

Intra-respondent taste heterogeneity in instantaneous panel surveys

Stephane Hess* John M. Rose†

July 26, 2007

Abstract

The vast majority of discrete choice modelling (DCM) applications are now estimated on Stated Preference (SP) data, including but not limited to the field of transport research. In SP data, each respondent is faced with multiple choice situations, and recognising this repeated choice nature of the data is a crucial modelling issue. With the increasing popularity of the Mixed Multinomial Logit (MMNL) model, most applications now rely exclusively on a random coefficients approach in dealing with the repeated choice nature of the data. Here in turn, the assumption is generally made that tastes vary across respondents, but not across observations for the same respondent. In this paper, we question this assumption and show that it is important to also allow for variation in tastes across replications for the same respondent.

1 Introduction

Stated preference (SP) data have proven useful in solving many transport studies related problems. For example, SP data have been used to examine the demand for a cycle-way network (e.g. [Ortúzar *et al.*, 2000](#)), to examine the benefits derived from various traffic calming measures (e.g. [Garrod *et al.*, 2002](#)), to study the influences on parking choice (e.g. [Hensher and King, 2001](#); [Hess and Polak, 2004](#)), and to establish the Value of Travel Time Savings (VTTS) of commuters and non-commuters (e.g. [Hensher, 2001a,b](#); [Hess *et al.*, 2005](#); [Axhausen *et al.*, 2007](#)).

*Institute for Transport Planning and Systems, ETH Zürich, stephane.hess@ivt.baug.ethz.ch,
Tel: +41(0)1 633 3196, Fax: +41(0)1 633 1057

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

Typically, SP experiments present sampled respondents with a number of different choice sets, each consisting of a finite set of alternatives defined on a number of attribute dimensions. Respondents are then asked to specify their preferred alternative in each choice situation. These responses can be used to estimate models of choice behaviour, which, depending on the type of experiment conducted, may allow for the estimation of the direct or cross elasticities (or marginal effects) of the alternatives as well as the marginal rates of substitution respondents are willing to make in trading between two attributes (i.e., willingness to pay measures, for example, VTTS).

Aside from dealing with issues of response quality (cf. [Louviere et al., 2000](#)), an important consideration in this context is how to represent the fact that each individual is faced with multiple choice situations in a given SP survey.

It is clearly a major assumption to treat the observations for the same individual as independent, i.e., postulating that any variations in tastes/behaviour are the same across choice situations for a given respondent as they are across respondents. In the context of long-term panel studies with Revealed Preference (RP) data, techniques such as auto-regressive parameters are typically used to account for learning and experience effects. However, in the context of SP surveys, we are dealing with an instantaneous panel, such that there is limited scope for learning and habit formation. Indeed, [Hensher and Greene \(2003, p.160\)](#) state that: “. . .correlation is not likely to be autoregressive for ‘instantaneous’ stated choices since it is not the product of a long period of accumulated experience commonly attributed to state dependence”.

Unobserved heterogeneity is often described as relating to “differences across decision-makers in the intrinsic preference for a choice alternative (preference heterogeneity) and/or in the sensitivity to characteristics of the choice alternatives (response heterogeneity)” ([Bhat and Castelar, 2002](#)). In this context, [Outwerslot and Rietveld \(1996\)](#) and [Abdel-Aty et al. \(1997\)](#) estimated models decomposing the error term of estimated models into an individual-specific effect which is distributed independently over individuals, and an observation-specific effect distributed independently over both individuals and observations. In both cases, it was found that it was the parameter estimates that were affected, and not the asymptotic standard errors when compared to more traditional discrete choice models estimated using the same data.

More recently, the increasing popularity of the Mixed Multinomial Logit (MMNL) model has meant that the majority of applications now account for the repeated choice nature of SP data with the help of a random coefficients specification that allows for a variation in tastes across respondents, but with constant tastes across replications for the same respondent (cf. [Train, 2003](#); [Hensher and Greene, 2003](#)).

Here, it should be said that there exists no theoretical argument that the preferences of individual respondents should be fixed across choice replications. Indeed, research evidence appears to suggest the contrary to be true. As such, it is known that respondents do, over the course of an SP experiment, experience both learning and cognitive burden effects (e.g. [Cooke and Mellers, 1995](#); [Brazell and Louviere, 1998](#); [Pullman *et al.*, 1999](#); [Ohler *et al.*, 2000](#); [Deshazo and Fermo, 2002](#); [Verlegh *et al.*, 2002](#); [Arentze *et al.*, 2003](#)). One way that these effects have been observed to manifest themselves is in terms of the response times of respondents completing SP choice tasks. [Haaiker *et al.* \(2000\)](#) and [Rose and Black \(2006\)](#) demonstrate that respondents often spend a far greater amount of time undertaking the first three to four choice tasks than they do for subsequent choice tasks. They show that failure to account for this may result in biased parameter estimates, due to the fact that most discrete choice models are designed to average over all choice tasks, even those that account for individual level heterogeneity.

A further argument against the assumption of within respondent preference homogeneity derives from the very nature of SP data. Unlike most RP surveys where market data is largely correlated in terms of the attribute levels experienced over time, SP surveys are designed to induce trading off by markedly changing the attribute levels experienced from one choice situation to another¹. Thus, why would a respondent's preferences not vary when faced by two very different choice situations? For example, compare a situation in which a respondent faces a 10 minute travel time relative to a travel cost of \$3 to a situation where the same respondent faces an alternative with a travel time of 20 minutes at a cost of \$1. A respondent may indeed be more or less sensitive to time or cost than other respondents, and this would support the case for treating the SP data as a panel. However, it is equally likely that the actual levels of the SP experiment may influence the degree to which preferences are likely to be more or less sensitive to the attributes of the experiment. This in turn would speak in favour of also allowing for within-respondent preference heterogeneity.

