# Robust discrete choice models with t-distributed kernel errors

Rico Krueger [*]     Michel Bierlaire [*]     Thomas Gasos [*]

Prateek Bansal [†]

May 18, 2021

[*]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

Inferences of robust behavioural and statistical models are insensitive to outlying observations resulting from aberrant behaviour, misreporting and misclassification. Standard discrete choice models such as logit and probit lack robustness to outliers due to their rigid kernel error distributions. In this paper, we analyse two robust alternatives to the multinomial probit (MNP) model. The two models belong to the family of robit models whose kernel error distributions are heavy-tailed t-distributions which moderate the influence of outlying observations. 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 model, the generalised multinomial robit (Gen-MNR) model, is more flexible than MNR, as it allows for distinct heavy-tailedness in each dimension of the kernel error distribution. For both models, we derive efficient Gibbs sampling schemes, which also allow for a straightforward inclusion of random parameters. In a simulation study, we illustrate the excellent finite sample properties of the proposed Bayes estimators and show that MNR and Gen-MNR produce more exact elasticity estimates if the choice data contain outliers through the lens of the non-robust MNP model. In a case study on transport mode choice behaviour, MNR and Gen-MNR outperform MNP by substantial margins in terms of in-sample fit and out-of-sample predictive accuracy. We also find that the benefits of the more flexible kernel error distributions underlying MNR and Gen-MNR are maintained in the presence of random heterogeneity.

*Keywords:* robustness, probit, robit, Bayesian estimation, discrete choice, outliers

# 1  Introduction

Random utility maximisation is by far the most widely adopted decision-making 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. In principle, the absolute level of utility is not identifiable, and only differences in utility matter (Train, 2009). Typically, utility differences are assumed to be either independent or identically logistically (logit kernel) or jointly Gaussian distributed (probit kernel).

Logit is particularly popular in practice due to its closed-form choice probabilities. However, logit suffers from two drawbacks. First, substitution patterns in logit are restricted by the independence of irrelevant alternatives property, which implies that the odds of choosing one alternative over another do not depend on a third option. Second, logit assumes that the random error terms are homoskedastic across alternatives, which implies that the same level of decision uncertainty applies to all alternatives in a choice set. Probit overcomes the limitations of logit by allowing for the estimation of a full error covariance, subject to identification restrictions (Train, 2009). Probit can capture any substitution patterns and accommodates heteroskedasticity across utility differences. Probit choice probabilities lack closed-form expressions, yet advances in computational power combined with progress in simulation techniques (Burgette and Nordheim, 2012, Hajivassiliou et al., 1996, Imai and Van Dyk, 2005, McCulloch and Rossi, 1994, Train, 2009) and analytical approximations (Bhat, 2011) make this drawback less and less important.

Both logit and probit are constrained by strong parametric assumptions. Whilst the Gaussian distribution has a symmetric bell shape with light tails, the logistic distribution is also symmetric and exhibits slightly heavier tails than the Gaussian distribution. As a consequence, logit and probit lack robustness to outliers in the response data (Benoit et al., 2016, Hausman et al., 1998). In principle, a model is considered robust if its inferences are insensitive to outlying observations (Gelman et al., 2013). Whereas an outlier in continuous data is an extreme data point, an outlier in discrete data is an observation that is unexpected through the lens of a non-robust model (Gelman and Hill, 2006). Outliers in discrete choice analysis may result from aberrant behaviour as well as misreporting and misclassification of the response variable. Hausman et al. (1998) show that ignoring outliers in logit and probit applications can result in biased and inconsistent parameter estimates.

In this paper, we analyse two robust alternatives to the multinomial probit (MNP) model. Both alternatives belong to the family of robit models whose kernel error distributions are heavy-tailed t-distributions which moderate the influence of outlying observations. In the first model, the multinomial robit (MNR) model, a generic degrees of freedom parameter controls the heavy-tailedness of the kernel error distribution. The second model, the generalised multinomial robit (Gen-MNR), is more flexible than MNR, as it allows for different marginal heavy-tailedness of the kernel error distribution. For both models, we devise Bayes estimators, which also allow for a straightforward inclusion of random parameters. 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 as well as their mixed counterparts

with random parameters in a case study on transport mode choice behaviour in London, UK.[1]

The remainder of the paper is organised as follows: First, we give some background on robustness in discrete choice analysis and develop a motivation for the methodological innovations presented in this paper (Section 2). Then, we present the mathematical formulations of the MNP, MNR and Gen-MNR models and also explain how random utility parameters can be incorporated into the models (Section 3). Next, we outline the estimation approaches and succinctly discuss the adopted data augmentation techniques (Section 4). In Sections 5 and 6, we present the simulation and case studies. Finally, we conclude and identify avenues for future research (Section 7).

# 2 Background

In discrete choice data, outlying observations may result from aberrant behaviour as well as misreporting and misclassification of responses. From the analyst's perspective, outlier responses are unexpectedly stochastic, i.e. the contribution of the random error term is exceptionally large. Choice behaviour can be considered aberrant from the analyst's perspective if the analyst possesses little information about the factors influencing choices or if the postulated decision-making paradigm (such as random utility maximisation) does not accurately represent the decision protocols governing some of the observed choices. Misreporting frequently occurs in survey data collections, when a respondent intentionally or unintentionally selects an incorrect option (Hausman et al., 1998, Paleti and Balan, 2019). Similarly, misclassification occurs when a multinomial response variable is incorrectly labelled because of measurement error or ambiguity about the exact nature of the chosen alternative. For example, misclassification is encountered in automated smartphone-based travel surveys, when a chosen transport mode is labelled as a different alternative than it should be (Liang et al., 2019, Vij and Shankari, 2015).

Even though misreporting and misclassification are well-known problems in discrete choice analysis (Hausman et al., 1998), the formulation of robust discrete choice models based on flexible kernel error distributions has received limited attention. Also, efforts to capture aberrant choice behaviour in logit mixtures through scale heterogeneity have proven futile, as scale and correlation in logit mixtures are not separately identifiable (Hess and Rose, 2012, Hess and Train, 2017).

Robust models can be formulated on the basis of heavy-tailed distributions which moderate the influence outlying data points. 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. 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

---

[1]In what follows, we restrict our empirical analysis of MNR and Gen-MNR to comparisons with MNP, as MNP is quite general compared to popular multinomial choice models based on the generalised extreme value family of distributions (Train, 2009). As explained above, MNP can capture any substitution patterns and also accommodates heteroskedasticity across alternatives.

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

Researchers have proposed various departures from standard kernel error distributions for discrete choice models (also see Paleti, 2019, for a review), including generalised extreme value (McFadden, 1978), heteroskedastic extreme value (Bhat, 1995), negative exponential (Alptekinoğlu and Semple, 2016, Daganzo, 1979), 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 advancements do not aim at jointly enhancing robustness, admitting flexible substitution patterns and accommodating heteroskedasticity.

We are aware of three studies which formulate robust discrete choice models based on heavy-tailed kernel error distributions. 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 datasets. In another study, Peyhardi (2020) formulates a MNR model in the context of generalised linear models and shows that MNR can help in identifying artificial aspects in the design of stated preference experiments. Furthermore, Benoit et al. (2016) devise a multinomial choice model, in which utility differences follow a symmetric and heavy-tailed multivariate Laplace distribution. Using simulated and real data, Benoit et al. (2016) show that estimates of their proposed model are less sensitive to outlying observations than MNP estimates.

We identify two research gaps in the formulation and estimation of robust multinomial choice models. First, the kernel error distributions of existing models lack flexibility. 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 all utility differences exhibit the same level of aberrance. Unlike the t-distribution, the Laplace distribution underlying the formulation of the multinomial choice model proposed by Benoit et al. (2016) cannot exhibit varying levels of heavy-tailedness, since the Laplace distribution does not have a third parameter controlling the heavy-tailedness of the distribution. Second, the estimation approaches employed in the studies by Dubey et al. (2020) and Peyhardi (2020) 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 robust multinomial choice models (i.e. lack of flexibility) by formulating a generalised multinomial robit (Gen-MNR) model, in which each dimension of the kernel error distribution has its own DOF parameter. To that end, we adopt the non-elliptically contoured t-distribution (Jiang and Ding, 2016) as kernel error distribution in the formulation of a multinomial choice model. To tackle the second limitation (i.e. computationally-expensive estimation), we develop optimisation-free Bayes estimators for both MNR and Gen-MNR.[2] In the construction of the estimators, we exploit the hierarchical normal mixture representation of the t-distribution. To bypass complex likelihood computations in the estimation of MNR and Gen-MNR, 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 allows for a straightforward inclusion of random utility parameters.

For completeness, we also discuss the possibility to formulate robust multinomial choice models within an error components mixed logit (EC-MXL) framework (Walker et al., 2007). Typically, EC-MXL models are formulated using independent normal error components. In theory, it is possible to replace these independent normal error components with error components which are jointly distributed according to a heavy-tailed distribution or a flexible non-parametric distribution. However, it is not obvious how such a model could be estimated. First, identification in EC-MXL models is not trivial, even with independent normal error components (Walker et al., 2007). Second, it is well known that logit mixtures with non-normally distributed random parameters estimated via standard methods such as maximum simulated likelihood are plagued by convergence issues (see Bhat, 2011, and the literature referenced therein). Furthermore, non-parametric mixing distributions (see Bansal et al., 2018, Vij and Krueger, 2017, for reviews) suffer from the curse of dimensionality and require many more parameters to be estimated than MNR and Gen-MNR.

In principle, it is also possible to formulate a mixed probit model with error components distributed according to a heavy-tailed distribution within the framework presented by Bhat and Lavieri (2018). However, introducing error components which follow a heavy-tailed distribution into mixed probit would require the estimation of a large number of parameters, including a full covariance matrix for the Gaussian kernel error distribution as well as scale and shape parameters for the heavy-tailed error component distributions. It is likely that theoretical and empirical identification in such a model would prove problematic, especially with cross-sectional data.[3]

In sum, unlike possible extensions of logit and probit mixtures with error components, MNR and Gen-MNR are parsimonious, tractable and computationally efficient methods

---

[2]In spite of advances in computational hardware, computational efficiency remains an important issue in the estimation of behavioural and statistical models. In fact, computational efficiency takes on heightened importance, as datasets contain increasingly larger numbers of observations, alternatives and predictors. It should also be kept in mind that the ability of estimate complex discrete choice models is not solely due to advances in computational hardware but rather due to carefully crafted estimation algorithms, which take advantage of modern technology.

[3]A similar argumentation holds for possible extensions of mixed probit with error components distributed according to a flexible semi-parametric distribution such the power series transformation distribution presented in Fosgerau and Mabit (2013). Such a model would include many more parameters than MNR and Gen-MNR, and identification is likely to prove difficult, especially with cross-sectional data.

for the robust analysis of multinomial choice data. Compared to MNP, MNR has one more parameter controlling the heavy-tailedness of the kernel error distribution. In Gen-MNR, the number of additional parameters compared to MNP is strictly less than the number of alternatives.

# 3 Model formulations

In this section, we first present the formulations of the MNP, MNR and Gen-MNR models. Then, we explain how the models can be extended to accommodate random taste heterogeneity.

## 3.1 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. $\mathbf{X}_i = \begin{bmatrix} \mathbf{X}_{i1} \\ \vdots \\ \mathbf{X}_{i,J-1} \end{bmatrix} = \begin{bmatrix} \mathbf{X}_{i1}^{\text{obs}} - \mathbf{X}_{iJ}^{\text{obs}} \\ \vdots \\ \mathbf{X}_{i,J-1}^{\text{obs}} - \mathbf{X}_{iJ}^{\text{obs}} \end{bmatrix}$, where $\mathbf{X}_{ij}^{\text{obs}}$ 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. $y_i(\boldsymbol{w}_i) = y_i(c\boldsymbol{w}_i)$.[4] Therefore, we must set the scale of the model. Following Burgette and Nordheim (2012), we impose a trace restriction on $\boldsymbol{\Sigma}$ with $\text{tr}(\boldsymbol{\Sigma}) = J - 1$. We implement the trace restriction using a constrained inverse Wishart prior (Imai and Van Dyk, 2005, Burgette and Nordheim, 2012), which has the following nonstandard form:

$$P(\boldsymbol{\Sigma}) \propto |\boldsymbol{\Lambda}|^{-(\rho+J)/2} \left[ \text{tr}(\boldsymbol{\Lambda}\boldsymbol{\Sigma}^{-1}) \right]^{-\rho(J-1)/2} \mathbf{1}\{\text{tr}(\boldsymbol{\Sigma}) = J - 1\}, \tag{3}$$

---

[4]For a treatment of the identification issue in the context of MNP estimation, the reader is also directed to Bunch (1991).

where $\mathbf{1}\{\cdot\}$ is an indicator returning one if the condition inside the braces is true and zero otherwise.

