812 (-4.3)	-0.454* (-2.8)	0.137 (0.3)	0.614 (1.6)
BR3, budget reminder and three goods in the package, dummy (0=no, 1=yes)	0.019 (0.1)	-0.258 (-1.1)		-0.327 (-0.8)	0.103 (0.3)
Age	-0.036 (-5.1)	-0.048 (-7.7)	-0.040 (-6.7)	-0.0002 (-0.01)	-0.019 (-1.9)
Household income in 1000 NOK	0.0004 (0.9)	0.001 (1.6)	0.001 (1.7)	0.0002 (0.4)	-0.001 (-1.4)
Gender, dummy (0=male, 1=female)	-0.081 (-0.3)	-0.493 (-2.7)	-0.401 (-2.2)	1.315 (2.4)	0.441 (1.3)
Number of km ('000) by car per year	0.003 (0.2)	0.014 (1.5)	0.025 (2.5)	-0.034 (-2.1)	-0.025 (-1.8)
Constant	3.248 (6.5)	2.631 (6.7)	1.944 (5.1)	2.763 (3.6)	2.814 (4.3)

* This parameter is only for a budget reminder, and does not depend on the number of goods in the package. (The separate valuation of travel time was made before the valuation of the package.)

Table 5 shows that the budget reminder is a significant explanatory variable only in the logit models for travel time valued in the package and travel time valued separately. In these two models we also observe that males in general and people with larger driving distance are more likely to have a WTP larger than

[5] Tobit models (without non-normality) give significant parameters for BR2 in models 1, 2 and 3. However, Tobit model 1 hides the fact that BR2 is significant (at 0.05 level) in the OLS models and not in the logit models, and Tobit models 2 and 3 hide the fact that BR2 is significant in the logit models and not in the OLS models. In addition, a Tobit model can hide the fact that respondents replying that their WTP is zero may do this as a protest against, for example, the context or other parts of the questionnaire. The different signs of the parameters for driving distance in the logit model and the OLS model both for traffic safety (model 4) and environment (model 5) illustrate this point.

zero. These results are the opposite of those in the models for traffic safety and environment, and may be a result of protest answers.

The models in Table 6 show that only BR-2 in the model of WTP for the total package is significantly different from zero. These models therefore confirm the above hypothesis. It is worth noting that in all OLS models the driving distance per year is an important explanatory variable. This may be interpreted as an element of use-value in the WTPs for the goods valued here. Both reduced travel time and traffic safety are goods that result in higher utility with increased road use. The importance of driving distance in the model for environmental improvement may be interpreted as an element of "the polluter pays principle", but it may also be a result of use-value because the drivers are the first to obtain benefit from reductions in environmental problems caused by road traffic. Household income is a significant explanatory variable in all OLS models except in the model of WTP for travel time valued separately. This may be a result of the importance of driving distance (i.e. use-value) for the valuation of travel time and that the valuation is made in a context where travel time is the only good available "for purchase". The fact that reduced travel time is a personal benefit more than a benefit for the household is confirmed by tests that show that personal income is a better, but still not significant, explanatory variable than household income in this model. The parameters for age show different signs in the various OLS models while the parameters for gender are negative in all these models, i.e. women have a lower WTP for these goods than men.

Table 6. OLS models that predict respondent WTP. T-statistics in parentheses.

	1) total package N=599	2) travel time in the package N=436	3) travel time valued separately N=424	4) traffic safety in the package N=391	5) environment in the package N=327
BR2, budget reminder and two goods in the package, dummy (0=no, 1=yes)	-0.331 (-3.1)	-0.083 (-0.6)	-0.025* (-0.2)	0.131 (0.8)	-0.054 (-0.3)
BR3, budget reminder and three goods in the package, dummy (0=no, 1=yes)	0.139 (1.1)	-0.056 (-0.4)		-0.110 (-0.7)	-0.162 (-1.1)
Age (ln = natural log)	-0.431 (-2.8)	-0.332 (-1.6)	-0.340 (-1.4)	0.230 (1.1)	0.255 (1.2)
Household income in 1000 NOK (ln = natural log)	0.311 (4.7)	0.244 (3.0)	0.084 (0.9)	0.248 (2.9)	0.325 (3.5)
Gender, dummy (0=male, 1=female)	-0.170 (-1.6)	-0.196 (-1.5)	-0.124 (-0.8)	-0.150 (-1.1)	-0.176 (-1.2)
Number of km ('000) by car per year (ln = natural log)	0.457 (7.1)	0.650 (7.7)	0.799 (8.0)	0.486 (5.2)	0.213 (2.5)
Constant	6.734 (10.8)	5.502 (6.9)	6.359 (6.8)	3.796 (4.7)	3.950 (4.5)
R^2 (adj.)	0.17	0.18	0.16	0.14	0.10

* This parameter is only for a budget reminder, and does not depend on the number of goods in the package. (The separate valuation of travel time was made before the valuation of the package.)

