

promoting access to White Rose research papers



Universities of Leeds, Sheffield and York
<http://eprints.whiterose.ac.uk/>

This is an author produced version of a paper published in **Environmental and Resource Economics**.

White Rose Research Online URL for this paper:

<http://eprints.whiterose.ac.uk/43606/>

Paper:

Hess, S and Rose, JM (2009) *Should reference alternatives in pivot design SC surveys be treated differently?* Environmental and Resource Economics, 42 (3). 297 - 317.

<http://dx.doi.org/10.1007/s10640-008-9244-6>

Should reference alternatives in pivot design SC surveys be treated differently?

Stephane Hess*[†] John M. Rose[‡]

November 3, 2008

Abstract

Analysts are increasingly making use of pivot style Stated Choice (SC) data in the estimation of choice models. These datasets often contain a reference alternative whose attributes remain invariant across replications for the same respondent. This paper presents evidence to suggest that respondents react differently to the attributes of these reference alternatives and those of purely hypothetical alternatives. While some such evidence exists in the existing literature, this paper goes further and details a number of different departures from a common treatment of the two types of alternatives, relating both to the observed part of utility and the error term.

1 Introduction

In comparison to other fields of study, the introduction of discrete choice models to the environmental economics literature is relatively new, with the first known application arising in the late 1980s. Almost three decades after [Luce \(1959\)](#) developed the logit model and thirteen years after [McFadden \(1974\)](#) demonstrated that the probabilities of logit models imply that the unobserved utility are distributed extreme value type I, [Bockstael et al. \(1987, 1991\)](#) employed discrete choice models to analyse recreational choice using revealed preference (RP) data. Since then, discrete choice models have been used extensively in environmental

*Institute for Transport Studies, University of Leeds, s.hess@its.leeds.ac.uk, Tel: +44 78 1082 7630

[†]Institute of Transport and Logistics Studies, The University of Sydney, stephaneh@itls.usyd.edu.au

[‡]Institute of Transport and Logistics Studies, The University of Sydney, johnr@itls.usyd.edu.au, Tel: +61 2 9351 0168, Fax: + 61 2 9351 0088

valuations particularly using stated choice (SC) data (e.g. [Adamowicz et al., 1994](#); [Layton, 2000](#); [Willis et al., 2005](#)).

As with discrete choice models, SC theory and methods are not new, even predating the use of discrete choice models. The first SC experiment is thought to have been conducted by [Thurstone \(1931\)](#) who, using a crude form of experimental design, attempted to estimate indifference curves by asking respondents to make choices between different combinations of coats, hats and shoes. However, criticism of SC methods has been around for almost as long as the methods themselves. Indeed, whilst Thurstone reported some success using this method, economists at the time derided the methodology on the basis of a lack of realism which they argued was likely to give rise to spurious results. In particular [Wallis and Friedman \(1942\)](#) stated:

“It is questionable whether a subject in so artificial an experimental situation could know what choices he would make in an economic situation; not knowing, it is almost inevitable that he would, in entire good faith, systematise his answers in such a way as to produce plausible but spurious results.”

and

“...for a satisfactory experiment it is essential that the subject give actual reactions to actual stimuli... Questionnaires or other devices based on conjectural responses to hypothetical stimuli do not satisfy this requirement. The responses are valueless because the subject cannot know how he would react.”

Such arguments against the use of SC data have not abated over time (e.g. [Camerer and Hogarth, 1999](#); [Diamond and Hausman, 1994](#); [List, 2001](#)). Yet despite these criticisms, SC techniques have progressed to become the dominant method for collecting and analysing data related to individual choice behaviour. In part, this is due to the limitations of RP data, such as restrictions on the alternatives available for modelling purposes as well as issues related to capturing data on non-chosen alternatives. The success in making SC methods more realistic may also partly account for the increased use of SC data. Studies showing that SC results may be similar to those obtained from RP models up to a scale also strengthen the case for using SC data (for a review of such literature, see [Louviere et al., 2000](#)).

Attempts to make SC choice tasks more realistic have taken many forms over the years. Making SC choice tasks more incentive-compatible (e.g. [Ding, 2007](#);

Rousseas and Hart, 1951) and individual customisation of SC choice tasks to respondent specific experiences (e.g. Rose et al., 2008; Train and Wilson, 2007) represent just two approaches researchers have taken to improve the realism of SC choice tasks. It is with the latter approach that this paper is concerned.

Within the literature, there has been a growing trend towards the promotion of SC experiments where the attributes of the alternatives are pivoted around the knowledge base of sampled respondents (for applications see for example Hensher and Greene, 2003; Hensher, 2004). The use of a respondent’s knowledge base to derive the attribute levels of the SC experiments is supported by a number of theories derived in behavioural and cognitive psychology, economics, case-based decisions theory and minimum-regret theory (cf. Starmer, 2000; Kahnemann and Tversky, 1979; Gilboa et al., 2002). In total, these theories support approaches that psychologically relate experiments to individual specific experiences and perceptions. For example, prospect theory (Kahnemann and Tversky, 1979) argues that individuals use decision heuristics when making choices and promotes the idea that the context in which a decision is made is an important determinant of the choice process, supporting the use of reference alternatives in SC choice tasks.

Typically, the relating of SC experiments to individual specific experiences can take one of three forms. Firstly, such experiments may involve the presence of a status quo alternative which is represented as a null alternative with the attributes and attribute levels of the alternative not shown as part of the experiment (cf. Campbell et al., 2008; Scarpa et al., 2007a; Willis et al., 2005) et al. A second form of these experiments involves respondents being shown alternatives with attribute levels based on their own experiences but not the exact levels as described (e.g. Hensher and Rose, 2007; Train and Wilson, 2007). A final form of these experiments involves the inclusion of one or more alternatives in the choice task being described with the exact levels representing each respondent’s recent experiences (e.g. Hensher and Greene, 2003; Hensher, 2004). It is this last representation that we wish to focus on in this research.

When faced with an SC choice task where one alternative is represented as an individual specific reference alternative, there exists the possibility that respondents may treat that alternative systematically different to other ‘hypothetical’ alternatives that they are presented with. This systematic difference may arise either as a result of an experienced reference alternative being compared with constructed, non-experienced, hypothetical alternatives, or from the fact that the reference alternative is held constant across choice tasks whereas the remaining alternatives are forced to vary by way of the experimental design. Theoretically, it is also possible for the hypothetical alternatives of SC experiments involving reference alternatives to be more highly correlated with each other than with the

reference alternative (e.g. [Train and Wilson, 2007](#)).

Numerous studies in the environmental economics literature have found empirical evidence supporting the above 'status quo' effects (see for example [Lehtonen et al., 2003](#); [Kontoleon and Yabe, 2003](#); [Mogas et al., 2006](#); [Scarpa et al., 2007a](#)), whilst [Scarpa et al. \(2005\)](#) have tested a number of different econometric model specifications using Monte Carlo Simulations and found error components models which account for the possible correlation structure across alternatives to be more robust to potential misspecification issues than other model forms.

The purpose of this paper is to examine different approaches to modelling SC experiments involving the inclusion of individual specific reference alternatives. An important reference in this context is the work of [Scarpa et al. \(2007b\)](#), who look at the inclusion of a constant for the *status quo* alternative, as well as dealing with the differences in error between this alternative and the hypothetical SC alternatives through a Nested Logit or Error Components specification.

