Approximation issues in simulation-based estimation of random coefficients models

Stephane Hess* Kenneth E. Train†

July 7, 2009

Abstract

This paper discusses issues of approximation in the estimation of mixed logit models that contain variation in tastes across respondents and possible additional variation across choices for each respondent. We identify several simplifications to the true log-likelihood function that have been utilised in past work and implemented in existing software. We examine the accuracy of these simplifications through Monte Carlo methods. For models with inter-respondent variation in tastes but no intra-respondent variation, we find that treating the multiple choices of each respondent as if they were choices by different individuals (i.e., treating the panel data as if they were cross-sectional) results in only a small loss of efficiency when sample sizes are sufficiently large. However, for models with intra- as well as inter-variation in tastes, simplifications of the log-likelihood function result in a large loss of efficiency, even with numerous observations. This latter result implies that the computationally intensive formula for the log-likelihood function, which fully accounts for the intra- and inter-respondent variation, needs to be used, at least until an accurate approximation is developed.

1 Introduction

In part as a result of improved estimation performance (cf. Bhat, 2001, 2003; Hess et al., 2006), researchers and practitioners are increasingly making use of the Mixed Multinomial Logit (MMNL) model (cf. McFadden and Train, 2000; Train, 2003) for the representation of random taste heterogeneity across respondents.

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

†Department of Economics, University of California, Berkeley, train@econ.berkeley.edu, Tel: +1 415 291 1023

At the same time, there has been an increasing use of datasets, such as those from Stated Preference (SP) experiments, that contain multiple choices for each respondent. In this context the question arises as to how the presence of multiple choices for each respondent should be accommodated during model estimation and by extension in the representation of taste heterogeneity.

The presence of multiple choices for each respondent can provide important additional information on preferences. However, it was also recognised early on that the fact that multiple observations stem from the same respondent needs to be recognised when interpreting model results. There have especially been concerns that treating multiple response data in the same way as single response (or cross-sectional) data can have biasing effects on the standard errors for the estimated parameters (cf. Ortúzar and Willumsen, 2001). As a result, 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).

These post-estimation correction approaches can be very useful in the case of closed form models such as Multinomial and Nested Logit where the likelihood function for repeated choice data is identical to the corresponding cross-sectional specification. However, in models making use of simulation-based estimation such as MMNL, it is in fact possible to account for the repeated choice nature of the data directly in the specification of the likelihood function, avoiding the need for post-estimation correction approaches. This fact was recognised in the work of Revelt and Train (1998) which is discussed in more detail in the following section.

The main aim of the work by Revelt and Train (1998) was to recognise the repeated choice nature of the data in the representation of random taste heterogeneity. Revelt and Train work on the assumption that tastes vary across respondents, but stay constant across choices for the same respondent. This specification thus makes the assumption that any taste heterogeneity in the data arises from inter-respondent variations as opposed to intra-respondent variations. In later work, Hess and Rose (2009) generalise this approach by relaxing the assumption of intra-respondent homogeneneity of tastes, thereby allowing both intra- and iter-respondent taste variation.

The specification of Revelt and Train (1998), hereafter referred to as the R-T specification, is the most common approach when using MMNL models on data with multiple choices for each respondent. The general conclusion has been that this specification leads to very significant improvements in model fit and a greater ability to retrieve taste heterogeneity. However, as discussed in detail in the following section, this approach requires an adaptation of the likelihood function to be used in model estimation. While this approach has now been implemented in the majority of estimation packages, this implementation has not been universal. As an example, it is the authors' understanding that the R-T formulation has not yet been implemented in ALogit[1], a package that is used especially widely in large-scale applications in a commercial context.

It has been suggested that an approximation to the panel approach could be used by adapting the simulation approach used in a cross-sectional framework (see e.g. Paag et al., 2001). However, the quality of approximation that this approach offers to the R-T specification is unknown, as discussed further in the following section. To some extent, the question could be asked why we should care about this, as the majority of packages now give users the possibility of using the R-T approach, and developers of other packages could be encouraged to also implement this approach. However, this implementation may be difficult in some frameworks and additionally, a large number of applications are still relying on existing versions of the aforementioned software package. Furthermore, as discussed recently by Hess and Rose (2009), approximation approaches could have additional benefits when estimating model structures that accommodate both inter and intra-respondent heterogeneity.

The first aim of this paper is therefore to test the ability of the alternative approaches to estimating the R-T specification. Secondly, we discuss the quality of approximation offered by different specifications to the structure developed by Hess and Rose (2009). From this point of view, this paper aims to answer two questions. Firstly, we discuss whether approximations to the R-T formulation can be used in the estimation of MMNL models on panel data, and secondly, whether such or other approximations may be useful in the estimation of the model structure developed by Hess and Rose (2009).

The remainder of this paper is organised as follows. The following section gives an overview of the various specifications. This is followed in Section 3 by a discussion of the empirical framework used in the analysis. Results of the analysis are summarised in Section 4, with conclusions presented in Section 5.

2 Methodology

2.1 Taste vary over respondents but are constant over replications for each respondent

We observe a sample of N decision makers, identified as n with $n = 1, \ldots, N$, where respondent n faces T_n choice situation. Let β give a vector of taste coefficients that is distributed across the population of decision makers, where we assume that β is the same for all choice situations faced by each decision maker.

