Bayesian estimator for Logit Mixtures with inter- and intra-consumer heterogeneity

Becker, Felix; Danaf, Mazen; Song, Xiang; Atasoy, Bilge; Ben-Akiva, Moshe

DOI
10.1016/j.trb.2018.06.007

Publication date
2018

Document Version
Accepted author manuscript

Published in
Transportation Research Part B: Methodological

Citation (APA)
Becker, F., Danaf, M., Song, X., Atasoy, B., & Ben-Akiva, M. (2018). Bayesian estimator for Logit Mixtures with inter- and intra-consumer heterogeneity. *Transportation Research Part B: Methodological*, *117*(Part A), 1-17. https://doi.org/10.1016/j.trb.2018.06.007

Bayesian Estimator for Logit Mixtures with Inter- and Intra-Consumer Heterogeneity

Felix Becker

Corresponding author

Research Assistant, ETH Zurich

Department of Civil, Environmental, and Geomatic Engineering

Stefano-Franscini-Platz 5, 8093 Zurich, Switzerland

Room HIL F 34.1

Email: fbecker@ethz.ch

Mazen Danaf

Research Assistant, Massachusetts Institute of Technology

Department of Civil and Environmental Engineering

Intelligent Transportation Systems Lab, Room 1-249

Email: mdanaf@mit.edu

Xiang Song

Research Assistant, Massachusetts Institute of Technology

Department of Civil and Environmental Engineering

Intelligent Transportation Systems Lab, Room 1-249

Email: bensong@mit.edu

Bilge Atasoy

Assistant Professor, TU Delft

Department of Maritime and Transport Technology

Email: b.atasoy@tudelft.nl

Moshe Ben-Akiva

Professor of Civil Engineering, Massachusetts Institute of Technology

Department of Civil and Environmental Engineering, Room 1-181

Email: mba@mit.edu

ABSTRACT

Estimating discrete choice models on panel data allows for the estimation of preference heterogeneity in the sample. While the Logit Mixture model with random parameters is mostly used to account for variation across individuals, preferences may also vary across different choice situations of the same individual. Up to this point, Logit Mixtures incorporating both inter- and intra-consumer heterogeneity are estimated with the classical Maximum Simulated Likelihood (MSL) procedure. The MSL procedure becomes computationally expensive with an increasing sample size and can be burdensome in the presence of a multi-modal likelihood function. We therefore propose a Hierarchical Bayes estimator for Logit Mixtures with both levels of heterogeneity. It builds on the Allenby-Train procedure, which considers only inter-consumer heterogeneity. To test the proposed procedures, we analyze how well the true patterns of heterogeneity are recovered in a simulation environment. Results from the Monte Carlo simulation suggest that falsely ignoring intra-consumer heterogeneity despite its presence in the data leads to biased estimates and a decreased goodness of fit. The latter is confirmed by a real-world example of explaining mode choices for GPS traces. We further show that the runtime of the proposed estimator is substantially faster than for the corresponding MSL estimator.

Keywords: Hierarchical Bayes, Mixed Logit, Logit Mixture, inter-consumer heterogeneity, intra-consumer heterogeneity, panel data

1 INTRODUCTION

Research about taste heterogeneity has traditionally focused on variation across respondents (inter-consumer). Nonetheless, some researchers (Bhat and Castelar 2002; Bhat and Sardesai 2006; Cherchi 2009; Hess and Rose 2009) emphasize the importance of considering varying preferences among different choice situations, also called menus, for one individual (intra-consumer). Accounting for variations among menus is especially important when the data are collected over a long period of time. Hess and Rose (2009) further argue that in a survey setting, individuals' preferences may alter during the course of time in which they complete the survey. For example, respondents who are in the learning phase tend to consider only a fraction of presented attributes.

In terms of estimating Logit Mixtures with inter- and intra-consumer heterogeneity, it has so far been proposed to use Maximum Simulated Likelihood (MSL) estimators (Bhat and Castelar 2002; Hess and Rose 2009). From the Bayesian perspective, the Allenby-Train procedure, a Hierarchical Bayes estimator for Logit mixtures with inter-consumer heterogeneity, is available. This procedure was first mentioned by Allenby in 1997 in tutorial notes at the Advanced Research Techniques forum (as cited by (Train 2009, 300)), and later generalized by (Train 2001). Furthermore, Dekker et al. (2016) used a Gibbs Sampler to estimate an integrated-choice latent variable (ICLV) model with inter-consumer preference heterogeneity and intra-consumer scale heterogeneity. Scale was represented as a function of individual- and menu-specific characteristics. A maximum approximate composite marginal likelihood estimator has been proposed to estimate inter- and intra-consumer heterogeneity with a Probit kernel (Bhat 2011; Bhat and Sidharthan 2011). Patil et al. (2017) further showed that MACML can outperform the Bayesian Markov Chain Monte Carlo (MCMC) approach for the multinomial probit.

Sampling from high-dimensional posterior distributions by applying MCMC methods has numerous advantages and benefits. As an example, Huber and Train (2001) accentuate that in some cases multiple local maxima exacerbate the search for the global maximum in the MSL. Regarding the estimation of variance-covariance matrices, they argue that it is computationally expensive to calculate the derivative of every element of the upper triangular matrix when using MSL. Drawing from the conditional posterior of a full variance-covariance matrix is less problematic. Section 4.2 further shows that the runtime for double mixtures is substantially shorter for MCMC than MSL. Furthermore, Train (2009, p. 283) points out that the posterior mean and standard deviation are similar to classical estimates and standard errors, provided an uninformative prior has been applied. This result enables a classical analysis of the results of Bayesian estimation where consistency and efficiency can be achieved under more relaxed conditions as compared to MSL (Train 2009, p. 283). Rossi and Allenby (2003) highlight another advantage of MCMC methods: both population- and individual-level parameters are produced in the estimation process. With MSL, post estimation Bayesian analysis is required to compute individual-parameters.

The development of a Hierarchical Bayes estimator to incorporate both inter- and intra-consumer heterogeneity is the main contribution of our paper.[1] The estimation procedure also provides the modeler with menu-level coefficients in addition to the already existing estimation of individual-level coefficients allowing for a valuable new application. In a system that is continuously learning from customers, menu-level coefficients of new choice situations can now be used to update existing individual-level coefficients. This idea is elaborated in (Danaf et al. 2017).

[1] The code is available on request.

The estimator is analyzed from three different perspectives. We simulate data in order to test the estimator's ability to recover the true parameters and its forecasting performance. We further compare the estimator to its MSL counterpart in terms of estimates and runtime. Lastly, we apply the estimator to transportation mode choice for GPS traces.

The remainder of the paper is broken down as follows: Section 2 describes the methodology behind the model formulation for Logit Mixtures with inter- and intra-consumer heterogeneity as well as the new estimator. Section 3 describes the framework to test the new estimator, and Section 4 presents the results. Discussion and conclusion follow in Sections 5 and 6.

2 METHODOLOGY

2.1 Model for Logit Mixtures with inter- and intra-consumer heterogeneity

The model used in this paper is assumed to have a logit kernel with a linear utility specification of choice j in menu m as shown in Eq. (1):

$$\mathrm{U_{jmn}} = \mathrm{X_{jmn}}\eta_{\mathrm{mn}} + \epsilon_{\mathrm{jmn}} \tag{1}$$

with $\mathrm{U_{jmn}}$ indicating individual n's unobserved utility of alternative j in menu m and $\mathrm{X_{jmn}}$ denoting alternative attributes. Note that each individual n is presented M_n menus and each menu m has J_{mn} alternatives. The error term ϵ_{jmn} follows the Gumbel distribution.

A model formulation for Logit Mixtures with only inter-consumer heterogeneity has three sets of parameters: the vector of sample-level parameters μ, the individual parameters ζ_n for every individual n, and the inter-consumer covariance matrix Ω_b. In order to account for intra-consumer heterogeneity, we add the menu-level parameters η_{mn} for every menu m of every individual n in the sample as well as the intra-consumer covariance matrix Ω_w.

We assume that ζ_n and η_{mn} are normally distributed as shown in Eqs. (2) and (3). Readers interested in varying the distributional assumptions of the parameters are referred to Train (2009, pp. 305–7).

$$\eta_{mn} \sim \mathcal{N}(\zeta_n, \Omega_w) \tag{2}$$

$$\zeta_n \sim \mathcal{N}(\mu, \Omega_b) \tag{3}$$

The probability not conditional on the hyperparameters η and ζ is presented in Eq. 4:

$$P(d_n|\mu, \Omega_b, \Omega_w) = \int_{\zeta_n} \prod_{m=1}^{M_n} \left[\int_{\eta_{mn}} \prod_{j=1}^{J_{mn}} P_j(\eta_{mn})^{d_{jmn}} h(\mathrm{d}\eta_{mn}|\zeta_n, \Omega_w) \right] f(\mathrm{d}\zeta_n|\mu, \Omega_b), \tag{4}$$

where d_{jmn} is equal to one if individual n chooses alternative j in menu m and zero otherwise, and:

$$P_j(\eta_{mn}) = \frac{exp\left(V_{jmn}(\eta_{mn})\right)}{\sum_{j'=0}^{J_{mn}} exp\left(V_{j'mn}(\eta_{mn})\right)} \tag{5}$$

$$h(d\eta_{mn}|\zeta_n, \Omega_w) \sim \mathcal{N}(\zeta_n, \Omega_w) \tag{6}$$

$$f(d\zeta_n|\mu, \Omega_b) \sim \mathcal{N}(\mu, \Omega_b) \tag{7}$$

In comparison to the model with inter-consumer heterogeneity, the integral over the menu-level coefficients (η_{mn}) is added.

2.2 Proposed Bayesian estimator for Logit Mixtures with inter- and intra-consumer heterogeneity

Before discussing the estimation procedure, we note that the Hierarchical Inverse-Wishart prior, which is introduced to Hierarchical Bayes for Logit Mixtures by Song et al. (2016), is omitted in the general description and is only referred to afterwards for clarity. We incorporate it into the subsequent model estimations because of its ability to mitigate the influence of the prior. Additionally, we present the possibility of restricting a subset of the parameters to be constant among menus or among individuals and menus after the generic case. Readers not familiar with Hierarchical Bayes for Mixed Logit are referred to Train (2009) for a more comprehensive introduction to models with only inter-consumer heterogeneity.

