Robust discrete choice models with t-distributed kernel errors

Rico Krueger [*] Prateek Bansal [†] Michel Bierlaire [*]
Thomas Gasos [*]

September 14, 2020

Report TRANSP-OR 200914
Transport and Mobility Laboratory
School of Architecture, Civil and Environmental Engineering
Ecole Polytechnique Fédérale de Lausanne
transp-or.epfl.ch

[*]École Polytechnique Fédérale de Lausanne (EPFL), School of Architecture, Civil and Environmental Engineering (ENAC), Transport and Mobility Laboratory, Switzerland, {rico.krueger,michel.bierlaire}@epfl.ch

[†]Transport Strategy Centre, Department of Civil and Environmental Engineering, Imperial College London, UK, {prateek.bansal@imperial.ac.uk}

Abstract

Models that are robust to aberrant choice behaviour have received limited attention in discrete choice analysis. In this paper, we analyse two robust alternatives to the multinomial probit (MNP) model. Both alternative models belong to the family of robit models, whose kernel error distributions are heavy-tailed t-distributions. The first model is the multinomial robit (MNR) model in which a generic degrees of freedom parameter controls the heavy-tailedness of the kernel error distribution. The second alternative, the generalised multinomial robit (Gen-MNR) model, has not been studied in the literature before and is more flexible than MNR, as it allows for alternative-specific marginal heavy-tailedness of the kernel error distribution. For both models, we devise scalable and gradient-free Bayes estimators. We compare MNP, MNR and Gen-MNR in a simulation study and a case study on transport mode choice behaviour. We find that both MNR and Gen-MNR deliver significantly better in-sample fit and out-of-sample predictive accuracy than MNP. Gen-MNR outperforms MNR due to its more flexible kernel error distribution. Also, Gen-MNR gives more reasonable elasticity estimates than MNP and MNR, in particular regarding the demand for under-represented alternatives in a class-imbalanced dataset.

Keywords: robustness, probit, robit, Bayesian estimation, discrete choice, transport mode choice

1 Introduction

Random utility maximisation is by far the most widely adopted decision paradigm in the formulation of discrete choice models. Random utility theory (McFadden, 1981) posits that a rational decision-maker chooses the option with the highest utility from a finite set of mutually-exclusive alternatives. Since the utility of an alternative depends both on observed factors as well as on factors that the analyst does not or cannot observe, the conditional indirect utility of an alternative contains a random error term. Typically, the random error terms of alternatives in a choice set are assumed to be either independent and identically Gumbel distributed (logit kernel) or jointly Gaussian distributed (probit kernel).

The Gumbel and the Gaussian distributions have restrictive shapes, which limit the explanatory and predictive powers of the resulting logit and probit choice models. Whereas the Gaussian distribution has a symmetric bell shape with light tails, the Gumbel distribution is right-skewed with a right tail that is slightly heavier than that of the Gaussian distribution. In recent years, researchers have explored various departures from standard kernel error distributions (see Paleti, 2019, for a review). These advancements include negative exponential (Alptekinoğlu and Semple, 2016), negative Weibull (Castillo et al., 2008), generalised exponential (Fosgerau and Bierlaire, 2009) and q-generalised reverse Gumbel (Chikaraishi and Nakayama, 2016) kernel error distributions, additive combinations of Gumbel and exponential error terms (Del Castillo, 2016), a class of asymmetric distributions (Brathwaite and Walker, 2018), copulas with Gumbel marginals (Del Castillo, 2020). However, these extensions do not aim at enhancing the robustness of choice models.

The concept of robustness is well established in statistics, with the notion that a robust model safeguards inferences against the influence of outliers and violations of modelling assumptions (e.g. Gelman et al., 2013). In discrete choice analysis, the need for robust models arises in various situations to address aberrant utility differences. For example, utility differences can be aberrant from the analyst's point-of-view, if the analyst possesses little information about the factors influencing choices. In this scenario, the contribution of the random disturbance to the conditional indirect utility can be relatively large for some observations. Utility differences also contain outliers if the postulated decision paradigm (such as random utility maximisation) does not accurately represent the decision protocols governing some of the observed choices. Furthermore, aberrant utility differences are a concern in class-imbalanced datasets, which are frequently encountered in non-experimental settings. This is because in class-imbalanced data, the utility differences involving under-represented alternatives are outliers relative to utility differences involving well-represented options.

Lange et al. (1989) advocate the use of the heavy-tailed t-distribution as a means to increase robustness in regression models. Compared to the Gaussian distribution, the t-distribution has one more parameter which controls the heavy-tailedness of the distribution to moderate outlying data points. In the context of generalised linear models, Liu (2004) proposes the binary robit model, which is built on a t-distribution with unknown degrees of freedom (DOF), as a robust alternative to logistic and probit regression models. Furthermore, Ding (2014) constructs a robust Heckman selection model using the t-distribution as kernel error distribution. Jiang and Ding (2016) formulate Heckman

selection and multivariate robit models based on t-distributions with different marginal DOF.

Robustness has received limited attention in multinomial choice analysis. Dubey et al. (2020) present the first multinomial robit (MNR) model, i.e. a multinomial choice model defined through a t-distributed kernel error with an estimable DOF. Dubey et al. (2020) make a strong empirical case to adopt the MNR model over the multinomial probit (MNP) model. First, the estimates of the MNP model are inconsistent, if the kernel errors in the data generating process are heavy-tailed. Second, the robustness of MNR results in superior in-sample fit and out-of-sample predictive ability for class-imbalanced data sets. In another study, Peyhardi (2020) formulates an MNR model in the context of generalised linear models and shows that MNR can help in identifying artificial aspects of the design of stated preference experiments.

We identify two research gaps in the formulation and estimation of MNR models. First, Dubey et al. (2020) and Peyhardi (2020) constrain the flexibility of the kernel error distribution by assuming that a single, generic DOF parameter controls the heavy-tailedness of the kernel error distribution. This modelling assumption implies that the same level of utility aberrance applies to all alternatives. Second, the estimation approaches employed in both studies are not scalable. Dubey et al. (2020) are unable to derive analytical gradients of the MNR model and thus rely on computationally-expensive numerical gradient approximations during the maximisation of the simulated log-likelihood of the model. Peyhardi (2020) estimates the DOF parameter by performing a grid search, which requires the model to be estimated at multiple values of the DOF parameter and suffers from the curse of dimensionality if the underlying kernel distribution had multiple DOF parameters. Besides, incorporating representations of unobserved heterogeneity is computationally expensive in both studies, as it necessitates an additional layer of simulation in the computation of the log-likelihood.

In this paper, we address the first limitation of existing MNR models (i.e. generic heavy-tailedness) by formulating a generalised MNR (Gen-MNR) model with alternative-specific DOF parameters. To that end, we adopt the non-elliptical contoured t-distribution (Jiang and Ding, 2016) as kernel error distribution. To tackle the second limitation (i.e. computationally-expensive estimation), we devise gradient-free Bayesian estimation approaches for both the MNR and the Gen-MNR models. In the construction of the Bayesian estimation approaches, we exploit the hierarchical normal mixture representation of the t-distribution. To bypass complex likelihood computations in the estimation of the MNR and the Gen-MNR models, we employ a combination of Bayesian data augmentation techniques used in the estimation of MNP models (Albert and Chib, 1993, McCulloch and Rossi, 1994) as well as of non-multinomial robit models (Ding, 2014, Jiang and Ding, 2016). Bayesian estimation also facilitates accommodating flexible semi-parametric representations of unobserved preference heterogeneity (Krueger et al., 2020).

We first use simulated data to investigate the properties of the proposed models and their estimation methods in terms of parameter recovery and elasticity estimates. Subsequently, we compare MNP, MNR and Gen-MNR in a case study on transport mode choice behaviour using revealed preference data from London, UK. In the real data application, we contrast willingness to pay and elasticity estimates as well as in-sample fit and out-of-sample predictive accuracy of the three models.

The remainder of the paper is organised, as follows: First, we present the mathematical

formulations of the MNP, MNR and Gen-MNR models (Section 2). Then, we outline the estimation approaches and succinctly discuss the adopted data augmentation techniques (Section 3). In Sections 4 and 5, we present the simulation and case studies, respectively. Finally, we conclude and identify avenues for future research (Section 6).

2 Model formulations

In this section, we present the formulations of the MNP, MNR and Gen-MNR models.

2.1 Multinomial probit (MNP)

We consider a standard random utility model in which an agent $i = 1, \ldots, N$ chooses from a set of J mutually exclusive alternatives. In principle, utility is not identified at an absolute level. Therefore, the MNP model is defined through a $J-1$-dimensional Gaussian latent variable vector $\boldsymbol{w}_i = \{w_{ij}, \ldots, w_{i,J-1}\}$ (McCulloch and Rossi, 1994). The elements of $\boldsymbol{w}_i$ correspond to the utility differences with respect to the base alternative J. The observed choice $y_i \in \{1, \ldots, J\}$ is assumed to arise from

$$y_i(\boldsymbol{w}_i) = \begin{cases} j & \text{if } \max(\boldsymbol{w}_i) = w_{ij} > 0 \\ J & \text{if } \max(\boldsymbol{w}_i) < 0, \end{cases} \quad \text{for } i = 1, \ldots, N. \tag{1}$$

The latent variable $\boldsymbol{w}_i$ is represented as

$$\boldsymbol{w}_i = \mathbf{X}_i\boldsymbol{\beta} + \boldsymbol{\varepsilon}_i \quad \text{with } \boldsymbol{\varepsilon}_i \sim N(\mathbf{0}, \boldsymbol{\Sigma}), \quad \text{for } i = 1, \ldots, N. \tag{2}$$

Here, $\mathbf{X}_i$ is a $(J-1) \times K$ matrix of differenced predictors, i.e. $\mathbf{X}_i = \begin{bmatrix} \mathbf{X}_{i1} \\ \vdots \\ \mathbf{X}_{i,J-1} \end{bmatrix} = \begin{bmatrix} \mathbf{X}_{i1}^{obs} - \mathbf{X}_{iJ}^{obs} \\ \vdots \\ \mathbf{X}_{i,J-1}^{obs} - \mathbf{X}_{iJ}^{obs} \end{bmatrix}$, where $\mathbf{X}_{ij}^{obs}$ is the observed attribute vector of alternative j for agent i. $\boldsymbol{\beta}$ is a K vector of taste parameters. $\boldsymbol{\Sigma}$ is a $(J-1) \times (J-1)$ covariance matrix. The latent variable representation (2) is not identified, because $\boldsymbol{w}_i$ can be multiplied by any positive scalar c without changing the likelihood (1), i.e. $y_i(\boldsymbol{w}_i) = y_i(c\boldsymbol{w}_i)$. Therefore, we must set the scale. We follow Burgette and Nordheim (2012) and impose a trace restriction on $\boldsymbol{\Sigma}$, i.e. $\text{tr}(\boldsymbol{\Sigma}) = J - 1$. To complete the specification of the MNP model, we place a normal prior on $\boldsymbol{\beta}$, i.e. $\boldsymbol{\beta} \sim N(\zeta_0, \mathbf{B}_0)$, and an Inverse-Wishart prior on $\boldsymbol{\Sigma}$, i.e. $\boldsymbol{\Sigma} \sim IW(\rho, \mathbf{S})$. Predictions under the Bayesian formulation of the MNP model can be sensitive to the selection of the base alternative J (Burgette and Nordheim, 2012).