[1] www.alogit.com

The distribution over decision makers is denoted $g(\beta \mid \Omega)$, where Ω are parameters of this distribution, such as the mean and variance. Let $P_{n,t}(i \mid x, \beta)$ denote the probability that respondent n chooses alternative i in choice situation t, conditional on β and the vector of attributes x. Let us further define $j_{n,t}$ to be the alternative chosen by respondent n in choice situation t, such that $P_{n,t}(j_{n,t} \mid x, \beta)$ gives the probability of the observed choice for respondent n in choice situation t, conditional on β and x.

As given by R-T, the LL function in this situation accounts for the assumption that tastes vary across respondents but stay constant across replications for the same respondent. This assumption is accommodated by carrying out the distribution across taste coefficients at the level of individual respondents, rather than individual choices. The LL function is:

$$\mathrm{LL}(\Omega) = \sum_{n=1}^{N} \ln\left(\int_{\beta}\left[\prod_{t=1}^{T_n}\left(P_{n,t}(j_{n,t} \mid x, \beta)\right)\right] g(\beta \mid \Omega)\,\mathrm{d}\beta\right). \tag{1}$$

This LL operates on *sequences* of choices rather than individual choices. Indeed, the location of the integral relative to the product across observations means that the logarithmic transformation from product to sum when moving from likelihood to log-likelihood now only applies to the product across respondents and not also to the product across choices.

Since the integrals do not take a closed form, the LL is approximated through simulation. The simulated LL is:

$$\mathrm{SLL}(\Omega) = \sum_{n=1}^{N} \ln\left(\frac{1}{R}\sum_{r=1}^{R}\left[\prod_{t=1}^{T_n}\left(P_{n,t}(j_{n,t} \mid x, \beta_{r,n})\right)\right]\right). \tag{2}$$

where $\beta_{r,n}$ is a draw from it density $g(\cdot)$ and R is the number of draws used in the simulation. Note that in this formulation the product over observations is calculated for each draw and averaged over draws, *prior* to taking the log. The SLL is the sum over decision makers of the log of the average of products.

There are two other estimators that have been utilised in this type of situation. The first procedure, which we call the cross-sectional (C-S) approach, is to treat the repeated choices by a given decision maker *as if* they were made by different respondents. The objective function (i.e, the equivalent to the LL function) is:

$$\mathrm{LL}(\Omega) = \sum_{n=1}^{N}\sum_{t=1}^{T_n} \ln\left(\int_{\beta} P_{n,t}(j_{n,t} \mid x, \beta)\, g(\beta \mid \Omega)\,\mathrm{d}\beta\right), \tag{3}$$

where the integration across the distribution of taste coefficients is carried out at the level of individual choices. This function is simulated as follows:

$$\mathrm{SLL}(\Omega) = \sum_{n=1}^{N} \sum_{t=1}^{T_n} \ln \left(\frac{1}{R} \sum_{r=1}^{R} P_{n,t}\left(j_{n,t} \mid x, \beta_{r,n,t}\right) \right). \tag{4}$$

where $\beta_{r,n,t}$ gives the r^{th} draw from $g(\cdot)$ for choice situation t for respondent n. Different draws are used for the T_n choice situations for respondent n, as well as for the N respondents.

If the parameters are identified by purely cross-sectional data (i.e., data consisting of only one observation per respondent), then the estimator based on this C-S approach is consistent for the estimation on panel data, for the following reason. First, maximising the log-likelihood based on one observation per respondent provides a consistent estimator. With T choices per respondent, we have T separate log-likelihoods, each of which provides a consistent estimator. The objective function in equation 3 is the sum of these T log-likelihood functions, and the sum of the objective functions of a set of consistent estimators provides another consistent estimator. Efficiency is reduced, relative to R-T, because the correlation over observations by a given respondent is not utilised in the estimation criterion.

A second approach, which can be used in packages that lack the R-T implementation, is to utilise the C-S formulation but, instead of taking different draws for each choice by a given respondent, using the same draws in all the choice situations for the same respondent (see e.g. Paag et al., 2001). The SLL under this approach is:

$$\mathrm{SLL}(\Omega) = \sum_{n=1}^{N} \sum_{t=1}^{T_n} \ln \left(\frac{1}{R} \sum_{r=1}^{R} P_{n,t}\left(j_{n,t} \mid x, \beta_{r,n}\right) \right). \tag{5}$$

From Equation 5, it can be seen that the only differences in comparison with Equation 4 lies in dropping the additional subscript t from the draws for β, where the same set of R draws is now reused in the simulation of all T_n choices for respondent n, thus leading to a requirement for NR draws, identical to the R-T approach.

This approach, which we will hereafter refer to as C-S-P to signify cross-sectional approximation to panel, attempts to accommodate the panel nature of the data by reusing the same draws across choices for a given respondent. It is possible that the approach accommodates some correlation across replications for a given individual through using the same draws in simulation, and this may increase efficiency relative to the C-S approach. However, the correlation over observations is a function of the number of draws, and should decrease as the number of draws rises. (We use the provisional words "is possible," "may," and "should", because the properties of this estimator have not been derived.)