In the basic form, Eq. (8) denotes the numerator of the joint posterior distribution:

$$K(\mu, \zeta_n \,\forall n, \eta_{mn} \,\forall mn, \Omega_w, \Omega_b | d_n \forall n)$$
$$\propto \prod_{n=1}^{N}\left[\prod_{m=1}^{M_n}\left[\prod_{j=1}^{J_{mn}}\left[P_j(\eta_{mn})^{d_{jmn}}\right] h(\eta_{mn}|\zeta_n, \Omega_w)\right] f(\zeta_n|\mu, \Omega_b)\right] k(\Omega_w)k(\mu)k(\Omega_b), \tag{8}$$

where:

$$\mathrm{k}(\mu) \sim \mathrm{N}(\mu_0, \mathrm{A}) \tag{9}$$

$$\mathrm{k}(\Omega_b) \sim \mathrm{IW}(\mathrm{T}, \mathrm{I_T}) \tag{10}$$

$$\mathrm{k}(\Omega_w) \sim \mathrm{IW}(\mathrm{T}, \mathrm{I_T}) \tag{11}$$

μ_0 represents the vector of means for the sample-level parameter's prior distribution and can be assigned arbitrary values, as A is a diagonal covariance matrix with diagonal values $a_{ii} \rightarrow \infty$, causing the prior to be diffuse. T depicts the number of unknown parameters, and I_T is the T-dimensional identity matrix.

Draws from the joint posterior are obtained by a five-layered Gibbs Sampler. In accordance with the concept of a Hierarchical Bayes estimator, the prior of the sample-level parameters is determined ex-ante and updated with individual-level parameters. The density of each individual parameter in the sample-distribution serves again as the prior for each individual parameter. The data used to update the individual parameters consist of the menu parameters. The density of the menu parameters in the distribution of the individual parameters is the prior for the menu-level parameters. Only the lowest level, the menu parameters, is updated using the likelihood of the collected data given the parameters.

Note that in the case of the Allenby-Train procedure the individual parameters are updated using the likelihood. Furthermore, a new layer for Ω_w, the covariance matrix accounting for intra-consumer heterogeneity, is introduced. The current Gibbs Sampler iteration is denoted by superscript i. The assignment of starting values is discussed in Appendix A.

Step I - μ:

The conditional posterior of the sample-level parameters is proportional to the right hand side of the term

$$K(\mu|\zeta_n\ \forall n, \eta_{mn}\ \forall mn, \Omega_w, \Omega_b) \propto f(\zeta_n \forall n|\mu, \Omega_b)k(\mu), \qquad (12)$$

which refers to a Bayesian update of a multivariate normal distribution. Based upon the fact that $k(\mu)$ is diffuse, the conditional posterior can be simplified to $\mathcal{N}\left(\bar{\zeta}^{i-1}, \frac{\Omega_b^{i-1}}{N}\right)$, with $\bar{\zeta}^{i-1} = \frac{1}{N}\sum_n \zeta_n^{i-1}$. A draw from this multivariate normal distribution is obtained by

$$\mu^i = \bar{\zeta}^{i-1} + \Psi^{i-1}\omega, \qquad (13)$$

where Ψ^{i-1} is the Cholesky factor of $\frac{\Omega_b^{i-1}}{N}$ and ω is a draw from the T-dimensional multivariate standard normal.

Step II- Ω_b:

The conditional posterior of Ω_b is shown on the right hand side of Eq. (14).

$$K(\Omega_b|\mu, \zeta_n\ \forall n, \eta_{mn}\ \forall mn, \Omega_w) \propto f(\zeta_n\ \forall n|\mu, \Omega_b)k(\Omega_b) \qquad (14)$$

With the Inverse-Wishart distribution being conjugate to the multivariate normal distribution, the closed form posterior is distributed Inverse-Wishart with T+N degrees of freedom and scale matrix $TI+N\bar{V}_b$, where:

$$\bar{V}_b = \frac{1}{N}\sum_{n=1}^{N}\left(\zeta_n^{i-1} - \mu^i\right)\left(\zeta_n^{i-1} - \mu^i\right)' \qquad (15)$$

A draw of Ω_b is obtained as:

$$\Omega_b^i = \left[\sum_{r=1}^{T+N} (\Gamma \upsilon_r)(\Gamma \upsilon_r)'\right]^{-1}, \quad (16)$$

where υ_r is a draw from the T-dimensional standard normal distribution for $r = 1, \ldots, T + N$, and Γ is the Cholesky factor of $[TI_T + N\overline{V}_b]^{-1}$.

Step III – Ω_w:

Drawing from the conditional posterior of the intra-consumer covariance matrix Ω_w, see Eq. (17), is similar to the previous step, as it is also considered to be distributed Inverse-Wishart. Each menu is assigned equivalent weight for the computation of Ω_w, as presented in Eq. (18). It is considered inappropriate to assign lower weights to menus of individuals for whom many menus are available.

$$K(\Omega_w|\mu, \zeta_n \, \forall n, \eta_{mn} \, \forall mn, \Omega_b) \propto h(\eta_{mn} \, \forall mn|\zeta_n \, \forall n, \Omega_w)k(\Omega_w) \quad (17)$$

The posterior's parameters are $T + M$ for the degrees of freedom and $TI_T + M\overline{V}_w$ for the scale matrix. M is the total number of menus in the data for all individuals, and:

$$\overline{V}_w = \frac{1}{M}\sum_{n=1}^{N}\sum_{m=1}^{M_n}\left(\eta_{mn}^{i-1} - \zeta_n^{i-1}\right)\left(\eta_{mn}^{i-1} - \zeta_n^{i-1}\right)' \quad (18)$$

After obtaining T + M draws of a T-dimensional standard normal distribution, labeled υ_s, $s = 1, \ldots, T + M$, the new draw of Ω_w is calculated as:

$$\Omega_w^i = \left[\sum_{s=1}^{T+M} (\Gamma \upsilon_s)(\Gamma \upsilon_s)'\right]^{-1}, \quad (19)$$

where Γ is the Cholesky factor of $[TI_T + M\overline{V}_w]^{-1}$.

Step IV – ζ_n:

The following operations are repeated for each individual $n = 1, \dots, N$. Despite the numerous repetitions, the computational complexity is manageable, because the terms that require matrix inversion are identical among all individuals with the same number of menus. The individual specific conditional posterior is proportional to Eq. (20). The product of the menu- and individual-level parameter's distribution is multiplied over all menus of individual n

$$K(\zeta_n|\mu, \eta_{mn}\ \forall mn, \Omega_b, \Omega_w) \propto \prod_{m=1}^{M_n} h(\eta_{mn}\ |\zeta_n\ \forall n, \Omega_w) f(\zeta_n|\mu, \Omega_b)\ \ n = 1, \dots, N \tag{20}$$

The conditional posterior distribution of ζ_n, can be denoted as $\mathrm{N}(\overline{\zeta_n}, \Sigma_{\zeta n})$, where

$$\overline{\zeta_n} = \left(\left[\Omega_b^i\right]^{-1} + M_n\ \left[\Omega_w^i\right]^{-1}\right)^{-1} \left(\left[\Omega_b^i\right]^{-1} \mu_i + M_n\left[\Omega_w^i\right]^{-1} \frac{1}{M_n} \sum_{m=1}^{M_n} \eta_{mn}^{i-1}\right), \tag{21}$$

and

$$\Sigma_{\zeta n} = \left(\left[\Omega_b^{i+1}\right]^{-1} + M_n\left[\Omega_w^{i+1}\right]^{-1}\right)^{-1}. \tag{22}$$

A draw form $\mathrm{N}(\overline{\zeta_n}, \Sigma_{\zeta n})$ is obtained by calculating $\zeta_n^i = \overline{\zeta_n} + \Psi_{\zeta_n}\omega$ where Ψ_{ζ_n} is the Cholesky factor of $\Sigma_{\zeta n}$ and ω is a draw from a T-dimensional standard normal.

Step V – η_{mn}:

The last step of the Gibbs Sampler is used to update the menu-level coefficients. The operation is executed for every menu $m = 1, \dots, M_n$ for every individual $n = 1, \dots, N$. The numerator of the conditional posterior of a menu-level coefficient is given in Eq. (23).

$$K(\eta_{mn}|\mu, \zeta_n, \Omega_b, \Omega_w) \propto \prod_{j=0}^{J_{mn}} \left[P_j(\eta_{mn})^{d_{jmn}}\right] h(\eta_{mn}\ \forall mn|\zeta_n, \Omega_w), \tag{23}$$

$$n = 1, 2, \dots, N, m = 1, 2, \dots M$$

As the posterior does not possess a closed form, a draw of η_{mn}^i is obtained by the following Metropolis-Hastings step:

The trial draw $\tilde{\eta}_{mn}^i$ is obtained as depicted in Eq. (24):

$$\tilde{\eta}_{mn}^i = \eta_{mn}^{i-1} + \sqrt{\rho}\Lambda_w \upsilon, \tag{24}$$

where Λ_w is the Cholesky factor of Ω_w, υ are T independent variables from $N(0,1)$, and ρ is a parameter of the jumping distribution, adjusted continuously in every iteration. Train (2006) chooses to decrease (increase) ρ by 10% in case less (more) than 30% of the trial menu-level coefficients have been accepted. The trial draw $\tilde{\eta}_{mn}^{i}$ is accepted if:

$$u \leq \frac{\prod_{j=0}^{J_{mn}} \left[P_j\left(\tilde{\eta}_{mn}^{i}\right)^{d_{jmn}}\right] \mathrm{h}\left(\tilde{\eta}_{mn}^{i} \middle| \zeta_{\mathrm{n}}, \Omega_{\mathrm{w}}\right)}{\prod_{j=0}^{J_{mn}} \left[P_j\left(\eta_{mn}^{i-1}\right)^{d_{jmn}}\right] \mathrm{h}\left(\eta_{mn}^{i-1} \middle| \zeta_{\mathrm{n}}, \Omega_{\mathrm{w}}\right)} \tag{25}$$

where u is a draw from the standard uniform distribution.