To construct the constrained inverse Wishart prior, we introduce the working parameter $\alpha$, which is not identified given the observed data $\mathbf{y}$, but is identified given $\{\mathbf{y}, \mathbf{w}\}$. We transform $\tilde{\mathbf{w}}_i = \alpha \mathbf{w}_i$ for $i = 1, \ldots, N$ so that $\tilde{\mathbf{w}}_i \sim N(\mathbf{X}_i \tilde{\boldsymbol{\beta}}, \tilde{\boldsymbol{\Sigma}})$ with $\tilde{\boldsymbol{\beta}} = \alpha\boldsymbol{\beta}$ and $\tilde{\boldsymbol{\Sigma}} = \alpha^2 \boldsymbol{\Sigma}$. We put an inverse Wishart prior on the intermediate, unidentified quantity $\tilde{\boldsymbol{\Sigma}}$ so that $\tilde{\boldsymbol{\Sigma}} \sim IW(\rho, \tilde{\boldsymbol{\Lambda}})$. After the transformation $\boldsymbol{\Sigma} = \tilde{\boldsymbol{\Sigma}}/\alpha^2$ with $\alpha^2 = \mathrm{tr}(\tilde{\boldsymbol{\Sigma}})/(J-1)$, the implied prior on the tuple $\{\boldsymbol{\Sigma}, \alpha^2\}$ is

$$P(\boldsymbol{\Sigma}, \alpha^2) \propto |\boldsymbol{\Lambda}|^{-(\rho+J)/2} e^{-\frac{\alpha_0^2}{2\alpha^2}\mathrm{tr}(\boldsymbol{\Lambda}\boldsymbol{\Sigma}^{-1})} (\alpha^2)^{-\rho(J-1)/2+1} \mathbf{1}\{\mathrm{tr}(\boldsymbol{\Sigma}) = J-1\}, \tag{4}$$

where $\alpha_0$ is some positive constant such that $\tilde{\boldsymbol{\Lambda}} = \alpha_0 \boldsymbol{\Lambda}$. The Jacobian of the transformation adds a factor proportional to $\alpha^{J(J-1)-2}$. The conditional distribution of $\alpha^2$ is

$$P(\alpha^2 | \boldsymbol{\Sigma}) \propto \alpha_0^2 \mathrm{tr}(\boldsymbol{\Lambda}\boldsymbol{\Sigma}^{-1})/\chi^2_{\rho(J-1)}. \tag{5}$$

Integrating (4) over (5) yields (3).

To complete the specification of the MNP model, we specify the prior $\boldsymbol{\beta} \sim N(0, \mathbf{B}_0^{-1})$. It is known that predictions under the Bayesian formulation of the MNP model can be sensitive to the selection of the base alternative $J$ (Burgette and Nordheim, 2012).

## 3.2   Multinomial robit (MNR)

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

$$\mathbf{w}_i = \mathbf{X}_i \boldsymbol{\beta} + \boldsymbol{\varepsilon}_i \quad \text{with } \boldsymbol{\varepsilon}_i \sim t(0, \boldsymbol{\Sigma}, \nu), \quad \text{for } i = 1, \ldots, N, \tag{6}$$

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

$$\boldsymbol{\varepsilon}_i \sim N(0, \boldsymbol{\Sigma}/q_i) \quad \text{with } q_i \sim \chi^2_\nu/\nu, \quad \text{for } i = 1, \ldots, N. \tag{7}$$

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 $\boldsymbol{\varepsilon}_i$ across different $i$. Figure 1 illustrates the relationship between the $\chi^2$-distribution (which controls the distribution of $\mathbf{q}$) and a t-distribution (which controls the distribution of $\boldsymbol{\varepsilon}$) with unit variance for different DOF $\nu$. For small $\nu < 30$, the t-distribution exhibits heavy tails. As $\nu$ approaches $\infty$, the t-distribution converges to the normal distribution. We use the same priors for $\boldsymbol{\beta}$ and $\boldsymbol{\Sigma}$ as in MNP and also introduce the working parameter $\alpha$. In addition, we put a Gamma prior on $\nu$ with $\nu \sim \mathrm{Gamma}(\alpha_0, \beta_0)$. 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.
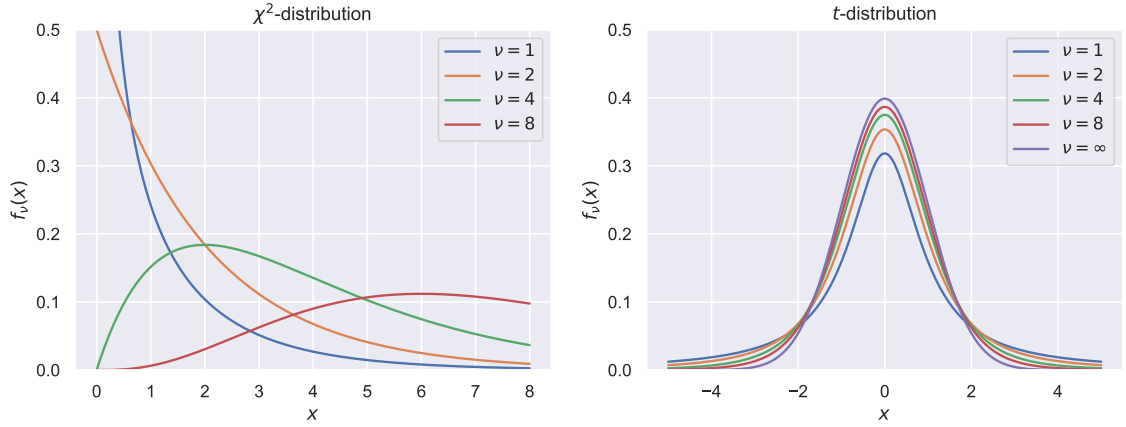
Figure 1: Relationship between $\chi^2$- and t-distributions for different degrees of freedom $\nu$

## 3.3 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-elliptically contoured t-distribution (NECT; Jiang and Ding, 2016). We have

$$\boldsymbol{w}_i = \mathbf{X}_i\boldsymbol{\beta} + \boldsymbol{\varepsilon}_i \quad \text{with } \boldsymbol{\varepsilon}_i \sim \text{NECT}_{\mathbf{p}}(\mathbf{0}, \boldsymbol{\Sigma}, \boldsymbol{\nu}), \quad \text{for } i = 1, \dots, N, \tag{8}$$

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

$$\boldsymbol{\varepsilon}_i = \mathbf{Q}_i^{-1/2}\boldsymbol{\Sigma}^{1/2}\mathbf{Z}_i, \quad \mathbf{Z}_i \sim \text{N}(\mathbf{0}, \mathbf{I}_{J-1}), \quad \text{for } i = 1, \dots, N, \tag{9}$$

where $\mathbf{Q}_i = \text{diag}(q_{i1}\mathbf{I}_{p_1}, \dots, q_{iM}\mathbf{I}_{p_M})$ is a $(J-1) \times (J-1)$ block-diagonal matrix with $q_{im} \sim \chi^2_{\nu_m}/\nu_m$ for $m = 1, \dots, M$. $\mathbf{I}_l$ 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 $\boldsymbol{\varepsilon} \sim \text{NECT}_{\mathbf{p}}(\mathbf{0}, \boldsymbol{\Sigma}, \boldsymbol{\nu})$, then $\varepsilon_j \sim t(0, \Sigma_{jj}, \nu_{m(j)})$, where $m(j)$ maps dimension j onto its associated DOF. In the rest of this work, we assume that $M = J-1$ without loss of generality. The Gen-MNR model uses the same prior distributions as the MNR model and also includes the working parameter $\alpha$. We let $\nu_j \sim \text{Gamma}(\alpha_0, \beta_0)$ for $j = 1, \dots, J-1$. 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.

## 3.4 Extension: Random taste parameters

MNP, MNR and Gen-MNR can be extended hierarchically to accommodate random taste heterogeneity. We refer to the resulting models as mixed MNP, MNR and Gen-MNR (henceforth, M-MNP, M-MNR, M-Gen-MNR). In M-MNP, the latent utility differences are represented as

$$\boldsymbol{w}_i = \mathbf{F}_i\boldsymbol{\beta} + \mathbf{D}_i\boldsymbol{\gamma}_i + \boldsymbol{\varepsilon}_i \quad \text{with } \boldsymbol{\varepsilon}_i \sim \text{N}(\mathbf{0}, \boldsymbol{\Sigma}), \quad \text{for } i = 1, \dots, N. \tag{10}$$

Here, $\mathbf{F}_i$ and $\mathbf{D}_i$ are matrices of differenced predictors pertaining to fixed, generic taste parameters $\boldsymbol{\beta}$ and distributed, individual-specific taste parameters $\boldsymbol{\gamma}_i$, respectively. $\mathbf{F}_i$ and $\mathbf{D}_i$ are $(J-1) \times K$ and $(J-1) \times L$ matrices, respectively. $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}_i$ are $K$ and $L$ vectors, respectively. We assume that $\boldsymbol{\gamma}_i \sim N(\boldsymbol{\eta}, \boldsymbol{\Omega})$ with $\boldsymbol{\Omega} = \mathrm{diag}(\omega_1, \ldots, \omega_L)$ for $i = 1, \ldots, N$. We let $\boldsymbol{\eta} \sim N(\mathbf{0}, \mathbf{C}_0)$ and $\omega_l \sim IG\left(\frac{\kappa}{2}, \kappa\lambda_l\right)$ with $\lambda_l \sim \mathrm{Gamma}\left(\frac{1}{2}, \frac{1}{\delta_l^2}\right)$ for $l = 1, \ldots, L$. IG denotes an inverse Gamma distribution, and $\lambda_l$ is a nuisance parameter. The induced prior on $\omega_l$ is a weakly-informative half-t prior (Gelman et al., 2006, Wand et al., 2011). The formulations of M-MNR and M-Gen-MNR are developed analogously.[5]

# 4   Inference and implementation details

For the estimation of the MNP, MNR and Gen-MNR models as well as their mixed counterparts, we employ Markov chain Monte Carlo methods in the form of Gibbs sampling (Robert and Casella, 2013). The sampling schemes for MNP, MNR and Gen-MNR are presented in Algorithms 1, 2 and 3, respectively. Algorithm 1 is based on the Gibbs samplers proposed by Burgette and Nordheim (2012) and Imai and Van Dyk (2005). Algorithms 2 and 3 are extensions of Algorithm 1. Whereas Algorithm 2 incorporates elements of the Gibbs sampler proposed by Ding (2014) for the robust Heckman selection model, Algorithm 3 incorporates elements of the Gibbs sampler proposed by Jiang and Ding (2016) for the multivariate robit model. Algorithm 4 describes a Gibbs sampler for the mixed MNP model. The algorithm is a straightforward extension of Algorithm 1. The estimation algorithms of mixed MNR and mixed Gen-MNR are developed analogously.[6] 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 $\mathbf{w}$ (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 $\mathbf{q}$. Data augmentation circumvents complex likelihood calculations in the estimation of the MNP, MNR and Gen-MNR models. This is because conditional on $\mathbf{w}$ and $\mathbf{q}$ (if applicable), the models reduce to standard Bayesian linear models.

Following Imai and Van Dyk (2005), we leverage marginal data augmentation (MDA; Van Dyk and Meng, 2001, Van Dyk, 2010) to increase the rate of convergence of the MCMC algorithms. The central idea of MDA is to marginalise out the working parameter $\alpha$, which is not identified given the observed data $\mathbf{y}$, but is identified given the augmented data $\{\mathbf{y}, \mathbf{w}\}$, in some of the conditional updates of the model parameters in order to improve mixing of the Markov chains. Imai and Van Dyk (2005) argue that

---

[5]In this paper, we restrict our attention to normal representations of unobserved heterogeneity. As MNP, MNR and Gen-MNR are embedded into a Bayesian inferential framework, more flexible semiparametric representations of unobserved heterogeneity can be incorporated into the models using the methods discussed in Krueger et al. (2020).

[6]To reduce the notational burden in the description of Algorithm 4, we assume that there is exactly one observation per individual. However, in our subsequent applications, we account for the panel structure of the data.

because $\alpha^2$ is assigned a distribution with positive variance in (5), the marginal model $\int P(\mathbf{y}, \mathbf{w}|\theta, \alpha) p(\alpha|\theta) d\alpha$ is more diffuse and thus mixes faster than the conditional model $P(\mathbf{y}, \mathbf{w}|\theta, \alpha)$.