2.2 Some tastes vary over people and choice situations and other tastes vary only over people

We next turn our attention to the situation described by Hess and Rose (2009), hereafter referred to as H-R. This specification allows for intra-respondent taste heterogeneity in addition to inter-respondent taste heterogeneity. Let β be decomposed as

$$\beta = \alpha + \gamma$$

where α is distributed across respondents but not over choices for a given respondent, while the vector γ is distributed over choices as well as respondents. That is, α captures inter-personal variation in tastes while γ captures intra-personal variation. The LL function[2] is given by:

$$LL(\Omega) = \sum_{n=1}^{N} \ln \left[\int_{\alpha} \left(\prod_{t=1}^{T_n} \left(\int_{\gamma} P_{n,t}(j_{n,t} \mid x, \alpha, \gamma)\, h(\gamma)\, \mathrm{d}\gamma \right) \right) g(\alpha)\, \mathrm{d}\alpha \right]. \quad (6)$$

In its most general form, the SLL for Equation 6 is given by:

$$SLL = \sum_{n=1}^{N} \ln \left[\frac{1}{R} \sum_{r=1}^{R} \left(\prod_{t=1}^{T_n} \frac{1}{K} \sum_{k=1}^{K} \left(P_{n,t}(j_{n,t} \mid \alpha_{r,n}, \gamma_{r,k,n,t}) \right) \right) \right]. \quad (7)$$

This simulation uses R draws from α for respondent n, along with $K\,T_n$ draws from γ for *each* draw of alpha. The total number of draws of γ for respondent n is therefore $R\,K\,T_n$ under this general specification.

The computational cost of estimating this model is very high, and thus far, it has not been implemented in any of the major packages. BIOGEME (Bierlaire, 2003) allows the user to estimate models combining inter-respondent and intra-respondent heterogeneity, but it is the authors' understanding that a simplification of the above formulation is used, with $K = 1$. The SLL thus takes the form:

$$SLL = \sum_{n=1}^{N} \ln \left[\frac{1}{R} \sum_{r=1}^{R} \left(\prod_{t=1}^{T_n} P_{n,t}(j_{n,t} \mid \alpha_{r,n}, \gamma_{r,n,t}) \right) \right]. \quad (8)$$

where the subscript for k on γ is not given since it is always 1. All simulation is performed at the level of respondents rather than individual choices. To

[2] Although conceptually very different, this model form is very similar in mathematical terms to the model developed by Bhat and Castelar (2002).

accommodate the intra-respondent variation in γ, separate draws are used for these parameters across choice situations, while the draws for α are reused across choices. We call this formulation with $K = 1$ the R-T-M specification, since it adapts the R-T specification to account for a mixture of inter- and intra-personal variation in tastes.

An alternative approach is to simulate each choice separately but use the same draws of γ in all the choices by a given respondent. This approach is analogous to C-S-P above but adapted for intra-personal variation in tastes. The objective function is

$$SLL = \sum_{n=1}^{N} \sum_{t=1}^{T_n} \ln \left[\frac{1}{R} \sum_{r=1}^{R} \left(P_{n,t} \left(j_{n,t} \mid \alpha_{r,n}, \gamma_{r,n,t} \right) \right) \right]. \tag{9}$$

This approach carries out all simulation at the level of individual choices, but the same draws of α are reused across choices for the same respondent. For γ, new draws are used in each choice situation. This approach can be used in software that relies entirely on cross-sectional approaches, such as ALogit. We call it C-S-M since it adapts C-S-P to account for a mixture of inter- and intra-personal variation in tastes.

Given the high computational cost of estimating models based on Equation 7, the use of the approximations in Equation 8 and 9 could potentially lead to significant savings. However, the quality of these approximations is unknown.

3 Empirical framework

To test the adequacy of the various approximations, an in-depth empirical analysis was carried out. The analysis is based on simulated data making use of a binary design with two attributes for each alternative, travel time (in minutes) and travel cost (in £). The dataset contains $5,000$ choices by 500 respondents. Two versions of the data were generated. The first version incorporates inter-personal variation in tastes but not intra-personal. This version is used to examine the methods R-T, C-S, and C-S-P. The second version incorporates both inter- and intra-personal variation in tastes. We use this version to examine H-R, C-S-M and R-T-M. In each case, the travel cost coefficient was kept fixed, and all heterogeneity was accommodated in the travel time coefficient.

The settings used in data generation are given in Table 1. Here, $\alpha_{\mathrm{TT},\mu}$ and $\alpha_{\mathrm{TC},\mu}$ give the mean values for the travel time and travel cost coefficients respectively. The standard deviation for the distribution across respondents are given by $\alpha_{\mathrm{TT},\sigma}$, with $\gamma_{\mathrm{TT},\sigma}$ giving the standard deviations for the variation across observations. The mean parameter $\gamma_{\mathrm{TT},\mu}$ is set to zero, with the combined tastes

Table 1: Settings used in data generation

	$\alpha_{\text{TT},\mu}$	$\alpha_{\text{TC},\mu}$	$\alpha_{\text{TT},\sigma}$	$\gamma_{\text{TT},\sigma}$
Case study 1	-0.2	-0.8	0.1	0
Case study 2	-0.2	-0.8	0.1	0.05

being given by $\beta_{\text{TT}} = \alpha_{\text{TT}} + \gamma_{\text{TT}}$. In each case, 15 different versions of the data were generated for the analysis, allowing us to test for stability and increasing the reliability of our conclusions.