2.3 Enhancements

In the steps above, all coefficients are distributed across individuals as well as menus. Nonetheless, it is also possible to account for coefficients that only vary among individuals or do not vary at all. For ease of presentation, the parameters are assigned to three different groups according to their maximum level of heterogeneity: no heterogeneity (1), inter-consumer heterogeneity (2), and inter- and intra-consumer heterogeneity (3). The elaboration of the case of only intra-consumer heterogeneity is omitted due to a smaller practical relevance. Steps II and IV would be omitted in that case.

Should the modeler decide that a subset of the parameters does not vary among individuals or menus (i.e. belongs to group (1)), a Metropolis Hastings step for the specific sample-level parameters can be employed as described by Train (2009). The conditional posterior is proportional to the term provided in Eq. (26), where $\mu_1$refers to the sample parameters for parameters without heterogeneity.

$$\begin{aligned} &K\left(\mu_1 \middle| \mu_{2,3}, \zeta_{2,3,\mathrm{n}}\ \forall \mathrm{n}, \eta_{\mathrm{mn}}\ \forall\ mn, \Omega_{\mathrm{w}}, \Omega_{\mathrm{b}}\right) \\ &\propto \prod_{n=1}^{N} \prod_{m=1}^{M_n} \prod_{j=1}^{J_{mn}} \left[P_j\left(\eta_{3,mn}, \zeta_{2,\mathrm{n}}, \mu_1\right)^{d_{jmn}}\right] k(\mu_1) \end{aligned} \tag{26}$$

The Metropolis Hastings step is performed in the same manner except for the fact that the acceptance rate cannot be calculated for one iteration; either all of the trial draws for the set of parameters belonging to group (1) are accepted or rejected. Therefore, after every hundredth iteration we evaluate the acceptance rate across the last one hundred iterations. The step-specific ρ is then increased by 2% if the acceptance rate is higher than 0.3 and vice versa.

In the case of parameters that vary among individuals but not among menus, it is essential to note that parameters of group (2) and (3) share the inter-consumer covariance matrix Ω_{b}. For this reason, steps I and II are jointly executed for both groups of parameters, while step IV needs to be split in two parts. In the first part, which refers to parameters of group (2), the sample-level multivariate normal distribution conditional on the parameters of group (3) is updated with the likelihood. In the second part, the respective distribution conditional on the parameters of group (2) is updated with the menu-level parameters.

In the first part, the conditional distribution is proportional to the term in Eq. (27). The prior distribution is conditional on the individual-level parameters of group (3), indicated by the respective subscript.

$$K(\zeta_{2,n}|\mu, \zeta_{3,n}, \eta_{mn}, \Omega_w, \Omega_b) \propto \prod_{m=1}^{M_n}\left[\prod_{j=1}^{J_{mn}}\left[P_j(\eta_{3,mn}, \zeta_{2,n}, \mu_1)^{d_{jmn}}\right] n(\zeta_{2,n}|\mu_{2,3}, \zeta_{3,n}, \Omega_b)\right], n = 1, \dots, N \tag{27}$$

The parameters of the conditional distribution are computed as denoted in Eqs. (28) and (29). Note that the subscript (x,y) of a covariance matrix refers to the submatrix whose rows are associated to group x and columns to group y.

$$\mu^i_{2,cond} = \mu^i_2 + \Omega^i_{b,2,3}\left[\Omega^i_{b,3,3}\right]^{-1}(\zeta^i_{3,n} - \mu^i_3) \tag{28}$$

$$\Omega^i_{b,2,cond} = \Omega^i_{b,2,2} - \Omega^i_{b,2,3}\left[\Omega^i_{b,3,3}\right]^{-1}\Omega^i_{b,3,2} \tag{29}$$

Due to the logit kernel the conditional posterior is again in a non-closed form which requires a Metropolis Hastings step that has the same structure as the one presented in step V of the generic procedure.

The adjusted conditional posterior of the group (3) individual-level parameters is presented in line with the appropriate parameters of the conditional distribution:

$$K(\zeta_n|\mu, \eta_{mn}\ \forall mn, \Omega_b, \Omega_w) \propto \prod_{m=1}^{M_n} h(\eta_{3,mn}|\zeta_{3,n}, \Omega_w) n(\zeta_{3,n}|\mu_{2,3}, \zeta_{2,n}, \Omega_b), \quad n = 1, \dots, N \tag{30}$$

$$\mu^i_{3,cond} = \mu^i_2 + \Omega^i_{b,3,2}\left[\Omega^i_{b,2,2}\right]^{-1}(\zeta^i_{2,n} - \mu^i_2) \tag{31}$$

$$\Omega^i_{b,3,cond} = \Omega^i_{b,3,3} - \Omega^i_{b,3,2}\left[\Omega^i_{b,2,2}\right]^{-1}\Omega^i_{b,2,3} \tag{32}$$

Another distinction of the methodology is the application of the Hierarchical Inverse-Wishart prior for the covariance matrices proposed by Huang and Wand (2013). It is introduced by Song et al. (2016) for Hierarchical Bayes Estimators for Logit Mixtures. The results of the latter paper indicate that the inflation of the variances observed by Balcombe et al. (2009) and Ben-Akiva et al. (2015) for the standard Allenby-Train procedure can be counteracted by using this prior structure. For this reason, our paper also considers the Hierarchical Inverse-Wishart prior. While Song et al. (2016) only show the adaption of the concept to step II, i.e. for the inter-consumer covariance matrix Ω_b , the adjustment is analog for Ω_w in step III.

Furthermore, we propose a block structure for the variance-covariance matrix. Suppose the model has T_2 coefficients with inter-consumer heterogeneity. Coefficients that are expected to be correlated can be grouped together in $1,\ldots,L$ blocks, with $1 \leq L \leq T_2$. This means that coefficients belonging to one block are correlated to each other without any restrictions, but are independent of the remaining $L-1$ blocks. The associated structure of Ω_b is displayed in Eq. (33).

$$\Omega_b = \begin{bmatrix} \Omega_{b,1} & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & \Omega_{b,L} \end{bmatrix} \quad (33)$$

In essence, each one of the conditional posteriors of the blocks is distributed Inverse-Wishart, see Eq. (34). Furthermore, V_l as well as the parameters of the prior distribution are block-specific, shown in Eqs. (35) to (37).

$$\Omega_{b,l}|\ldots \sim IW(p_l + N, \Phi_l + N\bar{V}) \quad (34)$$

$$\bar{V}_l = \frac{1}{N}\sum_{n=1}^{N}\left(\zeta_{2,3,n,l}^{i-1} - \mu_{2,3,l}^{i}\right)\left(\zeta_{2,3,n,l}^{i-1} - \mu_{2,3,l}^{i}\right)' \quad (35)$$

$$p_l = T_{2,3,l} \quad (36)$$

$$\Phi_l = T_{2,3,l} I_{(T_{2,3,l} \times T_{2,3,l})} \quad (37)$$

The structure requires the second and third steps to be executed independently for each block. The new draws of the single blocks constitute the global variance-covariance matrix, as denoted in Eq. (33). Since the prior distribution is still conjugate to the likelihood, the additional computation time is negligible. Nonetheless, the possibilities of restricting the variance-covariance matrix are limited from a modeler's perspective.

Readers interested in using distributions other than the normal distribution are referred to Train (2009), where the application of lognormal and triangulars is further explained.

3 EMPIRICAL FRAMEWORK

In this section, we evaluate our method using three different approaches. First, we simulate choice data and test whether true patterns of heterogeneity can be replicated. We also investigate the effects of misspecified models and the predictive performance on out of sample data. Next, we compare runtime and estimates to a Maximum Simulated Likelihood estimation. Finally, we test the addition of intra-consumer heterogeneity on a transportation mode choice example on GPS traces.

3.1 Monte Carlo Experiment: Effects of introducing intra-consumer heterogeneity

This section sets up a framework within which we test the proposed approach and identify scenarios where it outperforms models with only inter-consumer heterogeneity. We distinguish the baseline model from the Allenby-Train procedure by applying the Hierarchical Inverse-Wishart prior. The simulated data sets differ by sample size and level of intra-consumer heterogeneity. We consider alterations to the sample size as we aim to evaluate the benefit of collecting more data for this estimator. Furthermore, different levels of intra-consumer heterogeneity provide insight into how inter-consumer heterogeneity models behave when they erroneously ignore intra-consumer heterogeneity.

The simulated data sets are based on experiments in Ben-Akiva et al. (2015, p. 59). Each respondent must choose between three unlabeled cars with varying prices and attributes or reject all of the alternatives. The number of menus for each respondent is fixed at eight throughout the experiments. Table 1 shows the grapes' attributes and their associated levels.

Table 1: Attributes and the respective levels of the Grapes Data

Attribute	Symbol	Levels
Price	P	\$10,000 to \$30,000
Domestic car	D	Domestic (1) or Import (0)
Dark color	C	Dark (1) or Bright (0)
Size	L	Large (1) or Small (0)
Electric	E	Electric (1) or Non-eletric (0)

In the utility functions, see Eqs. (38) and (39), the disposable income is denoted as I_n and cancels out in the utility maximization. The three alternatives for the cars correspond to $j = 1,2,3$ and the reject option is indexed as $j = 0$. The subsequent tests require data sets with different heterogeneity structures. The utility denoted in (38) is in WTP-space and refers to the case of only inter-consumer heterogeneity. The ϵ_{jmn} are i.i.d. EV1 distributed and the parameters underlie a multivariate normal distribution. The subscript n demonstrates that the parameters are individual specific. In the scenarios with intra-consumer heterogeneity, depicted in (39), the parameters β_L and β_E are assigned menu specific parameters, while the scale parameter α, β_D, and β_C only vary among individuals. The parameters are also distributed multivariate normal on the intra-consumer level.

$$U_{jmn} \equiv I_n - P_{jmn} + D_{jmn}\,\beta_{D_n} + C_{jmn}\beta_{C_n} + L_{jmn}\beta_{L_n} + E_{jmn}\beta_{E_n} + \alpha_n\epsilon_{jmn} \tag{38}$$

$$U_{jmn} \equiv I_n - P_{jmn} + D_{jmn}\beta_{D_n} + C_{jmn}\beta_{C_n} + L_{jmn}\beta_{L_{mn}} + E_{jmn}\beta_{E_{mn}} + \alpha_n\epsilon_{jmn} \tag{39}$$

The artificial choices correspond to the alternative that maximizes the utility in a given menu. Data sets are simulated with sample sizes of 2000 and 4000 individuals. The theoretical true values as well as the sample values for the case of 2000 individuals for all intra-consumer heterogeneity scenarios are shown in Table 2. The covariances are shown in

Table 3.