The full conditional distribution of $\nu$ in the MNR model as well as the full conditional distributions of $\nu_j$ 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 the MCMC algorithms for all models considered in this paper in Julia (Bezanson et al., 2017).[7] Draws from the univariate truncated normal distribution are generated using a combination of inverse transform sampling and rejection sampling with a Rayleigh proposal (Botev and l'Ecuyer, 2016).

In the subsequent applications, the Gibbs samplers are executed with a single chain consisting of 100,000 draws including a warm-up period of 50,000 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).

Predictive choice distributions can be obtained via simulation at the estimated posterior means of the model parameters. For MNP and M-MNP, choice probabilities are calculated using the GHK simulator (Hajivassiliou et al., 1996, Train, 2009). For all other models, choice probabilities are obtained using frequency simulators (Lerman and Manski, 1981, Geweke et al., 1994).

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

## 5.1 Example I: Data generated according to MNR model

In the first example, data are generated according to the MNR model. We let $N = 10,000$ and $J = 4$. Furthermore, we set $\beta = (-1, 1, -1, 1, -1)^\top$, $\Sigma = \begin{bmatrix} 1.0 & 0.3 & 0.0 \\ 0.3 & 1.0 & 0.3 \\ 0.0 & 0.3 & 1.0 \end{bmatrix}$ and $\nu = 2$. Here, the first three predictors are alternative-specific constants. The remaining predictors are alternative-specific attributes. We draw $X_{ijk}^{obs} \sim U(0,5)$ for $i = 1, \ldots, N$, $j = 1, \ldots, J$ and $k = 4, 5$. 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 over the specification of the reference alternative.

The goal of Bayesian estimation is to infer the posterior distribution of the model parameters. The posterior distribution captures the knowledge and uncertainty about the dis-

---

[7]The estimation code is available at `https://github.com/RicoKrueger/robit`.

---

**Algorithm 1** Gibbs sampler for the MNP model

---

**Step 0:** Initialise parameters $w$, $\beta$, $\Sigma$, $\alpha$.
**for** $t = 1, \ldots, T$ **do**
    **Step 1:** Update $w$, $\alpha^2$.
    **for** $i = 1, \ldots, N$ **do**
        **for** $j = 1, \ldots, J - 1$ **do**
            Draw $w_{ij}|w_{i,-j}, \cdot$ as explained in Appendix A.1.
        **end for**
    **end for**
    Draw $\alpha^2|\cdot \sim \mathrm{tr}(\Lambda\Sigma^{-1})/\chi^2_{\rho(J-1)}$.
    Set $\tilde{w} = \alpha w$.
    **Step 2:** Update $\tilde{\beta}$, $\alpha^2$.
    Set $\hat{\mathbf{B}} = \left( \sum_{i=1}^{N} \mathbf{X}_i^\top \Sigma^{-1} \mathbf{X}_i + \mathbf{B}_0 \right)^{-1}$.
    Set $\hat{\beta} = \hat{\mathbf{B}} \left( \sum_{i=1}^{N} \mathbf{X}_i^\top \Sigma^{-1} \tilde{w}_i \right)$.
    Draw $\alpha^2|\cdot \sim \left( \sum_{i=1}^{N}(\tilde{w}_i - \mathbf{X}_i\hat{\beta})^\top \Sigma^{-1}(\tilde{w}_i - \mathbf{X}_i\hat{\beta}) + \hat{\beta}^\top \mathbf{B}_0\hat{\beta} + \mathrm{tr}(\Lambda\Sigma^{-1}) \right) /\chi^2_{(N+\rho)(J-1)}$.
    Draw $\tilde{\beta}|\cdot \sim N(\hat{\beta}, \alpha^2\hat{\mathbf{B}})$.
    **Step 3:** Update $\tilde{\Sigma}$.
    Set $\tilde{z}_i = \tilde{w}_i - \mathbf{X}_i\tilde{\beta}$.
    Draw $\tilde{\Sigma}|\cdot \sim IW \left( \rho + N, \tilde{\Lambda} + \sum_{i=1}^{N} \tilde{z}_i\tilde{z}_i^\top \right)$.
    Set $\alpha^2 = \mathrm{tr}(\tilde{\Sigma})/(J - 1)$.
    Set $\Sigma = \tilde{\Sigma}/\alpha^2$, $w = \tilde{w}/\alpha$, $\beta = \tilde{\beta}/\alpha$.
**end for**
**return** $\beta$, $\Sigma$

---

tribution of the model parameters, conditional on the evidence provided by the observed data. Figure 2 shows the marginal posterior distribution of the DOF parameter $\nu$ 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, because the true parameter value is well contained within the 95% central credible interval (i.e. the region in which the unknown parameter falls with 95% probability).[8] From Figures 4 and 5 in Appendix B.1, we conclude that Algorithm 2 also does an excellent job at recovering the remaining parameters $\beta$ and $\Sigma$.

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 $QL = \sum_{i=1}^{N} \sum_{j=1}^{J} (p_{nj} - \hat{p}_{nj})^2$, where $p_{nj}$ is the true choice probability simulated at the true parameter values, and where $\hat{p}_{nj}$ is the corresponding fitted choice probability. MNR offers the best fit to the data, closely followed by Gen-MNR. A possible explanation for the inability of Gen-MNR to perform exactly as well as MNR is that the estimation of multiple DOF parameters incurs a greater simulation error. Nonetheless, both MNR and Gen-MNR outperform the MNP by substantial margins.

---

[8]The bounds of the 95% central credible interval are given by the 2.5%- and 97.5%-quantiles of the posterior.

Finally, we contrast the elasticity estimates of the three models in several scenarios. In Table 2, we enumerate the aggregate arc elasticity estimates of the three models along with the corresponding true aggregate arc elasticities for those two scenarios in which we observe the most pronounced differences between the three models. In the first scenario reported in Table 2, we increase the first alternative-specific attribute of the first alternative by 10%. MNR and Gen-MNR produce direct aggregate arc elasticity estimates, which are closer to the truth than the direct aggregate arc elasticity estimates of MNP. In the considered scenario, the true direct aggregate arc elasticity is 1.68. While both MNR and Gen-MNR produce direct elasticity estimates of 1.70, MNP model a lower direct elasticity estimate of 1.60. In the second scenario reported in Table 2, we increase the first alternative-specific attribute of the third alternative by 10%. Again, we find that MNR and Gen-MNR produce less biased elasticity estimates than MNP. The true direct aggregate arc elasticity in the considered scenario is 1.67. While both MNR and Gen-MNR produce direct elasticity estimates of 1.65, MNP yields a lower direct elasticity estimate of 1.56.
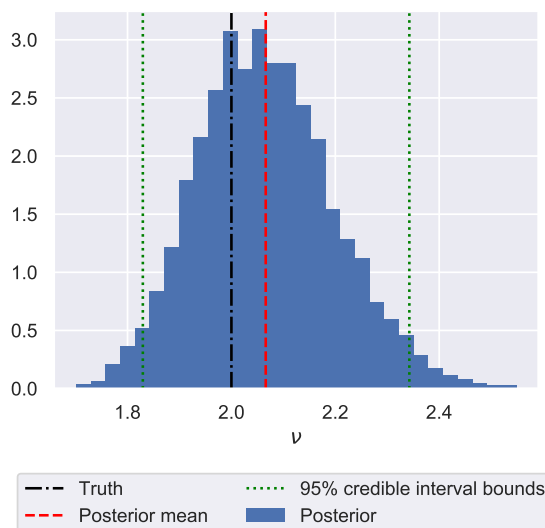


Figure 2: Estimated posterior distribution and true value of the degree of freedom parameter $\nu$ for the MNR model in simulation example I

| Model | Loss |
|---|---|
| MNP | 70.1 |
| MNR | 2.2 |
| Gen-MNR | 4.2 |

Table 1: Quadratic loss in simulation example I

| Scenario | | Model | Alt. 1 | Alt. 2 | Alt. 3 | Alt. 4 |
|---|---|---|---|---|---|---|
| $x_1^{\text{obs}}$ for j = 1 | increased by 10% | Truth | 1.68 | -0.29 | -0.31 | -0.35 |
| | | MNP | 1.60 | -0.26 | -0.29 | -0.33 |
| | | MNR | 1.70 | -0.28 | -0.30 | -0.34 |
| | | Gen-MNR | 1.70 | -0.28 | -0.31 | -0.34 |
| $x_1^{\text{obs}}$ for j = 3 | increased by 10% | Truth | -0.30 | -0.28 | 1.67 | -0.33 |
| | | MNP | -0.29 | -0.26 | 1.56 | -0.31 |
| | | MNR | -0.31 | -0.28 | 1.65 | -0.33 |
| | | Gen-MNR | -0.32 | -0.28 | 1.65 | -0.33 |

Table 2: Aggregate arc elasticities in simulation example I

## 5.2 Example II: Data generated according to Gen-MNR

In the second example, data are generated according to the Gen-MNR model. The data generating process is essentially same as in Example I, with the only difference we allow for different marginal heavy-tailedness by setting $\boldsymbol{\nu} = (5, 3, 1)^\top$.

Figure 3 shows the marginal posterior distributions of the DOF parameters $\nu_1$, $\nu_2$ and $\nu_3$ along with their corresponding true parameter values. It can be seen that Algorithm 3 performs well at recovering the DOF parameters of Gen-MNR, because the true parameter values are contained within the 95% central credible intervals. From Figures 6 and 7 in Appendix B.2, we further conclude that Algorithm 3 also does an excellent job at recovering $\boldsymbol{\beta}$ and $\boldsymbol{\Sigma}$.

Table 3 compares the in-sample fit of MNP, MNR and Gen-MNR in terms of the quadratic loss between the predicted and the true choice probabilities. As expected, Gen-MNR provides the best fit to the data, followed by MNR. Both MNR and Gen-MNR outperform MNP by substantial margins.

Again, we compare the elasticity estimates of the three models in several scenarios. In Table 4, we enumerate the aggregate arc elasticity estimates of the three models along with the corresponding true aggregate arc elasticities for those two scenarios in which we observe the most pronounced differences between the three models. Unsurprisingly, the two selected scenarios involve modifications to attributes of the third alternative, because the heavy-tailedness of the kernel error distribution underlying the considered data generating process is most pronounced in the dimension pertaining to the utility differences between the third alternative and the base alternative. In the first scenario reported in Table 4, we increase the first alternative-specific attribute of the third alternative by 10%. MNR and Gen-MNR yield direct aggregate arc elasticity estimates, which are much closer to the truth than the direct aggregate arc elasticity estimates of MNP. In the considered scenario, the true direct aggregate arc elasticity is 1.25. While MNR and Gen-MNR produce direct elasticity estimates of 1.22 and 1.24, MNP gives a considerably lower direct elasticity estimate of 1.08. In the second scenario reported in Table 4, we increase the second alternative-specific attribute of the third alternative by 10%. In this scenario, only Gen-MNR is able to produce an unbiased direct aggregate arc elasticity estimate. Gen-MNR gives a direct aggregate arc elasticity estimate $-0.65$, which is identical to the ground truth. By contrast, both MNP and MNR produce biased direct aggregate arc elasticity estimates of $-0.70$.
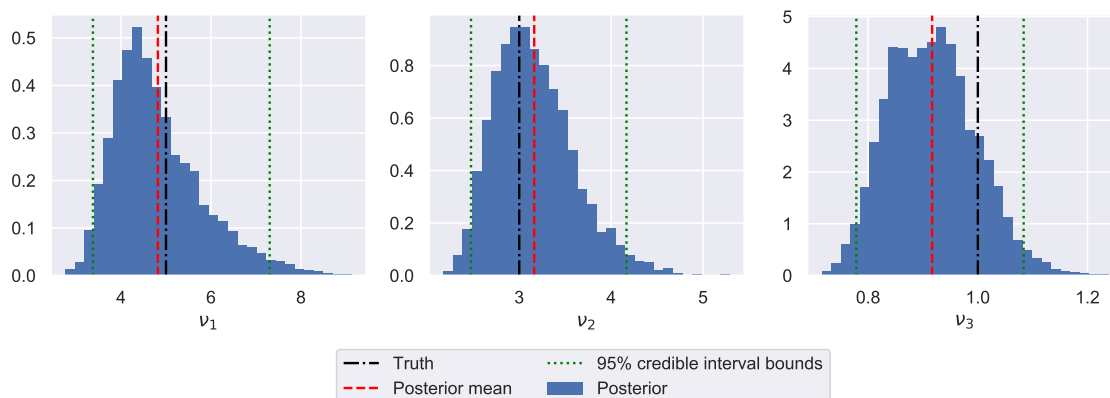
Figure 3: Estimated posterior distributions and true values of the degree of freedom parameters $\nu_1$, $\nu_2$ and $\nu_3$ for Gen-MNR in simulation example II

| Model | Loss |
|---------|------|
| MNP | 79.4 |
| MNR | 16.0 |
| Gen-MNR | 2.5 |

Table 3: Quadratic loss in simulation example II

| Scenario | | Model | Alt. 1 | Alt. 2 | Alt. 3 | Alt. 4 |
|----------|--|-------|--------|--------|--------|--------|
| $x_1^{obs}$ for j = 3 | increased by 10% | Truth | -0.34 | -0.29 | 1.25 | -0.34 |
| | | MNP | -0.31 | -0.26 | 1.08 | -0.30 |
| | | MNR | -0.34 | -0.28 | 1.22 | -0.33 |
| | | Gen-MNR | -0.35 | -0.29 | 1.24 | -0.35 |
| $x_2^{obs}$ for j = 3 | increased by 10% | Truth | 0.18 | 0.15 | -0.65 | 0.18 |
| | | MNP | 0.21 | 0.17 | -0.70 | 0.20 |
| | | MNR | 0.20 | 0.16 | -0.70 | 0.19 |
| | | Gen-MNR | 0.19 | 0.15 | -0.65 | 0.18 |

Table 4: Aggregate arc elasticities in simulation example II

# 6 Case study

In this section, we apply MNP, MNR and Gen-MNR as well as their mixed counterparts M-MNP, M-MNR and M-Gen-MNR in a case study on transport mode choice behaviour.