In this paper, we look in detail at the effects of the assumption of constant tastes across replications for the same individual in the case of instantaneous SP panels, and investigate whether this assumption should be relaxed?

The remainder of this paper is organised as follows. Section 2 presents the modelling approach used in the analysis in this paper. The empirical analysis conducted for this paper is described in Section 3. Finally, Section 4 presents the conclusions of the paper.

¹As an example, in RP data, a train travel time, whilst varying from day to day, will generally not vary as much as in the hypothetical choice situations constructed in SP data.

2 Methodology

In this section, we present the various model structures used in the analysis in Section 3. We start by introducing some general notation.

Let $U_{i,n,t}$ be the utility of alternative i (with $i = 1, \dots, I$) for respondent n (with $n = 1, \dots, N$) in choice situation t (with $t = 1, \dots, T_n$). Then we have that $U_{i,n,t} = V_{i,n,t} + \varepsilon_{i,n,t}$, where $V_{i,n,t}$ is the observed part of utility, and $\varepsilon_{i,n,t}$ is the usual type I extreme value term, distributed identically and independently across alternatives and choice situations. Further, let $V_{i,n,t} = \sum_{k=1}^K f_k(\beta_k, x_{k,i,n,t})$, where $x_{k,i,n,t}$ gives the value of an explanatory variable x_k of alternative i in choice situation t for respondent n . The value of this attribute acts on the utility $V_{i,n,t}$ through interaction with the taste coefficient β_k , with the functional form of this interaction given by $f_k(\cdot)$. With a linear formulation we would have that $f_k(\beta_k, x_{k,i,n,t}) = \beta_k x_{k,i,n,t}$.

Let $P_{n,t}(i | x, \beta)$ give the choice probability for alternative i in choice situation t of individual n , where x groups together all the attributes of all the alternatives in the choice set. Then, the log-likelihood (LL) function is given by:

$$LL(\beta) = \sum_{n=1}^N \sum_{t=1}^{T_n} \ln(P_{n,t}(j_{n,t} | x, \beta)), \quad (1)$$

where $j_{n,t}$ is the alternative chosen by respondent n in choice situation t , and where $\beta = \langle \beta_1, \dots, \beta_K \rangle$. In a model assuming an absence of random taste heterogeneity, such as the Multinomial Logit (MNL) model, we have that β stays identical across respondents.

In a model allowing for random taste heterogeneity, such as MMNL, the individual taste coefficients β_k are assumed to follow a certain random distribution across respondents, such that we have $\beta_{n,k} \sim g(\beta_k | \Omega_k)$, with Ω representing a set of parameters of the distribution of β_k , such as the mean and standard deviation in the case of a Normal distribution. From this, equation 1 now becomes:

$$LL(\Omega) = \sum_{n=1}^N \sum_{t=1}^{T_n} \ln(P_{n,t}(j_{n,t} | \Omega)), \quad (2)$$

which is conditional on a set of values for Ω , which groups together $\Omega_1, \dots, \Omega_K$, and where:

$$P_{n,t}(i | \Omega) = \int_{\beta_1} \dots \int_{\beta_K} P_{n,t}(i | x, \beta) g(\beta_1 | \Omega_1) \dots g(\beta_K | \Omega_K) d\beta_1 \dots d\beta_K, \quad (3)$$

or, for conciseness (and allowing for multivariate distributions²):

$$P_{n,t}(i | \Omega) = \int_{\beta} P_{n,t}(i | x, \beta) g(\beta | \Omega) d\beta. \quad (4)$$

The log-likelihood formulation in equation 2 relates to a cross-sectional formulation, making the assumption that the tastes vary across respondents as well as across observations for the same respondent. This is a direct result of the positioning of the integral. This assumption is changed by rewriting equation 2 into:

$$LL(\Omega) = \sum_{n=1}^N \ln(P_n(\Omega)), \quad (5)$$

where $P_n(\Omega)$ gives the probability of respondent n making the observed sequence of choices, and where:

$$P_n(\Omega) = \int_{\beta} \left[\prod_{t=1}^{T_n} (P_{n,t}(j_{n,t} | x, \beta)) \right] g(\beta | \Omega) d\beta. \quad (6)$$

The fact that the integration now takes place at the level of a sequence of choices, as opposed to individual choices, equates to an assumption that the tastes vary across respondents, but stay constant across observations for the same respondent.

The formulation in equation 5 and 6 is now the standard approach used when dealing with repeated choice data in discrete choice modelling. However, it has a shortcoming in that it assumes an absence of variation in tastes across choice situations for the same respondent.

Allowing for intra-personal variation in tastes in addition to inter-personal variation is conceptually straightforward, although it does increase the burden at the estimation and application stage. Here, we illustrate the concept on the basis of an example where there is only a single random taste coefficient. Extension to multivariate cases, with potential correlation between coefficients, is straightforward, and such examples are given in the applications in Section 3.