2.2 Multinomial robit (MNR)

The MNR model assumes a t-distributed kernel error for the latent variable $\boldsymbol{w}_i$, i.e.

$$\boldsymbol{w}_i = \mathbf{X}_i\boldsymbol{\beta} + \boldsymbol{\varepsilon}_i \quad \text{with } \boldsymbol{\varepsilon}_i \sim t(\mathbf{0}, \boldsymbol{\Sigma}, \nu), \quad \text{for } i = 1, \ldots, N, \tag{3}$$

where $\boldsymbol{\Sigma}$ is a $(J-1)\times(J-1)$ covariance matrix and ν is scalar degree of freedom (DOF). The t-distribution also has the following normal mixture representation (Ding, 2014):

$$\varepsilon_i \sim N(0, \boldsymbol{\Sigma}/q_i) \quad \text{with } q_i \sim \chi^2_\nu/\nu, \quad \text{for } i = 1, \ldots, N. \tag{4}$$

The latent variables $\mathbf{q} = \{q_1, \ldots, q_N\}$ allow for heavy-tailedness in the distribution of the kernel error by increasing the variability of ε_i across different i. Figure 1 illustrates the relationship between the χ^2-distribution (which controls the distribution of $\mathbf{q}$) and a t-distribution (which controls the distribution of ε) with unit variance for different DOF ν. For small $\nu < 30$, the t-distribution exhibits heavy tails. As ν approaches ∞, the t-distribution converges to the normal distribution. We use the same priors for $\boldsymbol{\beta}$ and $\boldsymbol{\Sigma}$ as in MNP. For identification, we also maintain the trace restriction on $\boldsymbol{\Sigma}$. We place a Gamma prior on ν with $\nu \sim \text{Gamma}(\alpha_0, \beta_0)$. Predictions under the Bayesian formulation of the MNR model can be sensitive to the selection of the base alternative in the same way as predictions under the Bayesian formulation of the MNP model.

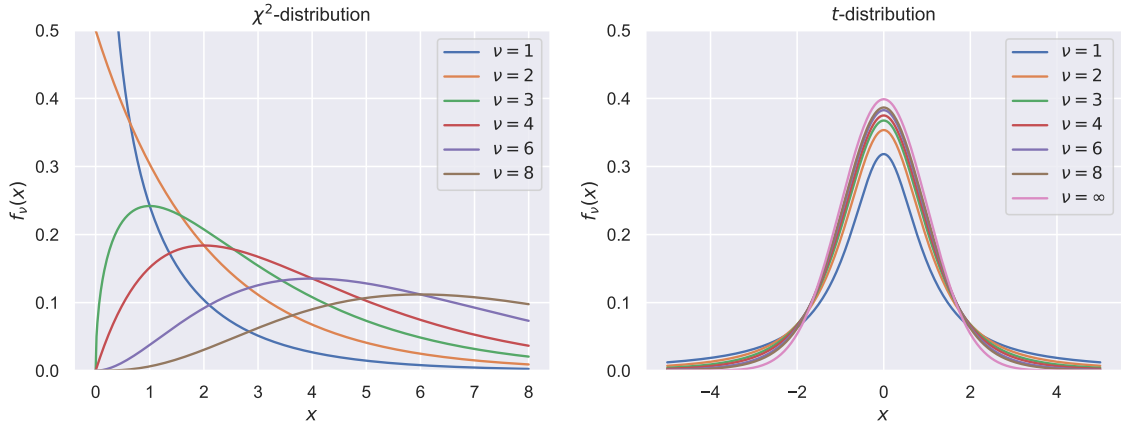

Figure 1: Relationship between χ^2- and t-distributions for different degrees of freedom ν

2.3 Generalised multinomial robit (Gen-MNR)

We generalise MNR by allowing for different marginal heavy-tailedness in the distribution of the latent variable $\boldsymbol{w}_i$. The Gen-MNR model assumes that the kernel error of $\boldsymbol{w}_i$ is drawn from a non-elliptical contoured t-distribution (NECT; Jiang and Ding, 2016). We have

$$\boldsymbol{w}_i = \mathbf{X}_i\boldsymbol{\beta} + \varepsilon_i \quad \text{with } \varepsilon_i \sim \text{NECT}_{\mathbf{p}}(0, \boldsymbol{\Sigma}, \boldsymbol{\nu}), \quad \text{for } i = 1, \ldots, N, \tag{5}$$

where $\boldsymbol{\Sigma}$ is a $(J-1)\times(J-1)$ covariance matrix and $\boldsymbol{\nu} = \{\nu_1, \ldots, \nu_S\}$ is S vector of DOF with $1 < S \leq J-1$. $\mathbf{p} = \{p_1, \ldots, p_S\}$ is a S vector giving the number of dimensions that are associated with each DOF ν_s. We have $p_s \in \mathbb{N}\setminus\{0\}$ and $\sum_{s=1}^{S} p_s = J-1$. The NECT distribution has the following normal mixture representation (Jiang and Ding, 2016):

$$\varepsilon_i = \mathbf{Q}_i^{-1/2}\boldsymbol{\Sigma}^{1/2}\mathbf{Z}_i, \quad \mathbf{Z}_i \sim N(0, \mathbf{I}_{J-1}), \quad \text{for } i = 1, \ldots, N, \tag{6}$$

where $\mathbf{Q}_i = \text{diag}(q_{i1}\mathbf{I}_{p_1}, \ldots, q_{iS}\mathbf{I}_{p_S})$ is a $(J-1)\times(J-1)$ block-diagonal matrix with $q_{is} \sim \chi^2_{\nu_s}/\nu_s$ for $s = 1, \ldots, S$. $\mathbf{I}_l$ is a $l \times l$ identity matrix. Each marginal component of a NECT-distributed random variable follows a univariate t-distribution with the respective

DOF, i.e. if $\varepsilon \sim \text{NECT}_p(0, \Sigma, \nu)$, then $\varepsilon_j \sim t(0, \Sigma_{jj}, \nu_{s(j)})$, where $s(j)$ maps dimension j onto its associated DOF. In the rest of this work, we assume that $S = J - 1$ without loss of generality. The Gen-MNR model uses the same prior distributions as the MNR model. We let $\nu_j \sim \text{Gamma}(\alpha_0, \beta_0)$ for $j = 1, \ldots, J-1$. We also maintain the trace restriction on Σ for identification. Predictions of the Gen-MNR model can be sensitive to the selection of the base alternative in the same way as predictions of the MNP and MNR models.

3 Inference

For the estimation of the MNP, MNR and Gen-MNR models, we employ Markov chain Monte Carlo methods in the form of Gibbs sampling (Robert and Casella, 2013). The sampling schemes for the three models are presented in Algorithms 1, 2 and 3.
Algorithm 1 corresponds to the Gibbs sampler proposed by Burgette and Nordheim (2012) without the marginal data augmentation scheme devised by Imai and Van Dyk (2005). Algorithm 2 is based on the Gibbs sampler proposed by Ding (2014) for the robust Heckman selection model, and Algorithm 3 is most closely related to the Gibbs sampler proposed by Jiang and Ding (2016) for the multivariate robit model.
All samplers involve data augmentation (Tanner and Wong, 1987) to facilitate their construction. The central idea of Bayesian data augmentation is to treat latent variables as unknown model parameters, which are imputed in additional sampling steps. Each of the samplers uses the data augmentation scheme developed by Albert and Chib (1993) and McCulloch and Rossi (1994) to impute the latent variable w (see Appendix A.1 for details). The samplers for the MNR and Gen-MNR models additionally incorporate the data augmentation schemes devised by Ding (2014) and Jiang and Ding (2016), respectively, to impute the latent variable q. Data augmentation circumvents complex likelihood calculations in the estimation of the MNP, MNR and Gen-MNR models. This is because conditional on w and q (if applicable), the models reduce to standard Bayesian linear models.
The full conditional distribution of ν in the MNR model as well as the full conditional distributions of ν_j and q_{ij} in the Gen-MNR model are nonstandard. To draw from these intricate distributions, we use Metropolised Independence samplers (Liu, 2008) with approximate Gamma proposals, as devised by Ding (2014) and Jiang and Ding (2016) (see Appendices A.2 and A.3 for details).
We implement Algorithms 1, 2 and 3 in Julia (Bezanson et al., 2017). In the subsequent applications, the Gibbs samplers are executed with a single chain consisting of 300,000 draws including a warm-up period of 200,000 draws. A thinning factor of 10 is applied to the post warm-up draws. Convergence is assessed with the help of the potential scale reduction factor (Gelman et al., 1992).

4 Simulation study

We conduct a simulation study consisting of two examples to investigate the properties of the proposed models and their estimation methods. The simulation study has two specific objectives. First, we aim to assess the ability of the proposed Gibbs samplers to recover model parameters in finite samples. Second, we aim to quantify the effects of ignoring

Algorithm 1 Gibbs sampler for the MNP model

Step 0: Initialise parameters β, Σ, w.
for $t = 1, \ldots, T$ **do**
  Step 1: Update w_{ij} given β, Σ, ν, $w_{i,-j}$.
  for $i = 1, \ldots, N$ **do**
    for $j = 1, \ldots, J-1$ **do**
      Draw $w_{ij}|\beta, \Sigma, w_{i,-j} \sim TN(\mu_{ij}, \tau_{ij}^2)$ as explained in Appendix A.1.
    end for
  end for
  Step 2: Update β given Σ, w.
  Set $\hat{B} = \left(\sum_{i=1}^{N} X_i^\top \Sigma^{-1} X_i + B_0\right)^{-1}$.
  Set $\hat{\beta} = \hat{B}\left(\sum_{i=1}^{N} X_i^\top \Sigma^{-1} w_i\right)$.
  Draw $\beta|\Sigma, w \sim N(\hat{\beta}, \hat{B})$.
  Step 3: Update Σ given β, w.
  Set $z_i = w_i - X_i\beta$, for $i = 1, \ldots, N$.
  Draw $\tilde{\Sigma}|\beta, w \sim IW\left(N + \rho, S + \sum_{i=1}^{N} z_i z_i^\top\right)$.
  Set $\alpha^2 = \mathrm{tr}(\tilde{\Sigma})/(J-1)$.
  Set $\Sigma = \tilde{\Sigma}/\alpha^2$.
  Set $w_i = (z_i + \alpha X_i\beta)/\alpha$.
end for
return β, Σ

non-normality and different marginal heavy-tailedness of the kernel error distribution on fit and elasticity estimates.