## 6.1 Data and utility specification

Revealed preference data for the case study are sourced from the London Passenger Mode Choice (LPMC) dataset, which was compiled 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). As the LPMC dataset is a revealed preference mode choice dataset based on a household travel survey, we conjecture that the data contain outlying observations due to aberrant behaviour, misreporting and misclassification.

In this case study, we restrict our analysis to home-based trips reported by individuals who are at least 12 years old. The original dataset contains 58,584 of such observations from 26,904 individuals. For our analysis, we consider 9,658 observations from 4,000 randomly selected individuals for model training. In addition, we randomly select one observation each from an additional set of 1,000 randomly selected individuals for out-of-sample validation. There are four mode choice alternatives, namely walking, cycling, transit and driving. In the training data, the observed market shares are 15.7%, 3.5%, 37.7% and 43.2%, respectively. In the test data, the observed market shares are 15.2%, 3.4%, 40.0% and 41.5%, respectively.

The considered specifications of the systematic utilities are 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).

In the mixed models, tastes with respect to in-vehicle travel time and traffic variability are assumed to be normally distributed. We performed a specification search with a primary focus on tractability to determine which taste parameters to treat as randomly distributed. Our main goal in this case study is to demonstrate that random parameters can be straightforwardly included in MNR and Gen-MNR. We do not seek to find a definitive utility specification for M-MNP, M-MNR and M-Gen-MNR.

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] | | | $\beta_{\text{cost}}$ | $\beta_{\text{cost}}$ |
| Out-of-vehicle time (ovtt) [hours] | $\beta_{\text{ovtt}}$ | $\beta_{\text{ovtt}}$ | $\beta_{\text{ovtt}}$ | |
| In-vehicle travel time (ivtt) [hours] | | | $\beta_{\text{ivtt}}^{(*)}$ | $\beta_{\text{ivtt}}^{(*)}$ |
| No. of transfers | | | $\beta_{\text{transfers}}$ | |
| Traffic variability (tv) | | | | $\beta_{\text{tv}}^{(*)}$ |
| **Individual- and context-specific attributes** | | | | |
| Female traveller | | $\beta_{\text{female, cycle}}$ | $\beta_{\text{female, transit}}$ | $\beta_{\text{female, drive}}$ |
| Traveller age $< 18$ years | | | $\beta_{\text{age}<18\text{ years, transit}}$ | $\beta_{\text{age}<18\text{ years, drive}}$ |
| Traveller age $\geq 65$ years | | | $\beta_{\text{age}\geq65\text{ years, transit}}$ | $\beta_{\text{age}\geq65\text{ years, drive}}$ |
| Travel during winter period (Nov–Mar) | | $\beta_{\text{winter, cycle}}$ | | |
| No. of household cars | | | | $\beta_{\text{cars, drive}}$ |

$^{(*)}$ Parameter assumed to be normally distributed in mixed models.

Table 5: Utility specifications considered in case study

## 6.2 Results

### 6.2.1 In-sample fit and out-of-sample predictive ability

Table 6 compares the in-sample fit and the out-of-sample predictive accuracy of the MNP, MNR and Gen-MNR models as well as their mixed counterparts in terms of the log-likelihood evaluated at the posterior means of the model parameters and the Akaike information criterion (AIC).

First, we consider the performance of MNP, MNR and Gen-MNR. Both MNR and Gen-MNR produce substantially higher log-likelihood values on the training data than MNP. Whereas MNP returns a log-likelihood of $-8295.6$, MNR and Gen-MNR give log-likelihood values of $-8099.7$ and $-8090.0$, respectively, on the training data. Remarkably, with just one more parameter, MNR outperforms MNP by nearly 200 log-likelihood points on the training data. Both the log-likelihood values and AIC indicate that MNR and Gen-MNR also exhibit better out-of-sample predictive than MNP. Having two more parameters than MNR, Gen-MNR produces marginally better log-likelihood values on the training and test data than MNR. Compared to MNR, Gen-MNR gives a lower AIC value on the training data and a negligibly higher AIC value on the training data. In other words, Gen-MNR is able to offset its higher complexity with superior in-sample fit.

Next, we consider the performance of M-MNP, M-MNR and M-Gen-MNR. M-MNR and M-Gen-MNR markedly outperform M-MNP in terms of in-sample fit and also exhibit superior out-of-sample predictive accuracy. With just one additional parameter, M-MNR outperforms M-MNP by more than 100 log-likelihood points on the training data. M-Gen-MNR exhibits better in-sample fit and out-of-sample predictive accuracy than M-MNR due to its more flexible kernel error distribution. Therefore, we conclude that the benefits of the more flexible kernel error distributions underlying the formulations of MNR and Gen-MNR are maintained in the presence of random heterogeneity.

Unsurprisingly, the mixed models produce higher log-likelihood values on the training data than their non-mixed counterparts, since the former account for the serial correla-

tion across choices made by the same individual. However, the log-likelihood values on test data indicate that the mixed models overfit the training data. The observation that the use of mixed random utility models can result in overfitting in applications to panel data is consistent with Cherchi and Cirillo (2010) and Zhao et al. (2020). Similarly, in cross-sectional settings, Wang et al. (2021) find that mixed random utility models offer poorer predictive performance than non-mixed random utility model. The poor predictive performance of the mixed models could be due to the normal heterogeneity distribution. However, recent research suggests that more flexible heterogeneity distributions may not result in substantial gains in predictive accuracy (Krueger et al., 2020).

| Model | No. of parameters | Log-lik. | | AIC | |
|---|---|---|---|---|---|
| | | Train | Test | Train | Test |
| MNP | 23 | -8295.6 | -962.0 | 16637.2 | 1970.1 |
| MNR | 24 | -8099.7 | -955.6 | 16247.4 | 1959.2 |
| Gen-MNR | 26 | -8090.0 | -954.3 | 16232.0 | 1960.6 |
| M-MNP | 25 | -7099.4 | -1055.3 | 14248.8 | 2160.6 |
| M-MNR | 26 | -6982.2 | -1028.1 | 14016.3 | 2108.2 |
| M-Gen-MNR | 28 | -6970.3 | -1027.2 | 13996.6 | 2110.3 |

Table 6: In-sample fit and out-of-sample predictive ability in case study

### 6.2.2 Parameter 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. The parameter estimates for M-MNP, M-MNR and M-Gen-MNR are reported in Appendix C.

First, we examine the estimates of the DOF parameters in MNR and Gen-MNR. For both models, we find evidence of heavy-tailedness in the kernel error distribution. For instance, the posterior mean of the generic DOF parameter $\nu$ of MNR is 2.230, which is indicative of substantial heavy-tailedness in comparison with a Gaussian kernel error distribution (see Figure 1). We detect pronounced and distinct marginal heavy-tailedness in Gen-MNR. Heavy-tailedness is most substantial for the utility differences involving the under-represented walking and cycling alternatives. The posterior mean of the utility differences associated with the walking and driving is 4.677, while the posterior mean of the utility differences associated with the cycling and driving is 1.192. By contrast, tails are only moderately heavy for the utility differences between the transit and the drive alternatives, since the posterior mean of the associated DOF parameter is 18.496. 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.

Next, we compare the estimates of the taste parameters $\beta$. Since the scale of $\beta$ 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). WTP indicators are scale-free and allow for a money-metric representation of sensitivities. We obtain the

posterior distribution of each WTP indicator by taking the ratio of the non-price coefficient of interest and the price coefficient at every posterior sample. Table 8 summarises the posterior distributions of the WTP indicators. We observe that the mean 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, the mean WTP values for reductions in out-of-vehicle time is 32.70 GBP/h, 37.03 GBP/h and 36.17 GBP/h, and the mean WTP value for reductions in in-vehicle time is 19.26 GBP/h, 22.14 GBP/h, 21.54 GBP/h in MNP, MNR and Gen-MNR, respectively. However, the locations of the credible intervals of the WTP indicators suggest that the posterior distributions of the WTP indicators overlap substantially, which in turn implies that the three models do not produce significantly different valuations of in-vehicle and out-of-vehicle travel time.[9] Also, the posterior distributions of the WTP indicators for reductions in transfers and traffic variability resemble each other closely, which suggests that the three models do not produce significantly different valuations of transfers and traffic variability.[10]

The models also provide insights into the influence of individual and context-specific attributes on mode choice propensities. For example, all models suggest that female travellers are relatively less likely to cycle and relatively more likely to use transit, compared to walking. Whereas MNP indicates a statistically significant relationship between female gender and the propensity to use the driving mode over walking, MNR and Gen-MNR do not suggest that this relationship is statistically significant, because the 95% central credible intervals of the respective parameters include zero. In all models, old age increases the propensities to use transit and the driving mode over walking. In addition, all models indicate that travel during the winter months reduces the propensity to cycle, compared to walking. According to all models, higher levels of car ownership increase the propensity to select the driving mode over walking.

---

[9]The WTP estimates for reductions in-vehicle travel time appear somewhat high compared to recommended appraisal values for the UK. According to Batley et al. (2019), the recommended appraisal values of the value of travel time is 11.21 GBP/h for commute trips, 5.12 GBP/h for all other non-work trips, and 18.23 GBP/h for employees' business trips. However, the values reported in Batley et al. (2019) are derived from an entirely different dataset than ours.

[10]As explained in Section 6.1, the attribute traffic variability is a measure of the travel time uncertainty for a given origin-destination pair. Travel time uncertainty is a fundamental determinant of travel behaviour and a pivotal quantity in the economic appraisal of transport projects (Carrion and Levinson, 2012). A traffic variability of zero indicates perfectly dependable travel times, and any value above zero indicates that travel times vary from day to day. For a given best-guess travel time, a higher traffic variability indicates less reliable travel times for the given origin-destination pair. Consequently, WTP for reductions in traffic variability can be interpreted as WTP for improvements in travel time dependability. For example, the Gen-MNR model suggests that on average, travellers are willing to pay 2.12 GBP for a 10% reduction in traffic variability.

