

## Heterogeneity in general multinomial choice models

Mauerer, Ingrid; Tutz, Gerhard

Veröffentlichungsversion / Published Version

Zeitschriftenartikel / journal article

### Empfohlene Zitierung / Suggested Citation:

Mauerer, I., & Tutz, G. (2023). Heterogeneity in general multinomial choice models. *Statistical Methods & Applications*, 32(1), 129-148. <https://doi.org/10.1007/s10260-022-00642-5>

### Nutzungsbedingungen:

Dieser Text wird unter einer CC BY Lizenz (Namensnennung) zur Verfügung gestellt. Nähere Auskünfte zu den CC-Lizenzen finden Sie hier:

<https://creativecommons.org/licenses/by/4.0/deed.de>

### Terms of use:

This document is made available under a CC BY Licence (Attribution). For more information see:

<https://creativecommons.org/licenses/by/4.0>



# Heterogeneity in general multinomial choice models

Ingrid Mauerer<sup>1</sup> · Gerhard Tutz<sup>2</sup>

Accepted: 18 April 2022 / Published online: 25 May 2022  
© The Author(s) 2022

## Abstract

Different voters behave differently at the polls, different students make different university choices, or different countries choose different health care systems. Many research questions important to social scientists concern choice behavior, which involves dealing with nominal dependent variables. Drawing on the principle of maximum random utility, we propose applying a flexible and general heterogeneous multinomial logit model to study differences in choice behavior. The model systematically accounts for heterogeneity that classical models do not capture, indicates the strength of heterogeneity, and permits examining which explanatory variables cause heterogeneity. As the proposed approach allows incorporating theoretical expectations about heterogeneity into the analysis of nominal dependent variables, it can be applied to a wide range of research problems. Our empirical example uses individual-level survey data to demonstrate the benefits of the model in studying heterogeneity in electoral decisions.

**Keywords** Categorical dependent variable · Heterogeneity · Multinomial logit model · Discrete choice analysis · Random utility maximization · Electoral decisions

## 1 Introduction

Many research questions in the social sciences are categorical and concern theoretical expectations about heterogeneous effects in choice behavior. Regression models for categorical dependent variables are well-established and widely applied in the social

---

✉ Ingrid Mauerer  
ingridmauerer@uma.es

Gerhard Tutz  
tutz@stat.uni-muenchen.de

<sup>1</sup> Faculty of Economics, University of Málaga, Campus El Ejido, 29013 Málaga, Spain

<sup>2</sup> Department of Statistics, LMU Munich, Akademiestraße 1, 80799 Munich, Germany

sciences to analyze research problems that involve two or more categories without an ordering structure (see, e.g., Agresti 2007; Long 1997; Tutz 2012). Applications of discrete choice models that incorporate heterogeneity are limited. One frequently applied model is the mixed logit model (MXL) (see, e.g., Greene et al. 2006; McFadden and Train 2000) and related models, which can be quite demanding to apply. For example, the researcher needs to decide on a distribution for the subject-specific heterogeneity to approximate the underlying behavioral process, and repeated measurements are necessary to identify the model.

The objectives of this contribution are:

1. to present and discuss a very flexible and general approach to account for heterogeneity in nominal dependent variables;
2. to investigate differences with alternative models allowing heterogeneity in choice behavior, both theoretically and empirically;
3. to apply it to empirical social science choice data to exemplify the types of insights that can be obtained from the approach.

Relying on the random utility maximization framework, we present the *General Heterogeneous Multinomial Logit Model* (GHMNL). The model builds on the standard multinomial logit model (MNL) (McFadden 1974), the most frequently applied approach to study choices among discrete alternatives.<sup>1</sup> As the MNL model, the GHMNL model is a classical discrete choice model that can handle both choice-specific and chooser-specific explanatory variables. In contrast to the MNL model, which ignores that the variance of the underlying latent traits can be chooser-specific, the GHMNL model accounts for such heterogeneous effects. It allows for systematically studying theoretical expectations about heterogeneity in choice behavior. The extension integrates a heterogeneity term into the systematic part of the utility function, which is linked to explanatory variables and permits accounting for behavioral tendencies in choice behavior without referring to latent variables. The heterogeneity term indicates the degree of distinctiveness of choice or the strength of heterogeneity in choice behavior and allows examining which explanatory variables cause heterogeneity. Compared to the MXL and related models, the GHMNL model comes with convenient properties, such as its closed-form solution for evaluating the outcome probabilities. It also frees researchers from making distributional assumptions for the random parameters and is computationally straightforward.

We apply the GHMNL model to electoral choice data and demonstrate its benefits in studying heterogeneity in issue voting behavior. The empirical application has several merits. Issue voting models typically contain both types of explanatory variables, choice-specific (issue proximities) and chooser-specific (socioeconomic voter attributes) ones. The political science literature on voter heterogeneity also provides several theoretical concepts why not all voters assign the same importance to issue considerations, including, for instance, platform divergence or political sophistication. We will demonstrate how the GHMNL model allows incorporating such theoretical expectations into the empirical modeling. Although we focus on

---

<sup>1</sup> We use the term MNL to refer to multinomial logit models that contain covariates that depend on the outcome categories and those that do not. The model is also known as conditional logit.

heterogeneity in electoral choices in our empirical application, we see great potential for applying the model to explore heterogeneous effects in all social science disciplines.