Table 2: True values for all scenarios – data with 2000 individuals

	Parameter		SD Inter		SD Intra							
Parameter	All scenarios		All scenarios		No Heterog.		Low Heterog.		Med. Heterog.		High Heterog.	
	Theo.	Sample	Theo.	Sample	Theo.	Sample	Theo.	Sample	Theo.	Sample	Theo.	Sample
$\ln(\alpha)$	-0.5	-0.491	0.3	0.299	0	0	0	0	0	0	0	0
β_D	1	1.005	0.4	0.402	0	0	0	0	0	0	0	0

β_C	0.9	0.898	0.3	0.299	0	0	0	0	0	0	0	0
β_L	2.5	2.489	1	0.988	0	0	0.5	0.497	1	0.995	2	1.990
β_E	1.5	1.523	0.5	0.518	0	0	0.25	0.250	0.5	0.499	1	0.999

The first level clearly refers to the case of no intra-consumer heterogeneity. Low, medium and high intra-consumer heterogeneity refer to 50%, 100% and 200% of the corresponding inter-consumer standard deviations.

Table 3: True values of the covariances for all scenarios - data with 2000 individuals

	Cov Inter		Cov Intra							
	All scenarios		No Heterog.		Low Heterog.		Med. Heterog.		High Heterog.	
	Theo.	Sample	Theo.	Sample	Theo.	Sample	Theo.	Sample	Theo.	Sample
β_L, β_E	0	≈ 0	0	0	-0.038	-0.038	-0.150	-0.153	-0.600	-0.611
β_D, β_C	0.072	0.074	0	0	0	0	0	0	0	0

Regarding the non-diagonal elements of the covariance matrices, only β_D and β_C are generated with a correlation on the inter-consumer level. β_L and β_E are correlated on the intra-consumer level. The true values of the data sets with 4000 individuals are shown in Appendix C.

We assume that the number of parameters and their distribution are known a priori. Models with inter-consumer heterogeneity are estimated with the same specification across data sets, meaning that these models are not specified correctly when intra-consumer heterogeneity is present. In terms of the inter- and intra-consumer heterogeneity models, we assume that only $\beta_L$and β_O are distributed on the intra-consumer level. Therefore, these models are specified with too many parameters when no intra-consumer heterogeneity is present in the data. In this paper the forecasting performance is assessed for the above models using out of sample data, which was generated with the same individual level coefficients and consists of eight new choice situations for each individual.

The MCMC estimations require us to determine other parameters in advance, including the starting values and the number of Gibbs Sampling or Metropolis Hastings algorithm iterations.

The total number of iterations is determined based on the complex inter-intra models; 400,000 iterations are used for each of the models, out of which 150,000 are discarded as burn-in. The number was not adjusted for the models not specified correctly, as a clear convergence to zero for superfluous parameters was not observed even with an increasing number of iterations. The modeler is advised to be suspicious in the presence of high autocorrelations and non-stationary Markov chains. Furthermore, diffuse priors allow for classical tests that might provide information on whether the parameter is significantly different from zero. The number of iterations is high compared to cases found in the literature. Ben-Akiva et al. (2015) stop the Gibbs Sampler after 200,000 iterations, whereas Train (2009) observes convergence after only 20,000 iterations for inter-consumer models.
Further settings such as target acceptance rates, starting values, thinning interval as well as the number of draws used for the simulation of the likelihood are discussed in Appendix A. The software and hardware used are described in Appendix B.

3.2 Monte Carlo Experiment: Hierarchical Bayes vs. Maximum Simulated Likelihood

Subsequently, runtime and estimates are compared to the Maximum Simulated Likelihood estimator. Due to slower runtimes of MSL, the *true* model is specified to be more parsimonious, see Eq. (40) .

$$U_{jmn} \equiv -P_{jmn} + L_{jmn}\beta_{L_{mn}} + \alpha\epsilon_{jmn} \quad (40)$$

β_L varies among individuals and menus and the model is specified without scale heterogeneity. Two datasets are generated with 500 and 2000 individuals and eight menus each. While the number of iterations for Hierarchical Bayes is set to 400,000, the draws for MSL are increased until the true parameters are replicated or the runtime becomes unreasonable. Both models use logit starting values.

The estimation routine for MSL is mostly coded in R. Based on the code of the CMC (2017), the draws are precomputed and the optimum of the likelihood is searched using the BFGS method, as implemented in the R-package maxLik (Henningsen and Toomet 2011). For the purpose of decreasing both runtime and memory usage, the likelihood calculation itself was rewritten and coded in C++.

3.3 Model estimation on empirical data: GPS traces

The effects of adding intra-consumer heterogeneity to the model specification are further investigated on a week-long mobility diary collected in the city of Basel, Switzerland (Becker, Ciari, and Axhausen, 2017). The sample consists of Free-Floating as well as Roundtrip Car Sharing users, and a control group. Readers interested in how the chosen alternative and the attributes of the non-chosen alternatives are determined are referred to (Becker, Ciari, and Axhausen 2017a). Within this work, the alternative car sharing is excluded due to its low modal split in the sample (1.8%). The remaining alternatives are car, public transit, bike, and walk. The variable costs for transit are adjusted to the season ticket ownership and the variable costs for the car are set to 0.268 CHF per km (TCS 2013). In addition, all trips with origin or destination not in Switzerland are excluded. The final dataset consists of 357 individuals and a total of 10202 menus**.** We test two different model specifications, as displayed in Eq. 41 and 42. They are distinguished by the maximum level of heterogeneity for β_{Time} (individual and menu).

$$U_{jmn} \equiv ASC_j - cost_{jmn} - \exp\left(\beta_{Time_n}\right) * Traveltime_{jmn} + \alpha\epsilon_{jmn} \quad (41)$$

$$U_{jmn} \equiv ASC_j - cost_{jmn} - \exp\left(\beta_{Time_{nm}}\right) * Traveltime_{jmn} + \alpha\epsilon_{jmn} \quad (42)$$

4 RESULTS

The results section is structured similarly to the previous section. First we refer to the Monte Carlo experiment comparing inter- and inter-intra-consumer heterogeneity models. Then we discuss the estimation results and runtimes of Hierarchical Bayes and Maximum Simulated Likelihood. Finally, we present the transportation mode choice case.

4.1 Monte Carlo Experiment: Effects of introducing intra-consumer heterogeneity

Within this section, we focus on the estimation results on data with 2000 individuals. Differences observed for models tested on data with 4000 individuals are mentioned if applicable. The respective tables are provided in Appendix C. Given that diffuse priors are used for the model estimation, a classical interpretation is chosen.

Table 4: Comparison goodness of fit - data with 2000 individuals

Scenario	No Heterog.		Low Heterog.		Med. Heterog.		High Heterog.	
Model	Inter	Inter- Intra	Inter	Inter- Intra	Inter	Inter- Intra	Inter	Inter- Intra
Null loglik	-34498.866	-34498.866	-34452.666	-34452.666	-34167.486	-34167.486	-33594.096	-33594.096
Final loglik	-11021.705	-11020.806	-11272.681	-11267.877	-12049.754	-11994.480	-14164.530	-13839.595
ρ^2	0.680	0.680	0.672	0.673	0.647	0.649	0.578	0.588
T	11	14	11	14	11	14	11	14
p-value LR-Test	0.615		0.022		0.000		0.000	

Table 4 summarizes the goodness of fit statistics for models estimated on data with 2000 individuals and all heterogeneity levels. In the case of no intra-consumer heterogeneity, the null hypothesis claiming that the intra-coefficients are all zero cannot be rejected at a p-value of 0.615. Therefore, we would exclude intra-coefficients if no intra-consumer heterogeneity is present in the data, and in each of the scenarios where intra-consumer heterogeneity is present, the null hypothesis can be rejected.

Table 5 Parameter estimates - data with inter- and different levels of intra-consumer heterogeneity (2000 individuals), model with only inter-consumer heterogeneity

Parameter	No Heterog.			Low Heterog.			Med. Heterog.			High Heterog.		
	Mean	Std Dev	Pct. Err.	Mean	Std Dev	Pct. Err.	Mean	Std Dev	Pct. Err.	Mean	Std Dev	Pct. Err.
$\ln(\alpha)$	-0.522	0.020	-	-0.489	0.019	-	-0.398	0.019	-	-0.134	0.017	-
β_D	1.013	0.022	0.8%	1.024	0.022	1.9%	1.015	0.023	1.0%	1.010	0.025	0.5%
β_C	0.884	0.020	1.6%	0.887	0.020	1.2%	0.877	0.022	2.4%	0.851	0.024	5.3%
β_L	2.457	0.034	1.3%	2.429	0.034	2.4%	2.377	0.034	4.5%	2.240	0.034	10.0%
β_E	1.505	0.024	1.2%	1.506	0.023	1.1%	1.479	0.024	2.9%	1.469	0.028	3.5%
$\ln(\alpha)$ SD Inter	0.318	0.028	-	0.312	0.028	-	0.296	0.027	-	0.203	0.033	-
β_D SD Inter	0.425	0.032	5.7%	0.426	0.035	5.8%	0.418	0.036	4.1%	0.400	0.047	0.6%
β_C SD Inter	0.280	0.041	6.2%	0.270	0.041	9.6%	0.307	0.045	2.7%	0.263	0.064	12.0%
β_L SD Inter	0.968	0.034	2.0%	0.978	0.035	1.0%	0.951	0.034	3.7%	0.893	0.036	9.6%
β_E SD Inter	0.501	0.032	3.4%	0.505	0.032	2.6%	0.513	0.033	1.0%	0.587	0.037	13.3%

In Table 5 the parameter estimates are presented based on models with only inter-consumer heterogeneity. With increasing levels of intra-consumer heterogeneity, the scale coefficient decreases in absolute terms, which indicates a lower explanatory power of the model. Furthermore, the β_L estimate decreases from 2.457 to 2.240 (8.83%), while β_E only declines from 1.505 to 1.469 (2.39%).