When simulating data, it is important to ensure a good range for the probabilities of chosen alternatives, where these probabilities should cover as much of the $0 - 1$ interval as possible (cf. Fosgerau and Nielsen, 2007). A summary across the 15 samples for each dataset is given in Table 2 where we can see that the data clearly meets these requirements. The specific scale assumptions made in the simulation of the data led to ρ^2 measures of around 0.44, which should be a sufficiently large weight for the observed part of utility to facilitate parameter recovery.

As stated above, the data for case study 1 were examined with three different methods

R-T The panel MMNL model, with SLL given by Equation 2.

C-S The cross-sectional MMNL model, with SLL given by Equation 4.

C-S-P The cross-sectional approximation to the panel model, with SLL given by Equation 5

and the data for case study 2 were similarly examined with three methods:

H-R The joint inter and intra-respondent heterogeneity model, with SLL given by Equation 7.

C-S-M The cross-sectional approximation to the H-R model, with SLL given by Equation 9.

R-T-M The panel approximation to the H-R model, with SLL given by Equation 8.

Each model was estimated on the 15 samples for each case study. To test for the effects of sample size, the estimation was carried out while gradually increasing the sample size, starting with the first 50 respondents, and increasing the

Table 2: Statistics on probabilities of chosen alternative in simulated datasets

	Case study 1	Case study 2
min P(chosen)	0.0071	0.0059
mean P(chosen)	0.7535	0.7557
max P(chosen)	0.9999	0.9999
ρ^2	0.4383	0.4415

sample size by 50 respondents at each step until the full sample size of 500 respondents (and $5,000$ choices) was reached. In the simulation-based estimation of the models, we made use of Halton draws (Halton, 1960), and after extensive pre-testing[3], we made use of 200 draws in each dimension. All models were coded and estimated in Ox 4.1 (Doornik, 2001).

4 Results

This section discusses the results from the two case studies. We will first look at the two case studies in turn before summarising the findings.

4.1 Case study 1

The results for the first case study are presented in two parts. Table 3 shows the mean results across the 15 subsamples for each of the 10 different sample sizes, using three indicators, namely estimation time (in seconds), the t-ratio for $\alpha_{\text{TT},\sigma}$, and the inter-respondent coefficient of variation (cv.) for the travel time coefficient, which has a *true* value of 0.5 (cf. Table 1). To additionally show the variation across the 15 subsamples for each sample size, Figure 1 shows Box plots for the coefficient of variation and the t-ratio for $\alpha_{\text{TT},\sigma}$.

Combining the evidence from Table 3 and Figure 1, a number of observations can be made. Firstly, we can see that there are essentially no differences across the three methods in terms of estimation time. Secondly, all three methods offer good performance in retrieving the correct degree of heterogeneity in the travel time coefficient. However, differences arise between the three methods when looking at stability across runs as well statistical significance of estimates. Indeed, we can see that the estimates for R-T are far more stable across the 15 runs in each

[3] The number of draws was increased up to $10,000$ draws, but no changes in results were observed beyond about 100 draws. To keep estimation times manageable in the face of the large number of models (900 in total), we settled on $R = 200$.

Table 3: Mean results for different numbers of draws across 15 subsamples for case study 1

	R-T			C-S			C-S-P		
Resp.	time (s)	t-rat. (σ)	cv.	time (s)	t-rat. (σ)	cv.	time (s)	t-rat. (σ)	cv.
50	4.93	4.24	0.44	4.60	1.76	0.43	4.61	1.74	0.42
100	10.33	6.78	0.49	9.17	3.05	0.49	9.93	3.06	0.49
150	15.88	8.16	0.48	14.74	3.55	0.47	15.71	3.55	0.48
200	20.44	9.54	0.49	17.72	4.26	0.50	21.23	4.27	0.50
250	28.32	10.70	0.49	23.97	4.95	0.49	26.67	4.97	0.50
300	34.36	11.61	0.48	27.64	5.19	0.48	30.08	5.21	0.48
350	35.95	12.65	0.48	30.44	5.66	0.48	35.29	5.68	0.48
400	42.11	13.51	0.48	34.80	6.03	0.47	43.77	6.05	0.47
450	47.63	14.28	0.48	40.61	6.37	0.48	46.86	6.39	0.48
500	51.33	15.16	0.49	47.15	6.68	0.48	52.70	6.70	0.48