In the present paper, we expand the investigation into potential differences in treatment across alternatives. In particular, we examine whether the parameter estimates of hypothetical choice tasks should be treated the same as those of reference alternatives. We further examine the error structures of these hypothetical versus real alternatives in SC tasks. Of particular interest are differences in willingness to pay estimates as well as the possible presence of non-linearity of preferences as attribute levels of the hypothetical alternatives in the experiment are either improved or made worse when compared to the reference alternative.

The context of the empirical data used for this study is a transportation problem. Whilst such a context is different to typical environmental data sets, the issues studied are not. Of primary concern is the marginal willingness to pay outcomes of the models, the underlying theories of which are invariant to differences between private and public goods. Moreover, such transport datasets have been used before to demonstrate these concepts within the environmental economics literature (see for example [Greene et al., 2005](#); [Hensher, 2006](#)).

The remainder of this paper is organised as follows. The next section looks at the various methodological issues arising with pivot design surveys. This is followed in Section 3 by an application illustrating some of these issues with the help of a typical pivot design dataset. Finally, Section 4 summarises the work and provides some directions for future research.

2 Methodological issues

We will base our discussion of the methodological and conceptual issues on the case where the choice set contains three alternatives, of which one is a reference

alternative corresponding to a recently chosen *real world* option. Consequently, the attributes of this alternative are kept constant across choice situations, while those of the remaining two alternatives vary across replications, with attribute levels being framed around those of the reference alternative. This is consistent with the data used in the application in Section 3 which in turn is representative of a general type of pivot design dataset.

Let $U_{i,n,t}$ give the utility of alternative i in choice situation t for respondent n . Then in our three alternative example:

$$\begin{aligned} U_{1,n,t} &= V_{1,n} + \epsilon_{1,n,t} \\ U_{2,n,t} &= V_{2,n,t} + \epsilon_{2,n,t} \\ U_{3,n,t} &= V_{3,n,t} + \epsilon_{3,n,t} \end{aligned} \tag{1}$$

where $V_{i,n,t}$ and $\epsilon_{i,n,t}$ refer to the observed and unobserved utility components respectively. The fact that the attributes of the reference alternative are kept constant across replications leads to the use of $V_{1,n}$ rather than $V_{1,n,t}$.

The discussion in this chapter looks separately at the effects that the inclusion of a reference alternative potentially has on the observed part of utility and on the unobserved part of utility. Given the focus in the recent literature (e.g. Scarpa et al., 2007b), we will begin with the latter.

2.1 Implications for unobserved part of utility

In the most basic choice modelling scenario, the error terms are distributed independently and identically (*iid*) across alternatives and observations (cf. Equation 1). Here, we will look at three potential departures in this context, namely correlation between errors for different alternatives, correlation in errors over replications and heteroscedasticity in errors.

2.1.1 Correlation between alternatives

The most common departure from an *iid* scenario is one in which the errors for some of the alternatives are correlated with each other, leading to heightened substitution between those alternatives. In the case of pivot style data, a case can clearly be made that two hypothetical alternatives are more similar to each other than a hypothetical alternative and the reference alternative, and as such could potentially have correlated error terms (cf. Scarpa et al., 2005). Such correlation can be accommodated quite easily through a nesting structure in a GEV model (McFadden, 1978) or through appropriately specified error components in a MMNL model (Walker, 2001).

2.1.2 Correlation across replications

Another point that deserves some attention is the fact that with pivot data, just as with other SC data, we have multiple observations for the same respondent. However, in pivot data, a further level of complication is introduced by the fact that one of the alternatives now stays constant across replications.

In any logit type model, the underlying extreme value terms $\varepsilon_{i,n,t}$ are distributed independently across observations. However, it is not clear whether this assumption is appropriate in the case of an alternative whose attributes do not change across observations. Indeed, some of the unobserved components of the utility for this alternative can clearly be assumed to be shared across observations. In fact, it could be argued that this should apply to the majority of the components in the unobserved part of utility. Possible exclusions include learning and fatigue effects, but these are generally accommodated through terms other than the extreme value terms.

It is not clear a priori what approach can be used to address this issue, or what the effects of not addressing it are on model estimates. If the underlying extreme value structure is to be retained¹, the correlation in the errors across alternatives needs to be represented through additional error terms. This is possible with the use of an Error Components Logit (ECL) formulation in which normally distributed random variates are added to the utility functions. However, to maintain homoscedasticity, error components with the same variance need to be added to the utility functions of all alternatives. However, with the correlation only applying for the first alternative, the formulation in Equation 1 would have to change to:

$$\begin{aligned} U_{1,n,t} &= V_{1,n} + \sigma \xi_{1,n} + \varepsilon_{1,n,t} \\ U_{2,n,t} &= V_{2,n,t} + \sigma \xi_{2,n,t} + \varepsilon_{2,n,t} \\ U_{3,n,t} &= V_{3,n,t} + \sigma \xi_{3,n,t} + \varepsilon_{3,n,t}, \end{aligned} \tag{2}$$

where $\xi_{1,n}$, $\xi_{2,n,t}$ and $\xi_{3,n,t}$ are normal variates with a mean of zero and a standard deviation of 1, where the multiplication by a common σ ensures homoscedasticity. Here, $\xi_{1,n}$ varies only across respondents, while $\xi_{2,n,t}$ and $\xi_{3,n,t}$ are also distributed within respondents. This results in the required correlation structure, but leads to significant estimation problems due to the placement of integrals at different locations in the likelihood function, as discussed by [Hess and Rose \(2007\)](#).

¹I.e., short of moving away from a model structure based on extreme value distributions.

2.1.3 Heteroscedasticity across alternatives

For two reasons, the assumption of identical error terms across alternatives may also be violated in the case of data with a reference alternative. Firstly, the variation in the attributes for some of the alternatives but not for the reference alternative may lead to a greater error for the former. Secondly, the difference between a *known* reference alternative and *unknown* SC alternatives may similarly lead to a greater error for the latter.

Such heteroscedasticity across alternatives is typically accommodated either in a heteroscedastic extreme value (HEV) model, as discussed by [Bhat \(1995\)](#) or through an appropriate ECL specification of the MMNL model, as discussed by [Walker \(2001\)](#). Both models can be difficult to estimate, and the identification issues for the ECL specification are non-trivial.

2.2 Implications for observed part of utility

In the observed part of utility, an attribute $x_{i,k}$ of alternative i is associated with a marginal utility coefficient $\beta_{i,k}$, where we have that:

$$V_{i,n,t} = \sum_{k=1}^K f_{i,k}(\beta_{i,k,n}, x_{i,n,t,k}, \Omega_{i,k}), \quad (3)$$

where $x_{i,n,t,k}$ gives the value of attribute k for alternative i as faced in choice situation t by respondent n . The function $f_{i,k}()$ interacts this attribute with the associate marginal utility coefficient $\beta_{i,k,n}$ and potentially also additional shape parameters contained in the vector $\Omega_{i,k}$.

In the formulation given in Equation 3 we made very few assumptions about the shape of the utility function. In most applications, a linear in parameters formulation is used, such that:

$$f_{i,k}(\beta_{i,k,n}, x_{i,n,t,k}, \Omega_{i,k}) = \beta_{i,k,n} x_{i,n,t,k}. \quad (4)$$

2.2.1 Alternative specific coefficients

Two further simplifications generally apply to Equation 4. The estimation of respondent specific marginal utility coefficients is not practical, such that $\beta_{i,k,n}$ is replaced by $\beta_{i,k}$ where taste heterogeneity is potentially included through interaction with socio-demographic attributes² or through a random coefficients formulation such as Mixed Multinomial Logit (MMNL) (cf. [Train, 2003](#)).