---

**Algorithm 2** Gibbs sampler for the MNR model

---

**Step 0:** Initialise parameters $\boldsymbol{w}$, $\mathbf{q}$, $\boldsymbol{\beta}$, $\boldsymbol{\Sigma}$, $\nu$.
**for** $t = 1, \ldots, T$ **do**
    **Step 1:** Update $\mathbf{q}$.
    **for** $i = 1, \ldots, N$ **do**
        Set $\mathbf{z}_i = \boldsymbol{w}_i - \mathbf{X}_i \boldsymbol{\beta}$.
        Draw $q_i|\cdot \sim \chi^2_{\nu+J-1} / \left( \mathbf{z}_i^\top \boldsymbol{\Sigma}^{-1} \mathbf{z}_i \right)$.
    **end for**
    **Step 2:** Update $\nu$.
    Calculate $\alpha^*$, $\beta^*$ as explained in Appendix A.2.
    Draw proposal $\nu' \sim \text{Gamma}(\alpha^*, \beta^*)$.
    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 (13) and (14), respectively.
    **Step 3:** Update $\boldsymbol{w}$, $\alpha^2$.
    **for** $i = 1, \ldots, N$ **do**
        **for** $j = 1, \ldots, J - 1$ **do**
            Draw $w_{ij}|\boldsymbol{w}_{i,-j}, \cdot \sim \text{TN}(\mu_{ij}, \tau^2_{ij})$ as explained in Appendix A.1.
        **end for**
    **end for**
    Draw $\alpha^2|\cdot \sim \text{tr}(\boldsymbol{\Lambda}\boldsymbol{\Sigma}^{-1})/\chi^2_{\rho(J-1)}$.
    Set $\tilde{\boldsymbol{w}} = \alpha \boldsymbol{w}$.
    **Step 4:** Update $\tilde{\boldsymbol{\beta}}$, $\alpha^2$.
    Set $\hat{\mathbf{B}} = \left( \sum_{i=1}^N q_i \mathbf{X}_i^\top \boldsymbol{\Sigma}^{-1} \mathbf{X}_i + \mathbf{B}_0 \right)^{-1}$.
    Set $\hat{\boldsymbol{\beta}} = \hat{\mathbf{B}} \left( \sum_{i=1}^N q_i \mathbf{X}_i^\top \boldsymbol{\Sigma}^{-1} \tilde{\boldsymbol{w}}_i \right)$.
    Draw $\alpha^2|\cdot \sim \left( \sum_{i=1}^N q_i(\tilde{\boldsymbol{w}}_i - \mathbf{X}_i\hat{\boldsymbol{\beta}})^\top \boldsymbol{\Sigma}^{-1}(\tilde{\boldsymbol{w}}_i - \mathbf{X}_i\hat{\boldsymbol{\beta}}) + \hat{\boldsymbol{\beta}}^\top \mathbf{B}_0\hat{\boldsymbol{\beta}} + \text{tr}(\boldsymbol{\Lambda}\boldsymbol{\Sigma}^{-1}) \right) / \chi^2_{(N+\rho)(J-1)}$.
    Draw $\tilde{\boldsymbol{\beta}}|\cdot \sim \text{N}(\hat{\boldsymbol{\beta}}, \alpha^2\hat{\mathbf{B}})$.
    **Step 5:** Update $\tilde{\boldsymbol{\Sigma}}$.
    Set $\tilde{\mathbf{z}}_i = \tilde{\boldsymbol{w}}_i - \mathbf{X}_i\tilde{\boldsymbol{\beta}}$.
    Draw $\tilde{\boldsymbol{\Sigma}}|\cdot \sim \text{IW}\left( \rho + N, \tilde{\boldsymbol{\Lambda}} + \sum_{i=1}^N q_i\mathbf{z}_i\mathbf{z}_i^\top \right)$.
    Set $\alpha^2 = \text{tr}(\tilde{\boldsymbol{\Sigma}})/(J-1)$.
    Set $\boldsymbol{\Sigma} = \tilde{\boldsymbol{\Sigma}}/\alpha^2$, $\boldsymbol{w} = \tilde{\boldsymbol{w}}/\alpha$, $\boldsymbol{\beta} = \tilde{\boldsymbol{\beta}}/\alpha$.
**end for**
**return** $\boldsymbol{\beta}$, $\boldsymbol{\Sigma}$, $\nu$

---

---
**Algorithm 3** Gibbs sampler for the Gen-MNR model
---

**Step 0:** Initialise parameters $w$, $q$, $\beta$, $\Sigma$, $\alpha$, $\nu$.

**for** $t = 1, \ldots, T$ **do**

    **Step 1:** Update $q$.

    **for** $i = 1, \ldots, N$ **do**

        **for** $j = 1, \ldots, J - 1$ **do**

            Calculate $\alpha^*$, $\beta^*$ 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 (19) and (20), respectively.

        **end for**

    **end for**

    **Step 2:** Update $\nu$.

    **for** $j = 1, \ldots, J - 1$ **do**

        Calculate $\alpha^*$, $\beta^*$ 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 $l(\nu_j)$ and $h(\nu_j)$ are defined in (13) and (14), respectively.

    **end for**

    **Step 3:** Update $w$, $\alpha^2$.

    **for** $i = 1, \ldots, N$ **do**

        **for** $j = 1, \ldots, J - 1$ **do**

            Draw $w_{ij}|w_{i,-j}, \cdot \sim \text{TN}(\mu_{ij}, \tau^2_{ij})$ as explained in Appendix A.1.

        **end for**

    **end for**

    Draw $\alpha^2|\cdot \sim \text{tr}(\Lambda\Sigma^{-1})/\chi^2_{\rho(J-1)}$.

    Set $\tilde{w} = \alpha w$.

    **Step 4:** Update $\tilde{\beta}$, $\alpha^2$.

    Set $Q_i = \text{diag}(q_i)$.

    Set $\hat{B} = \left(\sum_{i=1}^{N} X_i^\top Q_i^{1/2}\Sigma^{-1}Q_i^{1/2}X_i + B_0\right)^{-1}$.

    Set $\hat{\beta} = \hat{B}\left(\sum_{i=1}^{N} X_i^\top Q_i^{1/2}\Sigma^{-1}Q_i^{1/2}\tilde{w}_i\right)$.

    Draw $\alpha^2|\cdot \sim \left(\sum_{i=1}^{N}(\tilde{w}_i - X_i\hat{\beta})^\top Q_i^{1/2}\Sigma^{-1}Q_i^{1/2}(\tilde{w}_i - X_i\hat{\beta}) + \hat{\beta}^\top B_0\hat{\beta} + \text{tr}(\Lambda\Sigma^{-1})\right)/\chi^2_{(N+\rho)(J-1)}$.

    Draw $\tilde{\beta}|\cdot \sim N(\hat{\beta}, \alpha^2\hat{B})$.

    **Step 5:** Update $\tilde{\Sigma}$.

    Set $\tilde{z}_i = \tilde{w}_i - X_i\tilde{\beta}$.

    Draw $\tilde{\Sigma}|\cdot \sim IW\left(\rho + N, \tilde{\Lambda} + \sum_{i=1}^{N} Q_i^{1/2}z_iz_i^\top Q_i^{1/2}\right)$.

    Set $\alpha^2 = \text{tr}(\tilde{\Sigma})/(J-1)$.

    Set $\Sigma = \tilde{\Sigma}/\alpha^2$, $w = \tilde{w}/\alpha$, $\beta = \tilde{\beta}/\alpha$.

**end for**

**return** $\beta$, $\Sigma$, $\nu$

---

---

**Algorithm 4** Gibbs sampler for the mixed MNP model

---

**Step 0:** Initialise parameters $w, \beta, \gamma, \eta, \Omega, \lambda, \Sigma, \alpha$.

**for** $t = 1, \ldots, T$ **do**

    **Step 1:** Update $w, \alpha^2$.

    **for** $i = 1, \ldots, N$ **do**

        **for** $j = 1, \ldots, J - 1$ **do**

            Draw $w_{ij}|w_{i,-j}, \cdot \sim \text{TN}(\mu_{ij}, \tau_{ij}^2)$ as explained in Appendix A.1.

        **end for**

    **end for**

    Draw $\alpha^2|\cdot \sim \text{tr}(\Lambda\Sigma^{-1})/\chi_{\rho(J-1)}^2$.

    Set $\tilde{w} = \alpha w, \tilde{\gamma} = \alpha\gamma, \tilde{\eta} = \alpha\eta$.

    **Step 2:** Update $\tilde{\beta}, \alpha^2$.

    Set $\tilde{z}_i = \tilde{w}_i - D_i\tilde{\gamma}_i$.

    Set $\hat{B} = \left( \sum_{i=1}^{N} F_i^\top \Sigma^{-1} F_i + B_0 \right)^{-1}$.

    Set $\hat{\beta} = \hat{B} \left( \sum_{i=1}^{N} F_i^\top \Sigma^{-1} \tilde{z}_i \right)$.

    Draw $\alpha^2|\cdot \sim \left( \sum_{i=1}^{N} (\tilde{z}_i - F_i\hat{\beta})^\top \Sigma^{-1}(\tilde{z}_i - F_i\hat{\beta}) + \hat{\beta}^\top B_0\hat{\beta} + \text{tr}(\Lambda\Sigma^{-1}) \right) /\chi_{(N+\rho)(J-1)}^2$.

    Draw $\tilde{\beta}|\cdot \sim N(\hat{\beta}, \alpha^2\hat{B})$.

    **Step 3:** Update $\tilde{\gamma}, \alpha^2$.

    Set $\tilde{z}_i = \tilde{w}_i - F_i\tilde{\beta}$.

    Set $\hat{C}_i = (D_i^\top \Sigma^{-1} D_i + \Omega^{-1})^{-1}$.

    Set $\hat{\gamma}_i = \hat{C}(D_i^\top \Sigma^{-1} \tilde{z}_i + \Omega^{-1}\tilde{\eta})$.

    Draw $\alpha^2|\cdot \sim \left( \sum_{i=1}^{N} \left( (\tilde{z}_i - D_i\hat{\gamma}_i)^\top \Sigma^{-1}(\tilde{z}_i - D_i\hat{\gamma}_i) + (\hat{\gamma}_i - \tilde{\eta})^\top \Omega^{-1}(\hat{\gamma}_i - \tilde{\eta}) \right) + \text{tr}(\Lambda\Sigma^{-1}) \right) /\chi_{(N+\rho}^2$

    **for** $i = 1, \ldots, N$ **do**

        Draw $\tilde{\gamma}_i|\cdot \sim N(\hat{\gamma}_i, \alpha^2\hat{C}_i)$.

    **end for**

    Set $\gamma = \tilde{\gamma}/\alpha$.

    **Step 4:** Update $\eta$.

    Set $\hat{E} = (N\Omega^{-1} + C_0)^{-1}$.

    Set $\hat{\eta} = \hat{E} \left( \sum_{i=1}^{N} \Omega^{-1}\gamma_i \right)$.

    Draw $\eta|\cdot \sim N(\hat{\eta}, \hat{E})$.

    **Step 5:** Update $\Omega$.

    **for** $l = 1, \ldots, L$ **do**

        Draw $\omega_l|\cdot \sim \text{IG} \left( \frac{\kappa + N}{2}, \kappa\lambda_l + \frac{1}{2} \sum_{i=1}^{N}(\gamma_{il} - \eta_l)^2 \right)$

    **end for**

    Set $\Omega = \text{diag} \left( (\omega_1, \ldots, \omega_L) \right)$.

    **Step 6:** Update $\lambda$.

    **for** $l = 1, \ldots, L$ **do**

        Draw $\lambda_l|\cdot \sim \text{Gamma} \left( \frac{\kappa+1}{2}, \frac{1}{\delta_l^2} + \frac{\kappa}{\omega_l} \right)$.

    **end for**

    **Step 7:** Update $\tilde{\Sigma}$.

    Set $\tilde{z}_i = \tilde{w}_i - F_i\tilde{\beta} - D_i\tilde{\gamma}_i$.

    Draw $\tilde{\Sigma}|\cdot \sim \text{IW} \left( \rho + N, \Lambda + \sum_{i=1}^{N} \tilde{z}_i\tilde{z}_i^\top \right)$.

    Set $\alpha^2 = \text{tr}(\tilde{\Sigma})/(J - 1)$.

    Set $\Sigma = \tilde{\Sigma}/\alpha^2, w = \tilde{w}/\alpha, \beta = \tilde{\beta}/\alpha, \gamma = \tilde{\gamma}/\alpha$.

