Some lessons for working with repeated choice data

Stephane Hess*  John M. Rose†

November 30, 2006

Abstract

The vast majority of discrete choice modelling (DCM) applications are now estimated on Stated Preference (SP) data, including, but limited to in 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. Furthermore, the paper shows the benefits of accounting for correlation across replications for the same individual independently of the treatment of taste heterogeneity.

1 Introduction

Stated preference (SP) data have proven useful in solving many transportation related problems. For example, SP data have been used to examine the demand for a cycle-way network (e.g. Ortúzar et al., 2000a), 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; Axhausen et al., 2006).

*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”.

The unique nature of SP data has implications for the modelling to be undertaken on such data. Most econometric models fail to account for the repeated observations of SP data, treating each observation as if it were made by a different individual (cf. Ortúzar and Willumsen, 2001). Traditionally, this was thought to impact only upon the asymptotic standard errors of the model, and in order to correct for this, several authors resorted to bootstrap and jackknife methods (e.g. Cirillo et al., 2000; Ortúzar et al., 2000b).

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 Castelan, 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).

This trend means that any SP applications not allowing for random taste heterogeneity (i.e., making use of MNL, NL, . . .) often fail to account for the repeated choice nature of the data. Additionally however, 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. Haaijer 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 another\(^1\). 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

\(^{1}\) As 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.
to the attributes of the experiment. This in turn would speak in favour of also allowing for within-respondent preference heterogeneity.

In addition to the above concerns in relation to the assumption of within-respondent homogeneity, a separate problem arises with the over-reliance on the random coefficients formulation for dealing with repeated choice data. Firstly, it means that studies that do not make use of a random coefficients formulation now largely ignore the repeated choice nature of the data. Secondly, there is a possibility that not all of the effects of the correlation across replications for the same individual can be retrieved with the help of a random coefficients formulation. Not only does this potentially mean a loss of explanatory power, but there is possibly also a risk of confounding between serial correlation and random taste heterogeneity.

This paper deals with two separate but interrelated questions that arise from the above:

- What are the effects of the assumption of constant tastes across replications for the same individual in the case of instantaneous SP panels, and should this assumption be relaxed?

- What ways are there of dealing with the repeated choice nature of the data independently of a representation of random taste heterogeneity, but at the estimation stage (i.e., not using a posteriori correction)?

The remainder of this paper is organised as follows. Section 2 presents the various modelling approaches used in the analysis in this paper and discusses how they relate to each other in terms of gradually increasing flexibility. The empirical analysis conducted for this paper is described in Section 3. Finally, Section 4 presents the conclusions of the paper.

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 $i = 1, \ldots, I$) for respondent $n$ (with $n = 1, \ldots, N$) in choice situation $t$ (with $t = 1, \ldots, T_n$). Then we have that $U_{i,n,t} = V_{i,n,t} + \epsilon_{i,n,t}$, where $V_{i,n,t}$ is the observed part of utility, and $\epsilon_{i,n,t}$ 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 $\beta_k$, with the functional form
of this interaction given by $f_k()$. In this paper, we rely on a linear formulation, such that $f_k (\beta_k, x_{k,i,n,t}) = \beta_k x_{k,i,n,t}$.

Let $P_{n,t} (i \mid x, \beta)$ 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:

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

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

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

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

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

$$P_{n,t} (i \mid \Omega) = \int_{\beta_1} P_{n,t} (i \mid x, \beta) g (\beta_1 \mid \Omega_1) \ldots g (\beta_K \mid \Omega_K) d\beta_1 \ldots d\beta_K,$$

or, for conciseness (and allowing for multivariate distributions$^2$):

$$P_{n,t} (i \mid \Omega) = \int P_{n,t} (i \mid x, \beta) g (\beta \mid \Omega) d\beta.$$

The log-likelihood formulation in equation 2 relates to a cross-sectional formulation, making the assumption that the tastes vary across respondents as well.

