Paper:
Hess, S and Rose, JM (2009) *Allowing for intra-respondent variations in coefficients estimated on repeated choice data.* Transportation Research Part B: Methodological, 43 (6). 708 – 719.

http://dx.doi.org/10.1016/j.trb.2009.01.007

Allowing for intra-respondent variations in coefficients estimated on repeated choice data

Stephane Hess* John M. Rose†

January 26, 2009

Abstract

Partly as a result of the increasing reliance on Stated Choice (SC) data, the vast majority of discrete choice modelling applications are now estimated on data containing multiple observations for each respondent. At the same time there has been growing interest in the representation of unexplained heterogeneity in choice data, using random coefficients models such as Mixed Multinomial Logit (MMNL). The presence of multiple observations for each respondent can indeed be a great asset in the identification of such variations in tastes. However, in this paper, we question the validity of the common assumption that tastes vary across respondents but stay constant across repeated choices for the same respondent. We extend the existing framework for the MMNL analysis of *panel* data by allowing for intra-respondent heterogeneity on top of inter-respondent heterogeneity. An empirical analysis making use of a SC dataset for route choice confirms our hypotheses and shows that superior performance is obtained by our more general model.

Keywords: taste heterogeneity, panel data, repeated choice, stated choice, mixed logit

1 Introduction

The mixed multinomial logit model (MMNL) is fast becoming one of the most popular mathematical structures for the analysis of choice behaviour[1]. From a practical perspective, the MMNL model has two main advantages over other

*Institute for Transport Studies, University of Leeds, s.hess@its.leeds.ac.uk, Tel: +44 (0)113 34 36611, Fax: +44 (0)113 343 5334

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

[1] See for example McFadden and Train (2000), Hensher and Greene (2003) and Train (2003).

generally used discrete choice structures. Firstly, the model allows for the incorporation of *unexplained* preference heterogeneity, represented through a random distribution of marginal sensitivities across respondents (see for example Cardell and Dunbar, 1980; Ben-Akiva et al., 1993; Hess et al., 2005), a specification typically referred to as the random coefficients logit (RCL) model. Secondly, the model allows for a relaxation of the *IID* assumption of the MNL model, in what is typically referred to as the error components logit (ECL) model, discussed for example by Brownstone and Train (1999) and Walker (2001). This can be exploited with a view to incorporating heteroscedasticity across alternatives or heightened correlation between the errors of a subset of alternatives.

Although mathematically equivalent, the two versions of the MMNL model thus deal with two different sets of issues, and in the present paper, we are solely interested in the *RCL* specification. Random taste heterogeneity in a MMNL model is accommodated via the use of random parameters associated with attributes within the system of utility functions that make up the model. Random parameters represent unexplained preference heterogeneity insofar as the resulting distributions of parameters are dispersed in an unknown *random* manner over the sampled population. With regards to these random parameters, the model is capable of discerning whether or not heterogeneous preferences exist within the sampled population for various attributes, but not where any particular individual resides within a given parameter distribution. As a result, the model does not allow for a ready explanation as to why the differences in preferences exist over the sampled population, just that differences do exist. Whilst posterior analysis may locate the likely location of individual preferences based on their choices observed within the data (see for example Train, 2003; Hensher and Greene, 2003; Hess, 2008b), further analysis is required in order to indicate why preferences differ within the sample (see for example Hess, 2007).

As with other model structures, the majority of MMNL applications are now based on the use of Stated Choice (SC) data. An important characteristic of SC data is the presence of multiple responses for each respondent, something that is also the case with a small subset of Revealed Preference (RP) datasets, such as for example with travel diary data. It has long been recognised that treating repeated choice data in the same way as cross-sectional data, i.e. making an assumption of independence between choices for the same respondent, may not be appropriate. The primary concern has tended to be that such an assumption will result in biased standard errors for the estimated parameters (cf. Ortúzar and Willumsen, 2001). In this context, a dedicated body of research has looked at ways of correcting the standard errors after estimation, using techniques such as bootstrap and jackknife (e.g. Cirillo et al., 2000; Ortúzar et al., 2000), where the use of these approaches is independent of the assumed model structure, but where it has been used almost exclusively on models assuming an absence of unexplained taste heterogeneity.

In the context of the present paper, it can be argued that having information on multiple choices for the same respondent is a crucial asset in identifying variations in sensitivities in a sample population. Indeed, the discussions by Fosgerau and Nielsen (2007) suggest that with only a single observation for each respondent, it is difficult to distinguish between random taste heterogeneity and the *IID* extreme value terms in the model. In its most basic specification however, the MMNL model would indeed treat *panel* data in the same way as it would treat *cross-sectional* data, i.e. regarding separate observations from the same individual as if they came from different individuals. Aside from the above concerns on biased standard errors and the retrieval of taste heterogeneity, it was recognised that such an approach would also give an inadequate representation of taste heterogeneity, culminating in the work by Revelt and Train (1998).

The framework of Revelt and Train accommodates the repeated choice nature of the data by assuming that tastes vary across respondents in the sample, but stay constant across observations for the same respondent. In other words, this allows for inter-respondent heterogeneity but assumes intra-respondent homogeneity in tastes. In a footnote, Revelt and Train (1998) mention the possibility of generalising the framework to a situation where coefficients vary across choices, but state that "*our data consist of repeated choices within a survey, such that the assumption of β_n constant over choices seems reasonable*". The assumption of inter-respondent heterogeneity along with intra-respondent homogeneity is now commonplace in the MMNL literature, and in reviewing the transportation literature published over the past few years, the authors could not discern a single article that did not use the approach proposed by Revelt and Train (1998) when allowing for the repeated choice nature of the data.