Based on a brief review of the classical discrete choice model, we first derive the general heterogeneous multinomial choice model (GHMNL) and outline how it extends the standard MNL model. Next, we investigate the differences between the GHMNL and competing models. Then, we apply the GHMNL model to electoral choice data, followed by an empirical comparison with alternative models. Finally, we offer some concluding remarks.

## 2 The standard multinomial choice model

The multinomial logit model (MNL) is the most common model to study choice behavior (see, e.g., Hensher et al. 2015; Train 2009). One key feature of the MNL model limits our insights into heterogeneity in choice behavior. It ignores that the variances of the underlying latent traits can vary across decision makers. A brief review of the MNL model will help to show how the model we propose can account for heterogeneity.

Let  $Y_i \in \{1, \dots, J\}$  denote the dependent variable that consists of  $J$  unordered multiple categories for  $i \in \{1, \dots, n\}$  observations. Within the discrete choice framework, the categories represent  $J$  discrete, mutually exclusive, and finite alternatives of which decision makers choose one. The choice outcome can be a function of two types of explanatory variables: *choice-specific* and *chooser-specific* variables. The formers are specific for each category and take different values across both alternatives and choosers. They characterize the alternatives, such as price or distance in a classical mode choice situation. Chooser-specific variables contain characteristics of the decision makers, which vary over decision makers but are constant across alternatives, such as age or gender. Let  $z_{ijk}$ ,  $j \in \{1, \dots, J\}$ ,  $k \in \{1, \dots, K\}$  denote the choice-specific variables and  $s_{im}$ ,  $m \in \{1, \dots, M\}$  the chooser-specific covariates.

A common way to motivate a choice model is to consider the utilities associated with the alternatives as latent variables. Let  $U_{ij}$  denote an unobservable random utility that represents how attractive or appealing each alternative  $j \in \{1, \dots, J\}$  is for chooser  $i \in \{1, \dots, n\}$ . The decision makers are assumed to assess and compare each alternative and select the one that maximizes the random utility so that  $Y_i$  is linked to the latent variables by the *principle of maximum random utility*,

$$Y_i = j \Leftrightarrow U_{ij} = \max_{s \in \{1, \dots, J\}} U_{is}.$$

In a random utility framework, the utility is determined by  $U_{ij} = V_{ij} + \varepsilon_{ij}$ , where  $V_{ij}$  represents the systematic part of the utility, specified by explanatory variables and unknown parameters, whereas  $\varepsilon_{i1}, \dots, \varepsilon_{iJ}$  are independent and identically distributed (i.i.d.) random variables with distribution function  $F(\cdot)$ .

The systematic part of the utility function is specified as a linear predictor

$$V_{ij} = \beta_{j0} + \sum_{k=1}^K z_{ijk} \alpha_k + \sum_{m=1}^M s_{im} \beta_{jm} = \beta_{j0} + \mathbf{z}_{ij}^T \boldsymbol{\alpha} + \mathbf{s}_i^T \boldsymbol{\beta}_j, \quad (1)$$

where

- $\beta_{10}, \dots, \beta_{J0}$  are the alternative-specific constants.
- $\boldsymbol{\alpha}^T = (\alpha_1, \dots, \alpha_K)$  are the parameters<sup>2</sup> associated with the vector of choice-specific variables  $\mathbf{z}_{ij}^T = (z_{ij1}, \dots, z_{ijK})$ , which indicate the weight decision makers attach to each attribute  $k$  of the alternatives.
- $\boldsymbol{\beta}_j^T = (\beta_{j1}, \dots, \beta_{jM})$  is a coefficient vector that expresses how the chooser attributes contained in  $\mathbf{s}_i^T = (s_{i1}, \dots, s_{iM})$  determine the choice.

By assuming that  $\varepsilon_{i1}, \dots, \varepsilon_{iJ}$  are i.i.d. variables with distribution function  $F(x) = \exp(-\exp(-x))$ , which is known as the Gumbel or maximum extreme value distribution, one obtains the classical standard multinomial logit model (see McFadden 1974)

$$P(Y_i = j \mid \{\mathbf{z}_{ij}\}, \mathbf{s}_i) = \frac{\exp(V_{ij})}{\sum_{s=1}^J \exp(V_{is})} = \frac{\exp(\beta_{j0} + \mathbf{z}_{ij}^T \boldsymbol{\alpha} + \mathbf{s}_i^T \boldsymbol{\beta}_j)}{\sum_{s=1}^J \exp(\beta_{s0} + \mathbf{z}_{is}^T \boldsymbol{\alpha} + \mathbf{s}_i^T \boldsymbol{\beta}_s)}, \quad (2)$$

$j \in \{1, \dots, J\}$ . Since the chooser-specific variables  $\mathbf{s}_i$  are constant over the alternatives, not all of the corresponding coefficients are identifiable. The same applies to the constants. To identify the model, side constraints are needed. We will use the standard side constraint based on a reference alternative, whose coefficients are set to zero. We select the first alternative as reference and set  $\beta_{10} = 0$  and  $\boldsymbol{\beta}_1^T = (0, \dots, 0)$ .