**end for**

                        20

**return** $\beta, \eta, \Omega, \Sigma$

---

| 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%] |
| $\beta_{asc,\,cycle}$ | -2.139 | 0.123 | -2.410 | -1.928 | -3.348 | 0.359 | -3.883 | -2.448 | -2.050 | 0.169 | -2.366 | -1.737 |
| $\beta_{asc,\,transit}$ | -0.298 | 0.026 | -0.349 | -0.247 | -0.589 | 0.058 | -0.712 | -0.480 | -0.634 | 0.050 | -0.728 | -0.534 |
| $\beta_{asc,\,drive}$ | -0.841 | 0.057 | -0.943 | -0.723 | -1.522 | 0.122 | -1.774 | -1.276 | -1.742 | 0.086 | -1.896 | -1.564 |
| $\beta_{cost}$ | -0.060 | 0.009 | -0.078 | -0.043 | -0.106 | 0.017 | -0.140 | -0.075 | -0.110 | 0.016 | -0.142 | -0.080 |
| $\beta_{ovtt}$ | -1.938 | 0.115 | -2.125 | -1.691 | -3.859 | 0.311 | -4.522 | -3.206 | -3.911 | 0.171 | -4.212 | -3.549 |
| $\beta_{ivtt}$ | -1.142 | 0.084 | -1.304 | -0.971 | -2.310 | 0.212 | -2.726 | -1.871 | -2.329 | 0.133 | -2.589 | -2.072 |
| $\beta_{tv}$ | -1.314 | 0.096 | -1.496 | -1.128 | -2.228 | 0.187 | -2.593 | -1.849 | -2.285 | 0.106 | -2.492 | -2.085 |
| $\beta_{transfers}$ | -0.082 | 0.016 | -0.115 | -0.051 | -0.140 | 0.032 | -0.203 | -0.079 | -0.151 | 0.028 | -0.206 | -0.097 |
| $\beta_{female,\,cycle}$ | -0.447 | 0.073 | -0.595 | -0.310 | -1.580 | 0.346 | -2.270 | -0.896 | -1.882 | 0.568 | -2.876 | -1.002 |
| $\beta_{winter,\,cycle}$ | -0.189 | 0.064 | -0.314 | -0.067 | -0.489 | 0.210 | -0.908 | -0.103 | -0.465 | 0.166 | -0.782 | -0.166 |
| $\beta_{female,\,transit}$ | 0.058 | 0.018 | 0.024 | 0.095 | 0.086 | 0.028 | 0.033 | 0.143 | 0.097 | 0.027 | 0.045 | 0.151 |
| $\beta_{age<18\,years,\,transit}$ | 0.100 | 0.031 | 0.042 | 0.162 | 0.105 | 0.045 | 0.017 | 0.194 | 0.113 | 0.044 | 0.026 | 0.201 |
| $\beta_{age\geq65\,years,\,transit}$ | 0.182 | 0.028 | 0.130 | 0.238 | 0.270 | 0.043 | 0.190 | 0.361 | 0.260 | 0.041 | 0.179 | 0.338 |
| $\beta_{female,\,drive}$ | 0.056 | 0.026 | 0.006 | 0.107 | 0.016 | 0.043 | -0.069 | 0.102 | 0.073 | 0.043 | -0.010 | 0.155 |
| $\beta_{age<18\,years,\,drive}$ | -0.385 | 0.051 | -0.483 | -0.284 | -0.753 | 0.094 | -0.943 | -0.580 | -0.755 | 0.071 | -0.900 | -0.624 |
| $\beta_{age\geq65\,years,\,drive}$ | 0.182 | 0.037 | 0.110 | 0.257 | 0.286 | 0.062 | 0.168 | 0.413 | 0.283 | 0.063 | 0.165 | 0.408 |
| $\beta_{cars,\,drive}$ | 0.571 | 0.036 | 0.500 | 0.637 | 0.967 | 0.073 | 0.817 | 1.113 | 1.047 | 0.034 | 0.980 | 1.114 |
| $\Sigma_{walk-drive,\,walk-drive}$ | 0.735 | 0.091 | 0.569 | 0.912 | 1.185 | 0.122 | 0.933 | 1.403 | 1.633 | 0.062 | 1.500 | 1.741 |
| $\Sigma_{walk-drive,\,cycle-drive}$ | 0.051 | 0.131 | -0.254 | 0.278 | -0.527 | 0.313 | -0.952 | 0.273 | 0.295 | 0.114 | 0.089 | 0.555 |
| $\Sigma_{walk-drive,\,transit-drive}$ | 0.466 | 0.057 | 0.363 | 0.577 | 0.870 | 0.094 | 0.669 | 1.048 | 1.370 | 0.045 | 1.252 | 1.430 |
| $\Sigma_{cycle-drive,\,cycle-drive}$ | 1.883 | 0.133 | 1.628 | 2.129 | 1.091 | 0.198 | 0.724 | 1.513 | 0.158 | 0.086 | 0.051 | 0.382 |
| $\Sigma_{cycle-drive,\,transit-drive}$ | 0.378 | 0.097 | 0.124 | 0.518 | -0.290 | 0.282 | -0.653 | 0.429 | 0.262 | 0.088 | 0.092 | 0.450 |
| $\Sigma_{transit-drive,\,transit-drive}$ | 0.382 | 0.047 | 0.295 | 0.468 | 0.724 | 0.080 | 0.547 | 0.880 | 1.210 | 0.068 | 1.067 | 1.322 |
| $\nu$ | | | | | 2.230 | 0.239 | 1.772 | 2.717 | | | | |
| $\nu_{walk-drive}$ | | | | | | | | | 4.677 | 0.553 | 3.733 | 5.944 |
| $\nu_{cycle-drive}$ | | | | | | | | | 1.192 | 0.229 | 0.875 | 1.723 |
| $\nu_{transit-drive}$ | | | | | | | | | 18.496 | 8.192 | 9.191 | 40.983 |

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

| Attribute | Model | Mean | Std. dev. | [0.025% | 0.975%] |
|---|---|---|---|---|---|
| Out-of-vehicle travel time [GBP/h] | MNP | 32.70 | 4.58 | 25.34 | 43.75 |
| | MNR | 37.03 | 5.43 | 28.60 | 49.42 |
| | Gen-MNR | 36.17 | 5.30 | 27.85 | 48.62 |
| In-vehicle travel time [GBP/h] | MNP | 19.26 | 2.80 | 14.73 | 25.68 |
| | MNR | 22.14 | 3.26 | 17.07 | 29.36 |
| | Gen-MNR | 21.54 | 3.26 | 16.35 | 29.19 |
| Transfers [GBP/interchange] | MNP | 1.40 | 0.40 | 0.75 | 2.30 |
| | MNR | 1.36 | 0.41 | 0.66 | 2.26 |
| | Gen-MNR | 1.41 | 0.40 | 0.76 | 2.34 |
| Traffic variability [GBP] | MNP | 22.19 | 3.35 | 16.76 | 30.03 |
| | MNR | 21.41 | 3.38 | 16.13 | 29.11 |
| | Gen-MNR | 21.16 | 3.29 | 15.97 | 28.46 |

Table 8: Willingness-to-pay in case study

### 6.2.3 Elasticity estimates

Table 9 enumerates the aggregate arc elasticity estimates of MNP, MNR and Gen-MNR for various policy-relevant scenarios. In what follows, we highlight several noteworthy differences in the elasticity estimates produced by each of the three models. For example, the demand for walking is estimated to be more elastic to variations in walking time in MNR and Gen-MNR than in MNP. The aggregate arc elasticity of walking demand for a 10% reduction in walking time is $-1.77$ and $-1.78$ in MNR and Gen-MNR, respectively, but is only $-1.57$ in MNP. Innovations such as fast-moving walkways (e.g. Scarinci et al., 2017) offer widespread reductions in walking times. A policy to support such walkways is more compelling under the elasticity estimates of MNR and Gen-MNR. The three models also produce different elasticity estimates for changes in cycling travel times. Whereas MNP and Gen-MNR give elasticity estimates of $-0.97$ and $-0.86$, MNR suggests a considerably higher elasticity estimate of $-0.68$. 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) can result in widespread decreases in cycling travel times. Compared to MNP, Gen-MNR suggests a more moderate effect of such interventions on cycling demand. Likewise, Gen-MNR indicates a weaker effect than MNP but a stronger effect than MNR of changes in transit out-of-vehicle and in-vehicle travel times on cycling demand. In sum, the elasticity estimates reveal interesting differences in how MNP, MNR and Gen-MNR capture demand sensitivities. Our analysis of the elasticity estimates aims to illustrate how different kernel error distribution can produce different and potentially erroneous policy recommendations.

| Scenario | | | Model | Walk | Cycle | Transit | Drive |
|---|---|---|---|---|---|---|---|
| Cycling out-of-vehicle travel time | decreased by 10% | | MNP | 0.00 | -0.97 | 0.06 | 0.02 |
| | | | MNR | 0.00 | -0.68 | 0.02 | 0.03 |
| | | | Gen-MNR | 0.01 | -0.86 | 0.04 | 0.02 |
| Walking out-of-vehicle travel time | decreased by 10% | | MNP | -1.57 | 0.03 | 0.52 | 0.16 |
| | | | MNR | -1.77 | 0.04 | 0.58 | 0.16 |
| | | | Gen-MNR | -1.78 | 0.16 | 0.58 | 0.16 |
| Driving cost | increased by 10% | | MNP | 0.01 | 0.02 | 0.05 | -0.05 |
| | | | MNR | 0.01 | 0.02 | 0.05 | -0.04 |
| | | | Gen-MNR | 0.01 | 0.03 | 0.05 | -0.05 |
| Driving in-vehicle travel time | increased by 10% | | MNP | 0.06 | 0.12 | 0.30 | -0.30 |
| | | | MNR | 0.07 | 0.21 | 0.34 | -0.33 |
| | | | Gen-MNR | 0.07 | 0.15 | 0.33 | -0.33 |
| Driving traffic variability | decreased by 10% | | MNP | 0.14 | 0.19 | 0.38 | -0.40 |
| | | | MNR | 0.13 | 0.29 | 0.36 | -0.38 |
| | | | Gen-MNR | 0.13 | 0.19 | 0.35 | -0.38 |
| Transit fares | increased by 10% | | MNP | 0.08 | 0.10 | -0.13 | 0.08 |
| | | | MNR | 0.08 | 0.03 | -0.12 | 0.08 |
| | | | Gen-MNR | 0.09 | 0.07 | -0.13 | 0.07 |
| Transit in-vehicle travel time | decreased by 10% | | MNP | 0.19 | 0.29 | -0.40 | 0.26 |
| | | | MNR | 0.21 | 0.10 | -0.43 | 0.29 |
| | | | Gen-MNR | 0.21 | 0.19 | -0.43 | 0.29 |
| Transit out-of-vehicle travel time | decreased by 10% | | MNP | 0.34 | 0.34 | -0.45 | 0.24 |
| | | | MNR | 0.43 | 0.10 | -0.49 | 0.26 |
| | | | Gen-MNR | 0.43 | 0.21 | -0.50 | 0.26 |

Table 9: Aggregate arc elasticities for MNP, MNR and Gen-MNR in case study

# 7 Conclusion

Models which are robust to the influence of outlying observations due to aberrant behaviour, misreporting and misclassification 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. Both alternative models belong to the family of robit whose kernel error distributions are heavy-tailed t-distributions that moderate the influence of outlying observations. The first model is the multinomial robit (MNR) model, in which a single, generic degrees of freedom parameter controls the heavy-tailedness of the kernel error distribution. The second model, the generalised multinomial robit (Gen-MNR) model, is based on the non-elliptically contoured multivariate t distribution, which allows for distinct heavy-tailedness in each dimension of the kernel error distribution. For both models, we devise gradient- and optimisation-free Gibbs samplers, which also allow for a straightforward inclusion of random taste parameters.

We contrast MNP, MNR and Gen-MNR as well as their mixed counterparts with random parameters, M-MNP, M-MNR and M-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 produce more exact elasticity estimates if the kernel error distribution underlying the data generating process is heavy-tailed, i.e. if the choice data contain outliers through the lens of a non-robust MNP model. In the case study, we demonstrate that both MNR and Gen-MNR outperform MNP by considerable margins in terms of in-sample fit and out-of-sample predictive accuracy. Gen-MNR delivers the best in-sample fit and out-of-sample predictive accuracy due to its more flexible kernel error distribution. We also find that M-MNR and M-Gen-MNR offer substantially better in-sample fit and out-of-sample predictive accuracy than M-MNP. In other words, the benefits of the more flexible kernel error distributions underlying MNR and Gen-MNR persist in the presence of random heterogeneity.

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 analysis suggests several directions for future research. First, advances in Bayesian machine learning can be leveraged to improve the explanatory and predictive powers of MNR and Gen-MNR. For example, Dirichlet process mixtures can be used to accommodate flexible semi-parametric parametric representations of unobserved heterogeneity (Krueger et al., 2020). Similarly, the horseshoe prior (Carvalho et al., 2010) can be employed to perform variable selection in large predictor spaces. Furthermore, the systematic parts of the latent utility differences could be represented using Bayesian additive regression trees (BART), which automatically partition large predictor spaces to capture interaction effects and nonlinearities (Chipman et al., 2010, Kindo et al., 2016). As these extensions are rooted in the Bayesian inferential paradigm, they can be incorporated into MNR and Gen-MNR with relative ease. A second direction for future research is to use Bayesian modelling to automate aspects of the specification of MNR and Gen-MNR. For example, Burgette et al. (2020) propose a symmetric prior, which obviates the specification of a base alternative, for MNP kernel error covariances. Such symmetric priors can also be developed for MNR and Gen-MNR. Similarly, more diffuse prior formulations for the covariance of kernel error distributions of MNP, MNR and Gen-MNR can be investigated along the lines of Huang et al. (2013). In addition, a prior can be designed for Gen-MNR to automatically infer a parsimonious mapping of degrees of freedom parameters onto utility differences. Finally, a third direction for future research is to formulate robust discrete choice models based on skew-t-distributions, which can also capture asymmetric kernel 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.
MB: conception and design, manuscript editing, supervision.
TG: conception and design, method development and implementation, manuscript writing

and editing.
PB: conception and design, 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.