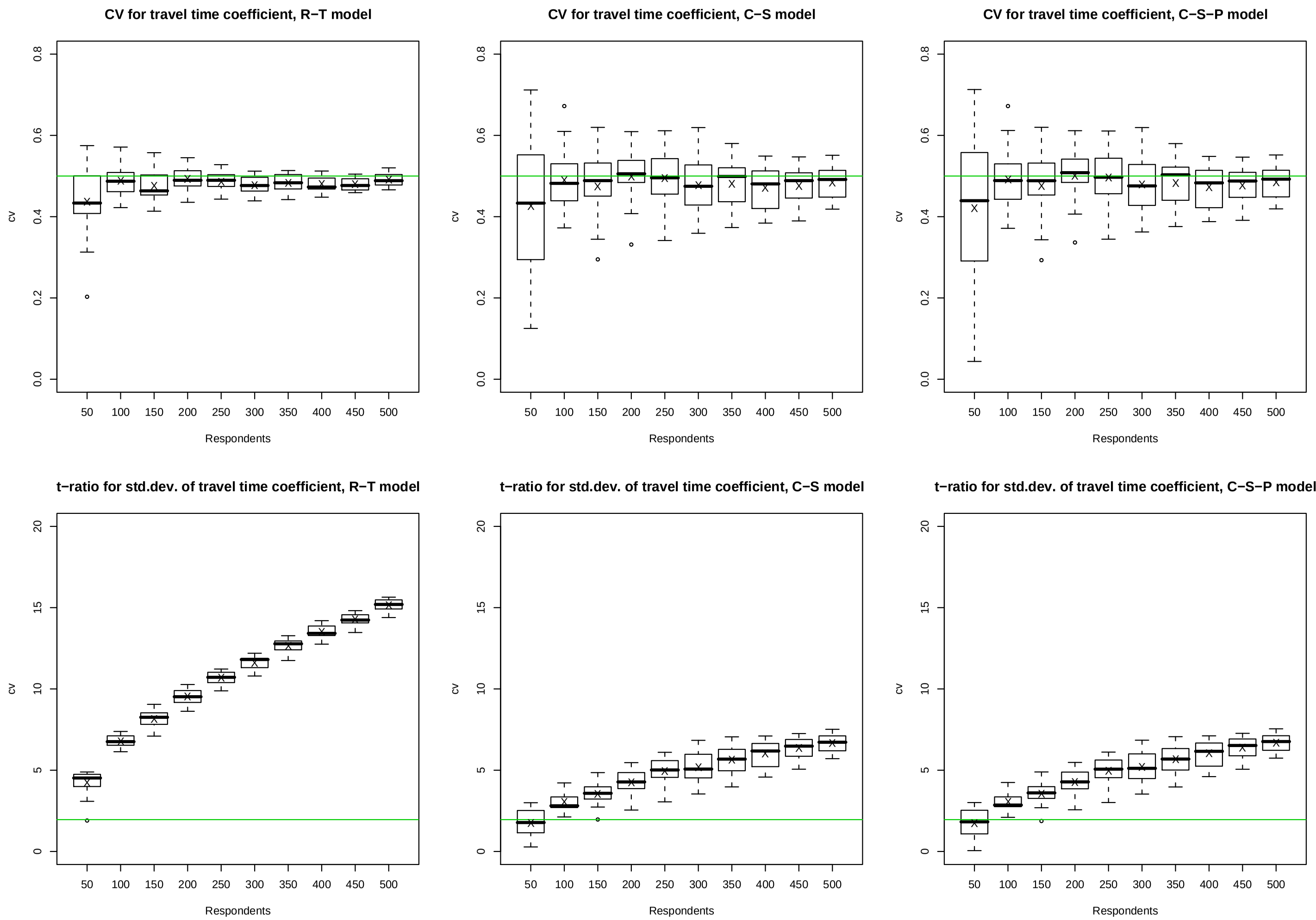

Figure 1: Summary of results for case study 1

subsample, even with small sample sizes. Additionally, increases in the sample size have a much stronger decreasing effect on the standard error of $\alpha_{\mathrm{TT},\sigma}$ in R-T than in the remaining two methods. This finding would suggest that C-S's loss in efficiency is relevant, even with samples of the size considered herein. Finally, and consistent with earlier discussions, C-S-P essentially produces exactly the same results as C-S, confirming earlier suspicions that the two models should be the same when estimated with a sufficiently high number of draws, i.e. showing that any attempt with C-S-P to regain some of the efficiency lost when moving from R-T to C-S is not particularly effective.

4.2 Case study 2

We now turn our attention to the results of the second case study, which contains both inter-respondent and intra-respondent variation in tastes. The results are again presented both in terms of the mean results across the 10 sample sizes (Table 4) and the variation across the 15 subsamples for each sample size (Figure 2 and Figure 3). In addition to those indicators already used in the presentation of the results for the first case study, we now also include the t-ratio for $\gamma_{\mathrm{TT},\sigma}$, i.e. the standard deviation for the intra-respondent travel time coefficient, and the associated coefficient of variation, which has a *true* value of 0.25 (cf. Table 1).

The results show that estimation times for R-T-M are slightly higher than those for C-S-M. However, the real differences arise when looking at H-R, where estimations take well over 100 times as long as with C-S-M and R-T-M, reaching on average 3.5 hours with the full sample size, compared to around 1.5 minutes for C-S-M and R-T-M. This finding is consistent with earlier discussions (cf. Section 2).

Turning our attention next to the retrieval of heterogeneity in the travel time coefficient, a number of observations can be made. Firstly, with the *true* value for $\mathrm{cv}_{\mathrm{inter}}$ being 0.5, we can see that good performance is obtained with H-R and also with R-T-M. C-S-M on the other hand underestimates the degree of inter-respondent heterogeneity and the estimates are accompanied by high standard errors. When looking at the stability of results across the 15 runs for each subsample (cf. Figure 2), we observe stable performance only for H-R and R-T-M, along with the expected reduction in standard errors as the sample size is increased.