---

$^2$All 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.
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:

\[ LL(\Omega) = \sum_{n=1}^{N} \ln \left( P_n(\Omega) \right), \tag{5} \]

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

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

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 two shortcomings that are addressed in this paper:

- it assumes an absence of variation in tastes across choice situations for the same respondent,
- and it assumes that all correlation between replications for the same individual is captured through the specification of random taste heterogeneity.

In this paper, we propose two extensions that address these shortcomings. We will now look at these in turn.

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 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 distribution is Normal, with mean \( \mu_{\beta_{x_k}} \) and standard deviation \( \sigma_{\beta_{x_k}} \), such that \( \beta_{x_k} \sim N \left( \mu_{\beta_{x_k}}, \sigma_{\beta_{x_k}} \right) \). 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 \( \beta_{n,x_k} \), 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 \( \beta'_{x_k} \) 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 \( \beta_{x_k} \), we can set the mean of the distribution of \( \gamma_{x_k} \) 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:

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

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 \( \Omega \) now contains parameters for the distributions of \( \beta_{x_k} \) and \( \gamma_{x_k} \).

Allowing for correlation across choice situations independently of random taste heterogeneity requires a different approach, where here, we make use of an error components formulation. Specifically, we are interested in creating correlation across replications for the same respondent, where this individual-specific effect is the same across all alternatives. Adding the same term to all utilities is clearly not possible for reasons of identification. However, adding an error component to all but one of the alternatives creates problems in its own. Indeed, while now identifiable, and while creating correlation across observations, this approach also creates correlation between only some of the alternatives, approximating a nesting structure. Furthermore, this approach would introduce heteroscedasticity. In this paper, we take a slightly different approach, which we illustrate here on the basis of a three-alternative choice set. Specifically, we introduce error components as follows:

\[
\begin{align*}
U_{1,n,t} &= V_{1,n,t} + \varepsilon_{1,n,t} + \sigma \xi_{1,n} \\
U_{2,n,t} &= V_{2,n,t} + \varepsilon_{2,n,t} + \sigma \xi_{2,n} \\
U_{3,n,t} &= V_{3,n,t} + \varepsilon_{3,n,t} + \sigma \xi_{3,n}
\end{align*}
\tag{8}
\]

where \( \xi_{1,n} \), \( \xi_{2,n} \) and \( \xi_{3,n} \) are \( N (0, 1) \) variates that are distributed independently across alternatives and across respondents, but not across observations for the same respondent. Clearly, this creates correlation across choice situations for the
same respondents, where the multiplication by a common $\sigma$ leads to the same level of correlation across the entire choice set. No correlation across alternative is introduced, and neither is heteroscedasticity.

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 for respondents undertaking non-commuting trips only. 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. Once recruited, a time and location was agreed upon for the survey to be undertaken using a face-to-face computer aided personal interview (CAPI). Quotas were imposed to insure a range of travel times over the sample; between 10 and 30 minutes, 31 to 60 minutes, and more than 61 minutes (capped at two hours). Trips of less than 10 minutes were excluded for both practical and theoretical reasons. From a practical perspective, it was felt that varying travel times and costs around a small base was not likely to produce levels which would be likely to induce a change of route in reality in the SP experiment. Secondly, within the Sydney context, for political reasons, shorter travel times are unlikely to attract road user charges, and hence represent an unrealistic situation to present to respondents as part of the survey.

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
Figure 1: An example of a state preference screen

was explained to them by the interviewer. An example choice situation (taken from a practice game) is shown in Figure 1.

The SP experiment was constructed using efficient experimental design methods. Efficient SP design techniques require first generating the design, and then assuming a set of prior parameter estimates, obtaining the expected asymptotic variance-covariance (AVC) matrix for the design using the prior parameter estimates. The design is then changed in some manner, and assuming the same prior parameter estimates, the AVC matrix of the new design is then calculated. In each instance, the design and related AVC matrix is stored. The design that minimises the AVC matrix under the parameter priors is then kept and used as the preferred SP design in the main study (see Bliemer and Rose 2006 or Ferrini and Scarpa (2006) for a review of efficient SP design methods). For this study, the parameter priors were identified from a literature review, ultimately being drawn from Hensher (2001a).