Bansal, P., Daziano, R. A., and Achtnicht, M. (2018). Extending the logit-mixed logit model for a combination of random and fixed parameters. *Journal of choice modelling*, 27:88–96.

Batley, R., Bates, J., Bliemer, M., Börjesson, M., Bourdon, J., Cabral, M. O., Chintakayala, P. K., Choudhury, C., Daly, A., Dekker, T., et al. (2019). New appraisal values of travel time saving and reliability in great britain. *Transportation*, 46(3):583–621.

Benoit, D. F., Van Aelst, S., and Van den Poel, D. (2016). Outlier-robust bayesian multinomial choice modeling. *Journal of Applied Econometrics*, 31(7):1445–1466.

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

Bhat, C. R. (1995). A heteroscedastic extreme value model of intercity travel mode choice. *Transportation Research Part B: Methodological*, 29(6):471–483.

Bhat, C. 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):923–939.

Bhat, C. R. and Lavieri, P. S. (2018). A new mixed mnp model accommodating a variety of dependent non-normal coefficient distributions. *Theory and Decision*, 84(2):239–275.

Botev, Z. I. and l'Ecuyer, P. (2016). Simulation from the normal distribution truncated to an interval in the tail. In *VALUETOOLS*.

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

Bunch, D. S. (1991). Estimability in the multinomial probit model. *Transportation Research Part B: Methodological*, 25(1):1–12.

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.

Burgette, L. F., Puelz, D., Hahn, P. R., et al. (2020). A symmetric prior for multinomial probit models. *Bayesian Analysis*.

Carrion, C. and Levinson, D. (2012). Value of travel time reliability: A review of current evidence. *Transportation research part A: policy and practice*, 46(4):720–741.

Carvalho, C. M., Polson, N. G., and Scott, J. G. (2010). The horseshoe estimator for sparse signals. *Biometrika*, 97(2):465–480.

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.

Cherchi, E. and Cirillo, C. (2010). Validation and forecasts in models estimated from multiday travel survey. *Transportation research record*, 2175(1):57–64.

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.

Daganzo, C. (1979). *Multinomial Probit*. Academic Press.

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.

Fosgerau, M. and Mabit, S. L. (2013). Easy and flexible mixture distributions. *Economics Letters*, 120(2):206–210.

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. et al. (2006). Prior distributions for variance parameters in hierarchical models (comment on article by browne and draper). *Bayesian analysis*, 1(3):515–534.

Gelman, A. and Hill, J. (2006). *Data analysis using regression and multilevel/hierarchical models*. Cambridge university press.

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

Geweke, J., Keane, M., and Runkle, D. (1994). Alternative computational approaches to inference in the multinomial probit model. *The review of economics and statistics*, pages 609–632.

Hajivassiliou, V., McFadden, D., and Ruud, P. (1996). Simulation of multivariate normal rectangle probabilities and their derivatives theoretical and computational results. *Journal of econometrics*, 72(1-2):85–134.

Hausman, J. A., Abrevaya, J., and Scott-Morton, F. M. (1998). Misclassification of the dependent variable in a discrete-response setting. *Journal of econometrics*, 87(2):239–269.

Hess, S. and Rose, J. M. (2012). Can scale and coefficient heterogeneity be separated in random coefficients models? *Transportation*, 39(6):1225–1239.

Hess, S. and Train, K. (2017). Correlation and scale in mixed logit models. *Journal of choice modelling*, 23:1–8.

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.

Huang, A., Wand, M. P., et al. (2013). Simple marginally noninformative prior distributions for covariance matrices. *Bayesian Analysis*, 8(2):439–452.

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.

Lerman, S. and Manski, C. (1981). On the use of simulated frequencies to approximate choice probabilities. *Structural analysis of discrete data with econometric applications*, 10:305–319.

Liang, X., Zhang, Y., Wang, G., and Xu, S. (2019). A deep learning model for transportation mode detection based on smartphone sensing data. *IEEE Transactions on Intelligent Transportation Systems*, 21(12):5223–5235.

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. (1978). Modeling the choice of residential location. *Transportation Research Record*, (673).

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.

Paleti, R. and Balan, L. (2019). Misclassification in travel surveys and implications to choice modeling: application to household auto ownership decisions. *Transportation*, 46(4):1467–1485.

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.

Train, K. E. (2009). *Discrete choice methods with simulation*. Cambridge University Press.

Van Dyk, D. A. (2010). Marginal markov chain monte carlo methods. *Statistica Sinica*, pages 1423–1454.

Van Dyk, D. A. and Meng, X.-L. (2001). The art of data augmentation. *Journal of Computational and Graphical Statistics*, 10(1):1–50.

Vij, A. and Krueger, R. (2017). Random taste heterogeneity in discrete choice models: Flexible nonparametric finite mixture distributions. *Transportation Research Part B: Methodological*, 106:76–101.

Vij, A. and Shankari, K. (2015). When is big data big enough? implications of using gps-based surveys for travel demand analysis. *Transportation Research Part C: Emerging Technologies*, 56:446–462.

Walker, J. L., Ben-Akiva, M., and Bolduc, D. (2007). Identification of parameters in normal error component logit-mixture (neclm) models. *Journal of Applied Econometrics*, 22(6):1095–1125.

Wand, M. P., Ormerod, J. T., Padoan, S. A., Frühwirth, R., et al. (2011). Mean field variational bayes for elaborate distributions. *Bayesian Analysis*, 6(4):847–900.

Wang, S., Mo, B., Hess, S., and Zhao, J. (2021). Comparing hundreds of machine learning classifiers and discrete choice models in predicting travel behavior: an empirical benchmark. *arXiv preprint arXiv:2102.01130*.

Zhao, X., Yan, X., Yu, A., and Van Hentenryck, P. (2020). Prediction and behavioral analysis of travel mode choice: A comparison of machine learning and logit models. *Travel behaviour and society*, 20:22–35.

# A  Gibbs sampling details

## A.1  Sampling $w$

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

$$w_{ij} \sim TN(\mu_{ij}, \tau_{ij}^2), \quad \text{for } i = 1, \ldots, N, j = 1, \ldots, J-1. \tag{11}$$

For MNP, $\mu_{ij} = X_{ij}^\top \beta + \Sigma_{j,-j} \Sigma_{-j,-j}^{-1}(w_{i,-j} - X_{i,-j}\beta)$ and $\tau_{ij}^2 = \Sigma_{jj} - \Sigma_{j,-j}\Sigma_{-j,-j}^{-1}\Sigma_{-j,j}$. For MNR, $\mu_{ij} = X_{ij}^\top \beta + \Sigma_{j,-j}\Sigma_{-j,-j}^{-1}(w_{i,-j} - X_{i,-j}\beta)$ and $\tau_{ij}^2 = (\Sigma_{jj} - \Sigma_{j,-j}\Sigma_{-j,-j}^{-1}\Sigma_{-j,j})/q_i$. For Gen-MNR, $\mu_{ij} = X_{ij}^\top \beta + Q_{ijj}^{-1/2}\Sigma_{j,-j}\Sigma_{-j,-j}^{-1}Q_{i,-j,-j}^{1/2}(w_{i,-j} - X_{i,-j}\beta)$ and $\tau_{ij}^2 = (\Sigma_{jj} - \Sigma_{j,-j}\Sigma_{-j,-j}^{-1}\Sigma_{-j,j})/q_{ij}$. Here, the index $-l$ denotes the vector without the $l$th 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 $\nu$

The full conditional distribution of $\nu$ is nonstandard. Ding (2014) shows that

$$p(\nu|\cdot) \propto \exp\left\{\frac{N\nu}{2}\log\left(\frac{\nu}{2}\right) - N\log\Gamma\left(\frac{\nu}{2}\right) + (\alpha_0 - 1)\log\nu - \xi\nu\right\}, \tag{12}$$

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 (12) using a Metropolised Independence sampler (Liu, 2008) with an approximate Gamma proposal. The shape parameter $\alpha^*$ and the rate parameter $\beta^*$ of the proposal density are obtained as follows. The log conditional density of $\nu$ up to an additive constant is

$$l(\nu) = \frac{N\nu}{2}\log\left(\frac{\nu}{2}\right) - N\log\Gamma\left(\frac{\nu}{2}\right) + (\alpha_0 - 1)\log\nu - \xi\nu. \tag{13}$$

The log density of the Gamma proposal is

$$h(\nu) = (\alpha^* - 1)\log\nu - \beta^*\nu. \tag{14}$$

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

$$l'(\nu) = \frac{N}{2}\left[\log\left(\frac{\nu}{2}\right) + 1 - \psi\left(\frac{\nu}{2}\right)\right] + \frac{\alpha_0 - 1}{\nu} - \xi, \quad h'(\nu) = \frac{\alpha^* - 1}{\nu} - \beta^*, \tag{15}$$

$$l''(\nu) = \frac{N}{2}\left[\frac{1}{\nu} - \frac{1}{2}\psi'\left(\frac{\nu}{2}\right)\right] + \frac{\alpha_0 - 1}{\nu^2}, \quad h''(\nu) = -\frac{\alpha^* - 1}{\nu^2}, \tag{16}$$

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 $\nu^*$ 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

$$\alpha^* = 1 - (\nu^*)^2 l^*, \quad \beta^* = -\nu^* l^*. \tag{17}$$

## A.3  Sampling $q_{ij}$

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

$$p(q_{ij}|\cdot) \propto \exp\left\{-\frac{q_{ij}u_{ij}}{2} - \sqrt{q_{ij}}c_{ij} + \frac{\nu_j - 1}{2}\log q_{ij}\right\},\tag{18}$$

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 (18) using a Metropolised Independence sampler (Liu, 2008) with an approximate Gamma proposal. The shape parameter $\alpha^*$ and the rate parameter $\beta^*$ of the proposal density are obtained as follows. For $\nu_j \leq 1$, we set $\alpha^* = 1$ and $\beta^* = \frac{u_{ij}}{2}$. For $\nu_j > 1$, $\alpha^*$ and $\beta^*$ 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

$$f(q_{ij}) = -\frac{q_{ij}u_{ij}}{2} - \sqrt{q_{ij}}c_{ij} + \frac{\nu_j - 1}{2}\log q_{ij}.\tag{19}$$

The log density of the Gamma proposal is

$$g(q_{ij}) = (\alpha^* - 1)\log q_{ij} - \beta^* q_{ij}.\tag{20}$$

The mode of (20) 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 (19) are

$$f'(q_{ij}) = -\frac{u_{ij}}{2} - \frac{c_{ij}}{2\sqrt{q_{ij}}} + \frac{\nu_j - 1}{2q_{ij}}, \quad f''(q_{ij}) = \frac{c_{ij}}{4\sqrt{q_{ij}^3}} - \frac{\nu_j - 1}{2q_{ij}^2}.\tag{21}$$

The mode of (19) 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

$$\alpha^* = 1 - (m_{ij}^*)^2 l_{ij}^*, \quad \beta^* = -m_{ij}^* l_{ij}^*.\tag{22}$$