²This can be accommodated in $f_{i,k}()$.

Moving away from a notation with respondent specific coefficients does not cause any more issues here than in a case using data without a reference alternative. However, the situation is less straightforward with the second simplification. Thus far, we have assumed that all marginal utility coefficients are alternative specific. In practice however, at least some of the attributes will be shared across alternatives, and generic coefficients will be used. As such, while, in a mode choice model, separate travel time coefficients may be used for different modal alternatives, a common travel time coefficient will be used for alternatives sharing the same mode. Such a situation for example arises when using data from an unlabelled route choice experiment.

In most situations, there would be no reason to suppose that the marginal utility for a shared attribute is not identical across alternatives. However, in the case of pivot style data, this may no longer be the case. Two main reasons exist for why the marginal utility valuations might differ between the reference alternative and the hypothetical alternatives. As a first example, the fact that one alternative is a *real world* alternative while the other alternatives are hypothetical may play a role. As such, respondents may react differently to the attributes of an alternative that they have actually chosen in real life, potentially repeatedly so³. An additional reason for potential differences in the marginal utility coefficients lies in the fact that the attributes of the reference alternative stay constant across replications while those of the hypothetical alternatives vary. Here, the variation in attributes for the hypothetical alternatives may potentially lead to differences in response.

On the basis of the above, it seems important for researchers to test for the possibility of alternative specific responses, by estimating separate coefficients at least for the reference alternative and for any remaining hypothetical alternatives⁴. As such, for our three alternative example, the base situation from

³As an example, in the case of commuters, the reference alternative may be the result of experience collected over many years.

⁴There is little reason to expect differences in the marginal sensitivities across the hypothetical alternatives.

Equation 1 would be adapted as follows:

$$\begin{aligned}
U_{1,n,t} &= \sum_{k=1}^K f_{R,k}(\beta_{R,k,n}, x_{1,n,k}, \Omega_{R,k}) + \epsilon_{1,n,t} \\
U_{2,n,t} &= \sum_{k=1}^K f_{S,k}(\beta_{S,k,n}, x_{2,n,t,k}, \Omega_{S,k}) + \epsilon_{2,n,t} \\
U_{3,n,t} &= \sum_{k=1}^K f_{S,k}(\beta_{S,k,n}, x_{3,n,t,k}, \Omega_{S,k}) + \epsilon_{3,n,t},
\end{aligned} \tag{5}$$

where the functional forms as well as the marginal utility and shape coefficients are specific to a given group, either the reference alternative (R) or the two hypothetical SC alternatives (S). In practice, such an approach is not used, and coefficients relating to the same attribute are generic across alternatives. Testing the validity of this assumption is one of the objectives of this study. Here, it should also be noted that the difference in coefficients across alternatives may extend from the fixed coefficient scenario to a random coefficients case, a point addressed below. Furthermore, there is a question as to whether the same functional form can be used for the utility functions across the two types of alternatives.

2.2.2 Zero values for attributes of reference alternative

Another issue that should be kept in mind when dealing with pivot style data is the possibility that some of the attributes may take on a zero value for the reference alternative. This can for example be the case in toll road studies where the respondent was observed to choose an untolled option on the reference trip. Here, the issue is not so much one of alternative specific coefficients, but a division of the sample population into two groups, depending on whether or not a non zero value was observed for the reference trip.

In modelling terms, and working with a linear formulation, this would equate to replacing $\beta_k x_k$ by $\beta_{k,A} x_k I(A) + \beta_{k,B} x_k I(B)$, where $I(A)$ is set to 1 only if the given respondent falls into the first group, with a corresponding formulation for $I(B)$. In the present context, the two groups would correspond to respondents who did or did not face a non-zero value for a given attribute in the reference trip. The grouping becomes increasingly complex with more attributes potentially taking on zero values for the reference alternative.

In this context, the differences may also again extend to the unobserved part of utility, making it worthwhile to test for differences in scale, this time between

respondents who did or did not have a zero value for a given attribute in the reference alternative.

2.2.3 Asymmetrical preferences

A final important point is the possibility that respondents may in fact evaluate the attributes of the hypothetical alternatives relative to those of the reference alternative, rather than working with absolute values. In this context, increases and decreases (gains and losses) may be evaluated asymmetrically, as discussed in detail by Hess et al. (2007) and also applied by Hensher (2007). With this specification, any attributes shared across alternatives will disappear from the observed utility function for the reference alternative, where this will generally reduce to a constant. In the observed utility functions for the hypothetical alternatives, and working with a linear formulation, the term $\beta_k x_{j,k}$ will be replaced by $\beta_k^{inc} \max(x_{j,k} - x_{R,k}, 0) + \beta_k^{dec} \max(x_{R,k} - x_{j,k}, 0)$, where $x_{R,k}$ and $x_{j,k}$ give the value for attribute k for the reference and concerned hypothetical alternative j respectively, and where β_k^{inc} and β_k^{dec} are separate coefficients for increases and decreases in attribute x_k relative to the reference alternative.

3 Empirical example

In this section, we present the results for an empirical application testing various hypotheses set out in Section 2. We first provide an introduction to the data used in the analysis before turning our attention to the actual results.

3.1 Data

The data used in this analysis were collected in Sydney in 2004 as part of a wider study to obtain estimates of Value of Travel Time Savings (VTTS) of car drivers in the Sydney metropolitan area. For this paper, we use only data collected for respondents undertaking non-commuting trips.

As part of the initial study, a sampling strategy was employed whereby only respondents who had recently taken a trip within a particular corridor where a new toll road is proposed to be built were eligible to be surveyed. Recruitment took place using a computer aided telephone interview (CATI) employing a stratified geographical sampling frame drawn from a wide catchment area.

As part of the survey task, respondents were asked information about a recent trip that they had undertaken and which could potentially have used the proposed toll road had it been in existence. This information was then used to frame the context of the SC experiment. Based on the actual trip attribute levels

Sydney Road System

Practice Game

Make your choice given the route features presented in this table, thank you.

	Details of Your Recent Trip	Road A	Road B
Time in free-flow traffic (mins)	50	25	40
Time slowed down by other traffic (mins)	10	12	12
Travel time variability (mins)	+/- 10	+/- 12	+/- 9
Running costs	\$ 3.00	\$ 4.20	\$ 1.50
Toll costs	\$ 0.00	\$ 4.80	\$ 5.60

If you make the same trip again, which road would you choose? Current Road Road A Road B

If you could only choose between the 2 new roads, which road would you choose? Road A Road B

For the chosen A or B road, HOW MUCH EARLIER OR LATER WOULD YOU BEGIN YOUR TRIP to arrive at your destination at the same time as for the recent trip. (note 0 means leave at same time) min(s) earlier later

How would you PRIMARILY spend the time that you have saved travelling?

Stay at home Shopping Social-recreational Visiting friends/relatives
 Got to work earlier Education Personal business Other

Back Next

Figure 1: An example of a stated choice screen

reported, respondents were given 16 choice scenarios, each with three alternative routes described by time spent in free flow (FF) and slowed down time (SDT) travel conditions, travel time variability (VAR), running (petrol) costs (RC) and toll costs (TOLL). In all cases, the first alternative shown presented the respondent with the attribute levels provided as part of their recent trip as reported (the reference alternative). The remaining two alternatives represented competing hypothetical routes. As such, the reference alternative remained invariant across the 16 choice situations with only the levels of the hypothetical SC alternatives varying. Before commencing, respondents were given an example game to practice with, which was explained to them by the interviewer. An example choice situation (taken from a practice game) is shown in Figure 1.