To generate the efficient experimental designs\(^3\), it was necessary not only to assume prior parameter estimates, but also to assume the attribute levels experienced by the respondents. Unlike most SC experiments where the attributes are

---

\(^3\)The designs used as part of the study are located at http://www.itls.usyd.edu.au/about_itls/staff/johnr.asp.
fixed for the sampled population, the linking of the attribute levels to a recently experienced trip for each respondent meant that the attribute levels for the experiment were not known in advance. For this reason, it became necessary to also assume attribute levels as well as prior parameter values when constructing the design. Table 1 shows the attribute levels and percentages used to construct each of the experimental designs. 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 16 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 and additional error components for correlation across replications.

A summary of the different model structures that were estimated is given in Table 2. Each of the models was estimated with and without the additional error components whose role was described in Section 2. As such, there are 8 different base structures, with gradually increasing complexity. 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.

---

### Table 1: Experimental design attribute levels and pivot percentages

<table>
<thead>
<tr>
<th>Trip segment</th>
<th>FF</th>
<th>SDT</th>
<th>VAR</th>
<th>TC</th>
<th>TOLL</th>
</tr>
</thead>
<tbody>
<tr>
<td>&lt; 30 mins</td>
<td>10</td>
<td>5</td>
<td>5</td>
<td>$2.00</td>
<td>$2.00</td>
</tr>
<tr>
<td>31-60 mins</td>
<td>30</td>
<td>15</td>
<td>8</td>
<td>$3.50</td>
<td>$2.00</td>
</tr>
<tr>
<td>&gt;60 mins</td>
<td>40</td>
<td>20</td>
<td>10</td>
<td>$5.00</td>
<td>$2.00</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Pivot Percentages</th>
<th>FF</th>
<th>SDT</th>
<th>VAR</th>
<th>TC</th>
<th>TOLL</th>
</tr>
</thead>
<tbody>
<tr>
<td>Level 1</td>
<td>-50%</td>
<td>-50%</td>
<td>5%</td>
<td>-50%</td>
<td>-100%</td>
</tr>
<tr>
<td>Level 2</td>
<td>-20%</td>
<td>-20%</td>
<td>10%</td>
<td>-20%</td>
<td>20%</td>
</tr>
<tr>
<td>Level 3</td>
<td>10%</td>
<td>10%</td>
<td>15%</td>
<td>10%</td>
<td>40%</td>
</tr>
<tr>
<td>Level 4</td>
<td>40%</td>
<td>40%</td>
<td>20%</td>
<td>40%</td>
<td>60%</td>
</tr>
</tbody>
</table>
Random taste heterogeneity

<table>
<thead>
<tr>
<th></th>
<th>Intra-respondent</th>
<th>Inter-respondent</th>
<th>Error components</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>independent</td>
<td>correlated</td>
<td>NO</td>
</tr>
<tr>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>-</td>
<td>-</td>
<td>-</td>
<td>Model 3</td>
</tr>
<tr>
<td>X</td>
<td>-</td>
<td>-</td>
<td>Model 5</td>
</tr>
<tr>
<td>-</td>
<td>X</td>
<td>-</td>
<td>Model 7</td>
</tr>
<tr>
<td>-</td>
<td>-</td>
<td>X</td>
<td>Model 9</td>
</tr>
<tr>
<td>X</td>
<td>-</td>
<td>X</td>
<td>Model 11</td>
</tr>
<tr>
<td>X</td>
<td>-</td>
<td>-</td>
<td>Model 13</td>
</tr>
<tr>
<td>-</td>
<td>X</td>
<td>-</td>
<td>Model 15</td>
</tr>
</tbody>
</table>

Table 2: Summary of estimated models

From Table 2, the relationship between the various models should be clear. Model 5 reduces to model 3, which in turn reduces to model 1. Similarly, model 9 reduces to model 7, which reduces to model 1. Model 11 reduces to model 7, while model 13 reduces to both models 11 and model 9. Finally, model 15 reduces to model 13 and 11. Subsequent reductions are implicit, and corresponding relationships exist in the case of the models making use of the additional error components. 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.