The standard MNL model presented in Eqs. (1) and (2) ignores that the variance of the underlying latent traits can be subject-specific so that the variances are not allowed to differ across decision makers. Previous research has shown that ignoring variance heterogeneity can yield biased estimates (see, e.g., Tutz 2021).

### 3 A general heterogeneous multinomial choice model

In this section, we present and discuss a general multinomial choice model, which we refer to as the *General Heterogeneous Multinomial Logit Model*, in short GHMNL, that accounts for variance heterogeneity in choice behavior. Models of this type have been considered before by Hensher et al. (1998), DeShazo and Fermo (2002), and Tutz (2021).<sup>3</sup> The present work goes beyond these contributions in the following

<sup>2</sup> For simplicity, we assume that the parameters  $\boldsymbol{\alpha}$  are identical for all alternatives, i.e.,  $\boldsymbol{\alpha}_1 = \dots = \boldsymbol{\alpha}_j := \boldsymbol{\alpha}$ . This simplification results in a so-called generic or global effect, which does not depend on the alternatives. See Mauerer et al. (2015b); Mauerer (2016); Thurner (2000) for relaxation of the assumption in the study of proximity voting in multiparty elections, and see Mauerer et al. (2015a) for a parameter selection procedure to systematically reduce the resulting model complexity.

<sup>3</sup> Hensher et al. (1998) call the model ‘parametrised heteroscedastic MNL’ (PHMNL). In DeShazo and Fermo (2002), it is referred to as the ‘heteroscedastic logit model’. The contribution by Tutz (2021) is restricted to global covariates that do not depend on the outcome categories and therefore does not incorporate choice-specific explanatory variables, which lay at the heart of discrete choice models as attributes of the choice alternatives are the utility sources in discrete choice models.

ways. We outline in detail and discuss the specification of the utility functions and the choice probabilities in the GHMNL model. We put particular emphasis on the interpretation of the heterogeneity term that is incorporated into the utility function and the estimation methods. In addition, we have written an R function that allows the user to fit the GHMNL model. Section A in the Supplementary Information describes the routines to implement the model.

### 3.1 Utility functions and choice probabilities

The GHMNL model extends the standard MNL model by adding a heterogeneity term to the systematic part of the utility function. For simplicity, let all the explanatory variables and the constants be collected in the alternative-specific vector  $\mathbf{x}_{ij}^T = (\mathbf{1}_j^T, \mathbf{0}, \dots, \mathbf{z}_{ij}^T, \dots, \mathbf{0})$ , where  $\mathbf{1}_j$  is the  $j$ th unit vector and  $\mathbf{0}$  is a vector of zeros. Then, the utility functions take the form

$$V_{ij} = \beta_{j0} + \mathbf{z}_{ij}^T \boldsymbol{\alpha} + \mathbf{s}_i^T \boldsymbol{\beta}_j = \mathbf{x}_{ij}^T \boldsymbol{\delta},$$

where  $\boldsymbol{\delta}^T = (\beta_{10}, \dots, \beta_{J0}, \boldsymbol{\alpha}^T, \boldsymbol{\beta}_1^T, \dots, \boldsymbol{\beta}_J^T)$ . To derive the GHMNL model, we assume that the latent utilities are given more generally by

$$U_{ij} = \mathbf{x}_{ij}^T \boldsymbol{\delta} + \sigma_i \varepsilon_j,$$

where  $\sigma_i$  is the standard deviation associated with decision maker  $i$ . The standard deviation is linked to explanatory variables by assuming  $\sigma_i = e^{-\mathbf{w}_i^T \boldsymbol{\gamma}}$ , where  $\mathbf{w}_i$  is a vector of chooser-specific covariates and  $\boldsymbol{\gamma}$  is a vector of parameters. As a result, the utilities  $V_{ij}$  in the GHMNL model are specified as

$$V_{ij} = \mathbf{x}_{ij}^T \boldsymbol{\delta} e^{\mathbf{w}_i^T \boldsymbol{\gamma}} = (\beta_{j0} + \mathbf{z}_{ij}^T \boldsymbol{\alpha} + \mathbf{s}_i^T \boldsymbol{\beta}_j) e^{\mathbf{w}_i^T \boldsymbol{\gamma}}, \tag{3}$$

where

- $\mathbf{s}_i$  is a vector of chooser-specific covariates, and  $\mathbf{z}_{ij}$  is a vector of alternative-specific covariates. As in the standard MNL model, the variables  $\mathbf{s}_i$  have alternative-specific effects and  $\mathbf{z}_{ij}$  global effects.
- $\mathbf{w}_i^T = (w_{i1}, \dots, w_{iL})$  is a vector of chooser-specific variables, which can be a subset of  $\mathbf{s}_i$ . It contains attributes of the decision makers that are supposed to cause heterogeneity in choice behavior. The parameter vector  $\boldsymbol{\gamma}^T = (\gamma_1, \dots, \gamma_L)$  indicates the strength of heterogeneity in choosing one alternative.