The inter-standard deviations are influenced by the introduction of intra-consumer heterogeneity. It is interesting to observe that β_L and β_E seem to change in opposite directions. Comparing the scenarios of no and high intra-consumer heterogeneity, the inter-standard deviation of β_L decreases by 7.75%, whereas the counterpart of β_E increases by 17.17%. The coefficients of variation $\left(\frac{\sigma}{\mu}\right)$ for β_L only increases by 1.11%, whereas the corresponding value of β_E increases by 20.15%.

Table 6: Covariance - data with inter- and different levels of intra-consumer heterogeneity (2000 individuals), model with inter-consumer heterogeneity

Parameter	No Heterog.			Low Heterog.			Med. Heterog.			High Heterog.		
	Mean	Std Dev	Pct. Err.	Mean	Std Dev	Pct. Err.	Mean	Std Dev	Pct. Err.	Mean	Std Dev	Pct. Err.
β_D, β_C Inter	0.069	0.017	6.8%	0.059	0.017	20.3%	0.057	0.019	23.0%	0.04	0.022	45.9%

We also observe that the estimated inter-covariance between β_C and β_D declines from 0.069 to 0.040 with an augmenting intra-consumer heterogeneity. In the case of no heterogeneity, the estimated correlation between β_C and β_D is 0.58 and therefore close to the true value of 0.6. If intra-consumer heterogeneity is increased to the level "high", a value of 0.38 is estimated.

For the subsequent results, the model specification incorporates intra-consumer heterogeneity. Contrary to the previous model specification, the sample level parameter estimates only change slightly among the various scenarios, even for the parameters that are directly affected (see table Table 7).

Table 7: Parameter estimates - data with inter- and different levels of intra-consumer heterogeneity (2000 individuals), model with inter- and intra-consumer heterogeneity

Parameter	No Heterog.			Low Heterog.			Med. Heterog.			High Heterog.		
	Mean	Std Dev	Pct. Err.	Mean	Std Dev	Pct. Err.	Mean	Std Dev	Pct. Err.	Mean	Std Dev	Pct. Err.
$\ln(\alpha)$	-0.543	0.020	-	-0.525	0.023	-	-0.529	0.028	-	-0.516	0.029	-
β_D	1.014	0.021	0.8%	1.027	0.022	2.2%	1.028	0.022	2.3%	1.039	0.025	3.3%
β_C	0.888	0.020	1.2%	0.893	0.020	0.6%	0.887	0.021	1.3%	0.893	0.023	0.7%
β_L	2.431	0.027	2.3%	2.458	0.036	1.2%	2.480	0.038	0.4%	2.514	0.051	1.0%
β_E	1.506	0.023	1.1%	1.513	0.023	0.7%	1.494	0.025	1.9%	1.495	0.030	1.8%
$\ln(\alpha)$ SD Inter	0.323	0.026	-	0.326	0.031	-	0.346	0.035	-	0.342	0.046	-
β_D SD Inter	0.434	0.031	7.9%	0.426	0.033	5.8%	0.429	0.035	6.6%	0.450	0.038	12.0 %
β_C SD Inter	0.282	0.040	5.8%	0.270	0.044	9.7%	0.304	0.047	1.7%	0.237	0.060	20.7 %
β_L SD Inter	0.983	0.020	0.4%	0.984	0.034	0.3%	0.966	0.039	2.2%	0.940	0.051	4.9%

β_E SD Inter	0.500	0.029	3.6%	0.502	0.026	3.2%	0.500	0.033	3.5%	0.518	0.045	0.0%
β_L SD Intra	0.022	0.027	-	0.458	0.080	8.0%	0.966	0.039	2.9%	1.953	0.070	1.8%
β_E SD Intra	0.158	0.080	-	0.188	0.091	24.8%	0.500	0.033	0.1%	0.988	0.062	1.1%

In terms of the inter-standard deviations, the changes are less prominent. Comparing the cases of no and high intra heterogeneity, β_L only decreases by 4.37%, whereas β_E increases by 3.6%, reaching the true value in the sample. β_C is again unstable among the scenarios and reaches the best estimate when estimated on data with medium intra-consumer heterogeneity. The over-estimation of the inter-standard deviation of β_D has a positive relationship with the level of intra-consumer heterogeneity. In the case of 4000 individuals, all estimates are closer to their true value except for β_D.

In terms of inter-consumer covariance, the decline from 0.063 to 0.044 from scenario 1 to 4 is accompanied by a decline in the inter-standard deviation of β_C and a decrease in the estimated correlation between β_C and β_S. While a correlation of 0.51 is estimated in the case of low heterogeneity, the value declines to 0.41 in the last scenario.

Intra-consumer standard deviations are shown at the bottom of Table 7. If the model is wrongly specified, meaning no intra-consumer heterogeneity is present in the data, the null hypothesis that the intra-standard deviation of β_L is different from zero cannot be rejected. Nonetheless, the same t-test for the intra-standard deviation β_E can be rejected with a p-value of 0.048. In this case, it is worthwhile to investigate the associated Markov Chains. The Markov chain of the intra-standard deviation of β_E has not converged at a rather high number of iterations (400,000). When intra-consumer heterogeneity is deliberately introduced during the generation of the data, only the β_O estimate in the case of low intra-consumer heterogeneity is far from the true value, with an underestimation of 24.8%. For the other scenarios and estimates, the true patterns of intra-consumer heterogeneity recover and provide a correct picture of the coefficient's variation among the menus.

Table 8: Covariances - data with inter- and different levels of intra-consumer heterogeneity (2000 individuals), model with inter- and intra-consumer heterogeneity

Parameter	No Heterog.			Low Heterog.			Med. Heterog.			High Heterog.		
	Mean	Std Dev	Pct. Err.	Mean	Std Dev	Pct. Err.	Mean	Std Dev	Pct. Err.	Mean	Std Dev	Pct. Err.
β_D, β_C Inter	0.063	0.017	15.1%	0.058	0.017	20.8%	0.056	0.019	23.5%	0.044	0.021	40.5%
β_L, β_E Intra	0.000	0.001	-	0.006	0.033	115%	-0.049	0.051	67.9%	-0.545	0.101	10.8%

The covariance estimation method correctly estimates that there is no inter-covariance in the first scenario. The estimate of the last scenario of -0.545 is close to the true value -0.6. Nevertheless, the estimates for the second and third scenarios are not significantly different from zero and the estimates are far from the true values. This problem alleviates with an increased sample size of 4000.

We note that the models have been run twice with a different seed. For the correctly specified models, the sample level parameter estimates deviate on average by 0.03% for the inter-consumer model and 0.13% for all inter-intra-consumer models, with an upper value of 0.39%. For SD Inter the corresponding values are

0.86% and 1.08%. The inter standard deviation of β_C in inter-intra-models deviates most in the high intra-heterogeneity scenario and amounts to 0.237 and 0.248. In the low intra-consumer heterogeneity case, the intra-standard deviation of β_E has not converged even after one million iterations. The estimate deviates by 42.9%. In the remaining scenarios the intra-standard deviations deviate by 3.8% on average. The inter covariance of β_S and β_C deviates by 2.95% for the inter-model and on average by 1.46% for the inter-intra-consumer models. In addition, a Gelman and Rubin Multiple Sequence Diagnostic (Gelman 1992) with five chains has been conducted for the Inter-Intra model on data with medium intra-consumer heterogeneity. The potential scale reduction factor is one for all parameters except for β_E SD Intra, for which it amounts to 1.02 with an upper confidence limit of 1.06.

In order to assess the forecasting performance of the two model specifications, choices are predicted on out of sample data that were generated with the same individual level coefficients. For models estimated with only inter-consumer heterogeneity, individual level estimates are used for prediction. For models that incorporate inter- and intra-consumer heterogeneity, 2000 menu-level parameters are drawn for each choice situation. The mean of the predicted probabilities is presented in Table 9. For illustration purposes the case of very high intra-personal heterogeneity (β_L SD Intra=10, β_E SD Intra=5) was added.

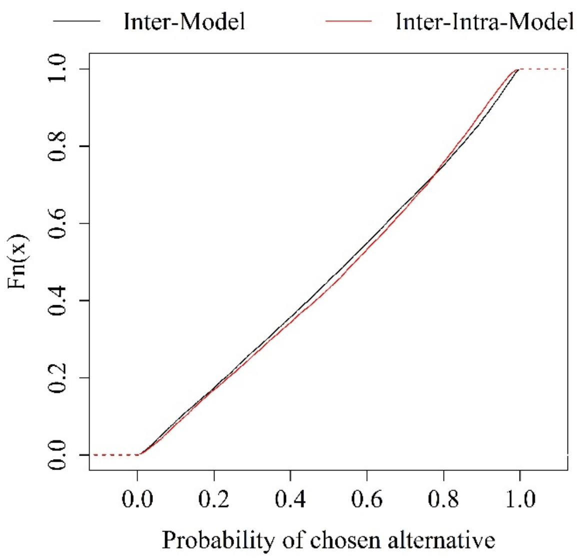
Table 9: Average predicted probability of chosen alternative

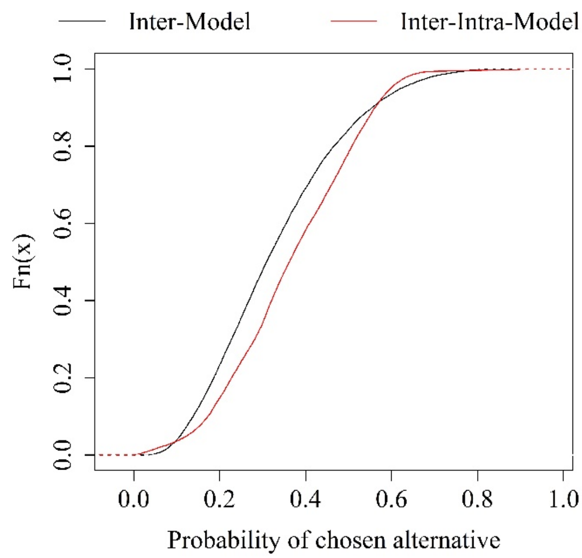
Model	No Heterog.	Low Heterog.	Med. Heterog.	High Heterog.	V. High Heterog.
Inter	0.657	0.646	0.615	0.535	0.330
Inter-Intra	0.654	0.646	0.616	0.540	0.366

The mean probability for every menu for the high and very high intra-consumer heterogeneity case is plotted in Figure 1. We see that the inter-model predicts more extreme probabilities than the inter-intra-consumer heterogeneity model.