It is important to be aware that these models lack important explanatory variables such as attitudes and actions related to environmental questions, if the respondent is annoyed, bothered or has illnesses affected by air pollution from road traffic, or if a family member has experienced a car accident. In addition these models lack variables concerning the interview situation that can determine "interviewer bias" and control questions that can determine whether the respondents have understood which goods are actually focused upon in this valuation study. The low R^2 values reported for the OLS models in Table 6, a common result from CV studies, is probably a result of this shortcoming of relevant explanatory variables.

4. Directions for further research

There remain many unanswered questions about how to present information and which substitutes (and complements) to value together in the valuation procedure, i.e. the question of which valuation context is correct. Macdonald and McKenney (1996), Hutchinson *et al.* (1995), Boxall *et al.* (1996) and Adamowicz *et al.* (1997) also mention the importance of correct choice set or context in valuation studies as a subject for further research. Brown *et al.* (1995), Macdonald and McKenney (1996), Hoevenagel and van der Linden (1993) and Whitehead and Blomquist (1999) all share the view that issues regarding how to present information about substitutes must be pursued if the CV method is to become a viable tool for estimating the economic value of public goods. By contrast, Neill (1995) concludes that something beyond a mere description of budgetary substitutes may be required if interaction effects are to be reflected in CV values. Another important conclusion by Neill (op.cit.) is that along a design spectrum ranging from the simple description of potential substitutes to the required valuation of these substitutes, there are any number of design alternatives. This is to say that substitution effects may be elicited with a design that does not require the full-blown valuation of all goods (see e.g. Jones 1997 for examples related to transport).

In addition to a context and information debate it is debatable whether the valuation of complex non-market goods (e.g. environmental impacts) should be undertaken by the CV method, or if other methods can better handle simultaneous valuation of the elements of a multiple impact policy. As possibly superior methods to CV, Conjoint Analysis or Stated Choice studies have been proposed and/or undertaken in many recent studies (Sælensminde and Hammer 1993,1994, McFadden 1994, Schkade and Payne 1994, Adamowicz 1995, Kahn 1995, Boxall *et al.* 1996, Roe *et al.* 1996, Adamowicz *et al.* 1998, Foster and Mourato 1998, Hanley *et al.* 1998 and Sælensminde 1999). Among other alternative methods, Decision Analysis or Multi Attribute Utility Theory (MAUT) are proposed by Gregory *et al.* (1993), Schkade and Payne (1994), Wenstøp *et al.* (1994) and Baron and Greene (1996), and Cognitive Questionnaire Design methods are proposed by Hutchinson *et al.* (1995).

It is possible that Conjoint Analysis/Stated Choice or other methods are preferable to the CV method, but if substitutes and complements are to be valued in the same choice set/context to reduce the embedding problem, the design of the study and the analysis will become more complex. Further, it is an open question if the respondents are capable of handling many and complex goods consistently. According to Sælensminde (1998) only 33 percent of the car travellers included in the current study chose consistently in the Stated Choice part of the value of time study.

The evidence suggesting that in most cases the conventional benefit-cost procedure overestimates the total value of a package of goods stands in contrast to Randall and Hoehn's (1996) results based on multi-stage budgeting. They show that incomplete multi-stage budgeting depresses the value of packages of prospects, and of components placed late in a sequence, more than it depresses independently estimated component values. One conclusion is that multi-stage budgeting is more important when one wishes to value large packages that encompass a larger part of the respondent's total budget. Multi-stage budgeting should therefore be included as an aspect in further research on interaction effects.