The model distinguishes between two types of effects: a *location effect* and a *heterogeneity effect*. The term  $\mathbf{x}_{ij}^T \boldsymbol{\delta}$  in Eq. (3) represents the location effect. It is also present in the standard MNL model and determines which alternative the chooser tends to prefer. The novel term  $\mathbf{w}_i^T \boldsymbol{\gamma}$  represents the heterogeneity effect that determines the impact of heterogeneity in choice behavior.

As the standard MNL model, the GHMNL model has a closed-form solution for evaluating the choice probabilities so that the utility functions  $V_{ij}$  are linked to the choice probabilities through a logistic response function,

$$P(Y_i = j \mid \{\mathbf{x}_{ij}\}, \mathbf{w}_i) = \frac{\exp(\mathbf{x}_{ij}^T \boldsymbol{\delta} e^{\mathbf{w}_i^T \boldsymbol{\gamma}})}{\sum_{s=1}^J \exp(\mathbf{x}_{is}^T \boldsymbol{\delta} e^{\mathbf{w}_i^T \boldsymbol{\gamma}})}, j \in \{1, \dots, J\}. \quad (4)$$

Alternatively, the relationship between the choice probabilities and the utility functions can be expressed in terms of odds:

$$\begin{aligned} \frac{P(Y_i = j \mid \{\mathbf{x}_{ij}\}, \mathbf{w}_i)}{P(Y_i = s \mid \{\mathbf{x}_{ij}\}, \mathbf{w}_i)} &= \exp\{(\mathbf{x}_{ij} - \mathbf{x}_{is})^T \boldsymbol{\delta} e^{\mathbf{w}_i^T \boldsymbol{\gamma}}\} \\ &= \exp\{[\beta_{j0} - \beta_{s0} + (\mathbf{z}_{ij} - \mathbf{z}_{is})^T \boldsymbol{\alpha} + \mathbf{s}_i(\boldsymbol{\beta}_j - \boldsymbol{\beta}_s)] e^{\mathbf{w}_i^T \boldsymbol{\gamma}}\}. \end{aligned}$$

### 3.2 Interpretation of the heterogeneity term

The essential novel term in the GHMNL model is the heterogeneity term. It is modeled by the factor  $e^{\mathbf{w}_i^T \boldsymbol{\gamma}}$  and represents the (inverse) standard deviation of the latent variables. The heterogeneity term can be understood as representing variance heterogeneity. However, it also allows for an interpretation without reference to latent variables, which are always elements used to build a model but cannot be observed. The heterogeneity term represents a specific choice behavior that permits accounting for behavioral tendencies that are not linked to particular alternatives:

- When  $\mathbf{w}_i^T \boldsymbol{\gamma} \rightarrow -\infty$ , one obtains  $P(Y_i = j \mid \{\mathbf{x}_{ij}\}, \mathbf{w}_i) = 1/J$ . In this extreme case, all alternatives have the same choice probabilities. It implies that the decision maker chooses an alternative at random because none of the covariates can systematically explain the choice. The chooser shows maximal heterogeneity.
- When  $\mathbf{w}_i^T \boldsymbol{\gamma} \rightarrow \infty$  and the condition  $\mathbf{x}_{ij}^T \boldsymbol{\delta} \neq 0$  holds at least for one  $j > 1$ , the probability for one of the  $j \in \{1, \dots, J\}$  alternatives approaches 1. In this case, the decision maker has a distinct preference, and shows minimal heterogeneity. Therefore, choosers with large  $\mathbf{w}_i^T \boldsymbol{\gamma}$ -values show less variability, they distinctly prefer specific alternatives.

Thus, the heterogeneity term  $\mathbf{w}_i^T \boldsymbol{\gamma}$  can be considered as an indicator of the degree of distinctness of choice or as a measure of heterogeneity in choice behavior. For small values of  $\mathbf{w}_i^T \boldsymbol{\gamma}$ , the difference between the choice probabilities becomes small. By contrast, the difference between a specific alternative and the remaining ones gets larger when  $\mathbf{w}_i^T \boldsymbol{\gamma}$  increases. As the heterogeneity term contains attributes of the decision makers, the model systematically accounts for heterogeneity in choice behavior across individuals. It allows examining which explanatory variables cause heterogeneous effects. For example, suppose  $w_i$  denotes age and  $\boldsymbol{\gamma}$  is positive. It would suggest that older decision makers have more clear-cut preferences than younger ones. The former tend to prefer specific alternatives, while younger decision

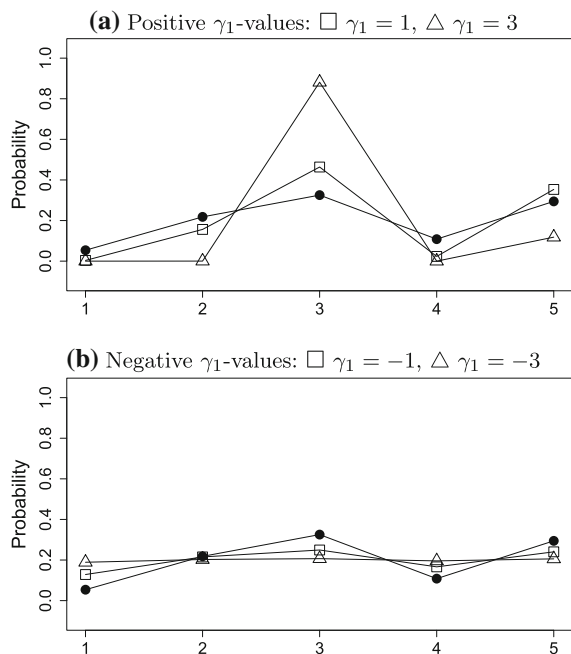
makers have less distinct preferences and show more heterogeneity in selecting one alternative.