For $\mathrm{cv}_{\mathrm{intra}}$, the *true* value is 0.25, and here we observe good performance by H-R, albeit with significantly higher standard errors than for the inter-respondent component. On the other hand, R-T-M is unable to retrieve any sensible measure of intra-respondent heterogeneity, while C-S-M actually overstates the degree of

Table 4: Mean results for different numbers of draws across 15 subsamples for case study 2

H-R

Resp.	time (s)	t-rat. (σ_{inter})	cv_{inter}	t-rat. (σ_{intra})	cv_{inter}
50	1,175	3.79	0.44	0.89	0.23
100	2,305	5.84	0.47	1.00	0.19
150	3,535	7.18	0.47	1.29	0.20
200	4,544	8.36	0.49	1.59	0.22
250	5,727	9.19	0.48	2.15	0.25
300	6,665	10.07	0.48	2.22	0.24
350	8,133	10.88	0.49	2.34	0.25
400	9,965	11.67	0.48	2.29	0.23
450	10,643	12.33	0.48	2.52	0.24
500	12,389	13.04	0.49	2.62	0.24

C-S-M

Resp.	time (s)	t-rat. (σ_{inter})	cv_{inter}	t-rat. (σ_{intra})	cv_{inter}
50	7.91	0.73	0.32	0.72	0.30
100	16.41	0.76	0.25	1.56	0.38
150	23.73	1.14	0.26	1.86	0.36
200	30.06	0.82	0.25	2.17	0.41
250	39.07	1.01	0.27	2.35	0.42
300	46.32	1.37	0.26	2.38	0.38
350	54.21	2.11	0.28	2.67	0.35
400	64.83	1.28	0.19	3.78	0.42
450	72.90	0.68	0.14	4.70	0.47
500	85.12	1.09	0.20	3.69	0.45

R-T-M

Resp.	time (s)	t-rat. (σ_{inter})	cv_{inter}	t-rat. (σ_{intra})	cv_{inter}
50	8.33	4.13	0.44	0.86	0.07
100	17.93	6.43	0.47	0.71	0.04
150	23.51	7.88	0.47	0.77	0.04
200	39.24	9.21	0.48	0.68	0.03
250	43.94	10.37	0.48	0.70	0.03
300	57.26	11.29	0.48	0.53	0.02
350	64.23	12.33	0.48	0.50	0.02
400	74.89	13.13	0.48	0.55	0.02
450	86.55	13.90	0.48	0.63	0.02
500	94.19	14.78	0.49	0.61	0.02