4.1 Example I: Data generated according to MNR model

In the first example, data are generated according to the MNR model. We let $N = 40,000$ and $J = 4$. We set $\beta = (1, -2, 1, 1, -1, 1, -1)^\top$ and $\Sigma = D\Omega D$, where $\Omega = \begin{bmatrix} 1.0 & 0.3 & 0.0 \\ 0.3 & 1.0 & 0.3 \\ 0.0 & 0.3 & 1.0 \end{bmatrix}$ and $D = \mathrm{diag}\left(\sqrt{\sigma^2}\right)$ with $\sigma^2 = (1.4, 0.8, 1.2)^\top$. Furthermore, we let $\nu = 2$. Here, the first three predictors are alternative-specific constants. The remaining predictors are alternative-specific attributes. We draw $X_{ijk}^{obs} \sim U(0, 2)$ for $i = 1, \ldots, N$, $j = 1, \ldots, J$ and $k = 4, \ldots, 7$. The fourth alternative is set as reference alternative in data generation and model estimation. For the sake of simplicity, we do not perform a search on the specification of the reference alternative. The simulated data intentionally exhibit significant class imbalance. The realised market shares of the choice alternatives are 42.2%, 3.3%, 40.4% and 14.1%.

Figure 2 shows the posterior distribution of the DOF parameter ν of the MNR model along with the corresponding true parameter value used in the generation of the data. It can be seen that Algorithm 2 performs well at recovering the DOF parameter of the MNR model. From Figures 4 and 5 in Appendix B.1, we can conclude that Algorithm 2 also does an excellent job at recovering the remaining parameters β and Σ.

Table 1 compares the in-sample fit of the MNP, MNR and Gen-MNR models in terms of the quadratic loss (QL), which is defined as $\text{QL} = \sum_{i=1}^{N}\sum_{j=1}^{J}(p_{nj} - \hat{p}_{nj})^2$, where p_{nj} is the true choice probability simulated at the true parameter values that $y_n = j$ is realised, and where $\hat{p}_{nj}$ is the corresponding fitted choice probability. The MNR model offers the best fit to the data. Interestingly, the Gen-MNR model with multiple DOF parameters performs slightly worse than the simpler MNR model. A possible explanation for the inability of the Gen-MNR model to perform as well as the MNR model is that the estimation of multiple DOF parameters incurs a greater simulation error. Nonetheless, both the MNR and Gen-MNR models outperform the MNP model by a substantial margin. Finally, we contrast the elasticity estimates of the three models by considering two scenarios. Table 2 enumerates the aggregate arc elasticities computed for each of the three models along with the corresponding true aggregate arc elasticities. In the first scenario, we manipulate the first alternative-specific attribute of the under-represented second alternative. We observe that the MNR and Gen-MNR models produce direct aggregate arc elasticities, which are much closer to the truth than the direct aggregate arc elasticities of the MNP model. For example, for a 10% increase in the considered attribute, the true direct aggregate arc elasticity is 1.08. Whilst the MNR and Gen-MNR models give direct elasticities of 1.04 and 1.02, respectively, the MNP model produces a substantially lower direct elasticity of 0.85. In the second scenario, we manipulate the first alternative-specific attributes of the well-represented first alternative. In this scenario, the models perform equally well at recovering the true aggregate arc elasticities.

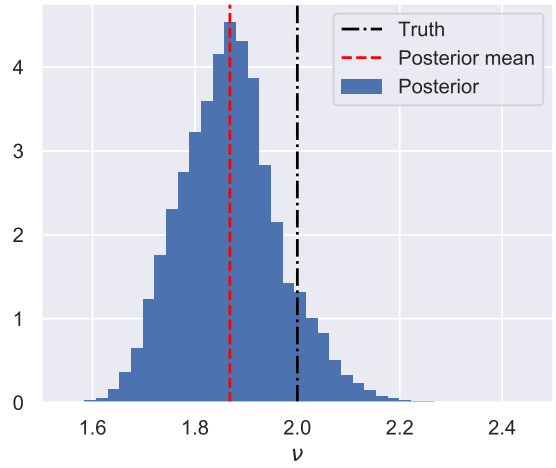

Figure 2: Estimated posterior distribution and true value of the degree of freedom parameter ν for the MNR model in simulation example I

Model	**Loss**
MNP	104.4
MNR	8.9
Gen-MNR	16.9

Table 1: Quadratic loss in simulation example I

Algorithm 2 Gibbs sampler for the MNR model

Step 0: Initialise parameters β, Σ, ν, w, q.
for $t = 1, \ldots, T$ **do**
  Step 1: Update w_{ij} given β, Σ, $w_{i,-j}$, q_i.
  for $i = 1, \ldots, N$ **do**
    for $j = 1, \ldots, J-1$ **do**
      Draw $w_{ij}|\beta, \Sigma, w_{i,-j}, q_i \sim TN(\mu_{ij}, \tau_{ij}^2)$ as explained in Appendix A.1.
    end for
  end for
  Step 2: Update q_{ij} given β, Σ, ν_j w_i.
  for $i = 1, \ldots, N$ **do**
    Set $z_i = w_i - X_i\beta$.
    Draw $q_i|\beta, \Sigma, w_i \sim \chi^2_{\nu+J-1} / \left(z_i^\top \Sigma^{-1} z_i\right)$.
  end for
  Step 3: Update β given Σ, w, q.
  Set $\hat{B} = \left(\sum_{i=1}^N q_i X_i^\top \Sigma^{-1} X_i + B_0\right)^{-1}$.
  Set $\hat{\beta} = \hat{B}\left(\sum_{i=1}^N q_i X_i^\top \Sigma^{-1} w_i\right)$.
  Draw $\beta|\Sigma, w, q \sim N(\hat{\beta}, \hat{B})$.
  Step 4: Update Σ given β, w, q.
  Set $z_i = w_i - X_i\beta$, for $i = 1, \ldots, N$.
  Draw $\tilde{\Sigma}|\beta, w, q \sim IW\left(N + \rho, S + \sum_{i=1}^N q_i z_i z_i^\top\right)$.
  Set $\alpha^2 = \mathrm{tr}(\tilde{\Sigma})/(J-1)$.
  Set $\Sigma = \tilde{\Sigma}/\alpha^2$.
  Set $w_i = (z_i + \alpha X_i\beta)/\alpha$.
  Step 5: Update ν given q.
  Calculate α^*, β^* as explained in Appendix A.2.
  Draw proposal $\nu' \sim \mathrm{Gamma}(\alpha^*, \beta^*)$.
  Accept the proposal with probability $\min\{1, \exp\left(l(\nu') - h(\nu') - l(\nu) + h(\nu)\right)\}$, where $l(\nu)$ and $h(\nu)$ are defined in (9) and (10), respectively.
end for
return β, Σ, ν

Algorithm 3 Gibbs sampler for the Gen-MNR model

Step 0: Initialise parameters $\boldsymbol{\beta}$, $\boldsymbol{\Sigma}$, $\boldsymbol{\nu}$, $\boldsymbol{w}$, $\mathbf{q}$.
for $t = 1, \ldots, T$ **do**
  Step 1: Update w_{ij} given $\boldsymbol{\beta}$, $\boldsymbol{\Sigma}$, $w_{i,-j}$, q_i.
  for $i = 1, \ldots, N$ **do**
    for $j = 1, \ldots, J-1$ **do**
      Draw $w_{ij}|\boldsymbol{\beta}, \boldsymbol{\Sigma}, w_{i,-j}, q_i \sim \text{TN}(\mu_{ij}, \tau_{ij}^2)$ as explained in Appendix A.1.
    end for
  end for
  Step 2: Update q_{ij} given $\boldsymbol{\beta}$, $\boldsymbol{\Sigma}$, $\boldsymbol{w}_i$, $\mathbf{q}_{i,-j}$.
  for $i = 1, \ldots, N$ **do**
    for $j = 1, \ldots, J-1$ **do**
      Calculate α^*, β^* as explained in Appendix A.3.
      Draw proposal $q'_{ij} \sim \text{Gamma}(\alpha^*, \beta^*)$.
      Accept the proposal with probability $\min\left\{1, \exp\left(f(q'_{ij}) - g(q'_{ij}) - f(q_{ij}) + g(q_{ij})\right)\right\}$, where $f(q_{ij})$ and $g(q_{ij})$ are defined in (15) and (16), respectively.
    end for
  end for
  Step 3: Update $\boldsymbol{\beta}$ given $\boldsymbol{\Sigma}$, $\boldsymbol{w}$, $\mathbf{q}$.
  Set $\hat{\mathbf{B}} = \left(\sum_{i=1}^N \mathbf{X}_i^\top \mathbf{Q}_i^{1/2} \boldsymbol{\Sigma}^{-1} \mathbf{Q}_i^{1/2} \mathbf{X}_i + \mathbf{B}_0\right)^{-1}$.
  Set $\hat{\boldsymbol{\beta}} = \hat{\mathbf{B}} \left(\sum_{i=1}^N \mathbf{X}_i^\top \mathbf{Q}_i^{1/2} \boldsymbol{\Sigma}^{-1} \mathbf{Q}_i^{1/2} \boldsymbol{w}_i\right)$.
  Draw $\boldsymbol{\beta}|\boldsymbol{\Sigma}, \boldsymbol{w}, \mathbf{q} \sim N(\hat{\boldsymbol{\beta}}, \hat{\mathbf{B}})$.
  Step 4: Update $\boldsymbol{\Sigma}$ given $\boldsymbol{\beta}$, $\boldsymbol{w}$, $\mathbf{q}$.
  Set $\boldsymbol{z}_i = \boldsymbol{w}_i - \mathbf{X}_i \boldsymbol{\beta}$, for $i = 1, \ldots, N$.
  Draw $\tilde{\boldsymbol{\Sigma}}|\boldsymbol{\beta}, \boldsymbol{w}, \mathbf{q} \sim IW\left(N + \rho, \mathbf{S} + \sum_{i=1}^N \mathbf{Q}_i^{1/2} \boldsymbol{z}_i \boldsymbol{z}_i^\top \mathbf{Q}_i^{1/2}\right)$.
  Set $\alpha^2 = \text{tr}(\tilde{\boldsymbol{\Sigma}})/(J-1)$.
  Set $\boldsymbol{\Sigma} = \tilde{\boldsymbol{\Sigma}}/\alpha^2$.
  Set $\boldsymbol{w}_i = (\boldsymbol{z}_i + \alpha \mathbf{X}_i \boldsymbol{\beta})/\alpha$.
  Step 5: Update ν_j given $\mathbf{q}_j$.
  for $j = 1, \ldots, J-1$ **do**
    Calculate α^*, β^* as explained in Appendix A.2.
    Draw proposal $\nu'_j \sim \text{Gamma}(\alpha^*, \beta^*)$.
    Accept the proposal with probability $\min\left\{1, \exp\left(l(\nu'_j) - h(\nu'_j) - l(\nu_j) + h(\nu_j)\right)\right\}$, where $l(\nu)$ and $h(\nu)$ are defined in (9) and (10), respectively.
  end for
end for
return $\boldsymbol{\beta}$, $\boldsymbol{\Sigma}$, $\boldsymbol{\nu}$