Let us assume that the sensitivity to a given attribute x_k varies randomly across respondents, with a distribution $g(\beta_{x_k})$. Let us further assume that this

²All the specifications set out here can be adapted easily to allow for correlation between individual taste coefficients, as is the case in several of the models estimated in Section 3.

distribution is Normal, with mean $\mu_{\beta_{x_k}}$ and standard deviation $\sigma_{\beta_{x_k}}$, such that $\beta_{x_k} \sim N(\mu_{\beta_{x_k}}, \sigma_{\beta_{x_k}})$. On top of this inter-personal variation, we want to allow for intra-personal variation, such that the value of the taste coefficient associated with x_k may deviate from the mean value for individual n , say β_{n,x_k} , across choice situations for that respondent. To this extent, we make use of an additional random distribution $h(\gamma_{x_k})$. The value for the combined taste coefficient β'_{x_k} in a given choice situation is then obtained as the sum of an individual-specific draw from $g(\beta_{x_k})$ and an observation-specific draw from $h(\gamma_{x_k})$. So we have with $\beta'_{x_k} = \beta_{x_k} + \gamma_{x_k}$. With the mean effects captured in the distribution of β_{x_k} , we can set the mean of the distribution of γ_{x_k} to zero. Again making use of a Normal distribution, we then have $\gamma_{x_k} \sim N(0, \sigma_{\gamma_{x_k}})$. The new log-likelihood function is then given by:

$$LL(\Omega) = \sum_{n=1}^N \ln \left[\int_{\beta_{x_k}} \left(\prod_{t=1}^{T_n} \left(\int_{\gamma_{x_k}} P_{n,t}(j_{n,t} | x, \beta'_{x_k}, \beta_f) h(\gamma_{x_k}) d\gamma_{x_k} \right) \right) g(\beta_{x_k}) d\beta_{x_k} \right]. \quad (7)$$

In this formulation, part of the integration is carried out outside the product over choice situations, relating to inter-respondent variation, with a remaining part carried out inside this product, relating to intra-respondent variation. The vector Ω now contains parameters for the distributions of β_{x_k} and γ_{x_k} , and the vector β_f contains any coefficients that are fixed across respondents.

It should be noted that this formulation is mathematically very similar to that employed by [Bhat and Castelar \(2002\)](#). However, the context is entirely different, where the work by [Bhat and Castelar \(2002\)](#) looks at the joint estimation on RP and SP data, and where the within respondent integration is used to account for scale differences.

Complications arise in the estimation of the above model structure, due to the positioning of the two separate layers of integration. In the absence of a closed-form solution, the term inside Equation 7 needs to be approximated through simulation.

The contribution by respondent n to the log-likelihood is given by:

$$LL_n(\Omega) = \ln \left[\int_{\beta_{x_k}} \left(\prod_{t=1}^{T_n} \left(\int_{\gamma_{x_k}} P_{n,t}(j_{n,t} | x, \beta'_{x_k}, \beta_f) h(\gamma_{x_k}) d\gamma_{x_k} \right) \right) g(\beta_{x_k}) d\beta_{x_k} \right]. \quad (8)$$

In simulating this term, we make use of R draws from $g(\beta_{x_k})$ and RT_n draws from (γ_{x_k}) , such that a separate set of R draws from (γ_{x_k}) is used for each observation for respondent n . With this in mind, the term in Equation 8 is approximated by:

$$SLL_n(\Omega) = \ln \left[\frac{1}{R} \sum_{r=1}^R \left(\prod_{t=1}^{T_n} \left(\frac{1}{L} \sum_{l=1}^L P_{n,t}(j_{n,t} \mid x, \beta'_{x_k,r,t}, \beta_f) \right) \right) \right], \quad (9)$$

where the draw $\beta'_{x_k,r,t}$ for the combined coefficient β'_{x_k} is obtained as the sum of an individual-specific draw $\beta_{x_k,r}$ and an observation-specific draw $\gamma_{x_k,l,t}$, and where the index for the inner summation uses the same upper limit as the outer summation, such that $R = L$.

Here, the positioning of the two summations over draws is crucial to adequately represent the model structure developed in this section. To our knowledge, this is not possible with existing estimation packages for MMNL. As an example, it is our understanding that in the widely used BIOGEME package (Bierlaire, 2003, 2005), the presence of any *panel* terms means that the integration over random terms is carried out outside the product over observations, independently of the presence of any *cross-sectional* terms. As such, in BIOGEME, the term in Equation 8 would be simulated as:

$$SLL_n(\Omega) = \ln \left[\frac{1}{R} \sum_{r=1}^R \left(\prod_{t=1}^{T_n} (P_{n,t}(j_{n,t} \mid x, \beta'_{x_k,r,t}, \beta_f)) \right) \right], \quad (10)$$

where the individual components of $\beta'_{x_k,r,t}$ are now both respondent-specific. While the approach in Equation 10 is easier (and cheaper) to estimate than the approach in Equation 9, it is not consistent with our proposed modelling methodology.

Given these issues with existing software, new code was developed in Ox 4.2 (Doornik, 2001), and this was used for the estimation of all models presented in this paper.

3 Empirical analysis

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

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 state preference screen

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 SP experiment. Based on the actual trip attribute levels 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 (TC) 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 (a *RP* alternative). The remaining two alternatives represented competing hypothetical routes (*SP* alternatives). As such, the *RP* alternative remained invariant across the 16 choice situations with only the levels of the *SP* 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

	Random taste heterogeneity			
	Intra-respondent		Inter-respondent	
	independent	correlated	independent	correlated
Model 1	-	-	-	-
Model 2	X	-	-	-
Model 3	-	X	-	-
Model 4	-	-	X	-
Model 5	-	-	-	X
Model 6	X	-	X	-
Model 7	X	-	-	X
Model 8	-	X	-	X

Table 1: Summary of estimated models

from a practice game) is shown in Figure 1.

The SP experiment was constructed using efficient experimental design methods. For a review of efficient SP 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.

3.2 Experimental framework

A total of 8 models were estimated in this analysis, ranging from a basic MNL model to a specification allowing for inter-respondent and intra-respondent taste heterogeneity, with correlation between random taste coefficients.