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.
\[ U_{j,n,t} = \delta_j + (\beta_{FF,n} + \gamma_{FF,n,t}) F_{F,j,n,t} + (\beta_{TC,n} + \gamma_{TC,n,t}) T_{C,j,n,t} + (\beta_{SDT,n} + \gamma_{SDT,n,t}) S_{DT,j,n,t} + (\beta_{TOLL,n} + \gamma_{TOLL,n,t}) T_{OLL,j,n,t} + \sigma \xi_{j,n} + \epsilon_{j,n,t} \]  

(9)

Some clarifications are required. The constant, \( \delta_j \), 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 7 to 16, with correlation between coefficients in models 9 and 10, and models 13 to 16. A Normal distribution is used in all models. In models 1 and 2, 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 11 to 16, 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 3 to 6, where the mean value is not constrained to zero. In models 5, 6, 15 and 16, we additionally allow for correlation between the four coefficients. The error component \( \sigma \xi_{j,n} \) is included for all models with even numbers, where \( \xi_{j,n} \) is specified to vary randomly across respondents (but not observations for the same respondents), following a Normal distribution with a mean of zero and a standard deviation of 1. The three terms \( \xi_{1,n} \), \( \xi_{2,n} \) and \( \xi_{3,n} \) are distributed independently. The estimation of the models presented in this paper was carried out using a mixture of Ox 4.2 (Doornik, 2001) and BIOGEME (Bierlaire, 2003).

3.3 Estimation results

Given the high number of models estimated in this study, it is not possible to presented 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 16 estimated models is given in Table 3, 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 parameters\(^4\).

\(^4\)Here, 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.
The first observation that can be made from Table 3 and Figure 2 relates to the improvements in model performance resulting from the inclusion of the additional error components discussed in Section 2. Across the 8 model structures, the improvements range from 86.14 units in log-likelihood to 107.63 units. Coming at the cost of just one additional estimated parameter ($\sigma$), these improvements are highly significant, indicating the presence of a significant level of correlation across observations for the same respondent.

The other significant observation to be made is the jump in model fit that is obtained when passing from a cross-sectional specification (models 1 – 6) to a specification that recognises the repeated choice nature of the dataset in the representation of random taste heterogeneity (models 9 – 16). This is most apparent when looking at Figure 2, and these improvements are independent of the inclusion or otherwise of the additional error components.

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, which are however smaller than those obtained with the inclusion of the error components or the recognition of the repeated choice nature in the representation of taste heterogeneity. As such, there are improvements in model fit when moving from models 3 and 4 to models 5 and 6. Similar improvements are observed when moving from models 7 and 8 to models 9 and 10. 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 models 13 and 14 to models 15 and 16. In the model without error components, there is an improvement in model fit, which is however not significant when taking into account the additional parameters. Finally, in the model without error components, there is in fact a drop in model fit when moving from model 14 to 16. 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 models 9 and 10 to models 11 and 12. When additionally comparing these fits to those obtained with models 7 and 8, 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.

In addition to the above comments, it can be noted that allowing for intra-personal variation does indeed lead to improvements in model performance, seemingly especially when allowing for correlation between inter-personal coefficients.

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. The model also incorporates the additional error component.

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 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 side. The simulation of the trade-offs was carried out on the basis of 100,000 respondent-specific draws and 1,600,000 observation specific draws. 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 trade-offs.

The plots show that, with a few exceptions (namely the trade-offs against travel cost in models 5 and 6), the inclusion of the additional error components has relatively little effect on the trade-offs. Other than 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., models 13 and 14. These results are presented in Table 4. 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,

---

5 An 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.

6 As 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.
Number of respondents 3,280 3,280
Number of observations 205 205
LL(0) -3,603.45 -3,603.45
LL(ˆβ) -2,026.89 -1,922.79
parameters 20 21
adj. ρ^2(0) 0.4320 0.4606