Bateman *et al.* (1997) conclude that non-additivity might be a common property of all goods, and that a revision of consumer theory might be a task for further research. In contrast, Smith (1993) concludes that it is important to move from a literature consisting of diverse valuation case studies to a more systematic set of benefit measures capable of being consistently aggregated or disaggregated. Smith further states that a long-term goal might be a methodology for valuing and quantifying non-market goods that is comparable to what is to be found in the theory and implementation of price and quantity indexes. The results from the theoretical and empirical studies of interaction effects reported in the current study, together with studies that confirm similarities between RP and SP valuations done in similar contexts, indicate that Smith's long-term goal may be a more preferable direction for future research than Bateman *et al.*'s suggestions.

5. Concluding remarks

The results in this paper show that a separate valuation procedure overstates the valuation of travel time, traffic safety and environmental impacts compared to a simultaneous valuation of these three non-market goods by an average of 47, 27 and 44 percent, respectively. A simultaneous valuation of travel time, traffic safety and environmental impacts is similar to the context used in CBAs and is consequently denoted "CBA context". These results demonstrate what is formalised in theoretical contributions and what to most economists is basic economic intuition: the value of a public good is dependent upon the context in which it is provided.

Another result is that a reminder of substitutes and budget constraints significantly reduces the value of a two-goods package but had no influence on the valuation of a three-goods package (the "CBA context"). This result suggests that such a reminder seems less necessary if the valuation procedure is more holistic. A comparison with the valuations of travel time and environmental impacts that are used in CBAs for road projects in Norway show that these are, respectively, 224 and 656 percent larger than the valuations from the current study that uses a

"CBA context". The conclusion from Hoehn and Randall (1989) that "too many proposals pass the benefit cost test" may therefore also be valid for CBAs done for Norwegian road projects. It is also shown that the relative valuations of the non-market goods included in CBAs are very different from the relative valuations of these goods found in "CBA context". The values used in CBAs seem to favour projects that improve travel time and/or environment problems at the expense of projects that improve traffic safety.

This study suggests that the correct valuation context may be that which is similar to the context in which the valuation results subsequently will be used. If the case is CBAs for road projects and one wishes to include non-market goods such as travel time, traffic safety, noise and air pollution, these non-market goods should be valued simultaneously. Only in that way can respondents pay sufficient attention to interaction effects between them.

This study does not suggest that the CV method is preferable to other Stated Preference methods. Elicitation methods that more explicitly value non-market goods relative to each other (like e.g. Stated Choice) may be better in measuring the effects of substitution and complementarity. The main focus in the current study has been on the context issue, and the study design has not been ideal for measuring the interaction effects between travel time, traffic safety and environmental impacts related to road traffic. The results should therefore only be interpreted as indications of the magnitudes of such effects.

References

Adamowicz, W. L. (1995) Alternative valuation techniques: A comparison and movement towards a synthesis. In Willis, K. and J. Corkindale, eds. (1995) *Environmental valuation: New perspectives*. Oxford: CAB International.

Adamowicz, W. L., J. Swait, P. Boxall, J. Louviere and M. Williams (1997) Perceptions versus Objective Measures of Environmental Quality in Combined Revealed and Stated Preference Models of Environmental Valuation, *Journal of Environmental Economics and Environment*, 32: 65-84.

Adamowicz, W. L., P. Boxall, M. Williams and J. Louviere (1998) Stated Preference approaches to measuring passive use values: Choice Experiments and Contingent Valuation. *American Journal of Agricultural Economics,* 80(1): 64-75.

Adamowicz, W. L., V. Bhardwaj and B. Macnab (1993) Experiments on the Difference between Willingness to Pay and Willingness to Accept, *Land Economics* 69(4): 416-27.

Arrow, K., R. Solow, P. R. Portney, E. E. Leamer, R. Radner and H. Schuman (1993) Report of the NOAA Panel on Contingent Valuation, Federal Register, January 15, vol. 58, no. 10, pp. 4601-4614.

Baron, J. and J. Greene (1996) Determinants of Insensitivity to Quantity in Valuation of Public Goods: Contribution, Warm Glow, Budget Constraints, Availability, and Prominence, *Journal of Experimental Psychology: Applied,* 2(2): 107-125.

Barro, S. C., M. J. Manfredo, T. C. Brown and G. L. Peterson (1996) Examination of the Predictive Validity of CVM Using an Attitude-Behavior Framework, *Society and Natural Resources*, 9: 111-124, 1996.

Bateman, I., A. Munro, B. Rhodes, C. Starmer and R. Sugden (1997) Does Part-Whole Bias Exist? An Experimental Investigation, *The Economic Journal*, 107 (March): 322-332.