In questioning the validity of the assumption of intra-respondent homogeneity, it is important to make a distinction between SC and RP data. The most plausible argument for the assumption of intra-respondent homogeneity in SC data is that such data represent a virtually instantaneous panel, which is consistent with the above argument by Revelt and Train. Unlike more traditional panel data captured over longer periods of time where external influences may impact upon a person's attitudes and preferences (e.g., influences of media and discussions with friends and family may sway an individual's thinking), such influences are unlikely to exist during a single survey sitting.

Nevertheless, there appear to be various reasons why the preferences of individual respondents may indeed vary across choice situations, even whilst taking part in a survey. For example, it is known that respondents do, over the course of a SC experiment, experience both learning and cognitive burden effects (e.g. Deshazo and Fermo, 2002; Arentze et al., 2003). One way that these effects have been observed to manifest themselves is in terms of the response times of individuals completing SC choice tasks (cf. Haaijer et al., 2000; Rose and Black, 2006). A possible consequence of such learning and boredom effects is that, in the first few choice situations, when respondents are still in the learning phase, they simplify the choice situations and maybe focus only on some of the attributes (essentially implying a zero coefficient for some attributes in some choice situations). After this initial phase, the respondents might be better able to process the full amount of information. However, later on, as fatigue and boredom set in, the respondents might again simplify the tasks and focus only on some of the attributes. Clearly, the assumption of constant marginal utility coefficients across choice situations would in this case not be justified[2].

In the context of learning and fatigue effects, there is also a possibility of a larger relative weight for the unobserved part of utility in some of the choice situations, leading to changes in the absolute values of the marginal utility coefficients (though not necessarily their ratios) across choice situations. Additionally, the very nature of SC data, with significant differences in the trade-offs presented across choice situations, may lead to complicated patterns of non-linearities in response that are difficult to retrieve in a non-random manner.

The situation with data on real world choices is slightly different given that, unlike SC data, RP panel data are normally collected over much longer time horizons than is the case with SC data where only a single *sitting* is used. As a result, such data may provide greater circumstance for the existence of intra-respondent variations in the estimated marginal utilities. One reason that intra-respondent variations in the marginal utilities estimated for respondents may be more likely to be observed in RP panel data is that over longer survey periods, individual specific circumstances have a greater probability of changing, with changes in circumstance, such as moving house, buying a new car, changes in economic circumstance, etc., likely to result in changes in preference[3]. Habit

[2]This highlights the important distinction to be made between variation in marginal utilities and variation in estimated marginal utility coefficients. While the fact that a respondent ignores an attribute in some of the choices does not imply a zero valuation, it does imply a zero marginal utility coefficient within a utility maximisation framework.

[3]As an example, Cirillo and Axhausen (2006) argue that time constraints may vary from the short to longer term and hence impact on respondent's values of time over these two time periods. Using data collected from a six week travel diary, they compare MMNL logit model results with and without allowing for the panel nature of the data. Whilst Cirillo and Axhausen find that the panel version of the model does indeed provide better statistical fit, their results also suggest the existence of intra-respondent variations in the estimated marginal utilities for cost and time parameters, albeit with a lower degree of heterogeneity than is the case for the inter-respondent variations.

formation and state dependence are also more likely to arise in the case of real world choice data.

This discussion has shown that the assumption of intra-respondent homogeneity in marginal utility coefficients may not always be justified. While not questioning the underlying reasoning for assuming that the majority of taste heterogeneity should be across respondents, this paper makes a case for the extension of the standard panel framework discussed by Revelt and Train (1998). Specifically, we introduce a version of the MMNL model that allows for both constant and non-constant marginal utilities across choices for the same respondent. This is a timely extension of existing methodology, not least due to recent results by Hess (2008a) which suggest that the cross-sectional and panel specifications are unable to produce consistent parameter estimates in the face of data that exhibits both inter-respondent and intra-respondent heterogeneity in tastes.

Our proposed model is a natural extension of the existing methodology that breaks free from the restrictive assumptions of both the cross-sectional formulation and the Revelt and Train panel formulation. It should be acknowledged that in an ideal situation, factors leading to intra-respondent heterogeneity (such as learning, fatigue or indeed non-linearities) should be dealt with directly in the observed part of utility, for example using lagged variables in a RP setting (e.g. Honore and Kyriazidou, 2000). However, just as with variations in sensitivities across respondents, there are limits to what can retrieved by a deterministic approach, again providing motivation for the development of a model approach that can account for any remaining *random* heterogeneity.

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

2 Methodology

In this section, we develop our new modelling framework that allows for the joint representation of inter-respondent and intra-respondent heterogeneity. We begin with a discussion of existing model structures.

2.1 Basic notation and MNL model

Let $U_{i,n,t}$ be the utility of alternative i for respondent n in choice situation t, where this consists of a *modelled* component $V_{i,n,t}$, commonly referred to as observed utility, and an unobserved component $\varepsilon_{i,n,t}$ such that:

$$U_{i,n,t} = V_{i,n,t} + \varepsilon_{i,n,t}, \tag{1}$$

where it is common practice to assume a linear relationship between attributes and tastes, such that:

$$U_{i,n,t} = \beta_{n,t} x_{i,n,t} + \varepsilon_{i,n,t}, \tag{2}$$

with $\beta_{n,t}$ giving a vector of taste coefficients[4] and $x_{i,n,t}$ giving a vector of attributes describing alternative i as experienced by respondent n in choice situation t.