Scenario		Truth				MNP				MNR				Gen-MNR			
		Alt. 1	**Alt. 2**	**Alt. 3**	**Alt. 4**	**Alt. 1**	**Alt. 2**	**Alt. 3**	**Alt. 4**	**Alt. 1**	**Alt. 2**	**Alt. 3**	**Alt. 4**	**Alt. 1**	**Alt. 2**	**Alt. 3**	**Alt. 4**
$x^{obs}_{n,2,4} \; \forall \, n = 1, \ldots, N$	increased by 5%	-0.03	1.05	-0.04	-0.05	-0.03	0.84	-0.03	-0.04	-0.03	1.01	-0.04	-0.04	-0.03	0.97	-0.04	-0.03
	increased by 10%	-0.03	1.08	-0.04	-0.05	-0.03	0.85	-0.03	-0.04	-0.03	1.04	-0.04	-0.05	-0.03	1.02	-0.04	-0.04
	increased by 25%	-0.04	1.16	-0.04	-0.06	-0.03	0.90	-0.03	-0.04	-0.04	1.12	-0.04	-0.05	-0.04	1.12	-0.05	-0.05
$x^{obs}_{n,1,4} \; \forall \, n = 1, \ldots, N$	increased by 5%	0.41	-0.27	-0.28	-0.38	0.41	-0.26	-0.28	-0.37	0.41	-0.27	-0.28	-0.38	0.41	-0.31	-0.28	-0.38
	increased by 10%	0.42	-0.29	-0.29	-0.39	0.41	-0.28	-0.29	-0.38	0.41	-0.30	-0.29	-0.39	0.41	-0.30	-0.29	-0.39
	increased by 25%	0.43	-0.31	-0.32	-0.42	0.43	-0.31	-0.31	-0.42	0.43	-0.31	-0.31	-0.42	0.43	-0.32	-0.31	-0.42

Table 2: Aggregate arc elasticities in simulation example I

Scenario		Truth				MNP				MNR				Gen-MNR			
		Alt. 1	**Alt. 2**	**Alt. 3**	**Alt. 4**	**Alt. 1**	**Alt. 2**	**Alt. 3**	**Alt. 4**	**Alt. 1**	**Alt. 2**	**Alt. 3**	**Alt. 4**	**Alt. 1**	**Alt. 2**	**Alt. 3**	**Alt. 4**
$x^{obs}_{n,2,4} \; \forall \, n = 1, \ldots, N$	increased by 5%	-0.04	1.14	-0.03	-0.06	-0.04	0.98	-0.03	-0.05	-0.04	1.07	-0.03	-0.06	-0.04	1.11	-0.03	-0.06
	increased by 10%	-0.04	1.16	-0.04	-0.07	-0.04	1.00	-0.03	-0.05	-0.04	1.11	-0.03	-0.07	-0.04	1.13	-0.03	-0.07
	increased by 25%	-0.05	1.25	-0.04	-0.08	-0.04	1.06	-0.03	-0.06	-0.05	1.19	-0.04	-0.08	-0.05	1.20	-0.04	-0.08
$x^{obs}_{n,1,4} \; \forall \, n = 1, \ldots, N$	increased by 5%	0.42	-0.37	-0.24	-0.43	0.42	-0.36	-0.24	-0.43	0.42	-0.36	-0.24	-0.43	0.42	-0.34	-0.24	-0.43
	increased by 10%	0.42	-0.37	-0.24	-0.44	0.42	-0.36	-0.24	-0.44	0.42	-0.36	-0.24	-0.44	0.42	-0.36	-0.24	-0.45
	increased by 25%	0.43	-0.39	-0.27	-0.48	0.44	-0.39	-0.27	-0.48	0.44	-0.39	-0.27	-0.47	0.44	-0.39	-0.26	-0.48

Table 3: Aggregate arc elasticities in simulation example II

4.2 Example II: Data generated according to Gen-MNR

In the second example, data are generated according to Gen-MNR. The data generating process is essentially same as before. The only difference is that we allow for different marginal heavy-tailedness by setting $\boldsymbol{\nu} = (5, 3, 1)^\top$. Furthermore, we let $\beta_2 = -1.8$ to induce a similar level of class imbalance as in the first example. The realised market shares in the simulated dataset are 42.5%, 40.3%, 3.3% and 13.4%. Again, alternative four is set as reference alternative in data generation and model estimation. For simplicity, no search over the specification of the reference alternative is performed.

Figure 3 shows the marginal posterior distributions of the DOF parameters ν_1, ν_2 and ν_3 along with their corresponding true parameter values. It can be seen that Algorithm 3 performs well at recovering the DOF parameters of Gen-MNR. From Figures 6 and 7 in Appendix B.2, we can conclude that Algorithm 3 also does an excellent job at recovering $\boldsymbol{\beta}$ and $\boldsymbol{\Sigma}$.

Table 4 compares the in-sample fit of MNP, MNR and Gen-MNR in terms of the quadratic loss. As expected, Gen-MNR provides the best fit to the data, followed by MNR. MNR and Gen-MNR outperform MNP by a significant margin.

Finally, Table 3 enumerates the aggregate arc elasticities of the three models along with their true counterparts for the exact same scenarios as in the first simulation example. In the first scenario, we manipulate the first alternative-specific attribute of the under-represented second alternative. We observe that MNR and Gen-MNR produce direct aggregate arc elasticities, which are closer to the truth than the corresponding estimates of MNP. For example, the true direct aggregate arc elasticity for a 10% increase in the considered attribute is 1.16. MNR and Gen-MNR produce direct aggregate arc elasticities of 1.10 and 1.13, respectively. By contrast, MNP returns a markedly lower direct aggregate arc elasticity of 1.00. Overall, the differences between MNR and Gen-MNR are minor. In the second scenario, we manipulate the first alternative-specific attribute of the well-represented first alternative. In this scenarios, all models perform equally well at recovering the true direct and indirect arc elasticities.

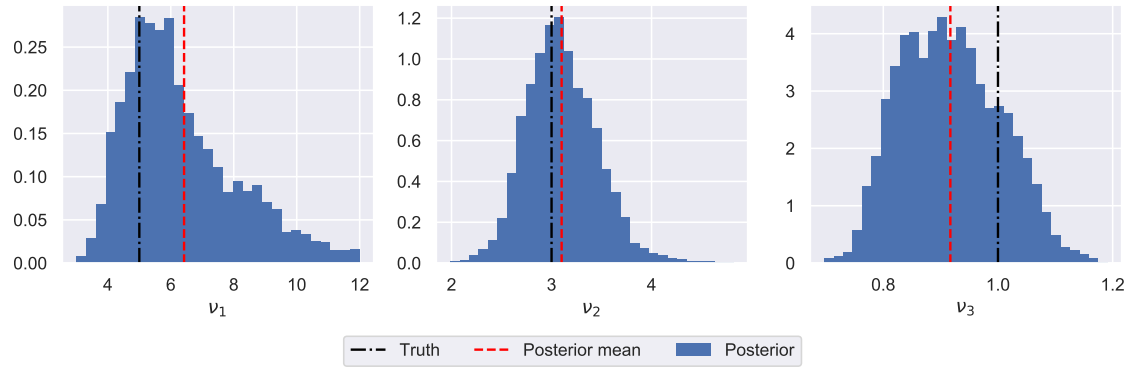

Figure 3: Estimated posterior distributions and true values of the degree of freedom parameters ν_1, ν_2 and ν_3 for Gen-MNR in simulation example II

Model	Loss
MNP	68.3
MNR	25.7
Gen-MNR	9.0

Table 4: Quadratic loss in simulation example II

5 Case study

We apply MNP, MNR and Gen-MNR in a case study on transport mode choice behaviour.

5.1 Data and utility specification

Revealed preference data for the case study are sourced from the London Passenger Mode Choice (LPMC) dataset, which was compiled and made publicly available by Hillel et al. (2018). The LPMC dataset consists of trip records from the London Travel Demand Survey, which was conducted from 2012 to 2015. For each trip record, Hillel et al. (2018) imputed tailored choice sets including the attributes of the chosen and the non-chosen alternatives using an online directions application programming interface. For more information about the LPMC dataset, the reader is directed to Hillel et al. (2018). In this case study, we restrict our analysis to home-based trips reported by individuals who are at least 12 years old. The resulting dataset comprises 58,584 observations. There are four mode choice alternatives, namely walking, cycling, transit and driving with observed market shares of 16.6%, 3.2%, 37.4% and 42.8%, respectively. We hold out 10% of the data corresponding to 5,858 observations for out-of-sample validation.
Each of the three models uses the same specification of the systematic utility, as shown in Table 5. The variable "traffic variability" is a measure of the driving travel time uncertainty for the given origin-destination pair. It is defined as the difference between the travel times in a pessimistic traffic scenario and in an optimistic one divided by the travel time in a typical, best-guess traffic scenario (see Hillel et al., 2018). The drive alternative is set as reference alternative in the estimation of all models. We performed a search over the specification of the reference alternative but found no substantive differences in parameter estimates, in-sample fit and out-of-sample predictive accuracy for different specifications of the reference alternative.

Variable	**Walk**	**Cycle**	**Transit**	**Drive**
Alternative-specific constants		$\beta_{\text{asc, cycle}}$	$\beta_{\text{asc, transit}}$	$\beta_{\text{asc, drive}}$
Alternative-specific attributes				
Cost [GBP]			β_{cost}	β_{cost}
Out-of-vehicle time (ovtt) [hours]	β_{ovtt}	β_{ovtt}	β_{ovtt}	
In-vehicle time (ivtt) [hours]			β_{ivtt}	β_{ivtt}
No. of transfers			$\beta_{\text{transfers}}$	
Traffic variability (tv)				β_{tv}
Individual- and context-specific attributes				
Female traveller		$\beta_{\text{female, cycle}}$	$\beta_{\text{female, transit}}$	$\beta_{\text{female, drive}}$
Traveller age $<$ 18 years		$\beta_{(\text{age}<18\text{ years})\vee(\text{age}\geq 65\text{ years}),\text{ cycle}}$	$\beta_{\text{age}<18\text{ years, transit}}$	$\beta_{\text{age}<18\text{ years, drive}}$
Traveller age $\geq$ 65 years		$\beta_{(\text{age}<18\text{ years})\vee(\text{age}\geq 65\text{ years}),\text{ cycle}}$	$\beta_{\text{age}\geq 65\text{ years, transit}}$	$\beta_{\text{age}\geq 65\text{ years, drive}}$
Travel during winter period (Nov–Mar)		$\beta_{\text{winter, cycle}}$		
No. of household cars				$\beta_{\text{cars, drive}}$

Table 5: Utility specification by alternative in MNP, MNR and Gen-MNR for the case study

5.2 Results

5.2.1 In-sample fit and out-of-sample predictive ability

Table 6 compares the in-sample fit and the out-of-sample predictive ability of MNP, MNR and Gen-MNR. We calculate Brier scores on the training and test data for each of the models. The Brier score (BS; Brier, 1950) is defined as $\text{BS} = \sum_{i=1}^{N} \sum_{j=1}^{J} (\mathbf{1}\{y_n = j\} - \hat{p}_{nj})^2$, where $\mathbf{1}\{y_n = j\}$ is an indicator, which equals one if the condition inside the braces is true and zero otherwise, and where $\hat{p}_{nj}$ is the predicted probability that $y_n = j$ is observed. The Brier score is a strictly proper scoring rule, because it is uniquely minimised by the true predictive choice probabilities (Gneiting and Raftery, 2007). Closely followed by MNR, Gen-MNR provides the best fit to the training data and exhibits superior out-of-sample predictive accuracy. Both MNR and Gen-MNR outperform MNP by a significant margin.