The SC experiment was constructed using efficient experimental design methods. For a review of efficient SC design methods, see [Bliemer and Rose \(2006\)](#) or [Ferrini and Scarpa \(2007\)](#). The final sample consisted of 205 effective interviews. For modelling purposes, this equates to 3,280 choice observations. For other recent applications using this dataset, see [Hess et al. \(2007\)](#) and [Hess and Rose \(2007\)](#).

3.2 Estimation results

We will now look in turn at the effects of allowing for the various phenomena discussed above. The ordering used reflects the discussion in Section 2, with 9 separate modelling approaches used as follows:

Model 1: base model

Model 2: correlation between error terms for hypothetical SC alternatives

Model 3: correlation across replications in error term for reference alternative

Model 4: heteroscedasticity between reference and hypothetical SC alternatives

Model 5: different scale for respondents with a zero reference value for some attributes

Model 6: different marginal sensitivities across reference and hypothetical alternatives

Model 7: different structural forms for observed utility function

Model 8: separate coefficients for respondents with a zero reference value for some attributes

Model 9: asymmetrical preference formation relative to attribute values for the reference alternative

The complexity in testing for these various phenomena varies significantly. In the following section, we will illustrate the effects of these phenomena in an empirical application. There is clearly some risk of confounding between the different phenomena, as discussed later on in the paper, but the actual distinguishing between effects is a secondary step, coming after the identification of potential individual effects.

A mixture of Biogeme ([Bierlaire, 2003](#)), NLogit ([Econometric Software, 2007](#)) and purpose written Ox code ([Doornik, 2001](#)) was used for the estimation of the models.

3.2.1 Base model

During the specification search for the base model, several non-linear transformations of the attributes were attempted (mainly with the help of Box-Cox transforms), but no significant levels of non-linearity were retrieved. To this extent, a

purely linear specification of the utility function was used. Specifically, we have that:

$$\begin{aligned}
V_{R,n} &= \delta_R + \beta_{\text{FFFF}}\text{FF}_{R,n} + \beta_{\text{SDT}}\text{SDT}_{R,n} + \beta_{\text{RCRC}}\text{RC}_{R,n} \\
&+ \beta_{\text{TOLL}}\text{TOLL}_{R,n} + \beta_{\text{VAR}}\text{VAR}_{R,n} \\
V_{S_1,n,t} &= \delta_{S_1} + \beta_{\text{FFFF}}\text{FF}_{S_1,n,t} + \beta_{\text{SDT}}\text{SDT}_{S_1,n,t} + \beta_{\text{RCRC}}\text{RC}_{S_1,n,t} \\
&+ \beta_{\text{TOLL}}\text{TOLL}_{S_1,n,t} + \beta_{\text{VAR}}\text{VAR}_{S_1,n,t} \\
V_{S_2,n,t} &= \beta_{\text{FFFF}}\text{FF}_{S_2,n,t} + \beta_{\text{SDT}}\text{SDT}_{S_2,n,t} + \beta_{\text{RCRC}}\text{RC}_{S_2,n,t} \\
&+ \beta_{\text{TOLL}}\text{TOLL}_{S_2,n,t} + \beta_{\text{VAR}}\text{VAR}_{S_2,n,t}.
\end{aligned} \tag{6}$$

Here, the observed utility for the reference alternative (R) is again kept constant across replications for the same respondent, while the utilities for the two hypothetical SC alternatives (S_1 and S_2) vary over choice situations. Alternative specific constants (ASC) are included for the first two alternatives (δ_R and δ_{S_1}), and all marginal utility coefficients are generic across alternatives, with the associated attributes labelled as in Section 3.1.

The estimation results for this model are summarised in Table 1. All estimated parameters are statistically significant, including the two alternative specific constants, where these indicate, all else being equal, choice inertia as well as a reading from left to right effect. The implied trade-offs from these models show higher willingness to pay (WTP) for slowed down time than for free flow time, with the valuations being lower relative to toll than relative to running cost⁵. However, the results also show that the former of these two differences is only significant at the 91% level.

3.2.2 Model with cross-alternative correlation in error terms

Attempts were made to estimate a NL model grouping together the two hypothetical SC alternatives. However, this model collapsed back to a MNL specification after having to impose a constraint on the nesting parameter due to earlier unacceptable estimates.

3.2.3 Model with cross-replication correlation in error term for reference alternative

Our next model allows for correlation across replications in the error terms for the reference alternative, with identically but cross-sectionally distributed error terms for the remaining two alternatives ensuring homoscedasticity.

⁵Standard errors of the ratios were obtained using simulation methods taking into account the asymptotic normal distribution of estimators (cf. [Armstrong et al., 2001](#)).

Table 1: Estimation results for base model

Observations	3,280	
Parameters	7	
LL($\hat{\beta}$)	-2,395.88	
adj. $\rho^2(0)$	0.3332	
	est.	asy. <i>t</i> -rat.
δ_R	0.2860	3.60
δ_{S_1}	0.1480	2.51
β_{FF}	-0.0813	-17.87
β_{SDT}	-0.0926	-17.12
β_{RC}	-0.3640	-12.82
β_{TOLL}	-0.4430	-28.22
β_{VAR}	-0.0087	-2.50
WTP indicators (AUD/hr)	pt. est.	asy. <i>t</i> -rat.
β_{FF}/β_{RC}	13.40	11.14
β_{SDT}/β_{RC}	15.26	10.72
β_{FF}/β_{TOLL}	11.01	16.64
β_{SDT}/β_{TOLL}	12.54	15.97
Diff. between WTP indicators		asy. <i>t</i> -rat.
β_{FF}/β_{RC} vs. β_{SDT}/β_{RC}		-1.71
β_{FF}/β_{TOLL} vs. β_{SDT}/β_{TOLL}		-1.73
β_{FF}/β_{RC} vs. β_{FF}/β_{TOLL}		2.30
β_{SDT}/β_{RC} vs. β_{SDT}/β_{TOLL}		2.28

The estimation results for this model are summarised in Table 2. Compared to the base model, we obtain an increase in LL by 70.84 units, which, at the cost of just one additional parameter, is highly significant. These gains are at least in part due to the fact that this model recognises the panel nature of the data, which also leads to an upwards correction of the standard errors, with the actual WTP measures remaining largely unaffected. The estimate for σ is highly significant, and suggests a correlation of 0.26 in the error terms for the reference alternative across replications for the same respondent. In a hypothetical panel MMNL model estimated in parallel⁶, this correlation measure drops to 0.21. This would suggest that the correlation picked up in the model in Table 2 is primarily due to the fact that we have multiple observations per respondent, although it is

⁶I.e. allowing for the same correlation effect for all alternatives, not just the reference alternative.