Under the further assumptions that the unobserved components are identically and independently distributed according to a type I extreme value distribution, and that the parameters β are fixed across the population and across choice situations, the probability that respondent n chooses alternative i in choice situation t is given by the Multinomial Logit (MNL) model (see McFadden, 1974), with:

$$P_{n,t}\left(i \mid \beta\right) = \frac{e^{V_{i,n,t}}}{\sum_{j=1}^{J} e^{V_{j,n,t}}}, \tag{3}$$

where J gives the number of alternatives faced by respondent n in choice situation t.

2.2 Traditional cross-sectional and panel formulations of MMNL

In a model allowing for random taste heterogeneity, such as MMNL, the vector of taste coefficients β is assumed to follow a certain random distribution in the sample, such that we have $\beta \sim f(\beta \mid \Omega)$, with Ω representing a set of parameters of the (multivariate) distribution of β.

Two main approaches exist in this context. The cross-sectional specification is the standard approach for data containing a single observation for each respondent. With this specification, all observations are treated as independent, meaning that, if used on repeated choice data, separate observations from the same respondent are treated in the same way as if they came from separate respondents. From a taste heterogeneity perspective, this means that sensitivities vary across choices for a given respondent in the same way that they vary across individual respondents[5]. The panel approach discussed by Revelt and Train (1998) on the other hand was designed specifically for repeated choice data. This specification explicitly takes the nature of the data into account by treating the sample

[4] For respondent n in choice situation t.

[5] It should again be noted that the work of Fosgerau and Nielsen (2007) suggests problems in identifying taste heterogeneity in a purely cross-sectional model, an observation confirmed in empirical tests by Hess (2008a) independently of whether the data come from cross-sectional or panel surveys.

in blocks of observations, one for each respondent. From a taste heterogeneity perspective, this means that the distribution of sensitivities is across respondents with tastes staying constant across observations for the same respondent.

The specific approach used in a given application is accommodated by adapting the specification of the log-likelihood function. With $j_{n,t}$ giving the choice for respondent n in choice situation t, the log-likelihood function in a MMNL model is given by:

$$LL(\Omega) = \ln\left[E\left(\prod_{n=1}^{N}\prod_{t=1}^{T_n} P_{n,t}\left(j_{n,t} \mid \beta_{n,t}\right)\right)\right], \tag{4}$$

where this is conditional on the parameters Ω of the distribution of β[6], N gives the total number of respondents, and T_n gives the number of choice situations for respondent n. The expectation in Equation 4 arises as the model assumes that (some of) the parameters are randomly distributed in some fashion over the sampled population.

If observations between any two respondents are assumed to be independent, then using the fact that $E(n_1 n_2) = E(n_1)E(n_2)$ in conjunction with the familiar log rule that $\ln[E(n_1)E(n_2)] = \ln[E(n_1)] + \ln[E(n_2)]$, the log-likelihood of the model becomes:

$$LL(\Omega) = \sum_{n=1}^{N} \ln\left[E\left(\prod_{t=1}^{T_n} P_{n,t}\left(j_{n,t} \mid \beta_{n,t}\right)\right)\right]. \tag{5}$$

If we further assume that each choice situation is assumed to be independent to all other choice situations, even if two choice situations relate to the same respondent, we arrive at the cross-sectional formulation of the MMNL model. Using the same mathematical principles as before, Equation 5 is then rewritten as:

$$LL(\Omega) = \sum_{n=1}^{N}\sum_{t=1}^{T_n} \ln\left[E\left(P_{n,t}\left(j_{n,t} \mid \beta_{n,t}\right)\right)\right], \tag{6}$$

where, when replacing expectation with integration, this equates to:

$$LL(\Omega) = \sum_{n=1}^{N}\sum_{t=1}^{T_n} \ln\left(\int_{\beta_{n,t}} P_{n,t}\left(j_{n,t} \mid \beta_{n,t}\right) f\left(\beta_{n,t} \mid \Omega\right) \mathrm{d}\beta_{n,t}\right). \tag{7}$$

[6]Where $\beta_{n,t}$ is the specific value of β for respondent n in choice situation t and where Ω is a population-level parameter.

In simulating the log-likelihood, this implies that the simulated draws are taken at the level of each individual choice task, and hence the parameters vary both within and between individual respondents[7].

Under the panel formulation of the MMNL model as outlined in Revelt and Train (1998), we no longer assume that choice tasks undertaken by the same respondent are independent. As a result, we have that $E(t_1 t_2) \neq E(t_1) E(t_2)$, and the simplification from Equation 5 to Equation 6 no longer applies. Instead, the log likelihood function now becomes:

$$LL(\Omega) = \sum_{n=1}^{N} \ln \left(\int_{\beta_n} \prod_{t=1}^{T_n} P_{n,t}(j_{n,t} \mid \beta_n) f(\beta_n \mid \Omega) \, \mathrm{d}\beta_n \right), \tag{8}$$

where $\beta_{n,t}$ has been replaced by β_n to illustrate that the tastes stay constant across choices. Unlike Equation 7, Equation 8 implies that the simulated draws are invariant within respondents. Further, whereas the probability in Equation 7 represents the individual probability that respondent n chooses alternative $j_{n,t}$ in choice situation t, Equation 8 shows that what is being modelled in the panel formulation of the MMNL model is actually the probability that respondent n is observed to make a *sequence of choices* over the T_n choice tasks, where, with the present modelling approach, the order of the sequence is of no importance.