	Brier score	
Model	**Train**	**Test**
MNP	22002.7	2436.8
MNR	21696.1	2404.9
Gen-MNR	21681.8	2401.9

Table 6: In-sample fit and out-of-sample predictive ability of MNP, MNR and Gen-MNR for the case study

5.2.2 Parameters estimates

Table 7 presents the estimates of the parameters of the MNP, MNR and Gen-MNR models. For each parameter, we report the posterior mean, the posterior standard deviation and the bounds of the 95% credible interval.

First, we examine the estimates of the DOF parameters in MNR and Gen-MNR. For both models, we find evidence of sizeable heavy-tailedness. For instance, the posterior mean of the generic DOF parameter ν in MNR is 2.031, which is indicative of substantial heavy-tailedness. We detect pronounced and distinct marginal heavy-tailedness in Gen-MNR. As expected, heavy-tailedness is most substantial for the utility differences involving the under-represented walking and cycling alternatives. The posterior means of the associated DOF parameters equal 2.583 and 4.315, respectively. By contrast, tails are only moderately heavy for the utility difference between transit and the drive alternatives, since the posterior mean of the associated DOF parameter is 10.647. The credible intervals of the three DOF parameters of Gen-MNR also do not overlap, which indicates that heavy-tailedness in each dimension of the kernel error distribution is statistically different. In sum, the estimates of the DOF parameters suggest that utility aberrance characterises a non-negligible proportion of the observed transport mode choices. The Gen-MNR model also provides several interesting behavioural insights which the MNR model fails to reveal. As the DOF parameter of the utility difference between the passive and frequently-chosen transit and drive alternatives is the largest of all DOF parameters, the utilities of these alternatives are comparatively less aberrant than the utilities of the other modes.

However, aberrance is noticeably stronger for utility differences involving the active and less frequently-chosen walking and cycling alternatives.
Next, we compare the estimates of the taste parameters $\boldsymbol{\beta}$. Since the scale of $\boldsymbol{\beta}$ is not necessarily the same in each of the three models, we contrast the sensitivities to alternative-specific attributes in terms of their implied willingness to pay (WTP), which is given by the ratio of a non-price coefficient of interest and the price coefficient. WTP is scale-free and allows for a money-metric representation of sensitivities. WTP for reductions in out-of-vehicle and in-vehicle time is slightly larger in MNR and Gen-MNR than in MNP. To be precise, WTP for a reduction in out-of-vehicle time is 33.8 GBP/h, 35.1 GBP/h and 37.0 GBP/h, and WTP for a reduction in in-vehicle time is 18.4 GBP/h, 19.9 GBP/h, 20.6 GBP/h in MNP, MNR and Gen-MNR, respectively. Interestingly, MNR gives a lower WTP for a reduction in traffic variability than the two other models. Whereas the implied value of a 10% reduction in traffic variability is 2.0 GBP in the MNP model and 1.9 GBP in Gen-MNR, the implied value of a 10% reduction in traffic variability is only 1.7 GBP in MNR. MNR also implies a lower transfer penalty. The implied transfer penalties in MNP and Gen-MNR are 1.4 GBP and 1.5 GBP, respectively, per transfer. By contrast, the implied transfer penalty in MNR is 1.2 GBP per transfer.
On the whole, the WTP estimates show that the three models can produce different economic valuations of non-price attributes. Strikingly, the WTP estimates for reductions in traffic variability and transfers of MNP and Gen-MNR resemble each other closely and differ noticeably from the corresponding WTP estimates of MNR.
The models also provide insights into the influence of individual- and context-specific attributes on mode choice propensities. Interestingly, there are no substantive differences between the three models. For example, all models suggest that female travellers are relatively less likely to cycle and relatively more likely to use transit. No gender differences in the propensity to use the driving mode are detected. Old age reduces the propensity to cycle but increases the propensities to use transit and the driving mode. Furthermore, travel during winter months reduces the propensity to cycle. Higher levels of car ownership increase the propensity to select the driving mode.

5.2.3 Elasticities

Table 8 enumerates aggregate arc elasticities for various policy-relevant scenarios. Our first observation is that the elasticities of demand for cycling in response to changes in the out-of-vehicle travel time of the cycling alternative differ markedly across the three models. Whereas MNP and MNR suggest that demand is inelastic, Gen-MNR indicates that demand is elastic. For example, the aggregate arc elasticity for a 10% decrease in the out-of-vehicle time of the cycling alternative is -0.86, -0.71 and -1.08 in MNP, MNR and Gen-MNR, respectively. A reduction of the out-of-vehicle travel time of the cycling alternative represents an important policy-relevant scenario. For instance, the construction of cycling superhighways (e.g. Rayaprolu et al., 2020) and stimulation of e-bike uptake (e.g. Dill and Rose, 2012) could result in wide-spread decreases of cycling travel times. Gen-MNR makes a stronger case for the effectiveness of these interventions than MNP and MNR.
The three models also produce different elasticities for changes to walking out-of-vehicle travel times. The demand for walking is estimated to be more elastic to changes in walking

time in MNR and Gen-MNR than in MNP. For example, the aggregate arc elasticity of walking demand for a 10% reduction in walking time is -1.56 in both MNR and Gen-MNR but is -1.35 in MNP. Besides, the demand for walking is estimated to be more elastic to changes in walking time in Gen-MNR than in MNP and MNR. For example, the aggregate arc elasticity of cycling demand for a 10% reduction in walking time is 0.30 in Gen-MNR, whereas it is 0.02 and 0.04 in MNP and MNR. Innovations such as fast-moving walkways (e.g. Scarinci et al., 2017) can encourage the use of sustainable transport modes. A policy to support such walkways would have been less compelling based on MNP estimates.

Furthermore, the estimates of the elasticities for changes in transit out-of-vehicle times differ noticeably across the three models. For instance, the elasticity of demand for cycling is approximately twice as high in MNP and Gen-MNR as in MNR. For a 10% reduction in transit out-of-vehicle travel time, the elasticity of cycling demand is 0.29 in both MNP and Gen-MNR, while it is only 0.13 in MNR. We make a similar observation regarding the estimates of the elasticities for changes in transit in-vehicle. The estimated elasticities of demand for cycling are roughly twice as high in MNP and Gen-MNR as in MNR. For a 10% reduction in transit in-vehicle travel time, the elasticity of cycling demand is 0.24 in MNP and 0.28 in Gen-MNR, while it is only 0.12 in MNR.

There are no noteworthy differences in elasticities for changes in transit fares, driving cost, driving in-vehicle travel time and driving travel time variability across the three models. All models suggest that demand is inelastic to changes in these variables, and the calculated aggregate arc elasticities have the expected signs.

In sum, the elasticity estimates reveal interesting differences between the three models, in particular regarding the elasticities of the demand for the under-represented walking and cycling alternatives. Overall, Gen-MNR produces the most plausible elasticity estimates. For example, Gen-MNR suggests that cycling demand is elastic in changes in cycling travel times. Also, Gen-MNR indicates that demand for cycling is sensitive to changes in walking times. The elasticity estimates of MNR are consistent with the ones of Gen-MNR in some scenarios (e.g. elasticity of cycling demand for changes in walking time) but differ starkly from the ones of MNP and Gen-MNR in other situations (e.g. elasticity of cycling demand for changes in transit out-of-vehicle travel time).

Parameter	MNP				MNR				Gen-MNR			
	Mean	**Std. dev.**	**[0.025%**	**0.975%]**	**Mean**	**Std. dev.**	**[0.025%**	**0.975%]**	**Mean**	**Std. dev.**	**[0.025%**	**0.975%]**
$\beta_{\text{asc, cycle}}$	-1.986	0.053	-2.089	-1.878	-2.887	0.183	-3.242	-2.563	-1.907	0.103	-2.098	-1.662
$\beta_{\text{asc, transit}}$	-0.296	0.011	-0.318	-0.275	-0.598	0.022	-0.640	-0.557	-0.587	0.020	-0.627	-0.548
$\beta_{\text{asc, drive}}$	-0.832	0.023	-0.877	-0.786	-1.654	0.051	-1.757	-1.563	-1.621	0.048	-1.733	-1.527
β_{cost}	-0.057	0.002	-0.062	-0.052	-0.118	0.006	-0.129	-0.106	-0.097	0.004	-0.106	-0.089
β_{ovtt}	-1.934	0.041	-2.006	-1.850	-4.136	0.119	-4.396	-3.962	-3.605	0.083	-3.786	-3.430
β_{ivtt}	-1.052	0.036	-1.124	-0.981	-2.348	0.086	-2.543	-2.200	-2.003	0.065	-2.132	-1.873
β_{tv}	-1.184	0.039	-1.260	-1.108	-2.049	0.075	-2.196	-1.906	-1.817	0.060	-1.939	-1.699
$\beta_{\text{transfers}}$	-0.080	0.008	-0.094	-0.065	-0.139	0.015	-0.167	-0.110	-0.142	0.012	-0.166	-0.117
$\beta_{\text{female, cycle}}$	-0.660	0.037	-0.734	-0.588	-1.850	0.193	-2.270	-1.499	-1.041	0.097	-1.258	-0.866
$\beta_{\text{winter, cycle}}$	-0.149	0.030	-0.207	-0.091	-0.342	0.085	-0.513	-0.178	-0.189	0.045	-0.282	-0.102
$\beta_{(\text{age}<18\text{ years})\vee(\text{age}\geq 65\text{ years}),\text{ cycle}}$	-0.519	0.049	-0.616	-0.425	-2.041	0.290	-2.641	-1.507	-1.025	0.121	-1.270	-0.800
$\beta_{\text{female, transit}}$	0.032	0.009	0.015	0.050	0.067	0.013	0.042	0.093	0.079	0.011	0.058	0.102
$\beta_{\text{age}<18\text{ years, transit}}$	0.048	0.014	0.020	0.075	0.002	0.020	-0.038	0.042	0.031	0.018	-0.004	0.065
$\beta_{\text{age}\geq 65\text{ years, transit}}$	0.167	0.013	0.142	0.192	0.232	0.019	0.196	0.269	0.219	0.017	0.188	0.252
$\beta_{\text{female, drive}}$	0.010	0.012	-0.014	0.034	-0.037	0.022	-0.079	0.006	0.029	0.018	-0.006	0.065
$\beta_{\text{age}<18\text{ years, drive}}$	-0.445	0.023	-0.490	-0.400	-0.989	0.047	-1.085	-0.903	-0.790	0.037	-0.863	-0.720
$\beta_{\text{age}\geq 65\text{ years, drive}}$	0.179	0.017	0.146	0.213	0.268	0.031	0.208	0.327	0.269	0.027	0.216	0.323
$\beta_{\text{cars, drive}}$	0.596	0.016	0.567	0.629	1.096	0.030	1.044	1.157	0.991	0.028	0.938	1.052
$\Sigma_{\text{walk}-\text{drive,walk}-\text{drive}}$	0.745	0.040	0.673	0.822	1.299	0.054	1.200	1.407	1.360	0.080	1.208	1.522
$\Sigma_{\text{walk}-\text{drive,cycle}-\text{drive}}$	-0.109	0.069	-0.248	0.028	-0.490	0.192	-0.849	-0.115	0.541	0.075	0.387	0.683
$\Sigma_{\text{walk}-\text{drive,transit}-\text{drive}}$	0.462	0.026	0.415	0.514	0.952	0.039	0.883	1.027	1.115	0.059	1.006	1.227
$\Sigma_{\text{cycle}-\text{drive,cycle}-\text{drive}}$	1.864	0.059	1.750	1.972	0.913	0.084	0.751	1.066	0.688	0.120	0.455	0.911
$\Sigma_{\text{cycle}-\text{drive,transit}-\text{drive}}$	0.321	0.048	0.225	0.409	-0.236	0.139	-0.513	0.006	0.405	0.062	0.288	0.536
$\Sigma_{\text{transit}-\text{drive,transit}-\text{drive}}$	0.390	0.021	0.353	0.432	0.788	0.033	0.729	0.849	0.952	0.053	0.858	1.053
ν					2.031	0.077	1.889	2.181				
$\nu_{\text{walk}-\text{drive}}$									4.315	0.221	3.897	4.761
$\nu_{\text{cycle}-\text{drive}}$									2.583	0.332	1.970	3.269
$\nu_{\text{transit}-\text{drive}}$									10.647	2.518	6.875	15.950