A summary of the different model structures that were estimated is given in Table 1. We start off with a basic MNL model, with no random taste heterogeneity. This is then followed by models allowing for random taste heterogeneity in a cross-sectional formulation, with or without correlation between individual taste coefficients. From here, we move to a specification that uses distribution of the taste coefficients across respondents rather than across observations, where again, in the more complicated models, we allow for correlation between taste coefficients. In the final sets of models, we allow jointly for inter-respondent and intra-respondent variation in tastes.

From Table 1, the relationship between the various models should be clear. Model 3 reduces to model 2, which in turn reduces to model 1. Similarly, model 5 reduces to model 4, which reduces to model 1. Model 6 reduces to model 4, while model 7 reduces to both models 6 and model 5. Finally, model 8 reduces to model 7 and 6. Subsequent reductions are implicit. The models with some inter-respondent variation in tastes are not directly comparable to the models

with intra-respondent variation only, as, in the former, we set the mean of any intra-agent variation to zero.

In the analysis presented in this paper, only four of the attributes were used, where the exclusion of trip time variability was based primarily on the fact that a large share of respondents indicated that they consistently ignored this attribute, and where very low levels of significance were consistently obtained for the associated coefficient.

In addition to the marginal utility coefficients associated with the remaining four attributes, two constants were included, associated with the *RP* alternative and the first *SP* alternative. The inclusion of alternative specific constants in the context of an unlabelled choice experiment was motivated by the fact that they allow us to capture inertia (i.e., choosing the *RP* alternative), as well as reading from left to right effects.

On the basis of this, the following utility specification was used for alternative j , where we show the specification for the most complicated model, with appropriate simplifications applying for other models.

$$\begin{aligned}
U_{j,n,t} = & \delta_j \\
& + (\beta_{FF,n} + \gamma_{FF,n,t}) FF_{j,n,t} \\
& + (\beta_{TC,n} + \gamma_{TC,n,t}) TC_{j,n,t} \\
& + (\beta_{SDT,n} + \gamma_{SDT,n,t}) SDT_{j,n,t} \\
& + (\beta_{TOLL,n} + \gamma_{TOLL,n,t}) TOLL_{j,n,t} \\
& + \varepsilon_{j,n,t}
\end{aligned} \tag{11}$$

Some clarifications are required. The constant, δ_j , is set to zero for $j = 3$, i.e., for the second of the two *SP* alternatives. The four main taste coefficients ($\beta_{FF,n}$, $\beta_{TC,n}$, $\beta_{SDT,n}$, and $\beta_{TOLL,n}$) are specified to vary randomly across respondents (using integration outside the product over observations) in models 4 to 8, with correlation between coefficients in models 5, 7 and 8. A Normal distribution is used in all models. In model 1, only a point value is estimated for the four coefficients, which is kept fixed across respondents and observations. The four additional taste coefficients ($\gamma_{FF,n,t}$, $\gamma_{TC,n,t}$, $\gamma_{SDT,n,t}$, and $\gamma_{TOLL,n,t}$) are used mainly for models 6 to 8, where their mean is set to zero, and where a Normal distribution is used. The coefficients are specified to vary across all observations, independently of the respondent. These coefficients are also estimated in models 2 and 3, where the mean value is not constrained to zero. In models 3 and 8, we additionally allow for correlation between the four coefficients.

Finally, it should be noted that in all models estimated here, a linear-in-attributes specification of the utility functions is used. This is based on pre-

liminary analyses that did not reveal consistent and significant non-linearities in response with the data at hand.

3.3 Estimation results

Given the high number of models estimated in this study, it is not possible to present detailed estimation results for each single model. Rather, we present an overview of the results across models, in conjunction with detailed results for the *recommended* model structure. Detailed results for all remaining models are available from the first author on request.

A summary of the final model fits for the 8 estimated models is given in Table 2, with a graphical representation in Figure 2. In each case, the adjusted $\rho^2(0)$ measure is presented alongside the final log-likelihood measure, to account for the differences across models in the number of estimated parameters³.

The first significant observation to be made is the jump in model fit that is obtained when passing from a cross-sectional specification (models 1 – 3) to a specification that recognises the repeated choice nature of the dataset in the representation of random taste heterogeneity (models 4 – 8). This is most apparent when looking at Figure 2.

The next observation relates to the effects of allowing for correlation between the various random taste coefficients, where it is worth mentioning that, in the models that allow for intra-respondent variation in addition to inter-respondent variation, no correlation was incorporated between the coefficients in these two groups. Overall, allowing for correlation leads to significant improvements in model fit. As such, there are improvements in model fit when moving from model 2 to model 3. Similar improvements are observed when moving from model 4 to 5 and from model 6 to 7. In fact, the only exception to this comes when allowing for correlation between the additional random coefficients for intra-respondent heterogeneity, i.e., when moving from model 7 to model 8. While there is an improvement in model fit, this is not significant when taking into account the additional parameters. These results suggest that allowing for correlation between the intra-respondent coefficients is not advisable with the present data. Finally, another indication of the effects of allowing for correlation is that there is a drop when moving from model 5 to model 6. When additionally comparing these fits to those obtained with model 4, it becomes apparent that allowing for intra-respondent variation only leads to significantly better model performance when also allowing for correlation between the inter-respondent coefficients.

³Here, it should be noted that the models presented here do include all coefficients for the appropriate formulation, with insignificant estimates not removed from the model.

Model	LL($\hat{\beta}$)	adj. $\rho^2(0)$
1	-2,399.29	0.3325
2	-2,321.13	0.3531
3	-2,310.39	0.3544
4	-2,053.29	0.4274
5	-2,033.27	0.4313
6	-2,049.21	0.4274
7	-2,026.89	0.4320
8	-2,023.00	0.4314

Table 2: Summary of model fits for estimated models

On the basis of the above discussion, the *recommended* model structure is one which allows for inter-personal as well as intra-personal variation in tastes, where correlation between coefficients however only exists for the former.