2.3 Generalised MMNL specification

The panel formulation of the MMNL model, as outlined in Revelt and Train (1998), represents the dominant form of the model used within the transportation literature today when dealing with repeated choice data. The cross sectional formulation of the model represents the opposite end of the spectrum in terms of the assumptions made regarding the distribution of marginal utilities within the sampled population. In this paper, we present a middle way between the two approaches. Specifically, we develop a generalised specification of the MMNL model that allows for two sets of parameters. In this generalised model, the elements in the first set of parameters vary only across respondents, while the elements in the second set of parameters vary across all observations. As a result, our model still maintains some inter-dependence between tastes across choice situations for the same respondent but additionally allows for differences in the marginal utilities across choices for a given individual. In the context of the earlier discussions, this would mean that the simplification from Equation 5 to

[7] Here, it can be seen that by dropping the integral, i.e. in the absence of random parameters, we arrive back at the log-likelihood function for the MNL model.

Equation 6 only applies in part, whereas in the panel formulation, it does not apply at all.

In our model, we let $\beta_{n,t}$ be a function of two terms, α_n and $\gamma_{n,t}$, where α_n varies over respondents with density $g(\alpha_n \mid \Omega_\alpha)$ and $\gamma_{n,t}$ varies over all choices with density $h(\gamma_{n,t} \mid \Omega_\gamma)$. The value for the combined taste coefficient in a given choice situation is then obtained as the sum of individual-specific draws from $g(\alpha_n \mid \Omega_\alpha)$ and observation-specific draws from $h(\gamma_{n,t} \mid \Omega_\gamma)$[8], i.e. $\beta_{n,t} = \alpha_n + \gamma_{n,t}$. The mean of the combined taste coefficients in $\beta_{n,t}$ is captured in α_n, such that we set the mean of the elements in $\gamma_{n,t}$ to zero and estimate only their spread or standard deviation. Under these assumptions, the log likelihood of the model is given by:

$$LL(\Omega) = \sum_{n=1}^{N} \ln \left[\int_{\alpha_n} \prod_{t=1}^{T_n} \left(\int_{\gamma_{n,t}} P_{n,t}(j_{n,t} \mid \beta_{n,t})\, h(\gamma_{n,t} \mid \Omega_\gamma)\, \mathrm{d}\gamma_{n,t} \right) g(\alpha_n \mid \Omega_\alpha)\, \mathrm{d}\alpha_n \right], \tag{9}$$

where there is now an integral inside the product over choices as well as outside[9]. The integral outside the product over choices accounts for inter-respondent heterogeneity, while the integral inside the product accounts for intra-respondent heterogeneity. Given the theoretical work by Fosgerau and Nielsen (2007) and the results by Hess (2008a), it should be noted that identification issues in relation to intra-respondent heterogeneity may also arise in this joint specification. However, at least in the application used in this paper, the degree of intra-respondent heterogeneity was so significant that variations could be retrieved for three out of the four coefficients (see Section 3).

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 9 needs to be approximated through simulation. Looking specifically at the contribution of respondent n to the likelihood (i.e. a single element in the summation over n in Equation 9), we make use of R draws from $g(\alpha \mid \Omega_\alpha)$ and $R\,T_n$ draws from $h(\gamma \mid \Omega_\gamma)$, such that a separate set of R draws from $h(\gamma \mid \Omega_\gamma)$ is used for each observation for respondent n. With this in mind, the subcomponent for respondent n in Equation 9 is approximated

[8]Here, it would also be possible to use a multiplicative approach instead of an additive approach. This would allow the degree of intra-respondent heterogeneity to be linked to the respondent-specific mean coefficient, but would further increase estimation complexity.

[9]Here, 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 looks at the joint estimation on Revealed Preference (RP) and SC data, and where the within respondent integration is used to account for scale differences.

by:

$$SLL_n(\Omega) = \ln\left[\frac{1}{R}\sum_{r=1}^{R}\left(\prod_{t=1}^{T_n}\left(\frac{1}{L}\sum_{l=1}^{L}P_{n,t}\left(j_{n,t} \mid \alpha_r, \gamma_{l,t}\right)\right)\right)\right], \tag{10}$$

where draws for the combined coefficient β are obtained as the sum of individual-specific draws α_r and observation-specific draws $\gamma_{l,t}$, 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[10], and for the purposes of the present paper, the *correct* simulation approach from equation 10 was used. Here, 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

This section presents the results of an empirical application making use of our proposed model structure to test for the presence of intra-respondent heterogeneity in repeated choice data. A SC dataset was chosen for the present analysis, partly for availability reasons. However, we also feel that any evidence of intra-respondent heterogeneity in this context would highlight the importance of the issue even more, given the high reliance on SC data in applied work, but also the fact that the assumption of intra-respondent homogeneity may be seen as very reasonable in this (instantaneous panel) context[11].

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 the valuation of travel time savings (VTTS) for car

[10] As an example, it is our understanding that in the widely used BIOGEME package (Bierlaire, 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. The draws used for any cross-sectional terms are observation-specific, thus at least partly recognising the difference between respondent-specific and observation-specific coefficients. However, recent work by Hess (2008a) shows that this approach fails to retrieve the correct parameter estimates, even with very high numbers of draws, casting doubts on its consistency with the proposed structure.

[11] See e.g. the reasoning by Revelt and Train (1998).

drivers in the Sydney metropolitan area in the context of new toll road developments. For this paper, we use only data collected for respondents undertaking non-commuting trips.

As part of the survey task, respondents were asked information about a recent trip that they had undertaken and which could potentially have used the proposed toll road had it been in existence. This information was then used to frame the context of the SC experiment. Based on the actual trip attribute levels 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). The SC experiment was constructed using efficient experimental design methods. For a review of efficient SC design methods, see Bliemer and Rose (2006) or Ferrini and Scarpa (2007).