Boxall, P. C., W. L. Adamowicz, J. Swait, M. Williams and J. Louviere (1996) A comparison of stated preference methods for environmental valuation, *Ecological Economics*, 18 (3): 243-253.

Boyle, K. J., F. R. Johnson, D. W. McCollum, W. H. Desvousges, R. W. Dunford, S. P. Hudson (1996) Valuing Public Goods: Discrete versus Continuous Contingent-Valuation Responses, *Land Economics*, 72 (3), pp 381-396

Boyle, K. J., H. F. MacDonald, H. Cheng and D. W. McCollum (1998) Bid design and yea saying in single-bounded, dichotomous-choice questions. *Land Economics* 74(1): 49-64.

Brown, T. C., S. C. Barro, M. J. Manfredo and G. L. Peterson (1995): Does Better Information About the Good Avoid the Embedding Effect? *Journal of Environmental Management,* 44: 1-10.

Carson, R. T. and R. C. Mitchell (1995) Sequencing and Nesting in Contingent Valuation Surveys, *Journal of Environmental Economics and Management*, 28: 155-173.

Carson, R. T., N. E. Flores, K. M. Martin and J. L. Wright (1996a) Contingent Valuation and Revealed Preference Methodologies: Comparing the Estimates for Quasi-Public Goods, *Land Economics*, 72(1): 80-99.

Carson, R. T., N. E. Flores and N. F. Meade (1996b) Contingent Valuation: Controversies and Evidence, Discussion Paper 96-36, Department of Economics, University of California, San Diego, November 1996.

Cummings, R. G., P. T. Ganderton and T. McGockin (1994) Substitution Effects in CVM Values. *American Journal of Agricultural Economics* 76: 205-214.

Dubourg, W. R., M. W. Jones-Lee and G. Loomes (1997) Imprecise Preferences and Survey Design in Contingent Valuation. *Economica* 64: 681-702.

Elvik, R. (1993) Økonomisk verdsetting av velferdstap ved trafikkulykker – Dokumentasjonsrapport. (Economic Valuation of Reduced Risk of Traffic Accidents.) Report no. 203/1993. Institute of Transport Economics, Oslo. (In Norwegian, with English summary.)

Foster, V. and S. Mourato (1998) Elicitation Format and Part-Whole Bias: Do Dichotomous Choice Contingent Valuation Surveys and Stated Preference Experiments Give the Same Results? Paper Presented at the First World Congress on Environmental and Resource Economics, Venice, June 25-27, 1998.

Frykblom, P. (1997) Hypothetical question modes and real Willingness to Pay, *Journal of Environmental Economics and Management,* 34(3): 275-287.

Green, D., K. E. Jacowitz, D. Kahneman and D. McFadden (1998) Referendum contingent valuation, anchoring, and willingness to pay for public goods. *Resource and Energy Economics* 20(2): 85-116.

Greene, W. H. (1993) *Econometric analysis*, Prentice-Hall International, 1993.

Gregory, R., S. Lichtenstein and P. Slovic (1993) Valuing Environmental Resources: A Constructive Approach, *Journal of Risk and Uncertainty*, 7:177-197, 1993.

Halvorsen, B. and K. Sælensminde (1998) Differences between Willingness-to-Pay estimates from Open-Ended and Discrete Choice Contingent Valuation Method: The effect of heteroscedasticity. *Land Economics* 74(2): 262-282.

Hanemann, W. M. (1991) Willingness to Pay and Willingness to Accept: How much can they differ? *American Economic Review,* 81: 636-647.

Hanley, N., R. E. Wright and V. Adamowicz (1998) Using Choice Experiments to value the environment. *Environmental end Resource Economics,* 11(3-4): 413-428.

Hoehn, J. P. (1991) Valuing the multidimensional impacts of environmental policy: Theory and methods. *American Journal of Agricultural Economics,* 73(2): 289-299.

Hoehn, J. P. and A. Randall (1989) Too many proposals pass the Benefit Cost test. *American Economic Review* 79(3): 544-551.

Hoehn, J. P. and J. B. Loomis (1993) Substitution Effects in the Valuation of Multiple Environmental Programs, *Journal of Environmental Economics and Management,* 25: 56-75.