Before moving on to the detailed presentation of the results for the *recommended* model, we will briefly look at the more substantive results for the different models, in the form of various trade-offs calculated from the four marginal utility coefficients. In the simulation of these trade-offs, the unbounded nature of the Normal distribution led to severe problems with outliers, giving an inflated range to the trade-offs, as well as a biased mean. These problems were so severe that the censoring approach advocated by [Hensher and Greene \(2003\)](#) had to be used, where this was applied at the level of the simulated trade-offs (rather than individual coefficients), and where, to maintain balance, an equal proportion of draws was removed to either side⁴. The simulation of the trade-offs was carried out on the basis of 100,000 respondent-specific draws and 1,600,000 observation specific draws⁵. Where appropriate, correlation was taken into account in the generation of the draws.

Four trade-offs were used in the analysis, showing the willingness to accept increases in running costs and road tolls in return for reductions in free flow and slowed down travel times. A graphical representation of the various trade-offs across models is given in [Figure 3](#), where we focus on the mean values of the

⁴An alternative approach would have been to work with draws conditioned on observed choices (cf. [Train, 2003](#); [Hess and Rose, 2006](#)). This is the topic on ongoing work, but is made more complicated by the presence of within-individual variation in tastes.

⁵As an illustration, in the models with no inter-respondent variation, 1,600,000 draws were used in the simulation, while, in the models with no intra-respondent variation, 100,000 draws were used in the simulation, with each draw being used 16 times (once per observation). In the models with combined inter-respondent and intra-respondent variation, the 1,600,000 intra-respondent draws were combined with 16 sets of the inter-respondent draws.

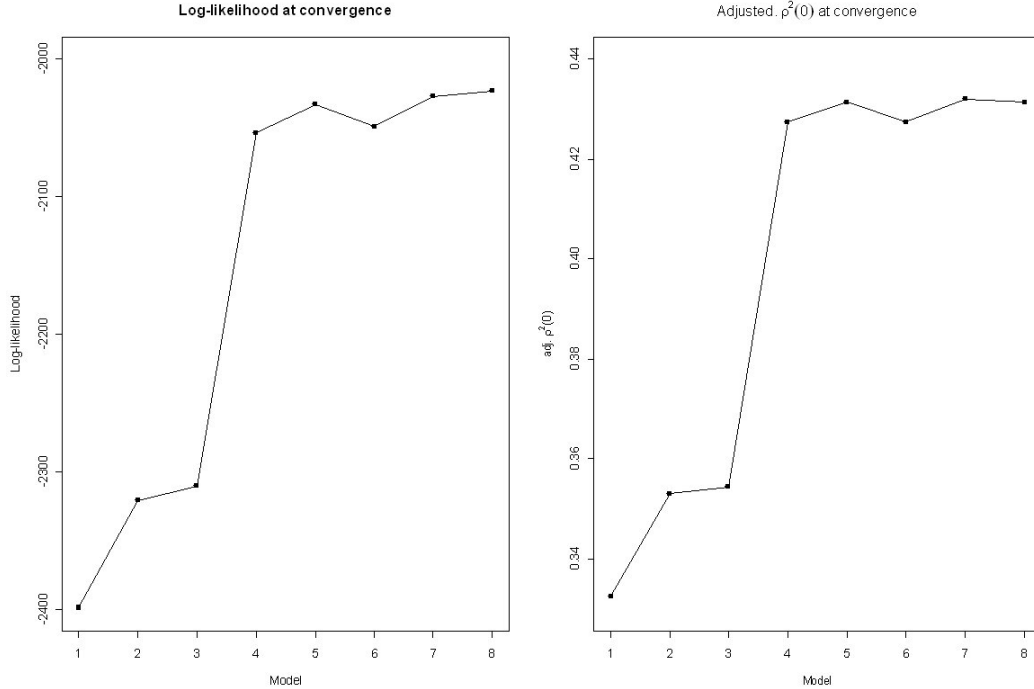


Figure 2: Model fit statistics for estimated models

trade-offs⁶.

The plots show that the main difference in the trade-offs seems to arise when moving from the cross-sectional models to the models recognising the repeated choice nature in the representation of random taste heterogeneity.

We now proceed with the description of the detailed estimation results for the recommended model structure, i.e., model 7. These results are presented in Table 3. Here, some explanations are required to supplement those from Section 3.1. The parameter $\beta_{FF,\mu}$ gives the mean value for the (Normal) distribution of the inter-respondent variation in the sensitivity to free flow time, with corresponding parameters for the remaining three attributes. The following 10 parameters relate to the Choleski transformation for multivariate Normals (cf. Train, 2003, pp.211-212). With η_1, η_2, η_3 and η_4 representing four independently distributed $N(0, 1)$ variates, draws from the distribution for the four inter-respondent distributed

⁶The standard deviations would be of little interest given the likely biasing impact of the censoring approach.