Table 2: Estimation results for model with cross-replication correlation in error term for reference alternative

Observations	3,280	
Parameters	8	
LL($\hat{\beta}$)	-2,325.04	
adj. $\rho^2(0)$	0.3530	
	est.	asy. <i>t</i> -rat.
δ_R	0.3140	2.38
δ_{S_1}	0.1710	2.52
β_{RC}	-0.4150	-9.12
β_{FF}	-0.0961	-11.33
β_{SDT}	-0.1040	-11.38
β_{TOLL}	-0.5140	-13.37
β_{VAR}	-0.0108	-1.92
σ	0.7570	12.56
WTP indicators (AUD/hr)	pt. est.	asy. <i>t</i> -rat.
β_{FF}/β_{RC}	13.89	7.10
β_{SDT}/β_{RC}	15.04	7.54
β_{FF}/β_{TOLL}	11.22	9.94
β_{SDT}/β_{TOLL}	12.14	9.67

not possible to completely discount the possibility that some of the correlation is caused by the invariant attributes nature of the reference alternative.

3.2.4 Model with heteroscedasticity across alternatives

To allow for heteroscedasticity across alternatives, an ECL formulation of the MMNL model was used, with independently distributed Normal variates with a mean of zero included in the utility functions (cf. Walker, 2001). The initial identification search showed that the most appropriate normalisation is to set the standard deviation for the error component for the reference alternative to zero. As such, only two error components were used, one for each of the two hypothetical SC alternative. With no observable difference between the two error components, the two standard deviations were constrained to the same value.

The results for this heteroscedastic model are summarised in Table 3. With one additional parameter compared to the base model, we obtain a statistically significant improvement in LL by 3.43 units. All estimated parameters are statistically significant, and the WTP indicators differ only marginally from the base

Table 3: Estimation results for model with heteroscedasticity across alternatives

Observations	3,280	
Parameters	8	
LL($\hat{\beta}$)	-2,392.45	
adj. $\rho^2(0)$	0.3340	
	est.	asy. <i>t</i> -rat.
δ_R	0.5340	2.58
δ_{S_1}	0.2480	2.21
β_{RC}	-0.5010	-4.76
β_{FF}	-0.1130	-4.92
β_{SDT}	-0.1260	-4.93
β_{TOLL}	-0.6010	-5.35
β_{VAR}	-0.0104	-2.23
σ_R	0	-
σ_S	1.5	2.36
WTP indicators (AUD/hr)	pt. est.	asy. <i>t</i> -rat.
β_{FF}/β_{RC}	13.53	10.30
β_{SDT}/β_{RC}	15.09	10.26
β_{FF}/β_{TOLL}	11.28	15.15
β_{SDT}/β_{TOLL}	12.58	15.03

model. With the variance of the unobserved part of utility for the reference alternative being fixed to $\frac{\pi^2}{6}$, the variance for the hypothetical SC alternative, at $\frac{\pi^2}{6} + 1.5^2$, is much higher. This strongly suggests the presence of heteroscedasticity⁷.

3.2.5 Models with different scale for respondents with zero values for reference alternative

In pivot designs, the possibility exists that some of the respondents are always presented with a zero value for one of the attributes of the reference alternative. In the present context, this would for example arise in the case of respondents who did not pay a toll on their reference trip. As a final step in our discussion of model error terms, we now look at allowing for differences in scale for such respondents.

In the data used here, 1 respondent had a zero value for the free flow time

⁷These results were confirmed through the estimation of a HEV model, with detailed results available on request.

Table 4: Estimation results for model with different scale for respondents with a zero toll for reference alternative

Observations	3,280		
Parameters	8		
LL($\hat{\beta}$)	-2,386.94		
adj. $\rho^2(0)$	0.3350		
	est.	t-rat.	
δ_R	0.2590	2.74	
δ_{S_1}	0.1810	2.52	
β_{RC}	-0.4480	-9.56	
β_{FF}	-0.0995	-11.31	
β_{SDT}	-0.1130	-11.59	
β_{TOLL}	-0.5170	-16.02	
β_{VAR}	-0.0098	-2.34	
α	0.7620	-3.18 [†]	
WTP indicators (AUD/hr)			
	pt. est.	t-rat.	
β_{FF}/β_{RC}	13.33	11.20	
β_{SDT}/β_{RC}	15.13	10.95	
β_{FF}/β_{TOLL}	11.55	16.31	
β_{SDT}/β_{TOLL}	13.11	16.02	

[†]: asymptotic t -ratio for α calculated with respect to 1

attribute (i.e. a fully congested trip), while 16 respondents had a zero value for the slowed down time attribute (i.e. a congestion-free trip). For either of these two attributes, no differences were observed between the two segments with a zero or non-zero value, in the observed or unobserved part of utility.

More promising however is the fact that 72 respondents, i.e. just over a third of the sample, did not pay a toll on their reference trip. It is a major assumption to suggest that these respondents behave identically to respondents who did pay a toll on their reference trip.

Generic coefficients were used across the two groups, but the utility functions for respondents who paid a toll were multiplied by a scale parameter (α). If the estimate of this scale parameter is larger than 1, it indicates a lower relative weight for the unobserved part of utility for these respondents.

The estimation results for the resulting model are summarised in Table 4. The model compares directly to the base model in Table 1. Here, the addition of one additional parameter (α) leads to a highly significant increase in log-

likelihood by 8.94 units. The actual value of α suggests that the variance of the unobserved part of utility is larger for respondents with a non-zero toll for the reference trip than for respondents with a zero toll. A very easy explanation arises for this in that respondents with a zero toll are far more likely to choose their reference trip than are respondents who did pay such a toll. While, in the overall sample, the market share for the reference alternative is 37.16%, it is much higher for respondents with a zero reference toll, at 58.42%, but is only 25.66% for respondents with a non-zero reference toll. This suggests that the behaviour of respondents with a zero reference toll is easier to model by being more deterministic (captive to reference alternative). The actual WTP indicators from this model are comparable to those from the base model.

3.2.6 Model with differences in marginal sensitivities

We now move on to various departures from the base model in the observed part of utility. Our first set of models move away from the assumption of equal response to attributes across all alternatives by estimating alternative specific taste coefficients. Here, no differences were identified between the two hypothetical SC alternatives, so that only two sets of coefficients were estimated, one set linked to the reference alternative (subscript R) and one set linked to the two hypothetical SC alternatives (subscript S). The estimation results for this model are summarised in Table 5.

In this model, all estimated coefficients remain statistically significant, with the exception of $\beta_{\text{VAR,S}}$, the coefficient for variability in the travel time for hypothetical SC alternatives. With 5 additional parameters, the model obtains an improvement in log-likelihood (LL) by 13.59 units over the base model, where this is significant above the 99% level of confidence.

Looking at the actual differences between coefficients, we observe that the difference between the coefficients for the two types of alternatives is significant above the 95% level for the two cost coefficients, while it is significant above the 90% level for the two travel time coefficients. The pattern of differences is bi-directional. For free flow time and tolls, the sensitivity is lower for the reference alternative than for the two hypothetical SC alternatives, with the converse applying for slowed down time and running costs. Finally, no significant differences are observed between $\beta_{\text{VAR,R}}$ and $\beta_{\text{VAR,S}}$.