In all cases, the first alternative shown presented the respondent with the attribute levels faced on their recent trip as reported. The remaining two alternatives represented competing hypothetical routes. The first alternative remained invariant across the 16 choice situations with only the levels of the remaining two alternatives varying. An example choice situation (taken from a practice game) is shown in Figure 1. The final sample consisted of 205 effective interviews, leading to an estimation sample of 3,280 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 there, we move to a specification that uses distribution of the taste coefficients across respondents rather than across observations (i.e. the Revelt and Train panel formulation), where again, in the more complicated model, 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 is a generalisation of model 2, which in turn is a generalisation of model 1. Similarly, model 5 is a generalisation of model 4, which is again a generalisation of model 1. Model 6 is a generalisation of both model 4 and model 2, while model

7 generalises both model 6 and model 5. Finally, model 8 generalises all other models. It should be noted that, with our specification, the models with some inter-respondent variation in tastes differ slightly from the models with intra-respondent variation only, as, in the former, we set the mean of any intra-agent variation to zero. From a practical point of view, the two are formally equivalent as it does not matter where the mean is added, such that likelihood ratio tests can still be used for comparisons.

Before proceeding, a few words should be said about correlation. Models 3, 5, 7 and 8 allow for correlation between randomly distributed coefficients[12]. Such an approach can for example allow for the (rather typical) situation in which respondents with high travel time sensitivity are more likely to have low cost sensitivity, with the converse being the case for respondents with low time sensitivity. On the other hand, allowing for correlated coefficients can also reduce the risk of confounding between heterogeneity in scale and heterogeneity in relative marginal utilities. Indeed, a situation in which all respondents have roughly similar relative sensitivities (e.g. VTTS) but different absolute sensitivities would still lead to estimates showing a high degree of heterogeneity in marginal utilities across respondents. When allowing the coefficients to be correlated, a high positive correlation for all pairs of coefficients would point towards scale differences rather than differences in relative sensitivities. It is useful to keep this in mind especially in the analysis of the results from models 7 and 8 given the earlier point about absolute sensitivities changing over the course of a SC experiment as a result of boredom, learning and fatigue.

We now turn our attention to model specification. In the analysis presented in this paper, only four of the attributes included in the SC survey were used, where the exclusion of travel time variability was based on the fact that a large share of respondents indicated that they had consistently ignored this attribute, and where very low levels of significance were 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 first two alternatives. 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 *current* option), 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

[12]No correlation is allowed for in the intra-respondent distributions in model 7.

appropriate simplifications applying for other models.

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

Some clarifications are required. The constant δ_j is set to zero for $j = 3$, i.e. for the second of the two hypothetical alternatives. The four main taste coefficients ($\alpha_{FF,n}$, $\alpha_{TC,n}$, $\alpha_{SDT,n}$, and $\alpha_{TOLL,n}$) 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. 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 in models 2, 3, 6, 7 and 8, where these coefficients are specified to vary across all observations, independently of the respondent. In models 6 to 8, the mean for the elements in β is captured by α, such that the means of the elements in γ are set to zero. Finally, in models 3 and 8, we additionally allow for correlation between the four coefficients.

With the scope of the present paper being firmly methodological rather than policy-oriented, Normal distributions are used in all models. This decision was taken for the sake of simplicity and ease of estimation, but does imply the usual problems with the computation of willingness to pay (WTP) indicators (cf. Hess et al., 2005). For this reason, no WTP indicators are computed in the present paper. Finally, it should also be noted that in all models estimated here, a linear-in-attributes specification of the utility functions is used. This is based on preliminary analyses that did not reveal consistent and significant non-linearities in response with the data at hand[13].

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 give an overview of the results across models, in conjunction with detailed results for two

[13] Along with standard specifications of non-linearity in response, we also tested for learning and fatigue effects by allowing the sensitivities to be different in the first few choice situations and in the last few choice situations. However, none of these differences proved to be significant.

selected models. Additional 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 parameters[14]. Finally, where appropriate, the p-values for LR tests are shown, with, as an example, p_2 giving the p-value for a LR test against model 2.

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. Here, there are statistically significant improvements in model fit when moving from model 2 to model 3, where the same applies 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 now only significant at the 74% level. These results suggest that allowing for correlation between the intra-respondent coefficients is not advisable with the present data when simultaneously allowing for inter-respondent variation in tastes. This result showing an absence of significant correlation would thus also indicate that the intra-respondent heterogeneity retrieved in models 7 and 8 is not a result of variations in scale.

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. Indeed, the gains in log-likelihood when moving from model 4 to model 6 are only significant at the 91% level, while the improvements for model 7 are significant at the highest levels in comparison with models 4, 5 and 6.

[14]It should be noted that the models presented here include all coefficients for the appropriate formulation, with insignificant estimates not removed from the models.

On the basis of the above discussion, the *recommended* model structure is model 7, i.e. the structure allowing for inter-personal as well as intra-personal variation in tastes, where correlation between coefficients however only exists for the former.

We now proceed with the description of the detailed estimation results for the recommended model structure. Here, Table 3 presents the results for model 7 alongside those of model 5, with the latter representing current practice in MMNL analyses. Here, some explanations are required to supplement those from Section 3.2. The parameter $\alpha_{FF,\mu}$ 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 table also presents the standard deviations for these coefficients (e.g. $\alpha_{FF,\sigma}$ for free flow time), along with the correlations between the four taste coefficients in their distribution across respondents. Both the standard deviations and correlations were calculated using a Cholesky transformation for multivariate Normals (cf. Train, 2003, pp.211-212), where standard errors were calculated analytically on the basis of the covariance matrix of the Cholesky matrix. Finally, the four different γ parameters relate to the standard deviations for the intra-respondent distributed coefficient values, where the mean value is zero.