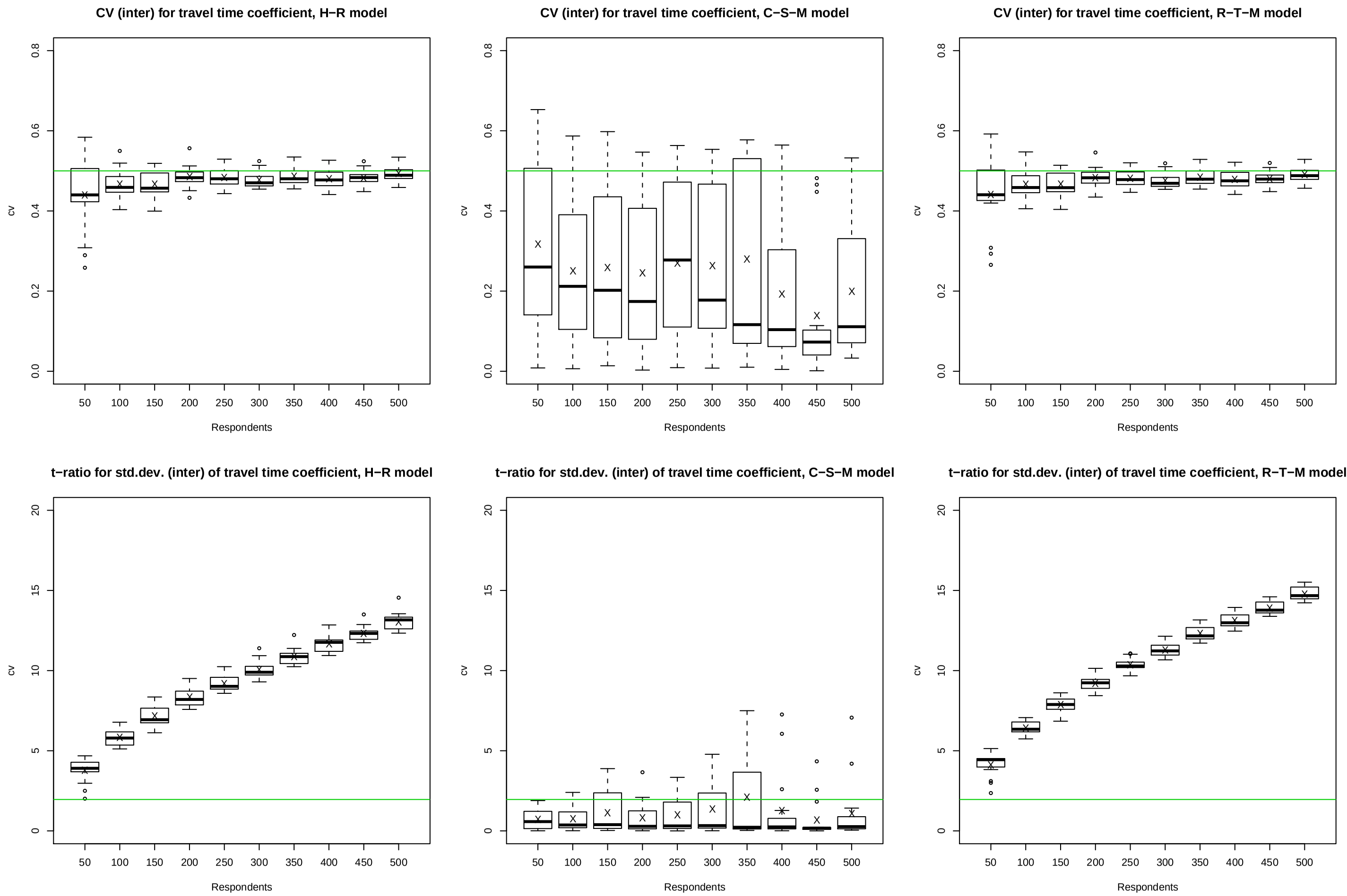

Figure 2: Summary of results for case study 2, part I

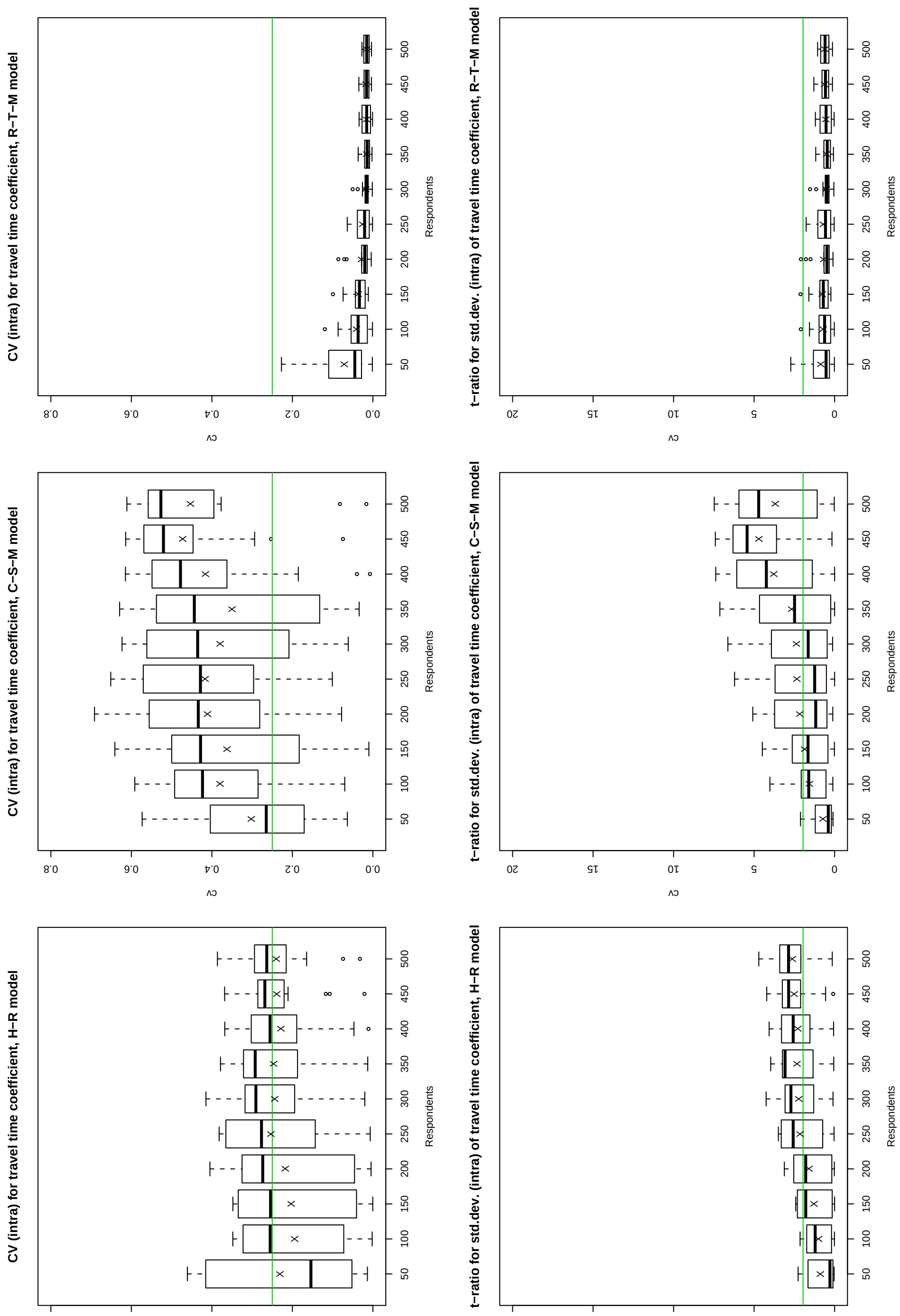

Figure 3: Summary of results for case study 2, part II

intra-respondent heterogeneity. As expected, the results across the 15 subsamples are far less stable than for the inter-respondent heterogeneity (cf. Figure 3), although there is an indication of increasing stability with sample size for H-R.