Table 7: Estimated parameters of MNP, MNR and Gen-MNR for the case study

Scenario		MNP				MNR				Gen-MNR			
		Walk	**Cycle**	**Transit**	**Drive**	**Walk**	**Cycle**	**Transit**	**Drive**	**Walk**	**Cycle**	**Transit**	**Drive**
Cycling out-of-vehicle travel time	decreased by 5%	-0.06	-0.87	0.06	0.04	-0.08	-0.70	0.05	0.05	-0.07	-1.06	0.07	0.04
	decreased by 10%	-0.03	-0.86	0.06	0.03	-0.04	-0.71	0.04	0.04	-0.02	-1.08	0.07	0.04
	decreased by 25%	-0.01	-0.83	0.06	0.03	-0.01	-0.71	0.03	0.04	0.00	-1.08	0.07	0.04
Walking out-of-vehicle travel time	decreased by 5%	-1.41	0.03	0.48	0.16	-1.63	0.06	0.56	0.15	-1.63	0.31	0.55	0.15
	decreased by 10%	-1.35	0.02	0.48	0.16	-1.56	0.04	0.57	0.15	-1.56	0.30	0.55	0.15
	decreased by 25%	-1.26	0.02	0.53	0.17	-1.42	0.04	0.62	0.17	-1.43	0.32	0.61	0.17
Transit out-of-vehicle travel time	decreased by 5%	0.25	0.31	-0.40	0.23	0.34	0.14	-0.44	0.25	0.33	0.32	-0.45	0.25
	decreased by 10%	0.28	0.29	-0.39	0.22	0.37	0.13	-0.44	0.24	0.37	0.29	-0.45	0.25
	decreased by 25%	0.29	0.28	-0.36	0.21	0.39	0.12	-0.40	0.23	0.38	0.27	-0.41	0.23
Transit in-vehicle travel time	decreased by 5%	0.10	0.27	-0.33	0.24	0.11	0.15	-0.37	0.27	0.10	0.31	-0.37	0.27
	decreased by 10%	0.12	0.24	-0.33	0.23	0.15	0.12	-0.36	0.26	0.14	0.28	-0.36	0.26
	decreased by 25%	0.13	0.22	-0.29	0.21	0.16	0.11	-0.33	0.24	0.15	0.25	-0.33	0.24
Transit fares	increased by 5%	0.14	0.09	-0.13	0.05	0.17	0.02	-0.15	0.06	0.17	0.05	-0.14	0.06
	increased by 10%	0.11	0.09	-0.13	0.06	0.13	0.03	-0.14	0.07	0.13	0.07	-0.13	0.06
	increased by 25%	0.09	0.11	-0.13	0.07	0.11	0.05	-0.14	0.07	0.11	0.09	-0.14	0.07
Driving cost	increased by 5%	0.07	0.01	0.05	-0.08	0.10	0.03	0.04	-0.08	0.09	0.00	0.04	-0.08
	increased by 10%	0.04	0.02	0.06	-0.07	0.05	0.05	0.05	-0.07	0.05	0.02	0.06	-0.07
	increased by 25%	0.02	0.02	0.07	-0.07	0.03	0.05	0.06	-0.07	0.03	0.03	0.06	-0.07
Driving in-vehicle travel time	increased by 5%	0.12	0.12	0.25	-0.28	0.15	0.18	0.29	-0.32	0.15	0.13	0.28	-0.32
	increased by 10%	0.09	0.13	0.27	-0.28	0.11	0.20	0.30	-0.32	0.10	0.16	0.29	-0.32
	increased by 25%	0.08	0.13	0.28	-0.30	0.09	0.22	0.31	-0.34	0.08	0.18	0.31	-0.34
Driving travel time uncertainty	decreased by 5%	0.06	0.19	0.31	-0.31	0.03	0.27	0.29	-0.28	0.03	0.24	0.30	-0.29
	decreased by 10%	0.09	0.18	0.31	-0.31	0.07	0.25	0.28	-0.28	0.07	0.22	0.28	-0.28
	decreased by 25%	0.10	0.17	0.29	-0.29	0.09	0.23	0.26	-0.26	0.09	0.20	0.27	-0.26

Table 8: Aggregate arc elasticities for MNP, MNR and Gen-MNR for the case study

6 Conclusion

Models that are robust to violations of modelling assumptions and safeguard inferences against aberrant choice behaviour have received limited attention in discrete choice analysis. In this paper, we present Bayesian formulations of two robust alternatives to the multinomial probit (MNP) model. These alternatives belong to the family of robit models whose kernel error distributions are heavy-tailed t-distributions. The first alternative is the multinomial robit model, in which a single, generic degrees of freedom (DOF) parameter controls the heavy-tailedness of the kernel error distribution. The second alternative is a generalised multinomial robit (Gen-MNR) model, whose kernel error distribution is a t-distribution with alternative-specific DOF parameters. The kernel error distribution of Gen-MNR is more flexible than the kernel error distribution of MNR, as it allows for different marginal heavy-tailedness. To the best of our knowledge, Gen-MNR has not been studied in the literature before. For both models, we devise scalable and gradient-free Gibbs samplers, which address the limitations of estimation approaches of existing robit choice models.

We contrast MNP, MNR and Gen-MNR in a simulation study and a case study on transport mode choice behaviour. The simulation study illustrates the excellent finite-sample properties of the proposed Bayes estimators. We also show that MNR and Gen-MNR yield more faithful elasticity estimates if the true data generating process involves a heavy-tailed kernel error distribution. In the case study, we demonstrate that both MNR and Gen-MNR outperform MNP by a significant margin in term of in-sample fit and out-of-sample predictive ability. More specifically, Gen-MNR delivers the best in-sample fit and out-of-sample predictive due to its more flexible kernel error distribution. Gen-MNR also produces more plausible elasticity estimates than MNP and MNR, in particular regarding the demand for under-represented alternatives in a class-imbalanced data set.

On the whole, our analysis suggests that Gen-MNR is a useful addition to the choice modeller's toolbox due to its robustness properties. In general, Gen-MNR should be preferred over the previously-studied MNR model because of its more flexible kernel error distribution. In practice, the non-elliptical contoured t-distribution used in the formulation of Gen-MNR can also be specified in a way such that one DOF parameter controls the heavy-tailedness of more than one marginal of the kernel error distribution. Analysts can exploit this feature of Gen-MNR to achieve more parsimonious model specifications.

Our work suggests several directions for future research. First, the hierarchical Bayesian modelling paradigm can be leveraged to accommodate flexible parametric and semi-parametric representations of unobserved taste heterogeneity (see Krueger et al., 2020) into the MNR and Gen-MNR models. Incorporating these representations only requires adding another layer to the proposed Gibbs sampling schemes. Second, flexible nonlinear specifications of the systematic utility can be incorporated into the MNR and Gen-MNR models to enhance their expressiveness and predictive abilities. For example, Kindo et al. (2016) propose a MNP model in which the systematic utilities are represented using the Bayesian additive regression trees (BART) model of Chipman et al. (2010). BART automatically partitions a large predictor space to capture interaction effects and nonlinearities. As BART has foundations in the Bayesian inferential paradigm, BART components can be incorporated into MNR and Gen-MNR with relative ease. Third, the proposed MNR and Gen-MNR models can be extended to skew-t-distributed kernel errors which

can also account for asymmetric error distributions (Kim et al., 2008, Lee and Mclachlan, 2014).

Author contribution statement

RK: conception and design, method development and implementation, data processing and analysis, manuscript writing and editing, supervision.
PB: conception and design, manuscript writing and editing.
MB: conception and design, manuscript editing, supervision.
TG: conception and design, method development and implementation, manuscript writing and editing.

References

Albert, J. H. and Chib, S. (1993). Bayesian analysis of binary and polychotomous response data. *Journal of the American statistical Association*, 88(422):669–679.

Alptekinoğlu, A. and Semple, J. H. (2016). The exponomial choice model: A new alternative for assortment and price optimization. *Operations Research*, 64(1):79–93.

Bezanson, J., Edelman, A., Karpinski, S., and Shah, V. B. (2017). Julia: A fresh approach to numerical computing. *SIAM review*, 59(1):65–98.

Brathwaite, T. and Walker, J. L. (2018). Asymmetric, closed-form, finite-parameter models of multinomial choice. *Journal of choice modelling*, 29:78–112.

Brier, G. W. (1950). Verification of forecasts expressed in terms of probability. *Monthly weather review*, 78(1):1–3.

Burgette, L. F. and Nordheim, E. V. (2012). The trace restriction: An alternative identification strategy for the bayesian multinomial probit model. *Journal of Business & Economic Statistics*, 30(3):404–410.