Figure 1: Distribution of probabilities of chosen alternative

(a) High intra-personal heterogeneity (b) Very high intra-personal heterogeneity





The difference in forecasting performance for both model specifications is limited for this data set. Readers interested in specific advantages for forecasting are referred to the Discussion section, where we elaborate on online updates.

4.2 Monte Carlo Experiment: Hierarchical Bayes vs. Maximum Simulated Likelihood

As described in Section 3.3, estimates and runtime of HB and MSL are compared on a simulated dataset. Table 10 shows the results for a dataset with 500 individuals with eight menus each. While the HB-estimates are close to the true sample estimates, the MSL estimate for β_L SD Intra has a large deviation to the true value. MSL is not able to recover the true intra-consumer standard deviation. Furthermore, the runtime for MSL for 1000 draws on each heterogeneity level is almost 17 times higher and amounts to more than one and a half days.

Table 10: Parameter estimates – HB vs MSL – 500 individuals

Method	True Values		HB		MSL		MSL	
Iterations	-		400K		-		-	
Draws Inter/Intra	-		-		500/500		1000/1000	
	Theo.	Sample	Mean	Std Dev	Est.	SE	Est.	SE
ln(α)	-0.500	-0.500	-0.512	0.026	-0.506	0.028	-0.506	0.028
β_L	2.500	2.539	2.514	0.056	2.505	0.056	2.505	0.055
β_L SD Inter	1.000	1.007	1.016	0.056	1.001	0.057	1.001	0.056
β_L SD Intra	0.500	0.500	0.493	0.085	0.422	0.132	0.421	0.133
Runtime	-		134 mins		664 mins		2246 mins	

In a further experiment, we increase the number of individuals to 2000. Both methods produce estimates that are close to the true values of the sample, as displayed in Table 11**Error! Not a valid bookmark self-reference.**. However, the estimate for the intra-consumer standard deviation still deviates by 1.8% and the runtime of Maximum Simulated Likelihood amounts to almost one and a half days.

Table 11: Parameter estimates – HB vs MSL – 2000 individuals

Method	True Values		HB		MSL	
Iterations	-		400K		-	
Draws Inter/Intra	-		-		500/500	
	Theo.	Sample	Mean	Std Dev	Est.	SE
ln(α)	-0.500	-0.500	-0.500	0.013	-0.499	0.014
β_L	2.500	2.539	2.536	0.028	2.534	0.028
β_L SD Inter	1.000	1.007	0.997	0.029	0.993	0.029
β_L SD Intra	0.500	0.500	0.503	0.053	0.491	0.062
Runtime	-		370 mins		2061 mins	

4.3 Model estimation on empirical data: GPS traces

As described in Section 3.3, a model considering only inter-consumer heterogeneity and a model considering both inter- and intra-consumer heterogeneity were estimated.

Table 12: Comparison Goodness of fit - GPS Traces

Model	Inter Model	Intra Model
Null loglik	-14142.975	-14142.975
Uncond. loglik	-9909.272	-9507.930
Nr parameters	5	6
ρ^2	0.299	0.328
p-value LR-Test	0.000	
Cond. Loglik	-9402.516	-6648.061
Mean P (cond.)	0.503	0.604
Nr. Indiv	357	
Nr. Menus	10202	

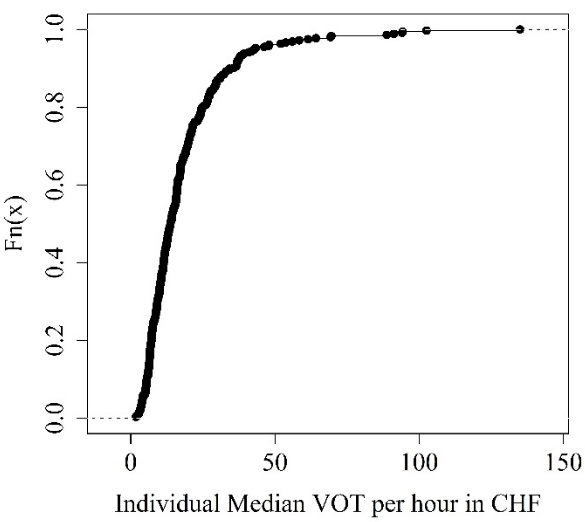
Table 12 shows that the addition of intra-consumer heterogeneity leads to an increase of 401.342 for the unconditional likelihood (see Eq. 4). The likelihood is calculated based on 2000 draws for every individual and menu. The difference for the conditional likelihood, which uses 7500 thinned draws of the deep level parameters of the MCMC chains, amounts to 2754.455.

Table 13: Parameter estimates – GPS Traces

Parameter	Inter-Model		Inter-Intra-Model	
	Mean	Std Dev	Mean	Std Dev
α	5.160	0.302	3.989	0.207
ASC_{PT}	11.883	0.792	14.724	0.831
ASC_{Bike}	8.049	0.631	10.135	0.658
ASC_{Walk}	13.378	0.917	16.817	0.961
β_{Time}	-0.360	0.071	-0.081	0.068
β_{Time} SD Inter	0.809	0.038	0.816	0.042
β_{Time} SD Intra			0.821	0.025

The parameter estimates are displayed in Table 13. All parameters are significant to the 5% level and their respective Markov chains have converged. The median VOT amounts to 8.11 CHF per hour in the inter-consumer model, and to 13.87 CHF per hour in the model also considering intra-consumer heterogeneity.

Figure 2: Distribution of Individual Median VOTs



The cumulative distribution function of the individual median VOTs is shown in Figure 2. For illustration purposes one VOT that amounts to 209 CHF/hr is excluded from the plot. While the number of individual VOTs close to zero is relatively low, 63.20% of the VOTS are between 10 and 50 CHF/hr. Only 0.56% of the VOTs are larger than 100 CHF/hr.

5 DISCUSSION

Selecting the appropriate model specification can be challenging when the data to be analyzed involves intra-personal heterogeneity. Based on our results presented in Section 4, modelers would be advised to investigate the Markov Chains and use enough draws for the computation of the unconditional likelihood. Section 4 also showed the consequences of falsely ignoring the intra-consumer heterogeneity level in the model specification. For both sample sizes, the population-level parameter estimates declined in willingness to pay space with an increasing level of intra-consumer heterogeneity, even though the true values did not. Interestingly this observation is not limited to the parameters that were specified to vary among the different menus. To conclude, the results indicate that the appropriate incorporation of intra-consumer heterogeneity avoids obtaining biased estimates. Apart from the parameter estimates, it is also noteworthy that the coefficient of variation increased for one of the parameters exposed to intra-consumer heterogeneity. This demonstrates that the intra-consumer heterogeneity is falsely detected as inter-consumer heterogeneity, leading to erroneous interpretations of the parameter's variation across respondents. It is crucial to know whether and how much the customer's preferences change among various situations or if subsets of the sample behave differently. Furthermore, the absolute value of the scale coefficient of inter-consumer models decreases in line with an increase of the intra-consumer heterogeneity level. This effect reveals the augmenting variance of the error term, with the model losing explanatory power. In contrast, we do not observe this effect for the inter- and intra-consumer heterogeneity models.

It is beneficial for the modeler to test for intra-consumer heterogeneity on panel data and to validate the test with an inspection of the Markov Chain. Not only is it possible that the parameter estimates are biased, but there may be a misleading picture of the parameter variation across individuals. The increased runtime of inter- and intra-consumer heterogeneity models can be justified based on these disadvantages. Although the offline forecasting did not show substantial advantages compared to inter-consumer heterogeneity models, the menu-level parameters can be used for online updates in a system that continuously collects data. Danaf et. al. (2017) show that the prediction performance of models using online updates is very similar to models estimated with the full Gibbs Sampler.

The comparison of MSL, for which both integrals were simulated with independent draws, and HB showed that the runtime for HB is substantially shorter than for MSL despite the precomputation of the draws and the additional MH step of a fixed parameter. For models that only incorporate inter-consumer heterogeneity, Train (2009) reports that runtimes for MSL and HB are comparable when all variables are distributed normal without correlations, yet the runtime more than doubles for HB if one variable does not vary among individuals. To summarize both the results of Train (2009) and this paper, HB is faster if a full covariance matrix is estimated and/or intra-consumer heterogeneity is added. Furthermore, the MSL-estimate for the intra-consumer standard deviation still deviates by 15.8% for 1000 draws on both levels and a sample size of 4000 menus.

The model results on GPS-traces showed that the addition of intra-personal heterogeneity can lead to substantial increases in the unconditional likelihood. Given that information is limited about trip characteristics and the number of people joining a trip is often available for GPS data, it is easy to justify that intra-consumer heterogeneity plays a role in explaining mode choices.

6 CONCLUSION AND OUTLOOK

In this paper, a Hierarchical Bayes estimator for Logit Mixtures with both inter- and intra-consumer heterogeneity is introduced and tested. By including parameter estimates for population-, individual-, and menu-levels, we provide a comprehensive picture of the variation of parameters among individuals and menus. In the Monte Carlo simulation, we show that disregarding the intra-consumer heterogeneity level in the specification leads to inconsistent parameter estimates and inflated coefficients of variation on the inter-consumer level. This error indicates that preference variation among the menus was mistaken as preference variation among individuals. The results of mode choice models on GPS-traces of inhabitants from the city of Basel, Switzerland further showed that the inclusion of intra-consumer heterogeneity can substantially improve the model fit.