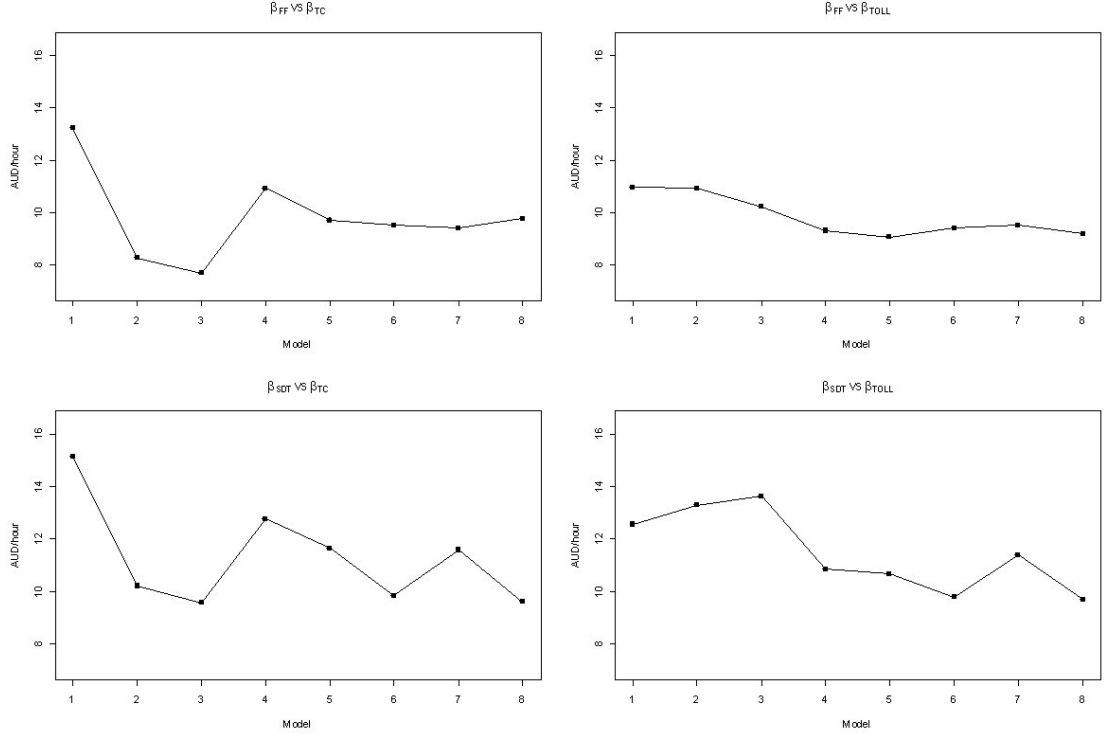


Figure 3: Mean values of trade-offs for various estimated models

coefficients are obtained as:

$$\begin{aligned}
 \beta_{FF} &= \beta_{FF,\mu} + s_{1,1}\eta_1 \\
 \beta_{TC} &= \beta_{TC,\mu} + s_{2,1}\eta_1 + s_{2,2}\eta_2 \\
 \beta_{SDT} &= \beta_{SDT,\mu} + s_{3,1}\eta_1 + s_{3,2}\eta_2 + s_{3,3}\eta_3 \\
 \beta_{TOLL} &= \beta_{TOLL,\mu} + s_{4,1}\eta_1 + s_{4,2}\eta_2 + s_{4,3}\eta_3 + s_{4,4}\eta_4
 \end{aligned} \tag{12}$$

With this notation, we have that $var(\beta_{FF}) = s_{1,1}^2$, $var(\beta_{TC}) = s_{2,1}^2 + s_{2,2}^2$, $var(\beta_{SDT}) = s_{3,1}^2 + s_{3,2}^2 + s_{3,3}^2$, and $var(\beta_{TOLL}) = s_{4,1}^2 + s_{4,2}^2 + s_{4,3}^2 + s_{4,4}^2$. The corresponding standard deviations for the four coefficients ($\beta_{FF,\sigma}$ for β_{FF}) are shown in the table, which also presents the correlations between individual taste coefficients.

The four γ parameters relate to the standard deviations for the intra-respondent distributed coefficient values, where the mean value is zero. Finally, Table 3 also

Number of respondents	3,280
Number of observations	205
LL(0)	-3603.45
LL($\hat{\beta}$)	-2026.89
parameters	20
adj. $\rho^2(0)$	0.4320

	est.	asy. t-rat.
δ_{RP}	0.2649	2.22
δ_{SP_1}	0.1976	2.20
$\beta_{FF,\mu}$	-0.1472	-8.80
$\beta_{TC,\mu}$	-0.7583	-9.73
$\beta_{SDT,\mu}$	-0.1668	-10.89
$\beta_{TOLL,\mu}$	-0.8468	-14.18
$s_{1,1}$	0.1222	8.55
$s_{2,1}$	-0.0269	-1.91
$s_{2,2}$	0.5029	7.51
$s_{3,1}$	0.0442	1.64
$s_{3,2}$	0.0121	1.00
$s_{3,3}$	0.0842	4.00
$s_{4,1}$	0.0499	1.06
$s_{4,2}$	0.3971	7.69
$s_{4,3}$	-0.2036	-4.34
$s_{4,4}$	0.3702	9.59
γ_{FF}	0.0565	1.68
γ_{TC}	0.3282	3.87
γ_{SDT}	0.0497	3.20
γ_{TOLL}	0.0057	0.12
$\beta_{FF,\sigma}$	0.1222	-
$\beta_{TC,\sigma}$	0.5036	-
$\beta_{SDT,\sigma}$	0.0959	-
$\beta_{TOLL,\sigma}$	0.5819	-
corr($\beta_{FF,\sigma}, \beta_{TC,\sigma}$)	-0.0535	-
corr($\beta_{FF,\sigma}, \beta_{SDT,\sigma}$)	0.4613	-
corr($\beta_{FF,\sigma}, \beta_{TOLL,\sigma}$)	0.0857	-
corr($\beta_{TC,\sigma}, \beta_{SDT,\sigma}$)	0.1012	-
corr($\beta_{TC,\sigma}, \beta_{TOLL,\sigma}$)	0.6768	-
corr($\beta_{SDT,\sigma}, \beta_{TOLL,\sigma}$)	-0.1817	-
	mean	std.dev.
β_{FF} vs. β_{TC}	9.39	5.44
β_{FF} vs. β_{TOLL}	9.52	5.44
β_{SDT} vs. β_{TC}	11.57	6.70
β_{SDT} vs. β_{TOLL}	11.38	7.26

Table 3: Detailed estimation results for model 7

presents the mean values and standard deviations for the four trade-offs calculated from the estimates, where these trade-offs were obtained using simulation with censoring of counter-intuitively signed trade-off values.