<table>
<thead>
<tr>
<th></th>
<th>est.</th>
<th>asy. t-rat.</th>
<th></th>
<th>est.</th>
<th>asy. t-rat.</th>
</tr>
</thead>
<tbody>
<tr>
<td>δ_{RP}</td>
<td>0.2649</td>
<td>2.22</td>
<td>0.2479</td>
<td>1.74</td>
<td></td>
</tr>
<tr>
<td>δ_{SP_1}</td>
<td>0.1976</td>
<td>2.20</td>
<td>0.2739</td>
<td>2.58</td>
<td></td>
</tr>
<tr>
<td>β_{FF,μ}</td>
<td>-0.1472</td>
<td>-8.80</td>
<td>-0.1815</td>
<td>-13.08</td>
<td></td>
</tr>
<tr>
<td>β_{TC,μ}</td>
<td>-0.7583</td>
<td>-9.73</td>
<td>-1.0315</td>
<td>-8.82</td>
<td></td>
</tr>
<tr>
<td>β_{SDT,μ}</td>
<td>-0.1668</td>
<td>-10.89</td>
<td>-0.1978</td>
<td>-12.30</td>
<td></td>
</tr>
<tr>
<td>β_{TOLL,μ}</td>
<td>-0.8468</td>
<td>-14.18</td>
<td>-1.2246</td>
<td>-13.91</td>
<td></td>
</tr>
<tr>
<td>s_{1,1}</td>
<td>0.1222</td>
<td>8.55</td>
<td>0.1888</td>
<td>10.51</td>
<td></td>
</tr>
<tr>
<td>s_{2,1}</td>
<td>-0.0269</td>
<td>-1.91</td>
<td>-0.0243</td>
<td>-1.61</td>
<td></td>
</tr>
<tr>
<td>s_{2,2}</td>
<td>0.5029</td>
<td>7.51</td>
<td>0.6586</td>
<td>8.56</td>
<td></td>
</tr>
<tr>
<td>s_{3,1}</td>
<td>0.0442</td>
<td>1.64</td>
<td>0.0716</td>
<td>4.56</td>
<td></td>
</tr>
<tr>
<td>s_{3,2}</td>
<td>0.0121</td>
<td>1.00</td>
<td>0.0440</td>
<td>2.56</td>
<td></td>
</tr>
<tr>
<td>s_{3,3}</td>
<td>0.0842</td>
<td>4.00</td>
<td>0.1031</td>
<td>7.09</td>
<td></td>
</tr>
<tr>
<td>s_{4,1}</td>
<td>0.0499</td>
<td>1.06</td>
<td>0.0514</td>
<td>0.82</td>
<td></td>
</tr>
<tr>
<td>s_{4,2}</td>
<td>0.3971</td>
<td>7.69</td>
<td>0.3806</td>
<td>7.47</td>
<td></td>
</tr>
<tr>
<td>s_{4,3}</td>
<td>-0.2036</td>
<td>-4.34</td>
<td>-0.3342</td>
<td>-7.39</td>
<td></td>
</tr>
<tr>
<td>s_{4,4}</td>
<td>0.3702</td>
<td>9.59</td>
<td>0.4705</td>
<td>8.58</td>
<td></td>
</tr>
<tr>
<td>γ_{FF,σ}</td>
<td>0.0565</td>
<td>1.68</td>
<td>0.1061</td>
<td>6.23</td>
<td></td>
</tr>
<tr>
<td>γ_{TC,σ}</td>
<td>0.3282</td>
<td>3.87</td>
<td>0.1262</td>
<td>2.09</td>
<td></td>
</tr>
<tr>
<td>γ_{SDT,σ}</td>
<td>0.0497</td>
<td>3.20</td>
<td>0.0548</td>
<td>4.79</td>
<td></td>
</tr>
<tr>
<td>γ_{TOLL,σ}</td>
<td>0.0057</td>
<td>0.12</td>
<td>0.0776</td>
<td>1.40</td>
<td></td>
</tr>
<tr>
<td>σ</td>
<td>-</td>
<td>-</td>
<td>1.2421</td>
<td>13.53</td>
<td></td>
</tr>
</tbody>
</table>