Figure 1 illustrates the behavioral tendencies the GHMNL model can uncover. It depicts the probabilities  $P(Y_i = j)$  for a five-choice situation based on a model with two covariates in the heterogeneity term  $w_i$ , one binary and one quantitative normally distributed explanatory variable. For the binary covariate, we consider the effect at value  $w_i^T = (1, 0)$ . The two panels depict the probabilities for different parameter values ( $\gamma_1$ ) in the heterogeneity term: panel (a) shows the effects for positive  $\gamma_1$ -values, panel (b) for negative  $\gamma_1$ -values. In both panels, the filled circles depict the base probabilities that result when no heterogeneity is present, that is, when  $\gamma_1 = 0$ , yielding the standard MNL model. Inspecting these base probabilities suggests that the chooser prefers alternative 3, and to a lesser extent alternative 5. Panel (a) shows that this pattern becomes more pronounced for increasing  $\gamma_1$ -values. Thus, the decision maker more distinctly prefers alternative 3 in the GHMNL model. By contrast, the pattern flattens for negative  $\gamma_1$ -values, as illustrated in panel (b). This indicates that the decision maker tends to choose an alternative at random and shows substantial heterogeneity in selecting one of the five alternatives.

### 3.3 Estimation

In the following, we outline how the parameters of the GHMNL model can be estimated. Let the choice outcome be represented as a vector  $y_i = (y_{i1}, \dots, y_{iJ})^T$  with  $y_{ij}$  taking the value 1 when alternative  $j \in \{1, \dots, J\}$  is chosen and 0 otherwise so

**Fig. 1** Illustration of the Heterogeneity Term in the GHMNL Model *Note:* Plots depict probabilities  $P(Y_i = j)$  for five alternatives  $j \in \{1, 2, 3, 4, 5\}$ . The model is based on one binary and one quantitative normally distributed covariate. Panel (a) shows the effects for positive  $\gamma_1$ -values, panel (b) for negative  $\gamma_1$ -values. Filled circles  $\bullet$  depict the base probabilities when no heterogeneity is present ( $\gamma_1 = 0$ ). The probability curves result by plugging the respective values into Eq. (4)





that one obtains  $y_i = (0, \dots, 0, 1, 0, \dots, 0)^T$  if  $Y_i = j$ . Let  $\pi_{ij} = P(Y_i = j \mid \{\mathbf{x}_{ij}\}, \mathbf{w}_i)$  denote the choice probabilities and  $\boldsymbol{\delta}^T = (\beta_{10}, \dots, \beta_{J0}, \boldsymbol{\alpha}^T, \boldsymbol{\beta}_1^T, \dots, \boldsymbol{\beta}_J^T, \boldsymbol{\gamma}^T)$  the overall parameter vector that collects all coefficients to be estimated.

Using the first alternative as reference, the kernel of the log-likelihood of the model presented in Eq. (4) is given by

$$\begin{aligned}
 l(\boldsymbol{\delta}) &= \sum_{i=1}^n \left\{ \sum_{j=2}^J y_{ij} \log \left( \frac{\pi_{ij}}{1 - \pi_{i2} - \dots - \pi_{iJ}} \right) + \log(1 - \pi_{i2} - \dots - \pi_{iJ}) \right\} \\
 &= \sum_{i=1}^n \left\{ \sum_{j=2}^J y_{ij} (\mathbf{x}_{ij}^T \boldsymbol{\delta}) e^{\mathbf{w}_i^T \boldsymbol{\gamma}} - \log \left( 1 + \sum_{s=2}^J \exp(\mathbf{x}_{is}^T \boldsymbol{\delta} e^{\mathbf{w}_i^T \boldsymbol{\gamma}}) \right) \right\}.
 \end{aligned}$$

For the maximization of the log-likelihood, we make use of the first derivatives, also known as score functions. They take the form