We now proceed with a detailed analysis of the results. No further discussion is required on model fit, where, as already mentioned above, model 7 shows a small but statistically significant improvement in fit over model 5. Turning to the actual model parameters, the significant and positive estimates for both alternative specific constants (in both models) 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. Here, the sensitivity to slowed down time is slightly higher than the sensitivity to free flow time, with the sensitivity to tolls being higher than the sensitivity to running costs. For all four attributes, we identify significant variations in sensitivity *across* respondents where the relative level of variation is highest for α_{FF} and lowest for α_{SDT}. In terms of correlation, we get the expected positive correlation between α_{FF} and α_{SDT}, along with positive correlation between α_{TC} and α_{TOLL}. There is also low negative correlation between α_{FF} and α_{TC} but this is only statistically significant in model 7. Correlation levels for other pairs of coefficients are relatively low and do not attain high levels of statistical significance.

Moving on to the parameters associated with intra-respondent variation, we obtain highly significant estimates for γ_{FF}, γ_{TC} and γ_{SDT}, while 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; the variation in sensitivities across respondents is larger than the variation across choices for the same respondent.

As a final step, we now conduct a comparison of the results across the two models. We first note that, on average, the estimates for model 7 are slightly larger than those in model 5, indicating a reduction in the relative weight of the error (compared to the *modelled* part of utility) when allowing for intra-respondent heterogeneity on top of inter-respondent heterogeneity. However, the rate of increase varies across the various common parameters, indicating differences in relative sensitivities as well as absolute sensitivities.

In this paper, we are clearly especially interested in the findings in terms of heterogeneity. Model 5 assumes intra-respondent homogeneity while model 7 allows for intra-respondent heterogeneity. The easiest way to process the estimates in this context is to look at the degree of heterogeneity in terms of the coefficients of variation, i.e. the ratio between the standard deviation and the mean parameter. The results of these calculations are summarised in Table 4, showing the degree of variation both across respondents and across choices for a given respondent. Starting with the common component, i.e. inter-respondent heterogeneity, we note that, with the exception of toll sensitivity, there is a reduction in the degree of inter-respondent heterogeneity when moving from model 5 to model 7. When noting that toll is the only attribute for which we could not retrieve any intra-respondent heterogeneity, this gives a strong indication that intra-respondent heterogeneity is confounded with inter-respondent heterogeneity in model 5, hence overstating the degree of variation in tastes *across* respondents.

Looking at the degree of intra-respondent heterogeneity, we can, as noted above, see that this is lower than the degree of inter-respondent heterogeneity. Nevertheless, the degree of intra-respondent degree is clearly not negligible. As an example, for the free flow travel time coefficient, the variation across choices for the same respondent is still almost half as high (46%) as the variation across respondents, with rates of 65% for the travel cost coefficient and 52% for the slowed down time coefficient.

As a final step, we can also compare the findings in terms of correlation across the two models. Aside from the earlier point about the correlation between α_{FF} and α_{TC} in model 7, we note a drop in the correlation between α_{FF} and α_{SDT} when moving from model 5 to model 7, along with an increase in the correlation between α_{TC} and α_{TOLL}.

4 Summary and conclusions

Given the high reliance on repeated choice data, be it from RP or SC surveys, 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 such 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 a result, any imprecise valuation attributable to incorrectly specified 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 accommodate random variations in sensitivities in the estimation of models on repeated choice data. Specifically, we have questioned the validity of the general assumption made in such work that all variations in tastes are restricted to be across respondents, i.e. enforcing intra-respondent homogeneity in sensitivities. We argue specifically that for a host of reasons, including non-linearities in response, learning and fatigue effects, thresholds and variations in scale across choice situations, the assumption of constant marginal utility coefficients may not be adequate. While some of these differences across choices may be restricted to difference in *estimated marginal utility coefficients* rather than differences in *actual marginal utilities*, a comprehensive modelling framework should still account for such differences.

The methodological contribution of the paper comes in the development of a model that allows for intra-respondent and inter-respondent taste heterogeneity at the same time. In an empirical analysis on SC data collected in Sydney, we find that the majority of heterogeneity in the marginal utility coefficients within the present data does indeed derive from variation 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 estimated sensitivity to the various attributes does indeed vary across choice situations for the same respondent. Furthermore, when not allowing for this intra-respondent heterogeneity, the degree of inter-respondent heterogeneity is overstated for three out of the four coefficients. 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 estimated marginal utility for each of the attributes.

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. Further results from other data sets are required before concrete conclusions can be drawn. This should include the use of simulated data in a systematic Monte Carlo study as well as data from RP surveys. Secondly, future applications should look at the implications in terms of WTP indicators, where this was not possible in the present study due to the reliance on the Normal distribution. Here, the use of alternative (and more flexible) distributions would also be interesting in testing whether the mean effects are more likely to be biased when not accounting for intra-respondent heterogeneity in such a case[15]

Finally, it should be noted again that in this paper, we have not aimed to examine in detail the reason for the violation of the assumption of constant marginal utility coefficients across replications. This should be the topic of a secondary stage in an analysis. Indeed, we recognise that the effects retrieved here may not solely be due to actual intra-respondent taste heterogeneity but may also be due to non-linearities in response[16], thresholds in sensitivities and various other factors discuss in Section 1, including variations in scale. 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

This paper is in parts based on work conducted while the first author was employed at ETH Zurich. Parts of the work were also 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. During the final stages of the research discussed in this paper, the first author was funded by a Leverhulme Early Career Fellowship. The authors would like to thank Andrew Daly, Mogens Fosgerau and Stefan Mabit for helpful discussions and comments.