<p>| | | | | | | | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>β_{FF,σ}</td>
<td>0.1222</td>
<td>-</td>
<td>0.1888</td>
<td>-</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>β_{TC,σ}</td>
<td>0.5036</td>
<td>-</td>
<td>0.6590</td>
<td>-</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>β_{SDT,σ}</td>
<td>0.0959</td>
<td>-</td>
<td>0.1330</td>
<td>-</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>β_{TOLL,σ}</td>
<td>0.5819</td>
<td>-</td>
<td>0.6932</td>
<td>-</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>corr(β_{FF,σ},β_{TC,σ})</td>
<td>-0.0535</td>
<td>-</td>
<td>-0.0369</td>
<td>-</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>corr(β_{FF,σ},β_{SDT,σ})</td>
<td>0.4613</td>
<td>-</td>
<td>0.5381</td>
<td>-</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>corr(β_{FF,σ},β_{TOLL,σ})</td>
<td>0.0857</td>
<td>-</td>
<td>0.0742</td>
<td>-</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>corr(β_{TC,σ},β_{SDT,σ})</td>
<td>0.1012</td>
<td>-</td>
<td>0.3110</td>
<td>-</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>corr(β_{TC,σ},β_{TOLL,σ})</td>
<td>0.6768</td>
<td>-</td>
<td>0.5459</td>
<td>-</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>corr(β_{SDT,σ},β_{TOLL,σ})</td>
<td>-0.1817</td>
<td>-</td>
<td>-0.1521</td>
<td>-</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>mean</th>
<th>std.dev.</th>
<th>mean</th>
<th>std.dev.</th>
</tr>
</thead>
<tbody>
<tr>
<td>β_{FF} vs. β_{TC}</td>
<td>9.39</td>
<td>5.44</td>
<td>9.46</td>
<td>5.41</td>
</tr>
<tr>
<td>β_{FF} vs. β_{TOLL}</td>
<td>9.52</td>
<td>5.44</td>
<td>8.51</td>
<td>4.77</td>
</tr>
<tr>
<td>β_{SDT} vs. β_{TOLL}</td>
<td>11.57</td>
<td>6.70</td>
<td>11.21</td>
<td>6.15</td>
</tr>
<tr>
<td>β_{SDT} vs. β_{TOLL}</td>
<td>11.38</td>
<td>7.26</td>
<td>9.89</td>
<td>6.19</td>
</tr>
</tbody>
</table>

Table 4: Detailed estimation results for models 13 and 14.
Figure 3: Mean values of trade-offs for various estimated models

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 \( \eta_1, \eta_2, \eta_3 \) and \( \eta_4 \) representing four independently distributed \( N(0,1) \) variates, draws from the distribution for the four inter-respondent distributed coefficients are obtained as:

\[
\begin{align*}
\beta_{FF} &= \beta_{FF,\mu} + s_{1,1}\eta_1 \\
\beta_{TC} &= \beta_{TC,\mu} + s_{2,1}\eta_1 + s_{2,2}\eta_2 \\
\beta_{SDT} &= \beta_{SDT,\mu} + s_{3,1}\eta_1 + s_{3,2}\eta_2 + s_{3,3}\eta_3 \\
\beta_{TOLL} &= \beta_{TOLL,\mu} + s_{4,1}\eta_1 + s_{4,2}\eta_2 + s_{4,3}\eta_3 + s_{4,4}\eta_4
\end{align*}
\]

(10)

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

The four \( \gamma \) parameters relate to the standard deviations for the intra-respondent distributed coefficient values, where the mean value if zero. Finally, the parameter \( \sigma \) multiplies the four alternative-specific \( N(0, 1) \) error components.

Finally, Table 4 also 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 in the two models. The first observation relates to the inclusion of the additional error components in model 14. Firstly, it can be seen that the estimation of \( \sigma \), leads a highly significant improvement in model fit by 104.1 units in log-likelihood. Secondly, there is no clear trend in the relative standard errors between the two models, which is consistent with earlier results by Ortúzar et al. (2000c). Overall, the parameters in model 14 obtain higher levels of significance than their counterparts in model 13.