$$\begin{aligned}
 \partial l(\boldsymbol{\delta}) / \partial \delta_t &= \sum_{i=1}^n \left\{ \sum_{j=2}^J y_{ij} x_{ijt} \delta_t e^{\mathbf{w}_i^T \boldsymbol{\gamma}} - \frac{\sum_{s=2}^J x_{ist} e^{\mathbf{w}_i^T \boldsymbol{\gamma}} \exp(\mathbf{x}_{is}^T \boldsymbol{\delta} e^{\mathbf{w}_i^T \boldsymbol{\gamma}})}{1 + \sum_{s=2}^J \exp(\mathbf{x}_{is}^T \boldsymbol{\delta} e^{\mathbf{w}_i^T \boldsymbol{\gamma}})} \right\}, \\
 \partial l(\boldsymbol{\delta}) / \partial \gamma_t &= \sum_{i=1}^n \left\{ \sum_{j=2}^J y_{ij} (\mathbf{x}_{ij}^T \boldsymbol{\delta}) e^{\mathbf{w}_i^T \boldsymbol{\gamma}} w_{it} - \frac{\sum_{s=2}^J \mathbf{x}_{is}^T \boldsymbol{\delta} e^{\mathbf{w}_i^T \boldsymbol{\gamma}} w_{it} \exp(\mathbf{x}_{is}^T \boldsymbol{\delta} e^{\mathbf{w}_i^T \boldsymbol{\gamma}})}{1 + \sum_{s=2}^J \exp(\mathbf{x}_{is}^T \boldsymbol{\delta} e^{\mathbf{w}_i^T \boldsymbol{\gamma}})} \right\}.
 \end{aligned}$$

As approximation of the covariance  $\text{cov}(\hat{\boldsymbol{\delta}})$ , we use the observed information  $-\partial^2 l(\hat{\boldsymbol{\delta}}) / \partial \boldsymbol{\delta} \partial \boldsymbol{\delta}^T$ .

### 4 Alternative approaches

A model that has been used to study heterogeneity in decision behavior is the mixed logit model (MXL) (see Greene et al. 2006; Hensher and Greene 2003; McFadden and Train 2000).<sup>4</sup> Two related models, which differ in parameterization, are the scale heterogeneity (S-MNL) model and the generalized multinomial logit model (G-MNL) (see Fiebig et al. 2010). A brief review and discussion of the MXL and related models will illustrate the advantages of applying the GHMNL model to account for heterogeneity in choice behavior.

#### 4.1 Mixed logit model

Following Greene et al. (2006), the MXL model can be derived from latent utilities

$$U_{ijt} = \mathbf{z}_{ijt}^T \boldsymbol{\alpha}_i + \varepsilon_{ijt},$$

where

- the additional index  $t$  refers to different choice occasions or tasks, and

<sup>4</sup> The model is also referred to as random parameters logit, mixed or heterogeneous multinomial logit, or hybrid logit model. We use the most popular term, mixed logit model.

- $\mathbf{z}_{ijt}$  is the full vector of explanatory variables, including attributes of the alternatives, socioeconomic characteristics of the decision makers, and the choice task itself.

Compared to the standard MNL model, the crucial extension in the MXL model is that the parameter vector  $\alpha_i$  is subject-specific so that the effects can vary across decision makers  $i$ . Assuming that the subject-specific effects are random and in part determined by an additional vector of covariates  $\mathbf{w}_i$ , the model becomes a mixed-effects model. The subject-specific effects are assumed to take the form

$$\alpha_i = \alpha + \Delta \mathbf{w}_i + \Sigma^{1/2} \mathbf{v}_i,$$

where

- $\Delta$  is a matrix of coefficients associated with the covariate vector  $\mathbf{w}_i$ ,
- $\mathbf{v}_i$  is a random vector of uncorrelated random variables with known variances,
- $\Sigma$  is a covariance matrix that determines the variance structure of the random term.

Maximum simulated likelihood estimates are obtained by maximizing the log-likelihood with respect to all the unknown parameters (see also Train 2009).

By allowing parameters to vary randomly over decision makers instead of assuming that they are the same for every chooser, the MXL model is very flexible and can account for a rather general form of heterogeneity. However, this flexibility comes with the cost of a large number of parameters, which might render estimates unstable without careful variable selection. Further drawbacks of the model are that one has to specify a specific distribution for the subject-specific random effects. The model parameters may not be identified without repeated measurements, that is, without having multiple choice observations  $t$  for the same chooser.

## 4.2 S-MNL and G-MNL models

Two models that are related to the MXL approach are the scale heterogeneity model (S-MNL) and the generalized multinomial logit model (G-MNL) (see Fiebig et al. 2010). Table 1 summarizes the definition of the utility functions and the parametrization of these models together with the standard MNL and the MXL model.

As the MXL model, both the S-MNL and the G-MNL models belong to the family of mixed-effects or mixture models. Compared to the MXL model, where the scale of the error term  $\lambda_i$  is normalized, the S-MNL model assumes that  $\lambda_i$  may vary across decision makers and follows a fixed distribution. For the estimation of parameters, the assumed distribution is essential. Scale heterogeneity is introduced by multiplying the utility weights  $\alpha$  with the subject-specific scaling factor  $\lambda_i$ , which scales  $\alpha$  up or down proportionately across chooser  $i$ . It is also possible that chooser-specific covariates  $\mathbf{w}_i$  enter the scale so that the scale may vary across decision makers according to their characteristics. As the S-MNL involves a simpler