Hoevenagel, R. and J. W. van der Linden (1993) Effects of different descriptions of the ecological good on willingness to pay values, *Ecological Economics*, 7: 223-238.

Hutchinson, W. G., S. M. Chilton and J. Davis (1995) Measuring non-use value of environmental goods using the contingent valuation method: Problems of information and cognitive questionnaire design methods, *Journal of Agricultural Economics*, 46 (1): 97-112.

Johansen, K. W. (1994) Verdsetting av spart reisetid – gjennomgang av kjørekostnadshåndbokas enhetspriser for reisetid på veg. Arbeidsdokument TP/0679/94, Transportøkonomisk institutt, Oslo.

Jones, P. (1997) Addressing the "packaging" problems in Stated Preference studies. Paper presented at the 25th European Transport Forum Annual Meeting, 1-5 September 1997.

Kahn, J. R. (1995) Square pegs and round holes: Can the economic paradigm be used to value the wilderness? *Growth and Change* 26: 591-610.

Kahneman, D. and J. L. Knetsch (1992) Valuing Public Goods: The Purchase of Moral Satisfaction, *Journal of Environmental Economics and Management*, 22, 57-70.

Kotchen, M. J. and S. D. Reiling (1999) Do reminders of substitutes and budget constraints influence contingent valuation estimates? Another comment. *Land Economics*, 75 (3): 478-482.

Layard, R. and Glaister, S. (1994) *Cost-Benefit Analysis*, Cambridge University Press.

Loomis, J., A. Gonzales-Caban and R. Gregory (1994) Do reminders of substitutes and budget constraints influence contingent valuation estimates? *Land Economics* 70(4): 499-506.

Loomis, J., A. Gonzales-Caban and R. Gregory (1995) Do reminders of substitutes and budget constraints influence Contingent Valuation estimates? Reply, *Land Economics*, 71(4): 544-45.

Loomis, J., M. Lockwood and T. DeLacy (1993) Some Empirical Evidence on Embedding Effects in Contingent Valuation of Forest Protection, *Journal of Environmental Economics and Management*, 24: 45-55.

Loomis, J., T. Brown, B. Lucero and G. Peterson (1996) Improving validity experiments of Contingent Valuation methods: Results of efforts to reduce the disparity of hypothetical and actual willingness to Pay, *Land Economics* 72(4): 450-61.

Macdonald, H. and D. McKenney (1996) Varying levels of information and the embedding problem in contingent valuation: the case of Canadian wilderness, *Canadian Journal of Forest Research*, 26 (7): 1295-1303.

Madden, P. (1991) A generalization of Hicksian-q substitutes and complements eith application to demand rationing. *Econometrica* 59(5): 1497-1508.

Magnussen, K. (1992) Valuation of reduced water pollution using the Contingent Valuation Method- testing for mental accounts and amenity misspecification. Chap. 10 in Navrud, S. (ed.) (1992) *Pricing the European Environment*, Scandinavian University Press, 1992

McFadden, D. (1994) Contingent Valuation and Social Choice, *American Journal of Agricultural Economics*, 76 (November): 689-708.

Mitchell, R. C. and R. T. Carson (1989) *Using Surveys to Value Public Goods: The Contingent Valuation Method*, Resources for the Future, The John Hopkins University Press, 1989.

Neill, H. R. (1995) The Context for Substitutes in CVM Studies: Some Empirical Observations, *Journal of Environmental Economics and Management*, 29: 393-397.

Neill, H. R., R. G. Cummings, P. T. Ganderton, G. W. Harrison and T. McGuckin (1994) Hypothetical Surveys and Real Economic Commitments, *Land Economics* 70(2): 145-54.

Nelson, P. S. and J. G. Towriss (1999) The monetary valuation of the environmental of road transport: A Stated Preference approach. Paper presented at European Transport Conference, PTRC, September 1999, Cambridge.

Noether, G. E. (1967) *Elements of Nonparametric Statistics.* The SIAM series of applied mathematics. John Wiley & Sons, Inc. New York 1967.

Penin, R. and P. Riera (1999) Integrated computation of time, morbidity and mortality values from Revealed Preference data. Paper presented at European Transport Conference, PTRC, September 1999, Cambridge.

Poe, G. L., M. P. Welsh and P. A. Champ (1997) Measuring the difference in mean willingness to pay when dichotomous choice contingent valuation responses are not independent, *Land Economics*, 73(2): 255-67.