We now proceed with a detailed analysis of the results. The significant and positive estimates for both alternative specific constants suggest the presence of inertia effects as well as reading left to right effects.

As expected, the mean values for the four taste coefficients are all negative, while they also attain high levels of statistical significance. All four coefficients have high standard deviation, with coefficients of variation ranging from 0.58 to 0.83. In terms of correlation, we get the expected high positive correlation between β_{FF} and β_{SDT} , along with positive correlation between β_{TC} and β_{TOLL} . Correlation levels for other pairs of coefficients are relatively small.

Next, we move to the parameters associated with intra-responder variation. Here, we obtain highly significant estimates for γ_{TC} and γ_{SDT} , while γ_{FF} is significant at the 91% level of confidence. There is no significant intra-personal variation for the sensitivity to toll. In each case, the level of intra-personal variation is lower than the level of inter-personal variation, which is consistent with intuition.

As a final step, we look at the four trade-offs calculated from the model estimates. Here, it can be seen that, as expected, free flow time is valued less negatively than slowed down time, where the relative variation in sensitivity across respondents (and observations) is similar for the two attributes. With these trade-offs, the effect of the censoring approach needs to be borne in mind, reducing the reliability of these results.

4 Summary & conclusions

Given the high reliance on SP data in the analysis of travel behaviour, a better understanding of how to estimate econometrical models from such data is crucial. This is particularly the case given that the outputs of models estimated on SP data have been used extensively in the past to help shape policy debate and determine transport related infrastructure projects. This situation will likely continue into the future. As such, any imprecise valuation attributable to incorrectly specified SP models carries more than a purely academic risk, with significant monetary or societal losses likely to accrue.

The purpose of this paper was to reexamine the question of how to address the fact that SP data contain multiple responses for each respondent. The issue is not new, having been acknowledged in the very first SP choice applications undertaken by [Louviere and Hensher \(1983\)](#); [Louviere and Woodworth \(1983\)](#).

However, it has only been in the last decade that the issue has become more widely discussed.

Over recent years, the standard approach for acknowledging the repeated choice nature of SP data has been in the context of random taste heterogeneity, with an assumption of inter-respondent variation in tastes, along with intra-respondent homogeneity in tastes. In this paper, we question the validity of this assumption of within-respondent homogeneity. Indeed, there appears to be only limited theoretical justification for treating SP data as within respondent panel data. Psychologically, one would expect individuals to inherently possess preferences for the attributes and attribute levels of alternatives within a well set-up choice experiment. However, whether these preferences are transient or fixed over choice replications is not immediately clear. Of course, we would anticipate an already time sensitive individual to remain thus over the course of an experiment, however, to assume that the precise degree of sensitivity to time is fixed over the entire course of the experiment represents an assumption that perhaps cannot be justified, as discussed at length in Section 1.

In this paper, we propose a model structure that can accommodate intra-respondent and inter-respondent heterogeneity jointly. The findings from our analysis suggest that the majority of taste heterogeneity within the present SP dataset does indeed derive from variation in tastes across individual respondents, giving the *panel* approach a significant advantage over a purely cross-sectional approach. However, there is also evidence to suggest the presence of some within respondent heterogeneity, such that the sensitivity to the various attributes does indeed vary across choice situations for the same respondent. Based on these findings, we conclude that in the present data, there exists for each respondent both an invariant component of marginal utility for each of the attributes across choice observations as well as a component which is choice situation specific. The invariant component may be thought of as representing a respondent's overall general marginal utility for each of the attributes.

Unlike traditional contingent valuation methods, SP choice models force respondents to handle multiple attributes and attribute levels when making their choices. As such, the second component of the marginal utility accounts for the fact that each hypothetical choice situation presented to a respondent relates different combinations of attribute levels, the various combinations of which may result in slight deviations around the marginal utilities for the attributes being modelled over all choice situations. Additionally, this component potentially accounts for learning and fatigue effects.

As with any piece of academic research, the work presented here has several limitations, and it is important to acknowledge this. Firstly, it should be noted that the results of this paper relate to a single data set. As such, results from

other data sets are required before concrete conclusions can be drawn. This should include, but not be limited to, the use of simulated data in a systematic Monte Carlo study.

Secondly, the points made by Fosgerau and Nielsen (2006) in relation to the difficulty of distinguishing between taste heterogeneity and other errors in the case of cross-sectional data need to be borne in mind. As such, there is a possibility that some of the intra-respondent variation retrieved here relates to factors other than traditional taste heterogeneity. This could for example include non-linearities in response, although it should again be noted that preliminary tests did not reveal consistent patterns of non-linear response. Other possible factors include for the example the presence of thresholds in preference formation. However, in closing it should be noted that while we may not with certainty be able to determine the exact source of the variations, our analysis presents clear evidence of deviations from the within-respondent homogeneity assumption typically made in random coefficients models.

Acknowledgements

Part of the work described in this paper was carried out during stays by the first author in the Institute of Transport and Logistics Studies at the University of Sydney, and in the Department of Civil and Environmental Engineering at the Massachusetts Institute of Technology, and a stay by the second author at the Pontificia Universidad Católica de Chile. The authors would like to thank Andrew Daly and Mogens Fosgerau for helpful discussions and comments.