**Table 1** Overview on Competing Approaches

	Utility Functions	Parametrization	Normalizations
MNL	$U_{ijt} = \mathbf{z}_{ijt}^T \boldsymbol{\alpha} + \epsilon_{ijt} / \lambda_i$	$\boldsymbol{\alpha}_i = \boldsymbol{\alpha}$	$\lambda_i = \lambda = 1;$ $\text{var}(\boldsymbol{\eta}_i) = 0$
MXL	$U_{ijt} = \mathbf{z}_{ijt}^T (\boldsymbol{\alpha} + \boldsymbol{\eta}_i) + \epsilon_{ijt}$	$\boldsymbol{\alpha}_i = \boldsymbol{\alpha} + \boldsymbol{\eta}_i$	$\lambda_i = \lambda = 1$
S-MNL	$U_{ijt} = \mathbf{z}_{ijt}^T (\boldsymbol{\alpha} \lambda_i) + \epsilon_{ijt}$	$\boldsymbol{\alpha}_i = \boldsymbol{\alpha} \lambda_i$	$\text{var}(\boldsymbol{\eta}_i) = 0$
G-MNL	$U_{ijt} = \mathbf{z}_{ijt}^T [\boldsymbol{\alpha} \lambda_i + \kappa \boldsymbol{\eta}_i + (1 - \kappa) \lambda_i \boldsymbol{\eta}_i] + \epsilon_{ijt}$	$\boldsymbol{\alpha}_i = \boldsymbol{\alpha} \lambda_i + \kappa \boldsymbol{\eta}_i + (1 - \kappa) \lambda_i \boldsymbol{\eta}_i$	

$\boldsymbol{\alpha}$ : mean utility weights,  $\lambda$ : scaling factor,  $\boldsymbol{\eta}_i$ : deviations from mean utility weights,  $\kappa \in [0, 1]$

parametrization and fewer coefficients, it describes the data more sparsely than the MXL model.

The G-MNL model assumes both coefficient and scale heterogeneity. It nests the MXL model and the S-MNL model. The MXL model results when the scale parameter  $\lambda_i = 1$ . Setting  $\text{Var}(\boldsymbol{\eta}_i) = 0$  yields the S-MNL model. The parameter  $\kappa^5$  in the G-MNL model indicates how the variance of the error term varies with the scale  $\lambda_i$ . There are two model variants. When  $\kappa$  approaches 1, the standard deviation of  $\boldsymbol{\eta}_i$  does not depend on the scaling of  $\boldsymbol{\alpha}$  (i.e.,  $\lambda_i$ ). When  $\kappa$  approaches 0, the standard deviation of  $\boldsymbol{\eta}_i$  is proportional to  $\lambda_i$ . Again, covariates can be included to explain the variation.

### 4.3 Comparing modeling approaches

The GHMNL and the alternative models (MXL, S-MNL, G-MNL) can be derived from latent utilities. The main difference between both approaches lies in the motivation of heterogeneity in choice behavior. In the GHMNL model, the variances of the latent utilities are allowed to vary across the characteristics of the decision makers. The alternative models assume that each chooser has its own parameters, which follow a fixed distribution. The GHMNL model also allows parameters to vary across choosers, but it does so in a more systematic and restrictive way. Here, the effect parameters associated with the alternative-specific covariates are  $\boldsymbol{\alpha} e^{\boldsymbol{w}_i^T \boldsymbol{\gamma}}$ . Under this specification, the covariates contained in  $\boldsymbol{w}_i$  modify the effects. Depending on the value of  $\boldsymbol{w}_i$ , the effect is strengthened or weakened. In addition, the same effect modification applies to all coefficients, which is a consequence of the derivation from the variances of the latent utilities. By contrast, the competing models allow for all sorts of parameter variation, including random variation and even a possible reversal of the sign of effects.

By allowing the effects to vary across decision makers, both approaches have in common that they assume a specific form of interaction. In the GHMNL model, an interaction between the variables  $\boldsymbol{x}_{ij}$  and  $\boldsymbol{w}_i$  is present because the linear term takes

<sup>5</sup> Note that in Fiebig et al. (2010) this parameter is denoted by  $\gamma$ . To ensure the uniqueness of the elements, we denote it by  $\kappa$ .

the form  $\mathbf{x}_{ij}^T \delta e^{w_i^T \gamma}$  (see Eq. 3). In the MXL model, for example, the interaction is included as the linear effect  $\mathbf{z}_{ijt}^T \boldsymbol{\alpha}_i$  contains the term  $\mathbf{z}_{ijt}^T \Delta \mathbf{w}_i$ . In both cases, the interaction can be seen as an interaction generated by effect modification. The effect of  $\mathbf{x}_{ij}$  (or  $\mathbf{z}_{ijt}$ ) is modified by  $\mathbf{w}_i$ , the latter variable is a so-called effect modifier.

Both approaches can be embedded into the general framework of varying-coefficient models (see, e.g., Fan and Zhang 1999; Hastie and Tibshirani 1993; Park et al. 2015). The varying-coefficients framework helps to see that identifiability problems arise if the variables  $\mathbf{z}_{ijt}$  and  $\mathbf{w}_i$  are not distinct. Guided by theoretical expectations about heterogeneity, the researcher applying the competing models might consider different variables in  $\mathbf{z}_{ijt}$  and  $\mathbf{w}_i$ . However, if the underlying theory does not provide such expectations, one faces the challenge of determining which explanatory variables are effect modifiers and which ones represent main effects. By contrast, the inclusion of the same set of variables in the location and the heterogeneity part does not cause any difficulties in the GHMNL model.