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 in both models. All four coefficients have high standard deviation, with coefficients of variation ranging from 0.58 to 0.83 in model 13 and from 0.57 to 1.04 in model 14. In terms of correlation, we get positive correlation between \( \beta_{FF} \) and \( \beta_{SDT} \), along with positive correlation between \( \beta_{TC} \) and \( \beta_{TOLL} \). Respondents with a high aversion to slowed down travel time have a lower sensitivity to toll, which is consistent with intuition. The positive correlation between \( \beta_{TC} \) and \( \beta_{SDT} \) on the other hand is surprising, and is higher in model 14. However, in the interpretation of these correlations, the higher standard errors for some of the Choleski factors need to be kept in mind.

Next, we move to the parameters associated with intra-respondent variation. Here, some interesting differences arise between model 13 and model 14. Indeed, other than for \( \gamma_{TC} \), the relative standard errors are much lower in model 14 than in model 13. Additionally, the relative values of the actual estimates are also higher in model 14, again with the exception of \( \gamma_{TC} \). In other words, including the additional error components in the model seems to, with the present data, increase the scope for retrieving intra-respondent variation in tastes. This also becomes apparent when comparing the model fit of 14 and 13 to that of models 10 and 9 respectively.

The final observation relates to the estimate for \( \sigma \), the factor multiplying the alternative-specific error components. The estimate is highly significant. The
covariance between the unobserved part of utility for alternative \( j \) in choice situation \( t_1 \) and \( t_2 \) is equal to \( \sigma^2 \), with the variance in each case being equal to \( \frac{\pi^2}{6} + \sigma^2 \). As such, the correlation between the unobserved parts of utility in separate choice situations for the same respondent is equal to 0.48. A calculation of this calculation across the different model structures shows that the correlation seems to increase with model complexity (cf. Table 5).

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. In model 14, the sensitivity to tolls is higher than sensitivity to travel cost. A similar observation cannot be made in model 13, where the results are inconsistent across \( \beta_{FF} \) and \( \beta_{SDT} \). Here, 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

<table>
<thead>
<tr>
<th>Model</th>
<th>Corr.</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>0.33</td>
</tr>
<tr>
<td>4</td>
<td>0.34</td>
</tr>
<tr>
<td>6</td>
<td>0.39</td>
</tr>
<tr>
<td>8</td>
<td>0.37</td>
</tr>
<tr>
<td>10</td>
<td>0.41</td>
</tr>
<tr>
<td>12</td>
<td>0.42</td>
</tr>
<tr>
<td>14</td>
<td>0.48</td>
</tr>
<tr>
<td>16</td>
<td>0.45</td>
</tr>
</tbody>
</table>

Table 5: Correlation in unobserved utility terms for given alternative across replications for same individual
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 turn, this trend has meant that applications making use of closed form models generally treat SP data as purely cross-sectional.

In this paper, we question the validity of relying solely on one or the other method. In one direction, we discuss the development of approaches that can be used to recognise the repeated choice nature of SP data independently of a specification of random taste heterogeneity. Here, we make use of error components that introduce correlation across choice situations for the same respondent, where the level of correlation is the same across all alternatives.

The other topic that is given slightly more consideration in this paper is the assumption of within individual homogeneity in tastes when making use of the typical panel formulation in a random coefficients model. 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 have estimated 16 different models offer varying treatments of the repeated choice nature of the dataset. Our findings can be summarised in three main points:

1. There is clear evidence to suggest that the majority of taste heterogeneity within the present SP dataset derives from variation in tastes across individual respondents, giving the panel approach a significant advantage over a purely cross-sectional approach.

2. There is however also evidence to suggest the presence of some within re-
spondent heterogeneity, such that the sensitivity to the various attributes
does indeed vary across choice situations for the same respondent.

3. Accounting for the correlation across replications for the same individual
independently of a treatment of random taste heterogeneity, in this case
with the help of error components, leads to highly significant improvements
in model fit.

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. Finally, the error components account for serial
correlation across replications that is independent of variation in tastes.

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.

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 for helpful discussions and comments.

References

data for studying the effect of advanced traffic information on drivers’ route


Brazell, 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.