[15]Even in the present application, the change in the mean parameters between model 5 and model 7 was not uniform across the four coefficients.

[16]It should again be noted that preliminary tests did not reveal consistent patterns of non-linear response, meaning that such non-linearities would have to be attribute-level specific.

References

Arentze, T., Borgers, A., Timmermans, H., DelMistro, R., 2003. Transport stated choice responses: effects of task complexity, presentation format and literacy. Transportation Research Part E 39 (3), 229–244.

Ben-Akiva, M., Bolduc, D., Bradley, M. A., 1993. Estimation of travel choice models with randomly distributed values of time. Transportation Research Record 1413, 88–97.

Bhat, C. R., Castelar, S., 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 36 (7), 593–616.

Bierlaire, M., 2005. An introduction to BIOGEME Version 1.4. biogeme.epfl.ch.

Bliemer, M. C. J., Rose, J. M., 2006. Designing Stated Choice Experiments: State-of-the-art. paper presented at the 11^{th} International Conference on Travel Behaviour Research, Kyoto, Japan.

Brownstone, D., Train, K., 1999. Forecasting new product penetration with flexible substitution patterns. Journal of Econometrics 89 (1-2), 109–129.

Cardell, S., Dunbar, F., 1980. Measuring the societal impacts of automobile downsizing. Transportation Research Part A 14 (5-6), 423–434.

Cirillo, C., Axhausen, K. W., 2006. Evidence on the distribution of values of travel time savings from a six-week diary. Transportation Research Part A 40 (5), 444–457.

Cirillo, C., Lindveld, K., Daly, A., 2000. Eliminating bias due to the repeated measurements problem in SP data. In: Ortúzar, J. de D. (Ed.), Stated Preference Modelling Techniques: PTRC Perspectives 4. PTRC Education and Research Services Ltd, London.

Deshazo, J. R., Fermo, G., 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., Scarpa, R., 2007. Designs with a-priori information for nonmarket valuation with choice-experiments: a monte carlo study. Journal of Environmental Economics and Management, 53 (3), 342–363.

Fosgerau, M., Nielsen, S. F., 2007. Deconvoluting preferences and errors: a semi-nonparametric model for binomial panel data. Working paper. Technical University of Denmark, Copenhagen.

Haaijer, R., Wagner, K., Wedel, M., 2000. Response latencies in the analysis of conjoint choice experiments. Journal of Marketing Research 37 (3), 376–382.

Hensher, D. A., Greene, W. H., 2003. The Mixed Logit Model: The State of Practice. Transportation 30 (2), 133–176.

Hess, S., 2007. Posterior analysis of random taste coefficients in air travel choice behaviour modelling. Journal of Air Transport Management 13 (4), 203–212.

Hess, S., 2008a. Approximation issues in simulation-based estimation of random coefficients models. paper presented at the European Transport Conference, Noordwijkerhout, The Netherlands.

Hess, S., 2008b. Conditional parameter estimates from Mixed Logit models: distributional assumptions and a free software tool. ITS working paper. Institute for Transport Studies, University of Leeds.

Hess, S., Bierlaire, M., Polak, J. W., 2005. Estimation of value of travel-time savings using mixed logit models. Transportation Research Part A 39 (2-3), 221–236.

Honore, B. E., Kyriazidou, E., 2000. Panel data discrete choice models with lagged dependent variables. Econometrica 68 (4), 839–874.

McFadden, D., 1974. Conditional logit analysis of qualitative choice behaviour. In: Zarembka, P. (Ed.), Frontiers in Econometrics. Academic Press, New York, pp. 105–142.

McFadden, D., Train, K., 2000. Mixed MNL Models for discrete response. Journal of Applied Econometrics 15, 447–470.

Ortúzar, J. de D., Roncagliolo, D. A., Velarde, U. C., 2000. Interactions and independence in stated preference modelling. In: Ortúzar, J. de D. (Ed.), Stated Preference Modelling Techniques: PTRC Perspectives 4. PTRC Education and Research Services Ltd, London.

Ortúzar, J. de D., Willumsen, L. G., 2001. Modelling Transport, 3rd Edition. John Wiley and Sons, Chichester.

Revelt, D., Train, K., 1998. Mixed Logit with repeated choices: households' choices of appliance efficiency level. Review of Economics and Statistics 80 (4), 647–657.

Rose, J. M., Black, I. G., 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.

Walker, J., 2001. Extended discrete choice models: Integrated framework, flexible error structures, and latent variables. Ph.D. thesis, MIT, Cambridge, MA.