Even if the model is correctly specified, the inter-correlation seems to be influenced by different levels of intra-consumer heterogeneity, which requires further investigation. Furthermore, the chain convergence is problematic for intra-standard deviations with a relatively small part worth.

For the implementations available, we show that Hierarchical Bayes has computational advantages to MSL. For a dataset of 16000 menus, the runtime is 5.5 times higher for MSL than HB.

Possible improvements to our method include further investigating trends and autocorrelation in the menu-level parameters. In conjunction with online-updates, this work could substantially increase the forecasting performance.

In addition, modeling burden can be reduced with available methods. Up to this point, modelers must determine the number of Gibbs Sampling iterations in advance and check whether the Markov chains have converged after the estimation. A method to determine the number of burn-in iterations by monitoring the convergence could eliminate this task. An example of this method is shown in the Gelman and Rubin Multiple Sequence Diagnostic, in which multiple chains are run and the within-chain is compared to the between-chain variance to determine whether the chain has converged (Gelman 1992).

We can also replace the general rule in the Allenby-Train procedure that only considers every tenth draw for the calculation of the posterior mean and variance. Link and Eaton (2012) emphasize that this so-called *thinning* is inefficient. However, the draws must be independent for the calculation of the standard deviation. This can be achieved by adjusting for the inherent order of the autocorrelation.

A final topic for future work involves finding the appropriate model specification in terms of the distribution, the maximum level of heterogeneity, and the correlation structure. Balcombe et al. (2009) point out that it is possible to compare non-nested sub-models resulting from a Bayesian analysis using the marginal likelihood. Combining the latter with a step-wise algorithm has the potential to drastically reduce the modeling effort.

ACKNOWLEDGEMENTS

This work is partially aligned with the TRIPOD: Sustainable Travel Incentives with Prediction, Optimization and Personalization research project sponsored by the U.S. Department of Energy Advanced Research Projects Agency-Energy (ARPA-E). It was awarded through the ARPA-E Traveler Response Architecture using Novel Signaling for Network Efficiency in Transportation (TRANSNET) program. We thank Romain Crastes dit Sourd for providing R-code for the MSL estimation. We further sincerely thank five anonymous reviewers for constructive criticism and insightful comments about the paper.

REFERENCES

Andersen, Laura Mørch. 2014. "Obtaining Reliable Likelihood Ratio Tests from Simulated Likelihood Functions." *PloS one* 9(10): e106136.

Balcombe, Kelvin, Ali Chalak, and Iain Fraser. 2009. "Model Selection for the Mixed Logit with Bayesian Estimation." *Journal of Environmental Economics and Management* 57(2): 226–37.

Becker, Henrik, Francesco Ciari, and Kay W Axhausen. 2017a. "Modeling Free-Floating Car-Sharing Use in Switzerland: A Spatial Regression and Conditional Logit Approach." *Transportation Research Part C: Emerging Technologies* 81(Supplement C): 286–99. http://www.sciencedirect.com/science/article/pii/S0968090X17301614.

Becker, Henrik, Francesco Ciari, and Kay Werner Axhausen. 2017b. "Measuring the Travel Behaviour Impact of Free-Floating Car-Sharing." In *96th Transportation Research Board Annual Meeting (TRB 2016)*,.

Ben-Akiva, Moshe, Daniel McFadden, and Kenneth Train. 2015. "Foundations of Stated Preference Elicitation: Choice-Based Conjoint Analysis, Consumer Choice Behavior, and Measurement of Consumer Welfare."

Bhat, Chandra R. 2011. "The Maximum Approximate Composite Marginal Likelihood (MACML) Estimation of Multinomial Probit-Based Unordered Response Choice Models." *Transportation Research Part B: Methodological* 45(7): 923–39. http://www.sciencedirect.com/science/article/pii/S019126151100049X.

Bhat, Chandra R, and Saul Castelar. 2002. "A Unified Mixed Logit Framework for Modeling Revealed and Stated Preferences: Formulation and Application to Congestion Pricing Analysis in the San Francisco Bay Area." *Transportation Research Part B: Methodological* 36(7): 593–616.

Bhat, Chandra R, and Rupali Sardesai. 2006. "The Impact of Stop-Making and Travel Time Reliability on Commute Mode Choice." *Transportation Research Part B: Methodological* 40(9): 709–30.

Bhat, Chandra R, and Raghuprasad Sidharthan. 2011. "A Simulation Evaluation of the Maximum Approximate Composite Marginal Likelihood (MACML) Estimator for Mixed Multinomial Probit Models." *Transportation Research Part B: Methodological* 45(7): 940–53. http://www.sciencedirect.com/science/article/pii/S0191261511000506.

Cherchi, Elisabetta. 2009. "A Mixed Logit Mode Choice Model for Panel Data:accounting for Different Correlation over Time Periods." *International Choice Modelling Conference, Harrogate*.

CMC. 2017. "CMC Choice Modelling Code for R." www.cmc.leeds.ac.uk.

Danaf, Mazen et al. 2017. "Personalized Recommendations Using Discrete Choice Models with Inter-and Intra-Personal Heterogeneity." In *Proceedings of the Fifth International Choice Modeling Conference*, Cape Town.

Dumont, Jeffrey, and Jeffrey Keller. 2015. *RSGHB: Functions for Hierarchical Bayesian Estimation: A Flexible Approach*. http://cran.r-project.org/package=RSGHB.

Gelman, Andrew. 1992. "Iterative and Non-Iterative Simulation Algorithms." *Computing Science and Statistics (Interface Proceedings)* 24: 433–38.

Henningsen, Arne, and Ott Toomet. 2011. "maxLik: A Package for Maximum Likelihood Estimation in {R}." *Computational Statistics* 26(3): 443–58. http://dx.doi.org/10.1007/s00180-010-0217-1.

Hess, Stephane, and John M Rose. 2009. "Allowing for Intra-Respondent Variations in Coefficients Estimated on Repeated Choice Data." *Transportation Research Part B: Methodological* 43(6): 708–19.

Hess, Stephane, and Kenneth E Train. 2011. "Recovery of Inter- and Intra-Personal Heterogeneity Using

Mixed Logit Models." *Transportation Research Part B: Methodological* 45(7): 973–90.

Huang, Alan, and M P Wand. 2013. "Simple Marginally Noninformative Prior Distributions for Covariance Matrices." *Bayesian Analysis* 8(2): 439–52.

Huber, Joel, and Kenneth Train. 2001. "On the Similarity of Classical and Bayesian Estimates of Individual Mean Partworths." *Marketing Letters* 12(3): 259–69. http://dx.doi.org/10.1023/A%3A1011120928698.

Link, William A, and Mitchell J Eaton. 2012. "On Thinning of Chains in MCMC." *Methods in Ecology and Evolution* 3(1): 112–15.

Matsumoto, Makoto, and Takuji Nishimura. 1998. "Mersenne Twister: A 623-Dimensionally Equidistributed Uniform Pseudo-Random Number Generator." *ACM Transactions on Modeling and Computer Simulation* 8(1): 3–30.

McCulloch, Robert, and Peter E Rossi. 1994. "An Exact Likelihood Analysis of the Multinomial Probit Model." *Journal of Econometrics* 64(1--2): 207–40.

Neiswanger, Willie, Chong Wang, and Eric Xing. 2013. "Asymptotically Exact, Embarrassingly Parallel MCMC." http://arxiv.org/pdf/1311.4780.

Newey, Whitney K, and Kenneth D West. 1987. "A Simple, Positive Semi-Definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix." *Econometrica* 55(3): 703–8.

Patil, Priyadarshan N et al. 2017. "Simulation Evaluation of Emerging Estimation Techniques for Multinomial Probit Models." *Journal of Choice Modelling* 23: 9–20. http://www.sciencedirect.com/science/article/pii/S1755534515300890.

Rossi, Peter E, and Greg M Allenby. 2003. "Bayesian Statistics and Marketing." *Marketing Science* 22(3): 304–28.

Sawtooth Software. 2009. "The CBC/HB System forHierarchical Bayes Estimation." http://www.sawtoothsoftware.com/products/advanced-analytical-tools/cbc-hierarchical-bayes-module/167-support/technical-papers/sawtooth-software-products/128-cbc-hb-technical-paper-2009 (March 26, 2016).

Song, Xiang et al. 2016. "Enhancements of Hierarchical Bayes Procedure for Logit Mixture: Working Paper." *Working Paper*.

TCS. 2013. "Ein Durchschnittsfahrzeug Kostet 76 Rappen pro Kilometer: Schnelle Und Einfache Berechnung Der Kilometerkosten Dank Dem TCS." *tcs.ch*.

Train, Kenneth E. 2001. "A Comparison of Hierarchical Bayes and Maximum Simulated Likelihood for Mixed Logit: Working Paper."

———. 2006. "Mixed Logit Estimation by Hierarchical Bayes." http://eml.berkeley.edu/Software/abstracts/.

———. 2009. *Discrete Choice Methods with Simulation*. 2nd ed. New York: Cambridge university press.

Appendix A: Settings for Hierarchical Bayes

It is necessary to set the target acceptance rates for the three Metropolis Hastings steps in the Gibbs Sampler. Train (2006) and Sawtooth Software (2009) use a value of 0.3. The value is kept to the industry standard of Hierarchical Bayes for Mixed Logit, even though further research could provide insight regarding the choice of the level and whether adjustments are useful for each of the three Metropolis Hastings steps.

Although starting values for the parameter estimates are regarded as critical by some authors (Ben-Akiva, McFadden, and Train 2015), an influence could not be observed when the number of iterations was set to 400,000 in this particular case. The starting values for the population-level parameters are therefore set to zero. However, if the variances are set to zero, then the sampling does not provide different values and the variance covariance matrix is not invertible. In accordance with (Dumont and Keller 2015), the variances are set to two. Furthermore, the user needs to determine the starting values for the three different ρ in the extended version if this is not done by the software. The starting values are similar to those in (Dumont and Keller 2015), where the starting value for the ρ of the parameters with no heterogeneity is 0.0001 and 0.1 for the MH-step of the parameters with inter-consumer heterogeneity. Due to the similar structure of the MH-step for the menu parameters, the starting value for the respective ρ is set to 0.1.