Some interesting differences also arise when looking at the implied WTP indicators. Here, the valuations of slowed down time remain higher than the valuations of free flow time, just as in the base model in Section 3.2.1. However, while for the hypothetical SC alternatives, the valuations relative to toll are still lower than those relative to running cost, the converse is now the case for the reference

Table 5: Estimation results for model with differences in marginal sensitivities

Observations	3,280		
Parameters	12		
LL($\hat{\beta}$)	-2,382.292		
adj. $\rho^2(0)$	0.3356		
	est.	asy. t -rat.	asy. t -rat. (Δ)
δ_R	0.3210	2.50	
δ_{S_1}	0.1510	2.54	
$\beta_{FF,R}$	-0.0734	-10.16	1.66
$\beta_{FF,S}$	-0.0833	-17.83	
$\beta_{SDT,R}$	-0.0993	-14.17	-1.75
$\beta_{SDT,S}$	-0.0915	-16.38	
$\beta_{RC,R}$	-0.4640	-8.30	-2.08
$\beta_{RC,S}$	-0.3620	-12.60	
$\beta_{TOLL,R}$	-0.3680	-12.65	2.82
$\beta_{TOLL,S}$	-0.4500	-26.81	
$\beta_{VAR,R}$	-0.0079	-2.19	0.55
$\beta_{VAR,S}$	-0.0150	-1.22	
WTP indicators (AUD/hr)	pt. est.	asy. t -rat.	
$\beta_{FF,R}/\beta_{RC,R}$	9.49	5.08	
$\beta_{SDT,R}/\beta_{RC,R}$	12.84	6.08	
$\beta_{FF,R}/\beta_{TOLL,R}$	11.97	8.06	
$\beta_{SDT,R}/\beta_{TOLL,R}$	16.19	9.64	
$\beta_{FF,S}/\beta_{RC,S}$	13.81	11.16	
$\beta_{SDT,S}/\beta_{RC,S}$	15.17	10.48	
$\beta_{FF,S}/\beta_{TOLL,S}$	11.11	16.69	
$\beta_{SDT,S}/\beta_{TOLL,S}$	12.2	15.35	
Diff. between WTP indicators	asy. t -rat.		
$\beta_{FF,R}/\beta_{RC,R}$ vs. $\beta_{FF,S}/\beta_{RC,S}$	-2.26		
$\beta_{SDT,R}/\beta_{RC,R}$ vs. $\beta_{SDT,S}/\beta_{RC,S}$	-1.12		
$\beta_{FF,R}/\beta_{TOLL,R}$ vs. $\beta_{FF,S}/\beta_{TOLL,S}$	0.68		
$\beta_{SDT,R}/\beta_{TOLL,R}$ vs. $\beta_{SDT,S}/\beta_{TOLL,S}$	2.83		

alternative. This comes as a result of the sharp drop in the sensitivity to tolls for the reference alternative.

Looking at the asymptotic t -ratios for the differences between WTP indicators, we can see that the WTP for reductions in free flow time relative to running cost is significantly higher for the hypothetical SC alternatives than for the refer-

ence alternative (the difference of AUD4.32 is significant at the 97% level), while the converse is the case when looking at the valuation of slowed down time relative to road tolls (the reference alternative valuation is AUD3.99 higher, with a confidence level of 99%). The differences in the remaining two WTP indicators are of lower absolute value, and are not significant at the usual levels of confidence.

3.2.7 Models with different structural forms for observed utility function

Various non-linear specifications of the utility functions were attempted for a model with group specific coefficients, based on the model from Section 3.2.6. However, just as for the base model in Section 3.2.1, no evidence of non-linear responses was found.

3.2.8 Separate model for respondents with zero values for reference alternative

As an analogue to the model in Section 3.2.5, we now look at differences in the observed utility functions for respondents with a zero toll for the reference alternative.

Given the differences in market shares for three alternatives in the two groups, it seems appropriate to also allow for differences in alternative specific constants. To this extent, separate models were estimated in the two groups, with results summarised in Table 6.

The results from the two separate models confirm the findings of the joint model in Section 3.2.5. The model performance in terms of the adjusted ρ^2 measure is far superior in the model for respondents with a zero reference toll than in the model for respondents with a non-zero reference toll⁸. We obtain a combined LL of $-2,380.40$ for the two models, meaning that, at the cost of 7 additional parameters compared to the base model, we obtain a statistically significant gain in LL by 15.48 units. As expected, the WTP indicators relative to toll are lower for respondents with a zero reference toll; these respondents have a higher aversion to road tolls. However, there is also a reduction in the WTP indicators relative to running costs, which suggests that respondents who chose an untolled option on their reference journey have a higher overall cost sensitivity, independent of the actual cost component.

⁸The adjusted ρ^2 measure for the combined model is 0.3355.

Table 6: Estimation results for separate models for respondents with a zero toll for reference alternative

	Zero toll for RP		Non-zero toll for RP	
Observations	1,152		2,128	
Parameters	7		7	
LL($\hat{\beta}$)	-710.57		-1669.83	
adj. $\rho^2(0)$	0.4330		0.2830	
	est.	t-rat.	est.	t-rat.
δ_R	0.1270	0.87	0.3290	3.33
δ_{S_1}	0.2440	1.93	0.1180	1.76
β_{RC}	-0.5510	-6.12	-0.3370	-11.48
β_{FF}	-0.1000	-9.63	-0.0765	-15.40
β_{SDT}	-0.1020	-8.6	-0.0894	-15.04
β_{TOLL}	-0.5480	-20	-0.4020	-16.71
β_{VAR}	-0.0127	-1.81	-0.0091	-2.27
WTP indicators (AUD/hr)	pt. est.	t-rat.	pt. est.	t-rat.
β_{FF}/β_{RC}	10.89	4.65	13.62	9.94
β_{SDT}/β_{RC}	11.11	4.87	15.92	9.56
β_{FF}/β_{TOLL}	10.95	9.93	11.42	12.24
β_{SDT}/β_{TOLL}	11.17	9.04	13.34	11.95

3.2.9 Model with asymmetrical preference formation

The final model structure we explore allows for the possibility that respondents treat the attribute levels of hypothetical SC alternatives not as absolute values, but relative to the values of the reference alternative. In this case, there is a possibility of asymmetrical preference formation, with respondents reacting differently to gains and losses relative to the attribute values for this reference trip. For an in-depth discussion of this topic, see [Hess et al. \(2007\)](#).

The utility function for the reference alternative now contains only a constant, while the utility functions for the two hypothetical SC alternatives contain two coefficients for each attribute, one associated with increases relative to the reference trip (e.g. $\beta_{FF,inc}$) and one associated with decreases (e.g. $\beta_{FF,dec}$). The estimation results for this model are summarised in [Table 7](#).

With 5 additional parameters compared to the base model, we obtain a statistically significant increase in LL by 19.42 units. Coefficients associated with reductions in attributes take on positive signs, with the converse being the case for increases. The only exception is the positive value for $\beta_{VAR,inc}$, where this is

Table 7: Estimation results for model with asymmetrical preference formation

Observations	3,280		
Parameters	12		
LL($\hat{\beta}$)	-2,376.464		
adj. $\rho^2(0)$	0.3372		
	est.	asy. <i>t</i> -rat.	asy. <i>t</i> -rat. (Δ) [†]
δ_R	0.2670	2.34	
δ_{S_1}	0.1590	2.72	
$\beta_{FF,dec}$	0.0736	10.13	1.48
$\beta_{FF,inc}$	-0.1050	-6.57	
$\beta_{SDT,dec}$	0.1180	12.92	
$\beta_{SDT,inc}$	-0.0313	-2.02	-3.99
$\beta_{RC,dec}$	0.3740	7.24	0.09
$\beta_{RC,inc}$	-0.3640	-4.88	
$\beta_{TOLL,dec}$	0.3320	10.56	3.99
$\beta_{TOLL,inc}$	-0.5060	-22.10	
$\beta_{VAR,dec}$	0.0077	2.18	-0.18
$\beta_{VAR,inc}$	0.0045	0.07	
WTP indicators (AUD/hr)	pt. est.	asy. <i>t</i> -rat.	
β_{FF}/β_{RC}	12.13	4.34	
β_{SDT}/β_{RC}	19.45	4.34	
β_{FF}/β_{TOLL}	8.73	9.61	
β_{SDT}/β_{TOLL}	13.99	11.93	
WTA indicators (min/AUD)	pt. est.	asy. <i>t</i> -rat.	
β_{FF}/β_{RC}	3.56	5.57	
β_{SDT}/β_{RC}	11.95	1.98	
β_{FF}/β_{TOLL}	3.16	5.83	
β_{SDT}/β_{TOLL}	10.61	2.02	
Diff. between WTP and WTA	asy. <i>t</i> -rat.		
β_{FF}/β_{RC}	-0.88		
β_{SDT}/β_{RC}	2.61		
β_{FF}/β_{TOLL}	-2.69		
β_{SDT}/β_{TOLL}	2.31		

[†]: sign differences taken into account in calculation of standard errors for differences

however not statistically significant. All other estimated parameters are significant at high levels of confidence.