References

- Abdel-Aty, M. A., R. Kitamura and P. P. Jovanis (1997) Using stated preference data for studying the effect of advanced traffic information on drivers' route choice, *Transportation Research Part C: Emerging Technologies*, **5** (1) 39–50.
- Arentze, T., A. Borgers, H. Timmermans and R. DelMistro (2003) Transport stated choice responses: effects of task complexity, presentation format and literacy, *Transportation Research Part E: Logistics and Transportation Review*, **39** (3) 229–244.
- Axhausen, K. W., S. Hess, A. König, G. Abay, J. J. Bates and M. Bierlaire (2007) *State of the art estimates of the Swiss value of travel time savings*, Transport Policy, forthcoming.

- Bhat, C. R. and S. Castelar (2002) A unified mixed logit framework for modeling revealed and stated preferences: formulation and application to congestion pricing analysis in the San Francisco Bay area, *Transportation Research Part B: Methodological*, **36** (7) 593–616.
- 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.
- Bierlaire, M. (2005) *An introduction to BIOGEME Version 1.4*, biogeme.epfl.ch.
- Bliemer, M. C. J. and J. M. Rose (2006) *Designing Stated Choice Experiments: State-of-the-art*, paper presented at the 11th International Conference on Travel Behaviour Research, Kyoto, Japan.
- Brazzell, J. D. and J. J. Louviere (1998) *Length effects in conjoint choice experiments and surveys: an explanation based on cumulative cognitive burden*, Department of Marketing, The University of Sydney, Sydney, Australia.
- Cooke, A. D. and B. A. Mellers (1995) Attribute range and response range: Limits of compatibility in multiattribute judgment, *Organizational Behavior and Human Decision Processes*, **63** (2) 187–194.
- Deshazo, J. R. and G. Fermo (2002) Designing choice sets for stated preference methods: the effects of complexity on choice consistency, *Journal of Environmental Economics and Management*, **44** (1) 123–143.
- Doornik, J. A. (2001) *Ox: An Object-Oriented Matrix Language*, Timberlake Consultants Press, London.
- Ferrini, S. and R. Scarpa (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.
- Fosgerau, M. and S. F. Nielsen (2006) *Deconvoluting preferences and errors: A semi-nonparametric model for binomial panel data*, Econometric Society European Meeting, Vienna, Austria.
- Garrod, G. D., R. Scarpa and K. G. Willis (2002) Estimating the benefits of traffic calming on through routes: A choice experiment approach, *Journal of Transport Economics and Policy*, **36** (2) 211–232.
- Haaijer, R., K. Wagner and M. Wedel (2000) Response latencies in the analysis of conjoint choice experiments, *Journal of Marketing Research*, **37**, 376–382.

- Hensher, D. A. (2001a) Measurement of the valuation of travel time savings, *Journal of Transport Economics and Policy*, **35** (1) 71–98.
- Hensher, D. A. (2001b) The valuation of commuter travel time savings for car drivers: Evaluating alternative model specifications, *Transportation*, **28**, 101–118.
- Hensher, D. A. and W. H. Greene (2003) The Mixed Logit Model: The State of Practice, *Transportation*, **30** (2) 133–176.
- Hensher, D. A. and J. King (2001) Parking demand and responsiveness to supply, pricing and location in the sydney central business district, *Transportation Research Part A: Policy and Practice*, **35** (3) 177–196.
- Hess, S., M. Bierlaire and J. W. Polak (2005) Estimation of value of travel-time savings using mixed logit models, *Transportation Research Part A: Policy and Practice*, **39** (2-3) 221–236.
- Hess, S. and J. W. Polak (2004) *Mixed Logit estimation of parking type choice*, paper presented at the 83rd Annual Meeting of the Transportation Research Board, Washington, DC.
- Hess, S. and J. M. Rose (2006) *Effects of distributional assumptions on conditional estimates from Mixed Logit models*, ITLS Working paper, Institute for Transport and Logistics Studies, University of Sydney, Sydney, Australia.
- Louviere, J. J. and D. A. Hensher (1983) Using discrete choice models with experimental design data to forecast consumer demand for a unique cultural event, *Journal of Consumer Research*, **10**, 348–361.
- Louviere, J. J., D. A. Hensher and J. Swait (2000) *Stated Choice Models: Analysis and Application*, Cambridge University Press, Cambridge.
- Louviere, J. J. and G. Woodworth (1983) Design and analysis of simulated consumer choice and allocation experiments: A method based on aggregate data, *Journal of Marketing Research*, **20**, 350–367.
- Ohler, T., A. Li, J. J. Louviere and J. Swait (2000) Attribute range effects in binary response tasks, *Marketing Letters*, **11** (3) 249–260.
- Ortúzar, J. de D., A. Iacobelli and C. Valeze (2000) Estimating demand for a cycle-way network, *Transportation Research Part A: Policy and Practice*, **34** (5) 353–373.

- Outwarslot, H. and P. Rietveld (1996) Stated choice experiments with repeated observations, *Journal of Transportation Economics and Policy*, **30**, 203–212.
- Pullman, M. E., K. J. Dodson and W. L. Moore (1999) A comparison of conjoint methods when there are many attributes, *Marketing Letters*, **10** (2) 1–14.
- Rose, J. M. and I. G. Black (2006) Means matter, but variance matter too: Decomposing response latency influences on variance heterogeneity in stated preference experiments, *Marketing Letters*, **17** (4) 295–310.
- Train, K. (2003) *Discrete Choice Methods with Simulation*, Cambridge University Press, Cambridge, MA.
- Verleg, P. W., H. N. Schifferstein and D. R. Wittink (2002) Range and number-of-levels effects in derived and stated measures of attribute importance, *Marketing Letters*, **13** (1) 41–52.