# B    Additional results for the simulation study

## B.1    Example I



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

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

## B.2 Example II
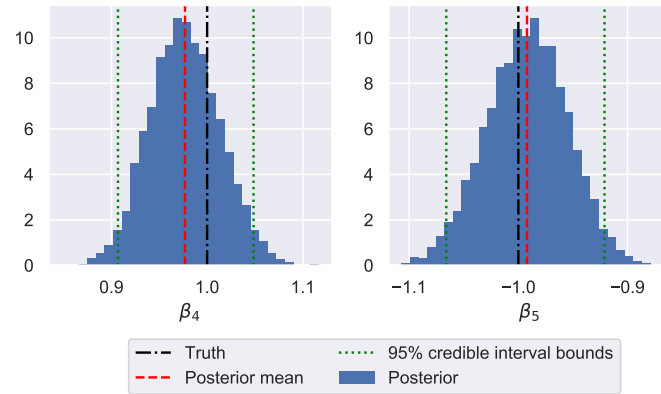


Figure 6: Estimated posterior distribution and true values of the taste parameters $\{\beta_4, \beta_5\}$ for the Gen-MNR model in simulation example II
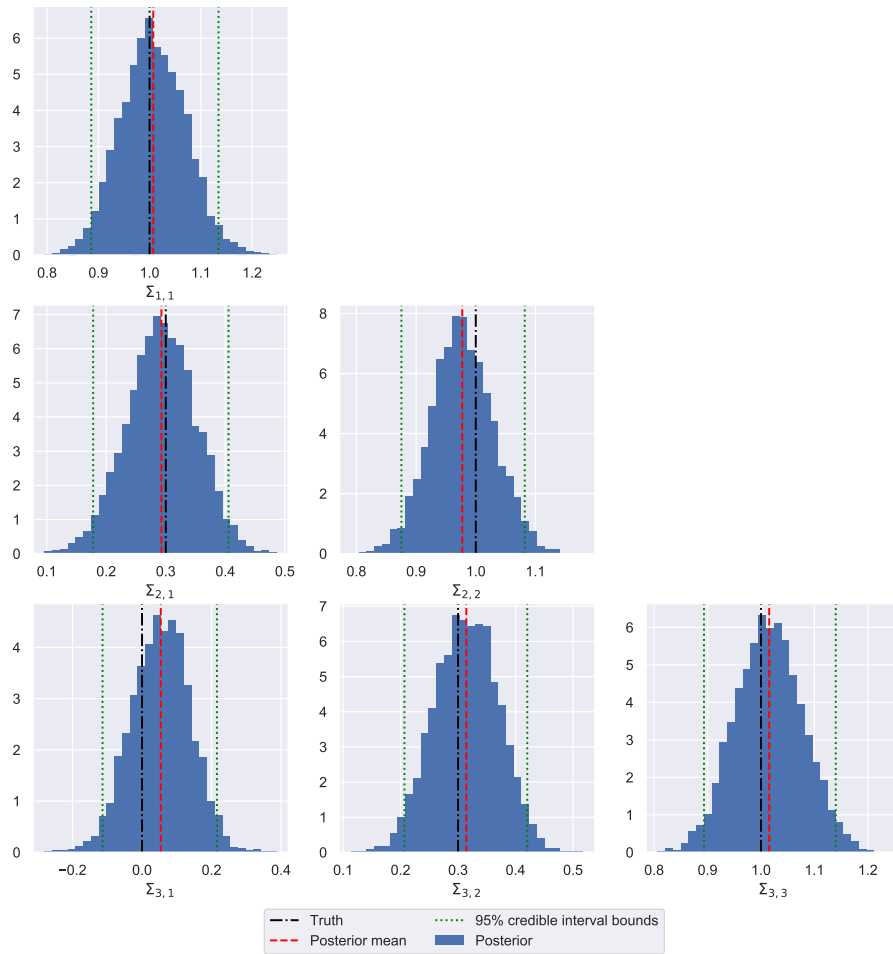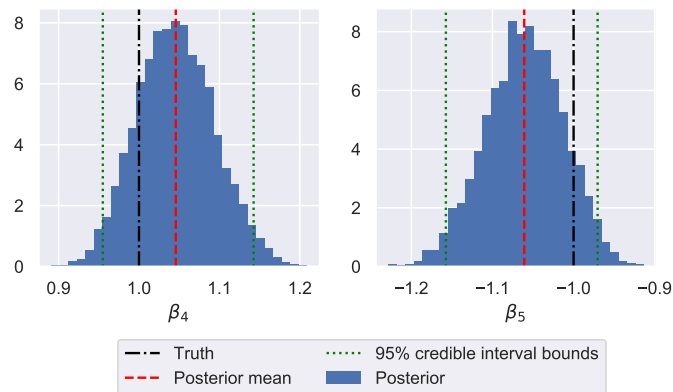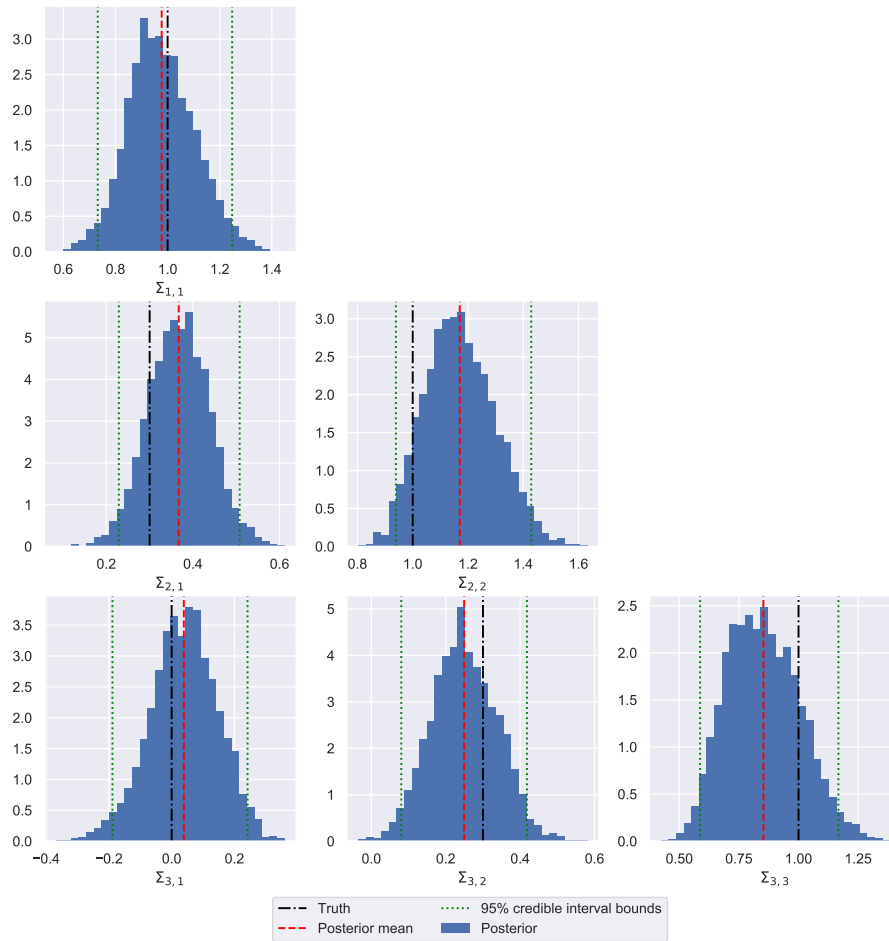
Figure 7: Estimated posterior distribution and true values of the unique elements of the covariance matrix $\Sigma$ for the Gen-MNR model in simulation example II

# C    Parameters estimates of M-MNP, M-MNR and M-Gen-MNR in case study

| Parameter | M-MNP | | | | M-MNR | | | | M-Gen-MNR | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Mean | Std. dev. | [0.025% | 0.975%] | Mean | Std. dev. | [0.025% | 0.975%] | Mean | Std. dev. | [0.025% | 0.975%] |
| $\beta_{asc,\,cycle}$ | -3.342 | 0.113 | -3.561 | -3.115 | -5.540 | 0.530 | -6.813 | -4.676 | -3.827 | 0.194 | -4.231 | -3.456 |
| $\beta_{asc,\,transit}$ | -0.684 | 0.122 | -0.956 | -0.480 | -1.413 | 0.186 | -1.822 | -1.127 | -0.890 | 0.181 | -1.274 | -0.528 |
| $\beta_{asc,\,drive}$ | -1.589 | 0.257 | -2.177 | -1.157 | -3.519 | 0.476 | -4.574 | -2.674 | -2.161 | 0.442 | -3.118 | -1.323 |
| $\beta_{cost}$ | -0.087 | 0.019 | -0.125 | -0.050 | -0.194 | 0.047 | -0.294 | -0.112 | -0.117 | 0.033 | -0.192 | -0.060 |
| $\beta_{ovtt}$ | -3.689 | 0.417 | -4.601 | -3.042 | -8.452 | 0.916 | -10.450 | -6.822 | -5.124 | 0.693 | -6.777 | -3.730 |
| $\beta_{transfers}$ | -0.153 | 0.045 | -0.247 | -0.071 | -0.368 | 0.099 | -0.574 | -0.156 | -0.215 | 0.080 | -0.361 | -0.076 |
| $\beta_{female,\,cycle}$ | -0.593 | 0.114 | -0.813 | -0.367 | -1.870 | 0.411 | -2.785 | -1.190 | -0.743 | 0.129 | -1.002 | -0.499 |
| $\beta_{winter,\,cycle}$ | -0.291 | 0.101 | -0.488 | -0.087 | -0.797 | 0.241 | -1.323 | -0.358 | -0.323 | 0.119 | -0.559 | -0.095 |
| $\beta_{female,\,transit}$ | 0.090 | 0.028 | 0.037 | 0.149 | 0.177 | 0.058 | 0.068 | 0.296 | 0.123 | 0.039 | 0.050 | 0.206 |
| $\beta_{age<18\,years,\,transit}$ | 0.145 | 0.055 | 0.046 | 0.260 | 0.269 | 0.104 | 0.068 | 0.480 | 0.169 | 0.074 | 0.036 | 0.325 |
| $\beta_{age\geq65\,years,\,transit}$ | 0.236 | 0.061 | 0.130 | 0.372 | 0.481 | 0.101 | 0.309 | 0.701 | 0.302 | 0.082 | 0.143 | 0.464 |
| $\beta_{female,\,drive}$ | 0.155 | 0.048 | 0.067 | 0.259 | 0.363 | 0.098 | 0.184 | 0.578 | 0.230 | 0.078 | 0.111 | 0.406 |
| $\beta_{age<18\,years,\,drive}$ | -0.450 | 0.097 | -0.663 | -0.285 | -0.980 | 0.170 | -1.341 | -0.666 | -0.618 | 0.131 | -0.887 | -0.384 |
| $\beta_{age\geq65\,years,\,drive}$ | 0.175 | 0.062 | 0.064 | 0.308 | 0.391 | 0.128 | 0.139 | 0.654 | 0.238 | 0.096 | 0.068 | 0.445 |
| $\beta_{cars,\,drive}$ | 0.796 | 0.130 | 0.575 | 1.094 | 1.723 | 0.219 | 1.314 | 2.204 | 1.043 | 0.215 | 0.625 | 1.485 |
| $\eta_{ivtt}$ | -1.801 | 0.332 | -2.464 | -1.149 | -4.736 | 1.104 | -7.534 | -3.169 | -2.709 | 0.661 | -4.215 | -1.672 |
| $\eta_{ltv}$ | -2.765 | 0.457 | -3.696 | -1.999 | -5.975 | 0.828 | -7.771 | -4.604 | -3.583 | 0.795 | -5.154 | -2.162 |
| $\omega_{ivtt}$ | 3.889 | 0.438 | 3.098 | 4.785 | 7.894 | 0.928 | 6.295 | 9.678 | 4.735 | 0.666 | 3.408 | 6.064 |
| $\omega_{tv}$ | 4.676 | 0.831 | 3.354 | 6.483 | 10.001 | 1.283 | 7.700 | 12.676 | 6.085 | 1.372 | 3.551 | 8.773 |
| $\Sigma_{walk-drive,\,walk-drive}$ | 0.275 | 0.094 | 0.145 | 0.523 | 0.594 | 0.123 | 0.402 | 0.906 | 0.258 | 0.103 | 0.081 | 0.497 |
| $\Sigma_{walk-drive,\,cycle-drive}$ | -0.466 | 0.115 | -0.695 | -0.225 | -0.604 | 0.186 | -0.915 | -0.252 | -0.377 | 0.139 | -0.637 | -0.084 |
| $\Sigma_{walk-drive,\,transit-drive}$ | 0.134 | 0.046 | 0.066 | 0.250 | 0.287 | 0.064 | 0.184 | 0.446 | 0.185 | 0.061 | 0.072 | 0.308 |
| $\Sigma_{cycle-drive,\,cycle-drive}$ | 2.579 | 0.141 | 2.218 | 2.776 | 2.142 | 0.160 | 1.722 | 2.390 | 2.552 | 0.152 | 2.224 | 2.836 |
| $\Sigma_{cycle-drive,\,transit-drive}$ | 0.058 | 0.137 | -0.164 | 0.353 | 0.061 | 0.172 | -0.252 | 0.346 | -0.039 | 0.188 | -0.468 | 0.274 |
| $\Sigma_{transit-drive,\,transit-drive}$ | 0.146 | 0.050 | 0.076 | 0.270 | 0.264 | 0.045 | 0.190 | 0.373 | 0.190 | 0.059 | 0.076 | 0.308 |
| $\nu$ | | | | | 2.596 | 0.411 | 2.013 | 3.562 | | | | |
| $\nu_{walk-drive}$ | | | | | | | | | 2.768 | 0.367 | 2.185 | 3.651 |
| $\nu_{cycle-drive}$ | | | | | | | | | 20.470 | 10.701 | 8.083 | 54.315 |
| $\nu_{transit-drive}$ | | | | | | | | | 14.940 | 10.611 | 4.378 | 41.846 |

Table 10: Estimated parameters of M-MNP, M-MNR and M-Gen-MNR in case study