Castillo, E., Menéndez, J. M., Jiménez, P., and Rivas, A. (2008). Closed form expressions for choice probabilities in the weibull case. *Transportation Research Part B: Methodological*, 42(4):373–380.

Chikaraishi, M. and Nakayama, S. (2016). Discrete choice models with q-product random utilities. *Transportation Research Part B: Methodological*, 93:576–595.

Chipman, H. A., George, E. I., McCulloch, R. E., et al. (2010). Bart: Bayesian additive regression trees. *The Annals of Applied Statistics*, 4(1):266–298.

Del Castillo, J. (2016). A class of rum choice models that includes the model in which the utility has logistic distributed errors. *Transportation Research Part B: Methodological*, 91:1–20.

Del Castillo, J. (2020). Choice probabilities of random utility maximization models when the errors distribution is a polynomial copula with gumbel marginals. *Transportmetrica A: Transport Science*, 16(3):439–472.

Dill, J. and Rose, G. (2012). Electric bikes and transportation policy: Insights from early adopters. *Transportation research record*, 2314(1):1–6.

Ding, P. (2014). Bayesian robust inference of sample selection using selection-t models. *Journal of Multivariate Analysis*, 124:451–464.

Dubey, S., Bansal, P., Daziano, R. A., and Guerra, E. (2020). A generalized continuous-multinomial response model with a t-distributed error kernel. *Transportation Research Part B: Methodological*, 133:114–141.

Fosgerau, M. and Bierlaire, M. (2009). Discrete choice models with multiplicative error terms. *Transportation Research Part B: Methodological*, 43(5):494–505.

Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A., and Rubin, D. B. (2013). *Bayesian data analysis*. CRC press.

Gelman, A., Rubin, D. B., et al. (1992). Inference from iterative simulation using multiple sequences. *Statistical science*, 7(4):457–472.

Gneiting, T. and Raftery, A. E. (2007). Strictly proper scoring rules, prediction, and estimation. *Journal of the American statistical Association*, 102(477):359–378.

Hillel, T., Elshafie, M. Z., and Jin, Y. (2018). Recreating passenger mode choice-sets for transport simulation: A case study of london, uk. *Proceedings of the Institution of Civil Engineers-Smart Infrastructure and Construction*, 171(1):29–42.

Imai, K. and Van Dyk, D. A. (2005). A bayesian analysis of the multinomial probit model using marginal data augmentation. *Journal of econometrics*, 124(2):311–334.

Jiang, Z. and Ding, P. (2016). Robust modeling using non-elliptically contoured multivariate t distributions. *Journal of Statistical Planning and Inference*, 177:50–63.

Kim, S., Chen, M.-H., and Dey, D. K. (2008). Flexible generalized t-link models for binary response data. *Biometrika*, 95(1):93–106.

Kindo, B. P., Wang, H., and Peña, E. A. (2016). Multinomial probit bayesian additive regression trees. *Stat*, 5(1):119–131.

Krueger, R., Rashidi, T. H., and Vij, A. (2020). A dirichlet process mixture model of discrete choice: Comparisons and a case study on preferences for shared automated vehicles. *Journal of Choice Modelling*, 36:100229.

Lange, K. L., Little, R. J., and Taylor, J. M. (1989). Robust statistical modeling using the t distribution. *Journal of the American Statistical Association*, 84(408):881–896.

Lee, S. and Mclachlan, G. J. (2014). Finite mixtures of multivariate skew t-distributions: some recent and new results. *Statistics and Computing*, 24(2):181–202.

Liu, C. (2004). Robit regression: a simple robust alternative to logistic and probit regression. *Applied Bayesian Modeling and Casual Inference from Incomplete-Data Perspectives*, pages 227–238.

Liu, J. S. (2008). *Monte Carlo strategies in scientific computing*. Springer Science & Business Media.

McCulloch, R. and Rossi, P. E. (1994). An exact likelihood analysis of the multinomial probit model. *Journal of Econometrics*, 64(1-2):207–240.

McFadden, D. (1981). Econometric models of probabilistic choice. *Structural analysis of discrete data with econometric applications*, 198272.

Paleti, R. (2019). Discrete choice models with alternate kernel error distributions. *Journal of the Indian Institute of Science*, pages 1–10.

Peyhardi, D. J. (2020). Robustness of student link function in multinomial choice models. *Journal of Choice Modelling*, 36:100228.

Rayaprolu, H. S., Llorca, C., and Moeckel, R. (2020). Impact of bicycle highways on commuter mode choice: A scenario analysis. *Environment and Planning B: Urban Analytics and City Science*, 47(4):662–677.

Robert, C. and Casella, G. (2013). *Monte Carlo statistical methods*. Springer Science & Business Media.

Scarinci, R., Markov, I., and Bierlaire, M. (2017). Network design of a transport system based on accelerating moving walkways. *Transportation Research Part C: Emerging Technologies*, 80:310–328.

Tanner, M. A. and Wong, W. H. (1987). The calculation of posterior distributions by data augmentation. *Journal of the American statistical Association*, 82(398):528–540.

A Gibbs sampling details

A.1 Sampling w

To update w, we iteratively sample from univariate truncated normal distributions. We have

$$w_{ij} \sim \mathrm{TN}(\mu_{ij}, \tau_{ij}^2), \quad \text{for } i = 1, \ldots, N, j = 1, \ldots, J-1. \tag{7}$$

For MNP, we have $\mu_{ij} = \mathbf{X}_{ij}^\top \boldsymbol{\beta} + \boldsymbol{\Sigma}_{j,-j}\boldsymbol{\Sigma}_{-j,-j}^{-1}(w_{i,-j} - \mathbf{X}_{i,-j}\boldsymbol{\beta})$ and $\tau_{ij}^2 = \boldsymbol{\Sigma}_{jj} - \boldsymbol{\Sigma}_{j,-j}\boldsymbol{\Sigma}_{-j,-j}^{-1}\boldsymbol{\Sigma}_{-j,j}$. For MNR, we have $\mu_{ij} = \mathbf{X}_{ij}^\top \boldsymbol{\beta} + \boldsymbol{\Sigma}_{j,-j}\boldsymbol{\Sigma}_{-j,-j}^{-1}(w_{i,-j} - \mathbf{X}_{i,-j}\boldsymbol{\beta})$ and $\tau_{ij}^2 = (\boldsymbol{\Sigma}_{jj} - \boldsymbol{\Sigma}_{j,-j}\boldsymbol{\Sigma}_{-j,-j}^{-1}\boldsymbol{\Sigma}_{-j,j})/q_i$. For Gen-MNR, we have $\mu_{ij} = \mathbf{X}_{ij}^\top \boldsymbol{\beta} + \mathbf{Q}_{ijj}^{-1/2}\boldsymbol{\Sigma}_{j,-j}\boldsymbol{\Sigma}_{-j,-j}^{-1}\mathbf{Q}_{i,-j,-j}^{1/2}(w_{i,-j} - \mathbf{X}_{i,-j}\boldsymbol{\beta})$ and $\tau_{ij}^2 = (\boldsymbol{\Sigma}_{jj} - \boldsymbol{\Sigma}_{j,-j}\boldsymbol{\Sigma}_{-j,-j}^{-1}\boldsymbol{\Sigma}_{-j,j})/q_{ij}$. Here, the index $-l$ denotes the vector without the lth element. For all models, the constraint on w_{ij} is $w_{ij} \geq \max\{0, w_{i,-j}\}$, if $y_{ij} = j$; $w_{ij} < 0$, if $y_{ij} = J$; $w_{ij} \leq \max\{0, w_{ij'}\}$, if $y_{ij} = j' \neq j$.

A.2 Sampling ν

The full conditional distribution of ν is nonstandard. Ding (2014) shows that

$$p(\nu|\cdot) \propto \exp\left\{\frac{N\nu}{2}\log\left(\frac{\nu}{2}\right) - N\log\Gamma\left(\frac{\nu}{2}\right) + (\alpha_0 - 1)\log\nu - \xi\nu\right\}, \tag{8}$$

where $\xi = \beta_0 + \frac{1}{2}\sum_{i=1}^N q_i - \frac{1}{2}\sum_{i=1}^N \log q_i$. $\Gamma(x)$ denotes the Gamma function. Ding (2014) proposes to sample from (8) using a Metropolised Independence sampler (Liu, 2008) with an approximate Gamma proposal. The shape parameter α^* and the rate parameter β^* of the proposal density are obtained as follows. The log conditional density of ν up to an additive constant is

$$l(\nu) = \frac{N\nu}{2}\log\left(\frac{\nu}{2}\right) - N\log\Gamma\left(\frac{\nu}{2}\right) + (\alpha_0 - 1)\log\nu - \xi\nu. \tag{9}$$

The log density of the Gamma proposal is

$$h(\nu) = (\alpha^* - 1)\log\nu - \beta^*\nu. \tag{10}$$

The first and second derivates of $l(\nu)$ and $h(\nu)$ are

$$l'(\nu) = \frac{N}{2}\left[\log\left(\frac{\nu}{2}\right) + 1 - \psi\left(\frac{\nu}{2}\right)\right] + \frac{\alpha_0 - 1}{\nu} - \xi, \quad h'(\nu) = \frac{\alpha^* - 1}{\nu} - \beta^*, \tag{11}$$

$$l''(\nu) = \frac{N}{2}\left[\frac{1}{\nu} - \frac{1}{2}\psi'\left(\frac{\nu}{2}\right)\right] + \frac{\alpha_0 - 1}{\nu^2}, \quad h''(\nu) = -\frac{\alpha^* - 1}{\nu^2}, \tag{12}$$

where $\psi(x)$ and $\psi'(x)$ are the di- and trigamma functions, respectively. The mode of $h(\nu)$ is $\frac{\alpha^*-1}{\beta^*}$ and the corresponding curvature is $\frac{(\beta^*)^2}{\alpha^*-1}$. We numerically find the mode ν^* of $l(\nu)$ and its corresponding curvature $l^* = l''(\nu^*)$. Ultimately, we match the modes and the corresponding curvatures of $l(\nu)$ and $h(\nu)$ to obtain

$$\alpha^* = 1 - (\nu^*)^2 l^*, \quad \beta^* = -\nu^* l^*. \tag{13}$$

A.3 Sampling q_{ij}

The full conditional distribution of q_{ij} is nonstandard. Jiang and Ding (2016) show that

$$p(q_{ij}|\cdot) \propto \exp\left\{-\frac{q_{ij}u_{ij}}{2} - \sqrt{q_{ij}}c_{ij} + \frac{\nu_j - 1}{2}\log q_{ij}\right\}, \tag{14}$$

where $u_{ij} = \nu_j + (\boldsymbol{\Sigma}^{-1})_{jj}(w_{ij} - \mathbf{X}_{ij}^\top \boldsymbol{\beta})^2$ and $c_{ij} = (w_{ij} - \mathbf{X}_{ij}^\top \boldsymbol{\beta}) \sum_{j' \neq j} \left(\sqrt{q_{ij'}(\boldsymbol{\Sigma}^{-1})_{jj'}}(w_{ij} - \mathbf{X}_{ij}^\top \boldsymbol{\beta})\right)$. Jiang and Ding (2016) propose to sample from (14) using a Metropolised Independence sampler (Liu, 2008) with an approximate Gamma proposal. The shape parameter α^* and the rate parameter β^* of the proposal density are obtained as follows. For $\nu_j \leq 1$, we set $\alpha^* = 1$ and $\beta^* = \frac{u_{ij}}{2}$. For $\nu_j > 1$, α^* and β^* are obtained through matching the modes and the corresponding curvatures of the target and the proposal densities. The log conditional density of q_{ij} up to an additive constant is

$$f(q_{ij}) = -\frac{q_{ij}u_{ij}}{2} - \sqrt{q_{ij}}c_{ij} + \frac{\nu_j - 1}{2}\log q_{ij}. \tag{15}$$