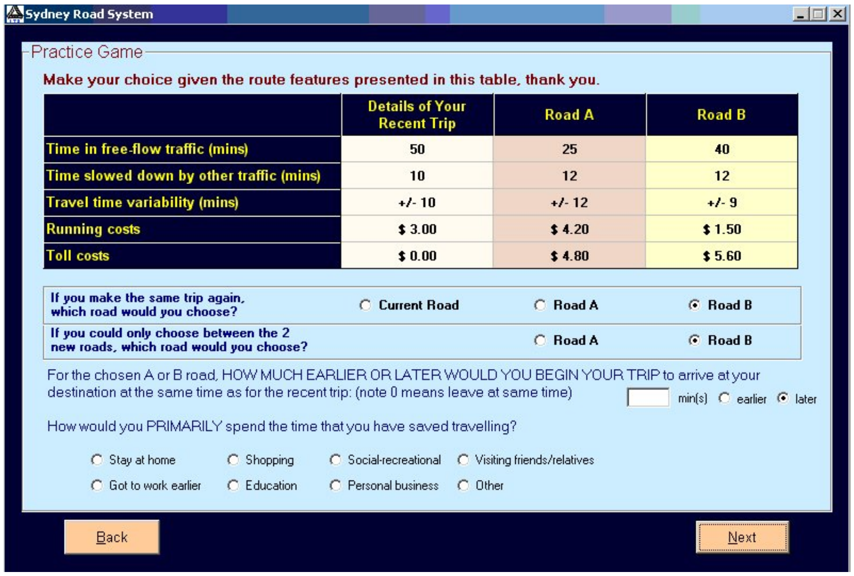

Figure 1: An example of a Stated Choice screen

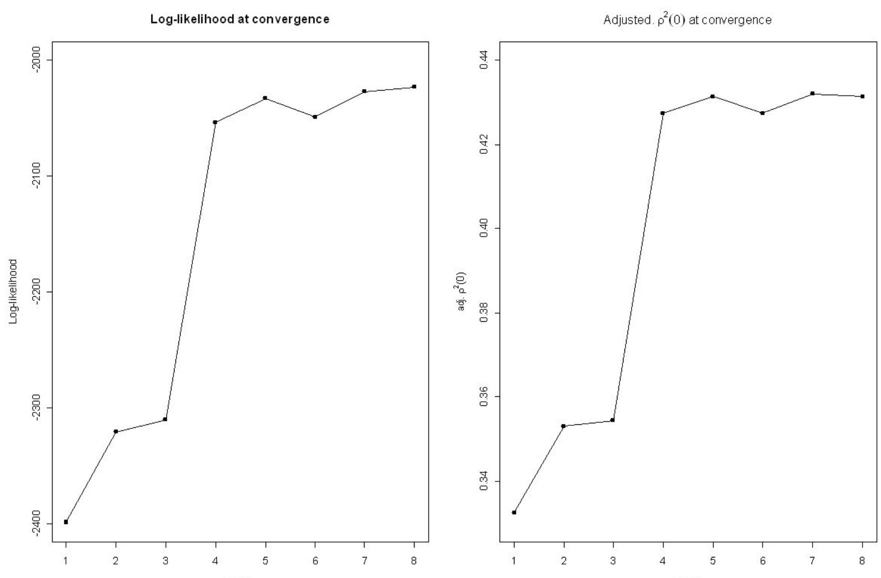

Figure 2: Model fit statistics for estimated models

Table 1: Summary of estimated models

	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 2: Summary of model performance for estimated models

Model	LL	par	adj. ρ^2	p_1	p_2	p_3	p_4	p_5	p_6	p_7
1	-2,399.29	6	0.3325	-	-	-	-	-	-	-
2	-2,321.13	10	0.3531	0	-	-	-	-	-	-
3	-2,310.39	16	0.3544	0	0		-	-	-	-
4	-2,053.29	10	0.4274	0	-	-	-	-	-	-
5	-2,033.27	16	0.4313	0	-	-	0	-	-	-
6	-2,049.21	14	0.4274	0	0	-	0.09	-	-	-
7	-2,026.89	20	0.4320	0	0	-	0	0.01	0	-
8	-2,023.00	26	0.4314	0	0	0	0	0.02	0	0.26

Table 3: Detailed estimation results for model 5 and model 7

	Model 5		Model 7	
LL($\hat{\beta}$)	-2033.27		-2026.89	
parameters	16		20	
adj. $\rho^2(0)$	0.4313		0.4320	

	est.	asy. t-rat.	est.	asy. t-rat.
δ_1	0.2121	3.06	0.2649	3.45
δ_2	0.1723	2.40	0.1976	2.40
$\alpha_{FF,\mu}$	-0.1324	-10.95	-0.1472	-10.72
$\alpha_{TC,\mu}$	-0.6918	-10.14	-0.7583	-12.13
$\alpha_{SDT,\mu}$	-0.1462	-13.00	-0.1668	-12.32
$\alpha_{TOLL,\mu}$	-0.7924	-15.08	-0.8468	-15.53
$\alpha_{FF,\sigma}$	0.1180	10.48	0.1220	10.30
$\alpha_{TC,\sigma}$	0.5182	9.04	0.5037	9.10
$\alpha_{SDT,\sigma}$	0.0924	9.74	0.0959	9.09
$\alpha_{TOLL,\sigma}$	0.5434	10.76	0.5819	11.91
corr(α_{FF},α_{TC})	-0.0282	-0.87	-0.0534	-2.14
corr(α_{FF},α_{SDT})	0.5443	4.15	0.4610	2.53
corr(α_{FF},α_{TOLL})	0.0405	0.32	0.0858	1.02
corr(α_{TC},α_{SDT})	0.2010	1.27	0.1013	0.77
corr(α_{TC},α_{TOLL})	0.5893	5.34	0.6768	12.59
corr(α_{SDT},α_{TOLL})	-0.1559	-1.39	-0.1818	-1.71
γ_{FF}	-	-	0.0565	2.53
γ_{TC}	-	-	0.3282	3.64
γ_{SDT}	-	-	0.0497	3.38
γ_{TOLL}	-	-	0.0057	0.11

Table 4: Coefficients of variation in models 5 and 7

	intra-respondent		inter-respondent	
	Model 5	Model 7	Model 5	Model 7
FF	-	0.38	0.89	0.83
TC	-	0.43	0.75	0.66
SDT	-	0.30	0.63	0.57
TOLL	-	0.01	0.69	0.69