Looking in detail at the response to increases and decreases, we observe a perfectly symmetrical response in the case of running costs, i.e. an increase in running cost by AUD1 is valued as negatively as a reduction in running cost by AUD1 is valued positively. For free flow time, the difference between increases and decreases (after taking into account the sign differences) is only significant at the 86% level but suggests that increases in free flow time are valued more negatively than reductions are valued positively.

For slowed down time and toll costs, the differences between increases and decreases are highly significant, where, by coincidence, the value of the asymptotic t -ratio is identical. For toll costs, there is clear evidence that increases are valued more negatively than decreases, which is consistent with intuition, and which is also reflected in the high aversion by respondents with a zero reference toll to move to a tolled SC alternative. However, for slowed down time, the surprising observation is made that decreases are valued far more positively than increases are valued negatively. A potential explanation for this could be in the way respondents trade off free flow time and slowed down time, in that an increase in slowed down time was in general associated with a reduction in free flow time, where the magnitude of the latter was greater. So respondents can be seen to accept increases in slowed down time in return for reductions in free flow time.

It should also be said that the asymmetries are at least in part a result of the design, where gains and losses were not presented in a symmetrical fashion. As such, very few increases in variability were presented, explaining the low significance level of $\beta_{\text{VAR,inc}}$. Furthermore, for road tolls, any reductions presented were always a 100% reduction, in which case the marginal gains of a reduction in tolls (per AUD) could indeed be expected to be lower, as observed. Independently of the reasons, the analysis presents clear evidence of asymmetries and suggests that a symmetrical specification is not appropriate.

With separate coefficients estimated for increases and decreases, two different trade-offs can be calculated, namely the willingness to pay (WTP) for reductions in travel time, and the willingness to accept (WTA) increases in travel time in return for reductions in travel cost. Here, the asymmetry for the marginal utility coefficients translates into similar asymmetry patterns for the actual trade-offs, where these are significant with the exception of the trade-offs between free flow time and running costs.

4 Summary, conclusions and directions for future research

This paper has discussed the issue of what special precautions might need to be taken when estimating models on pivot style data, where the choice sets include a reference alternative with invariant attributes across replications. The theoretical discussions in the paper identify a number of possible phenomena that may act on the observed or unobserved utility components.

The application conducted on a representative pivot style dataset collected in Sydney confirms the theoretical suspicions. As such, different advanced model specifications offer statistically significant gains in model fit when compared to the base model. In particular, the models suggest:

- the existence of correlation in the error terms across replications for the reference trip;
- the existence of differences in the variance of error terms between the reference alternative and the hypothetical SC alternatives;
- the presence of differences in the variance of unobserved utility components for respondents who did not pay a toll on the reference trip;
- the presence of differences in the marginal sensitivities between the reference alternative and the hypothetical SC alternatives;
- the presence of differences in the marginal sensitivities for respondents who did not pay a toll on the reference trip; and
- the prevalence of asymmetrical preference formation around the attribute values of the reference alternative.

A summary of all the estimated models is given in Table 8. With the exception of the model allowing for cross-alternative correlation and the model allowing for different structural forms, all other models outperform the base model. By far the biggest improvement is obtained by the model allowing for cross-replication correlation in the error for the first alternative. The remaining five models give quite similar improvements in fit over the base model. There is clearly a distinct possibility of some confounding across models, such that not all of the above factors may act to the same degree on the data. Further work is required to establish the extent of confounding, through specification of an appropriate joint model. Another area for further research is the investigation of these phenomena in a context where a reference alternative is included but the attribute levels for

Table 8: Summary of model performance

Model description	par.	LL($\hat{\beta}$)	adj. ρ^2
Linear MNL model	7	-2,395.88	0.3332
Cross-alternative correlation	collapses back to linear MNL		
Cross-replication correlation	8	-2,325.04	0.3530
Heteroscedasticity across alternatives	8	-2,392.45	0.3340
Differences in scale for respondents with zero toll	8	-2,386.94	0.3350
Differences in marginal sensitivities	12	-2,382.29	0.3356
Different structural forms	collapses back to linear MNL		
Differences in valuations for respondents with zero toll	14	-2,380.40	0.3355
Asymmetrical preference formation	12	-2,376.46	0.3372

the hypothetical alternatives are not derived from those of the current alternative. Furthermore, the implications of these differences on appraisal and forecasting deserve further attention, but the scale differences between some of the estimated models give an indication of potentially significant differences in forecasts⁹.

In summary, even though the evidence in this paper is limited to one dataset, the findings do clearly suggest that special care is required when estimating models on pivot style data. While not per se questioning the validity of pivot style approaches, analysts should, when attempting to exploit the realism benefits of such designs, also be aware of these potential methodological complications. With the increasing popularity of such data, whether for forecasting or for WTP implications, these warnings are very timely indeed. Finally, it should be noted that the issues raised in this paper potentially also play a role in non-pivot datasets that include a reference alternative, and clearly also in the case of surveys including a mixture of existing and hypothetical alternatives¹⁰.

5 Acknowledgements

The majority of this work was carried out during a visit by the first author to the Institute of Transport and Logistics Studies at the University of Sydney, funded by a Faculty of Economics and Business Visiting Scholar Grant.

⁹Notwithstanding the usual difficulties of forecasting from SC data in terms of the calibration of constants.

¹⁰We thank an anonymous reviewer for raising this issue.

References

- Adamowicz, W., Louviere, J. J., Williams, M., 1994. Combining stated and revealed preference methods for valuing environmental attributes. *Journal of Environmental Economics and Management* 26, 271–292.
- Armstrong, P., Garrido, R. A., Ortúzar, J. de D., 2001. Confidence interval to bound the value of time. *Transportation Research Part E: Logistics and Transportation Review* 37 (1), 143–161.
- Bhat, C. R., 1995. A heteroscedastic extreme value model of intercity travel mode choice. *Transportation Research Part B: Methodological* 29 (6), 471–483.
- Bierlaire, M., 2003. BIOGEME: a free package for the estimation of discrete choice models. Proceedings of the 3rd Swiss Transport Research Conference, Monte Verità, Ascona.
- Bliemer, M. C. J., Rose, J. M., 2006. Designing Stated Choice Experiments: State-of-the-art. paper presented at the 11th International Conference on Travel Behaviour Research, Kyoto, Japan.
- Bockstael, N. E., McConnell, K. E., Strand, I. E., 1991. Recreation. In: Braden, J. B., Kolstad, C. D. (Eds.), *Measuring the Demand for Environmental Quality*. North Holland, Amsterdam.
- Bockstael, N. E., Strand, I. E., Hanemann, W. M., 1987. Time and the recreational demand model. *American Journal of Agricultural Economics* 69, 293–302.
- Camerer, C. F., Hogarth, R. M., 1999. The effects of financial incentives in experiments: A review and capital-labor-production framework. *Journal of Risk and Uncertainty* 19 (1-3), 7–42.
- Campbell, D., Hutchinson, W. G., Scarpa, R., 2008. Incorporating discontinuous preferences into the analysis of discrete choice experiments. *Environmental and Resource Economics*, forthcoming.
- Diamond, P. A., Hausman, J. A., 1994. Contingent valuation: Is some number better than no number? *Journal of Economic Perspectives* 8 (4), 45–64.
- Ding, M., 2007. An incentive aligned mechanism for conjoint analysis. *Journal of Marketing Research* XLIV, 214–223.