4.3 Summary of results

With the wealth of results presented in the two case studies, a summary of the observations seems appropriate.

In the first case study, the R-T model performs as expected, offering accurate and stable recovery of the inter-respondent heterogeneity, while also showing the benefits of increasing the sample size (rapid drop in standard errors). The C-S model is also able to retrieve most of the heterogeneity, but offers less stable performance, along with higher associated standard errors and a less important impact of increases in sample size. As expected on the basis of our theoretical discussions, the C-S-P model is no different from the C-S model.

In the second case study, it becomes apparent that there are significant issues with confounding between intra-respondent and inter-respondent taste heterogeneity that neither the C-S-M nor the R-T-M model are able to deal with. Indeed, the R-T-M model is able to retrieve the inter-respondent heterogeneity almost perfectly but fails in recovering any intra-respondent heterogeneity. On the other hand, the C-S-M model is severely hampered by the confounding between the two types of heterogeneity, resulting in an overestimation of intra-respondent heterogeneity and an under-estimation of inter-respondent heterogeneity, with both being affected by high standard errors. The H-R approach, although computationally far more expensive, succeeds in retrieving both types of heterogeneity with a high degree of precision, albeit with more uncertainty for the intra-respondent component.

5 Conclusions

This paper has discussed issues of approximation in the estimation of Mixed Multinomial Logit models on repeated choice data such as typically found in SP surveys. Specifically, we have discussed three different approximations; an approximation to the Revelt and Train (1998) panel formulation based on a cross-sectional model, an approximation to the joint intra and inter-respondent heterogeneity model of Hess and Rose (2009) based on a Revelt and Train panel formulation, and an approximation to the joint intra and inter-respondent heterogeneity model of Hess and Rose based on a cross-sectional model.

Aside from a scientific interest in this issue, this analysis was motivated by the fact that some of these approximations are used in applied research (e.g. Paag et al., 2001), while others are implemented in existing software[4]. Additionally, there is a strong interest in the approximation approaches to the H-R model (Equation 8 and Equation 9) given the high cost of estimating this model on the basis of the SLL in Equation 7.

The findings from this research, which is based on a comprehensive Monte Carlo analysis, suggest somewhat different results for models with inter-respondent variation only from those with intra- and inter-respondent variation. For models with inter-respondent variation (i.e, the specification of Revelt and Train 1998), it seems that treating the panel data as if they were cross-sectional results in fairly accurate estimates provided that the sample size is sufficiently high (and the parameters are identified by cross-sectional data, as in our study.) There is of course a loss of efficiency, but the loss does not seem to be great for the sample sizes used in our study. However, for models with both intra- and inter-respondent variation, approximations to the formula given by Hess and Rose (2009) do not seem to perform well. This result is unfortunate since the H-R specification is highly computer intensive. However, the result is not unexpected, given the inherent difficulty of distinguishing the two types of variation. A potentially fruitful area for future reasearch is to develop an efficient implementation of the H-R specification.

Acknowledgements

The first author acknowledges the financial support by the Leverhulme Trust, in the form of a *Leverhulme Early Career Fellowship*. The authors would like to thank Stefan Mabit and Mogens Fosgerau for comments on an earlier version of this paper.

References

Bhat, C. R., 2001. Quasi-random maximum simulated likelihood estimation of the mixed multinomial Logit model. Transportation Research Part B 35 (7), 677–693.

Bhat, C. R., 2003. Simulation estimation of mixed discrete choice models using randomized and scrambled Halton sequences. Transportation Research Part B 37 (9), 837–855.

Bhat, C. R., Castelar, S., 2002. A unified mixed logit framework for modeling revealed and stated preferences: formulation and application to congestion

[4]The R-T-M implementation is used in BIOGEME.

pricing analysis in the San Francisco Bay area. Transportation Research Part B 36 (7), 593–616.

Bierlaire, M., 2003. BIOGEME: a free package for the estimation of discrete choice models. Proceedings of the 3^{rd} Swiss Transport Research Conference, Monte Verità, Ascona.

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.

Doornik, J. A., 2001. Ox: An Object-Oriented Matrix Language. Timberlake Consultants Press, London.

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.

Halton, J., 1960. On the efficiency of certain quasi-random sequences of points in evaluating multi-dimensional integrals. Numerische Mathematik 2, 84–90.

Hess, S., Rose, J. M., 2009. Allowing for intra-respondent variations in coefficients estimated on repeated choice data. Transportation Research Part B 43 (6), 708–719.

Hess, S., Train, K., Polak, J. W., 2006. On the use of a Modified Latin Hypercube Sampling (MLHS) method in the estimation of a Mixed Logit model for vehicle choice. Transportation Research Part B 40 (2), 147–163.

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

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

Paag, H., Daly, A., Rohr, C., 2001. Predicting the use of the copenhagen harbour tunnel. In: Hensher, D. A. (Ed.), Travel Behaviour Research: the Leading Edge. Pergamon, Oxford, pp. 627–646.

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.

Train, K., 2003. Discrete Choice Methods with Simulation. Cambridge University Press, Cambridge, MA.