In addition, it is common practice to set a thinning interval for the draws from the conditional posteriors. As the draws are based on the previous iteration or are exactly the same in case the trial values have not been accepted, the draws are autocorrelated. This prohibits the calculation of standard errors without any adjustments. Train (2009) and Ben-Akiva, McFadden, and Train (2015) circumvent this issue by considering only every tenth draw of the Gibbs Sampling for the estimation of the parameters (thinning). Despite the fact that this procedure is inefficient (Link and Eaton 2012), it is sufficient to account for significant autocorrelations up to the tenth lag. Since Markov chains are not autocorrelated up to high lags, the standard thinning method is chosen. However, it is important to note that the calculation of standard errors based on highly autocorrelated Markov chains requires methods like the Newey West standard error (Newey and West 1987), as mentioned in (McCulloch and Rossi 1994).

Another crucial point is the simulation of reliable likelihood values. Andersen (2014) indicated that asymmetric draws might lead to inconsistent likelihood-ratio tests, meaning that the likelihood of the restricted model is higher than the one of the unrestricted model. Nonetheless, the use of antithetic draws leads to high computational times for models with high dimensionality. For this reason, the stability of the likelihood was evaluated depending on the number of Halton draws, similar to (Hess and Train 2011). In this case, stability was reached after 2,000 draws on the inter-consumer level and 2,000 draws on the intra-consumer level. New draws on the intra-consumer level are obtained for each one of the draws on the inter-consumer level rather than reusing the draws.

Appendix B: Software and Hardware

In order to meet specific requirements for the output files and to have the flexibility to adapt the code to new improvements of the MCMC estimation, the software used for the estimations was implemented in R. However, parts of the code are based on the work of (Dumont and Keller 2015), whose code is based on the Matlab code of Train (2006). The estimations are carried out under R version 3.2.2 and the default R-random number generator Mersenne-Twister of Matsumoto and Nishimura (1998). For hardware, an

Ubuntu server with 24 x Intel(R) Xeon(R) CPU @ 2.00 GHz and 16 GB Ram was available. The computation time of inter-models on data with 2000 individuals took about 3.33 hours (0.03 seconds per Gibbs Sampling iteration). The respective times for inter-intra models are 15.5 hours and 0.14 seconds. Parallelization approaches like the one from Neiswanger, Wang, and Xing (2013) have not been implemented up to this point but promise runtime reductions.

Appendix D: Simulation experiments on data with 4000 individuals

Table D.1: True values for all scenarios – data with 4000 individuals

	Parameter		SD Inter		SD Intra							
Parameter	All scenarios		All scenarios		No Heterog.		Low Heterog.		Med. Heterog.		High Heterog.	
	Theo.	Sample	Theo.	Sample	Theo.	Sample	Theo.	Sample	Theo.	Sample	Theo.	Sample
$\ln(\alpha)$	-0.5	-0.491	0.3	0.300	0	0	0	0	0	0	0	0
β_S	1	1.010	0.4	0.407	0	0	0	0	0	0	0	0
β_C	0.9	0.903	0.3	0.302	0	0	0	0	0	0	0	0
β_L	2.5	2.483	1	0.979	0	0	0.5	0.501	1	1.003	2	2.005
β_O	1.5	1.502	0.5	0.509	0	0	0.25	0.250	0.5	0.500	1	1.000

Table D.2: True values of the covariances for all scenarios - data with 4000 individuals

	Cov Inter		Cov Intra							
	All scenarios		No Heterog.		Low Heterog.		Med. Heterog.		High Heterog.	
	Theo.	Sample	Theo.	Sample	Theo.	Sample	Theo.	Sample	Theo.	Sample
β_L, β_O	0	≈ 0	0	0	-0.0375	-0.038	-0.15	-0.152	-0.600	-0.607
β_S, β_C	0.072	0.075	0	0	0	0	0	0	0	0

Table D.3: Comparison goodness of fit - data with 4000 individuals

Scenario	No Heterog.		Low Heterog.		Med. Heterog.		High Heterog.	
Model	Inter	Inter- Intra	Inter	Inter- Intra	Inter	Inter- Intra	Inter	Inter- Intra
Null loglik	-69118.060	-69118.060	-68868.140	-68868.140	-68191.690	-68191.690	-66564.180	-66564.180
Final loglik	-22425.438	-22424.289	-23189.071	-23175.145	-24739.137	-24625.692	-28747.980	-28119.973
ρ^2	0.675	0.675	0.663	0.663	0.637	0.639	0.568	0.577
K	11	14	11	14	11	14	11	14
p-value LR-Test	0.513		0.000		0.000		0.000	

Table D.4: Parameter estimates - data with inter- and different levels of intra-consumer heterogeneity (4000 individuals), model with only inter-consumer heterogeneity

Parameter	True Values		No Heterog.		Low Heterog.		Med. Heterog.		High Heterog.	
	Theo.	Sample	Est.	SE	Est.	SE	Est.	SE	Est.	SE
$\ln(\alpha)$	-0.5	-0.491	-0.502	0.014	-0.445	0.013	-0.361	0.013	-0.136	0.012
β_S	1	1.010	1.003	0.015	0.999	0.016	0.987	0.016	0.965	0.018
β_C	0.9	0.903	0.904	0.015	0.900	0.015	0.884	0.016	0.854	0.016
β_L	2.5	2.483	2.484	0.025	2.452	0.024	2.372	0.024	2.184	0.024
β_O	1.5	1.502	1.501	0.017	1.493	0.017	1.485	0.017	1.455	0.019

Table D.5: Standard deviations for inter-consumer heterogeneity - data with inter and different levels of intra-consumer heterogeneity (4000 individuals), model with only inter-consumer heterogeneity

Parameter	True Values		No Heterog.		Low Heterog.		Med. Heterog.		High Heterog.	
	Theo.	Sample	Est.	SE	Est.	SE	Est.	SE	Est.	SE
$\ln(\alpha)$	0.3	0.300	0.317	0.021	0.280	0.021	0.259	0.023	0.189	0.026
β_S	0.4	0.407	0.419	0.024	0.447	0.024	0.458	0.025	0.466	0.031
β_C	0.3	0.302	0.350	0.029	0.320	0.031	0.348	0.031	0.302	0.038
β_L	1	0.979	1.004	0.024	0.979	0.024	0.941	0.025	0.901	0.025
β_O	0.5	0.509	0.507	0.023	0.512	0.023	0.522	0.024	0.551	0.027

Table D.6: Covariances for inter-consumer heterogeneity - data with inter- and different levels of intra-consumer heterogeneity (4000 individuals), model with only inter-consumer heterogeneity

Parameter	True Values		No Heterog.		Low Heterog.		Med. Heterog.		High Heterog.	
	Theo.	Sample	Est.	SE	Est.	SE	Est.	SE	Est.	SE
β_S, β_C	0.072	0.075	0.064	0.013	0.061	0.014	0.054	0.016	0.039	0.017

Table D.7: Parameter estimates - data with inter- and different levels of intra-consumer heterogeneity (4000 individuals), model with inter- and intra-consumer heterogeneity

Parameter	True Values		No Heterog.		Low Heterog.		Med. Heterog.		High Heterog.	
	Theo.	Sampl e	Est.	SE	Est.	SE	Est.	SE	Est.	SE
$\ln(\alpha)$	-0.5	-0.491	-0.515	0.016	-0.490	0.016	-0.495	0.018	-0.512	0.021
β_S	1	1.010	1.004	0.015	1.003	0.016	0.996	0.016	1.007	0.017
β_C	0.9	0.903	0.903	0.015	0.903	0.015	0.892	0.015	0.890	0.016
β_L	2.5	2.483	2.489	0.026	2.487	0.026	2.463	0.028	2.432	0.034
β_O	1.5	1.502	1.500	0.015	1.493	0.016	1.480	0.018	1.474	0.021

Table D.8: Standard deviations for inter-consumer heterogeneity - data with inter and different levels of intra-consumer heterogeneity (4000 individuals), model with inter- and intra-consumer heterogeneity

Parameter	True Values		No Heterog.		Low Heterog.		Med. Heterog.		High Heterog.	
	Theo.	Sample	Est.	SE	Est.	SE	Est.	SE	Est.	SE
$\ln(\alpha)$	0.3	0.300	0.324	0.021	0.302	0.022	0.320	0.024	0.327	0.035
β_S	0.4	0.407	0.418	0.024	0.447	0.024	0.459	0.025	0.480	0.027
β_C	0.3	0.302	0.353	0.026	0.323	0.030	0.342	0.030	0.312	0.036
β_L	1	0.979	1.000	0.025	0.980	0.024	0.953	0.027	0.969	0.035
β_O	0.5	0.509	0.505	0.017	0.508	0.019	0.508	0.025	0.507	0.030

Table D.9: Covariances for inter-consumer heterogeneity - data with inter- and different levels of intra-consumer heterogeneity (4000 individuals), model with only inter-consumer heterogeneity

Parameter	True Values		No Heterog.		Low Heterog.		Med. Heterog.		High Heterog.	
	Theo.	Sample	Est.	SE	Est.	SE	Est.	SE	Est.	SE
β_S, β_C	0.072	0.075	0.064	0.013	0.061	0.014	0.052	0.015	0.038	0.018

Table D.10: Standard deviations for intra-consumer heterogeneity - data with inter and different levels of intra-consumer heterogeneity (4000 individuals), model with inter- and intra-consumer heterogeneity

Parameter	No Heterog.		Low Heterog.		Med. Heterog.		High Heterog.	
	Est.	SE	Est.	SE	Est.	SE	Est.	SE
β_L	0.264	0.056	0.555	0.051	1.012	0.040	1.921	0.047
β_O	0.101	0.040	0.181	0.088	0.382	0.049	0.967	0.041

Table D.11: Covariances for intra-consumer heterogeneity - data with inter- and different levels of intra-consumer heterogeneity (4000 individuals), model with inter- and intra-consumer heterogeneity

Parameter	No Heterog.		Low Heterog.		Med. Heterog.		High Heterog.	
	Est.	SE	Est.	SE	Est.	SE	Est.	SE
β_L, β_O	-0.009	0.013	-0.026	0.026	-0.167	0.036	-0.555	0.069