The log density of the Gamma proposal is

$$g(q_{ij}) = (\alpha^* - 1)\log q_{ij} - \beta^* q_{ij}. \tag{16}$$

The mode of (16) and its corresponding curvature are $\frac{\alpha^*-1}{\beta^*} = m_{ij}^*$ and $\frac{(\beta^*)^2}{\alpha^*-1} = l_{ij}^*$, respectively. The first and second derivatives of (15) are

$$f'(q_{ij}) = -\frac{u_{ij}}{2} - \frac{c_{ij}}{2\sqrt{q_{ij}}} + \frac{\nu_j - 1}{2q_{ij}}, \quad f''(q_{ij}) = \frac{c_{ij}}{4\sqrt{q_{ij}^3}} - \frac{\nu_j - 1}{2q_{ij}^2}. \tag{17}$$

The mode of (15) is $m_{ij}^* = \left(\frac{\frac{c_{ij}}{2} + \sqrt{\left(\frac{c_{ij}}{2}\right)^2 + u_{ij}(\nu_j - 1)}}{\nu_j - 1}\right)^{-2}$, and the corresponding curvature is $l_{ij}^* = f''(m_{ij}^*)$. After matching the modes and corresponding curvatures of the log target and the log proposal densities, we obtain

$$\alpha^* = 1 - (m_{ij}^*)^2 l_{ij}^*, \quad \beta^* = -m_{ij}^* l_{ij}^*. \tag{18}$$

B Additional results for the simulation study

B.1 Example I

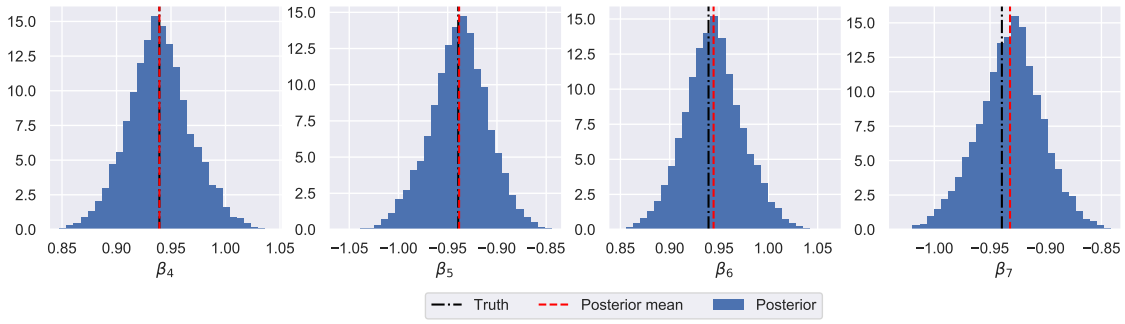

Figure 4: Estimated posterior distribution and true values of the taste parameters $\{\beta_4, \beta_5, \beta_6, \beta_7\}$ for MNR in simulation example I

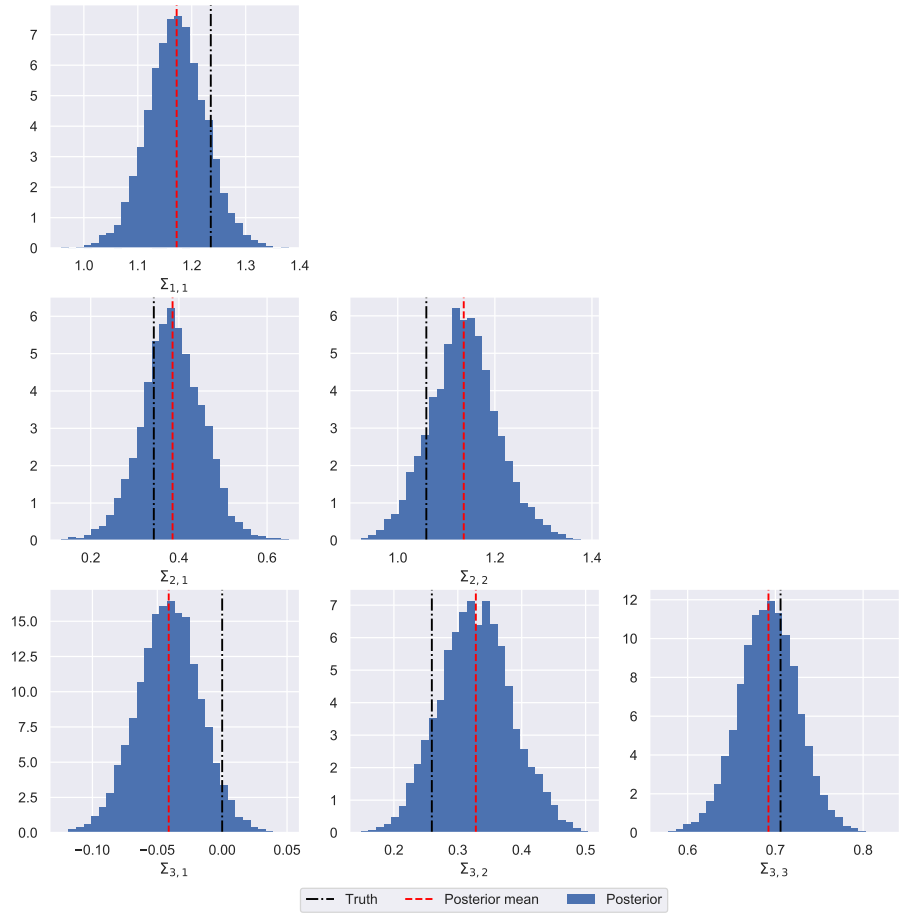

Figure 5: Estimated posterior distribution and true values of the unique elements of the covariance matrix $\boldsymbol{\Sigma}$ for MNR in simulation example I

B.2 Example II

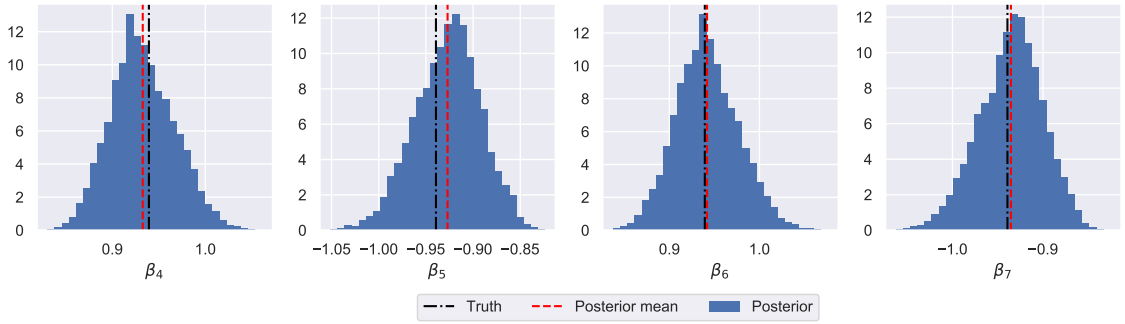

Figure 6: Estimated posterior distribution and true values of the taste parameters $\{\beta_4, \beta_5, \beta_6, \beta_7\}$ for the Gen-MNR model in simulation example II

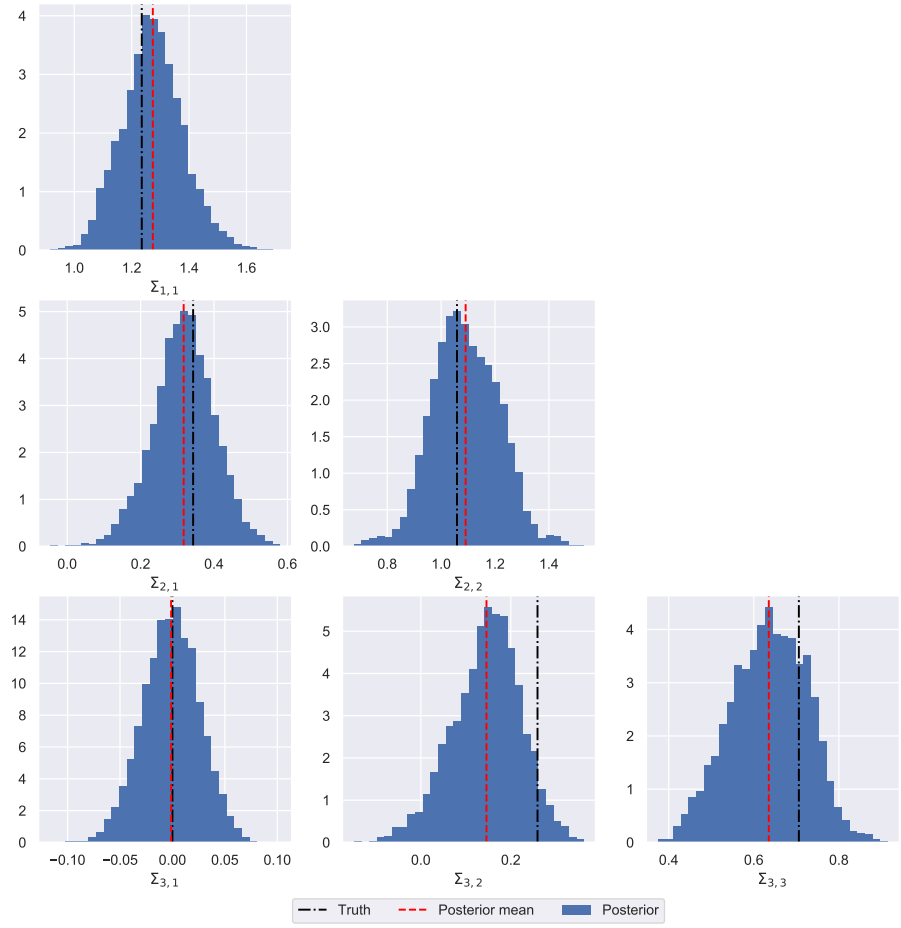

Figure 7: Estimated posterior distribution and true values of the unique elements of the covariance matrix $\boldsymbol{\Sigma}$ for the Gen-MNR model in simulation example II