PRA (1995) Håndbok 140. Konsekvensanalyser. Del IIb. Metodikk for beregning av prissatte konsekvenser – Brukerveiledning for EFFEKT 5, Statens vegvesen, Vegdirektoratet, Oslo.

Ramjerdi, F., L. Rand, I. A. F. Sætermo and K. Sælensminde (1997) The Norwegian value of time study, Report 379/1997, Institute of Transport Economics, Oslo.

Randall, A. (1997) The NOAA Panel Report: A Beginning or the End of an Era? *American Journal of Agricultural Economics*, 79(5): 1489-1494.

Randall, A. (1998) Beyond the crucial experiment: mapping the performance characteristics of Contingent Valuation, *Resource and Energy Economics*, 20: 197-206.

Randall, A. and J. P. Hoehn (1996) Embedding in Market Demand Systems, *Journal of Environmental Economics and Management*, 30: 369-380.

Randall, A., J. P. Hoehn and G. S. Tolley (1981) The Structure of Contingent Markets: Some Experimental Results. Paper presented at the Annual Meeting of the American Economic Association, Washington, D.C., December 1981.

Roe, B., K. J. Boyle and M. F. Teisl (1996) Using Conjoint Analysis to Derive Estimates of Compensating Variation, *Journal of Environmental Economics and Management*, 31(2): 145-159.

Samples, K. C. and J. R. Hollyer (1990) Contingent Valuation of Wildlife Resources in the Presence of Substitutes and Complements, in *Economic Valuation of Natural Resources: Issues, Theory and Applications*, edited by R. L. Johnson and G. V. Johnson. Boulder, CO: Westview Press, 1990.

Schkade, D. A. and J. W. Payne (1994) How People Respond to Contingent Valuation Questions: A Verbal Protocol Analysis of Willingness to Pay for an Environmental Regulation. *Journal of Environmental Economics and Management* 26: 88-109.

Smith, V. K. (1993) Nonmarket valuation of environmental resources: An interpretive appraisal, *Land Economics*, 69(1): 1-26.

Sælensminde, K. (1995) Verdsetting av reisetid. Resultater fra andre runde av hovedundersøkelsen, september-oktober 1995. (The value of travel time. Results from the second wave in the main study, september-october 1995.) Working paper TP/0931/1995, Institute of Transport Economics, Oslo. (In Norwegian.)

Sælensminde, K. (1998) The impact of inconsistencies on the valuation of travel time in Stated Choice studies. Paper presented at the first World Congress of Environmental and Resource Economists in Venice, June 25-27, 1998.

Sælensminde, K. (1999) Stated Choice valuation of urban traffic air pollution and noise. *Transportation Research D,* 4: 13-27.

Sælensminde, K. and F. Hammer (1993) Samvalganalyse som metode for verdsetting av miljøgoder. (Conjoint Analysis as a method of assessing environmental benefits.) Report no. 184/1993, Institute of Transport Economics, Oslo. (In Norwegian, with English summary.)

Sælensminde, K. and F. Hammer (1994) Verdsetting av miljøgoder ved bruk av samvalganalyse. (Assessing environmental benefits by means of Conjoint Analysis.) Report no. 251/1994, Institute of Transport Economics, Oslo. (In Norwegian, with English summary.)

Wardman, M. (1988) A comparison of Revealed Preference and Stated Preference Models of Travel Behaviour, *Journal of Transport Economics and Policy*, January 1988: 71-91.

Wardman, M. (1998) The Value of Travel Time. A Review of British Evidence, *Journal of Transport Economics and Policy*, 32(3): 285-316.

Welsh, M. P. and G. L. Poe (1998) Elicitation Effects in Contingent Valuation: Comparisons to a Multiple Bounded Discrete Choice Approach, *Journal of Environmental Economics and Management,* 36: 170-185.

Wenstøp, F., A. J. Carlsen, O. Bergland and P. Magnus (1994) Valuation of Environmental Goods with Expert Panels, Research Report 1994/1, BI Research Center, Oslo.

Whitehead, J. C. and G. C. Blomquist (1991) A Link Between Behavior, Information and Existence Value, *Leisure Sciences,* 13: 97-109.

Whitehead, J. C. and G. C. Blomquist (1995) Do reminders of substitutes and budget constraints influence Contingent Valuation estimates? Comment, *Land Economics*, 71(4): 541-43.