- Doornik, J. A., 2001. *Ox: An Object-Oriented Matrix Language*. Timberlake Consultants Press, London.
- Econometric Software, 2007. *Nlogit 4.0*. Econometric Software, New York and Sydney.
- Ferrini, S., Scarpa, R., 2007. Designs with a-priori information for nonmarket valuation with choice-experiments: a monte carlo study. *Journal of Environmental Economics and Management*, 53, 342–363.
- Gilboa, I., Schmeidler, D., Wakker, P., 2002. Utility in case-based decision theory. *Journal of Economic Theory* 105, 483–502.
- Greene, W. H., Hensher, D. A., Rose, J. M., 2005. Using classical simulation based estimators to estimate individual willingness to pay values: A mixed logit case study of commuters. In: Scarpa, R., Alberini, A. (Eds.), *Applications of Simulation Methods in Environmental and Resource Economics*. Springer, Dordrecht, The Netherlands, pp. 17–34.
- Hensher, D. A., 2004. Accounting for stated choice design dimensionality in willingness to pay for travel time savings. *Journal of Transport Economics and Policy* 38 (2), 425–446.
- Hensher, D. A., 2006. How do respondents handle stated choice experiments? - attribute processing strategies under varying information load. *Journal of Applied Econometrics* 21, 861–878.
- Hensher, D. A., 2007. Joint estimation of process and outcome in choice experiments and implications for willingness to pay. *Journal of Transport Economics and Policy*, forthcoming.
- Hensher, D. A., Greene, W. H., 2003. The Mixed Logit Model: The State of Practice. *Transportation* 30 (2), 133–176.
- Hensher, D. A., Rose, J. M., 2007. Development of commuter and non-commuter mode choice models for the assessment of new public transport infrastructure projects: A case study. *Transportation Research Part A: Policy and Practice* 41 (5), 428–443.
- Hess, S., Rose, J. M., 2007. Intra-respondent taste heterogeneity in instantaneous panel surveys. paper to be presented at the 11th triennial WCTR conference, July 2007, Berkeley, CA.

- Hess, S., Rose, J. M., Hensher, D. A., 2008. Asymmetrical preference formation in willingness to pay estimates in discrete choice models. *Transportation Research Part E: Logistics and Transportation Review* 44 (5), 847–863.
- Kahnemann, D., Tversky, A., 1979. Prospect theory: an analysis of decisions under risk. *Econometrica* 47 (2), 263–291.
- Kontoleon, A., Yabe, M., 2003. Assessing the impacts of alternative 'opt-out' formats in choice experiment studies: Consumer preferences for genetically modified content and production information in food. *Journal of Agricultural Policy and Resources* 5, 1–43.
- Layton, D. F., 2000. Random coefficient models for stated preference surveys. *Journal of Environmental Economics and Management* 40 (1), 21–36.
- Lehtonen, E., Kuuluvainen, J., Pouta, E., Rekola, M., Li, C. Z., 2003. Non-market benefits of forest conservation in southern finland. *Environmental and Social Policy* 6, 195–204.
- List, J. A., 2001. Do explicit warnings eliminate the hypothetical bias in elicitation procedures? evidence from field auctions for sportscards. *American Economic Review* 91 (5), 1498–1507.
- Louviere, J. J., Hensher, D. A., Swait, J., 2000. *Stated Choice Models: Analysis and Application*. Cambridge University Press, Cambridge.
- Luce, R., 1959. *Individual choice behavior: a theoretical analysis*. J.Wiley and Sons, New York.
- McFadden, D., 1974. Conditional logit analysis of qualitative choice behaviour. In: Zarembka, P. (Ed.), *Frontiers in Econometrics*. Academic Press, New York, pp. 105–142.
- McFadden, D., 1978. Modelling the choice of residential location. In: Karlqvist, A., Lundqvist, L., Snickars, F., Weibull, J. W. (Eds.), *Spatial Interaction Theory and Planning Models*. North Holland, Amsterdam, Ch. 25, pp. 75–96.
- Mogas, J., Rierab, P., Bennett, J., 2006. A comparison of contingent valuation and choice modelling with second-order interactions. *Journal of Forest Economics* 15, 5–30.
- Rose, J. M., Bliemer, M. C. J., Hensher, D. A., Collins, A. C., 2008. Designing Efficient Stated Choice Experiments Involving Respondent Based Reference Alternatives. *Transportation Research Part B: Methodological*, forthcoming.

- Rousseas, S. W., Hart, A. G., 1951. Experimental verification of a composite indifference map. *Journal of Political Economy* 59, 288–318.
- Scarpa, R., Campbell, D., Hutchinson, W. G., 2007a. Benefit estimates for landscape improvements: sequential bayesian design and respondents' rationality in a choice experiment. *Land Economics* 83 (4), 617–634.
- Scarpa, R., Ferrini, S., Willis, K. G., 2005. Performance of error component models for status-quo effects in choice experiments. In: Scarpa, R., Alberini, A. (Eds.), *Applications of Simulation Methods in Environmental and Resource Economics*. Springer, Dordrecht.
- Scarpa, R., Willis, K. G., Acutt, M., 2007b. Valuing externalities from water supply: Status quo, choice complexity and individual random effects in panel kernel logit analysis of choice experiments. *Journal of Environmental Planning and Management* 50 (4), 449–466.
- Starmer, C., 2000. Developments in non-expected utility theory: the hunt for a descriptive theory of choice under risk. *Journal of Economic Literature* XXXVIII, 332–382.
- Thurstone, L., 1931. The indifference function. *Journal of Social Psychology* 2, 139–167.
- Train, K., 2003. *Discrete Choice Methods with Simulation*. Cambridge University Press, Cambridge, MA.
- Train, K., Wilson, W. W., 2007. Estimation on stated-preference experiments constructed from revealed-preference choices. *Transportation Research Part B: Methodological*, forthcoming.
- Walker, J., 2001. *Extended discrete choice models: Integrated framework, flexible error structures, and latent variables*. Ph.D. thesis, MIT, Cambridge, MA.
- Wallis, W. A., Friedman, M., 1942. The empirical derivation of indifference functions. In: Lange, O., McIntyre, F., Yntema, T. O. (Eds.), *Studies in Mathematical Economics and Econometrics in memory of Henry Schultz*. University of Chicago Press, Chicago, IL, pp. 175–189.
- Willis, K. G., Scarpa, R., Acutt, M., 2005. Accessing water company customer preferences and willingness to pay for service improvements: a stated choice analysis. *Water Resources Research* 41, 1–11.