Whitehead, J. C. and G. C. Blomquist (1999) Do reminders of substitutes and budget constraints influence contingent valuation estimates? Reply to another comment. *Land Economics*, 75 (3): 483-484.

Yoo, K. E. and N. Ashford (1998) An application of a comparative analysis of Revealed Preference and Stated Preference methods to the air transport choice problem. *Transport Planning and Technology*, 21: 287-307.

Appendix. Selected text and questions from the questionnaire

THE DATA WAS COLLECTED DIRECTLY BY LAPTOP COMPUTERS SO THE X-TERMS WERE EASILY AND QUICKLY REPLACED BY A SPECIFIC AMOUNT OF MONEY AND THE Y-TERMS WERE REPLACED BY A SPECIFIC AMOUNT OF TIME FOR DIFFERENT RESPONDENTS.

Reminder of substitutes and budget constraints

ABOUT HALF OF THE RESPONDENTS WERE GIVEN THE FOLLOWING BUDGET REMINDER:

When you decide how much you are willing to pay for the transport improvements offered, you should remember that there may be other transport improvements, and also other goods and services than transport improvements, that you also would benefit spending your money on.

Questions for valuation of reduced travel time on a specific car journey

The car journey you described took Y minutes and the cost was X NOK.
Imagine that this journey can be done in (Y*0.75) minutes.

Question 0:
What is the maximum amount you are willing to pay for the journey with reduced travel time? (I.e. at what price are these two journeys of equal value to you?)
Journey 1: Y minutes and X NOK.
Journey 2: (Y*0.75) minutes and ? NOK.

Questions for valuation of reduced travel time on all car journeys

Question I:
How many kilometres do you drive each year?

Question II:
Please give an estimate of how many kilometres of total driving distance per year result from local journeys of fewer than 50 kilometres?

You stated earlier that you are willing to pay (ANSWER Q. 0) NOK to reduce the travel time by (Y*0.25) minutes (i. e. approximately XX NOK/hour) on your journey from A to B.

If you pay this amount per hour to reduce the travel time by 25 % on all your LOCAL car journeys, that would add up to XXX NOK per year for a reduction of YY hours. (Assumption: Average speed of 40 km/h on all your local car journeys. (ANSWER Q. II) km per year.)

Question III:
The car journey from A to B was perhaps not representative for your local car journeys. If you had the opportunity, would you pay:
1) an equal amount (i. e. XXX NOK per year)
2) less than XXX NOK per year, or
3) more than XXX NOK per year
to reduce the travel time by 25 % on all your LOCAL car journeys? *1 / 2 / 3*

IF ANSWER 2) OR 3) IN QUESTION III:
Question IV:
How much are you willing to pay per year to reduce the travel time by 25 % on all your LOCAL car journeys?

Question V:
How much are you willing to pay per year to reduce the travel time by 25 % (a total of YYY hours per year) on all your LONGER (more than 50 km) car journeys?
(Assumption: Average speed of 60 km/h on all your longer car journeys. (ANSWER Q. I - ANSWER Q. II) km per year.)

Questions for valuation of a package including reduced travel time, increased traffic safety and a reduction in environmental problems caused by road traffic

THE RESPONDENTS WERE DIVIDED INTO THREE GROUPS WHICH RECEIVED PACKAGES WITH DIFFERENT CONTENT. GROUP 1 RECEIVED A PACKAGE OF REDUCED TRAVEL TIME AND INCREASED TRAFFIC SAFETY. GROUP 2 RECEIVED A PACKAGE OF REDUCED TRAVEL TIME AND REDUCTION IN ENVIRONMENTAL PROBLEMS. GROUP 3 RECEIVED REDUCED TRAVEL TIME, INCREASED TRAFFIC SAFETY AND REDUCTION IN ENVIRONMENTAL PROBLEMS IN THEIR PACKAGE.

TO CONSERVE SPACE, ONLY THE TEXT GIVEN TO GROUP 3 IS SHOWN HERE. THE TEXTS GIVEN TO GROUP 1 AND 2 CAN BE FOUND BY DELETING THE PARTS RELATING TO ENVIRONMENTAL PROBLEMS AND TRAFFIC SAFETY, RESPECTIVELY.

When you only had the opportunity to pay for a reduction in travel time, you were willing to pay (ANSWER Q.IV + ANSWER Q.V) NOK per year. (This amount included both your local and longer car journeys.)

Now you will also have the opportunity to pay for increased traffic safety and for a reduction in the environmental problems connected to road traffic.

Traffic safety - some background information:
(Source: Statistics Norway.)

Because of traffic accidents in Norway approximately 350 people are killed and 12 000 injured each year. (Average for the last 10 years.)

Traffic and environment - some background information:
(Source: Norwegian Pollution Control Authority (NPCA).)

NOX and dust from road traffic are the major sources for local air pollution in Norway. Approximately 700 000 people live in areas with air so polluted that it may endanger their health.

Approximately 1 million people are exposed to noise from road traffic of more than 55 dBA, which is the limit set by NPCA. 260 000 people are highly annoyed by noise from road traffic in their homes.

«A package of efforts»:
Now we want you to think over how much you are willing to pay for «a package of efforts» that simultaneously gives you reduced travel time, increased traffic safety and better environment.

Assumptions:
Reduced travel time and increased traffic safety are accomplished by efforts connected to the shape of the road and new road projects.

New fuels replacing present day fuels will reduce environmental problems.
No restrictions are made on people's car-use to achieve the sketched improvements.

Imagine that you AT THE SAME TIME can achieve:
1) 25 % reduction in time.
2) 50 % reduction in the risk for traffic accidents. (The number of killed and injured people is reduced by 175 and 6000, respectively, each year.)
3) 50 % reduction in the environmental problems caused by road traffic. (The number of people exposed to excessive concentrations of NOX and dust is reduced by 350 000. The number of people exposed to noise levels over the critical limit is reduced by 500 000.)

Question VI:
What is the maximum amount of money you would be willing to pay each year to achieve these improvements?

Decomposition of the package

You are consequently willing to pay (ANSWER Q. VI) NOK to achieve:
1. 25 % reduced travel time.
2. 50 % reduction in the risk of traffic accidents.
3. 50 % reduction in environmental problems caused by road traffic.

Question VII:
If you, as in this case, have the opportunity to pay to achieve other transport improvements in addition to reduced travel time, how much of the total amount of (ANSWER Q. VI) NOK will you then say is related to the reduction in travel time?
(You can now change the amount of (ANSWER Q. IV + ANSWER Q. V) NOK which you were willing to pay when you had the possibility to pay to achieve reduced travel time only.)

Question VIII:
How much of the amount of (ANSWER Q. VI - ANSWER Q. VII) NOK that is left would you say is related to the achievement of a 50 % reduction in the risk for traffic accidents?

Question IX:
Is the amount of (ANSWER Q.VI - ANSWER Q.VII - ANSWER Q. VIII) NOK that is finally left what you are willing to pay per year to achieve a 50 % reduction in environmental problems caused by road traffic? (If your answer is "no", you can correct your answers to the previous questions.)

Yes / No

Sist utgitte TØI rapporter

Tittel	TØI Rapport
Transportytelser i Norge 1946-1999	487/2000
Trafikkutviklingen på Stavanger lufthavn Sola 2000-2010	486/2000
Gjesteundersøkelsen januar - april 2000	485/2000
Alternativ finansiering av kollektivtransport i by. Samfunnsøkonomiske konsekvenser av alternative finansieringspakker for Kristiansand	484/2000
Risiko i veitrafikken 1997-1998.	483/2000
Produktivitetsutviklingen i riksvegferjenæringen 1988 - 1996	482/2000
Modellverktøy for transporter i norsk utenrikshandel	480/2000
Prioriteringsverktøy for gang- og sykkeltiltak - premisser og veiledning	479/2000
Ny kollektivtransport på landsbygda	478/2000
Metoder i reisevaneforskningen. En diskusjon av metodiske problemer knyttet til reisevaneundersøkelser	477/2000
Mid-term Review of the Tanzania Road Sector Programme - Final Report	476/2000
Gjestestatistikk 1999	475/2000
Nyttekostnadsanalyse av kollektivtiltak. Eksempler	474B/2000
Nyttekostnadsanalyse av kollektivtiltak. Veileder	474A/2000
Nyttekostnadsanalyse av kollektivtiltak. Kort oversikt	474/2000

tøi

This report can be ordered from
Institute of Transport Economics
P.O.Box 6110 Etterstad • N-0602 Oslo • Tel: +47 22 57 38 00