In sum, the benefits of the GHMNL model as compared to the competing models are:

- Whereas the MXL, the S-MNL, and the G-MNL model can account for a rather general and unspecific form of heterogeneity without further motivation, the heterogeneity term in the GHMNL model can uncover specific behavioral tendencies. It provides an indicator of the degree of distinctness of choice and measures the strength of heterogeneity in choice behavior.
- The GHMNL is much sparser in terms of the number of parameters involved and therefore avoids that estimates render unstable without careful variable selection.
- It allows for a closed-form of the log-likelihood without the need to use simulation methods to obtain choice probabilities, which makes the GHMNL model computationally straightforward.
- The researcher does not need to decide on a specific and appropriate distribution for the random parameters to approximate the underlying behavioral process.
- The GHMNL model avoids identifiability problems and works without repeated measurements.

## 5 Application: proximity voting and heterogeneous electorates

The empirical application uses individual-level survey data on electoral choices to study heterogeneity in voting behavior. Our voter choice model follows the classical proximity model, where the main source of voter utility is the ideological proximity to the parties (see Davis et al. 1970; Downs 1957). Accordingly, proximity voting approaches expect that voter  $i$  casts a ballot for the party  $j$  that offers policy platforms closest to the voter's most preferred positions on  $K$  different policy issues.

We draw on the 2017 German parliamentary election study (Roßteutscher et al. 2018) and analyze heterogeneity in voter choice for one of the six major German parties in 2017: the Christian-Democratic Parties (CDU/CSU), the Social-Democratic

Party (SPD), the Liberal Party (FDP), the Greens, the Left, and the Alternative for Germany (AfD). The election study contains three policy issues (immigration, taxes, climate change) on which the respondents positioned themselves and the parties on eleven-point scales. Using voter-specific self-placements and perceptions of party placements, the choice-specific variables  $z_{ijk}$  in Eq. (3) contain the absolute proximity between each voter  $i$  and party  $j$  on each policy issue  $k$ . We use a separate data set for each issue, including only respondents with no missing values on the respective issue scales. The three data sets are quite large. The number of observations varies from 1807 to 1251.

The empirical application proceeds as follows. Based on previous research on heterogeneity in proximity voting, the first part examines one central source of heterogeneity: platform divergence. In the second part, we present the results of a fully-specified voter choice that also accounts for nonpolicy considerations in the voting calculus. The section closes with an empirical comparison of the GHMNL model and competing approaches. Section B in the Supplementary Information contains a detailed description of the measurement and coding of all considered variables. Section C reports the estimates of competing models.

## 5.1 Sources of heterogeneity in proximity voting

It has become accepted wisdom that not all voters follow issue considerations in the same way in making electoral decisions. The debate about heterogeneous electorates has a long tradition in the proximity voting literature. One example is the classic article on voter heterogeneity by Rivers (1988), stating that different subgroups of voters apply different choice criteria when voting. Another one is the issue public hypothesis by Converse (1964), postulating that the population can be divided into issue publics, each consisting of voters who intensively care about particular issues. Several concepts, conditions, or sources of heterogeneity have been proposed as to why we should expect systematic individual-level differences in the impact of issue considerations on voting. The concept of issue importance is the most frequently discussed source of heterogeneity (see, e.g., Edwards III et al. 1995; RePass 1971). If issues are considered individually salient to voters, then voters are expected to assign these issues a greater weight in the voting-decision process. A large research body has also argued that heterogeneity in issue voting is the result of differences in political sophistication or awareness (see, e.g., Gerber et al. 2015; Luskin 1987; MacDonald et al. 1995). In the following, we empirically examine another theoretical source of heterogeneity in spatial voting: platform divergence.

### 5.1.1 Platform divergence

A central condition that must be met so that issues determine voter choice is substantial divergence in offered party positions. Accordingly, voters who see clear differences between parties' policy proposals are expected to rely more strongly on issue attitudes when casting their ballots than those perceiving similar party stands (e.g., Alvarez and Nagler 2004; Weißels and Schmitt 2008). We employ a subject-

specific measure to examine whether platform divergence causes heterogeneity in the impact of issue considerations on party choice. We use the individually perceived range of party positions to identify the degree of platform divergence. The measure is constructed as follows: we first identified the two parties perceived to take the most extreme positions on both ends of the issue scales for each voter and issue. Then, we computed the absolute difference between these party positions. This results in eleven-point scales, where 0 indicates minimum platform divergence (i.e., all parties are perceived to offer the same position) and 10 maximum platform divergence (i.e., voters perceive the party positions to be spread across the entire issue scale).

### 5.1.2 Empirical models: platform divergence

We specify a separate model for each of the three policy issues to examine whether voters exhibit heterogeneous reactions to issues due to platform convergence:

$$V_{ij} = (\beta_{j0} + \alpha \cdot z_{ij})e^{w_i^T \gamma}. \quad (5)$$

In each model, the location term in Eq. (5) contains the party-specific constants and issue proximity. To identify the constants, we use the CDU as the reference party. In the heterogeneity term, we consider the concept of platform divergence. Since the heterogeneity term affects the complete location term and the considered source of heterogeneity is specific to each issue, the issue-by-issue model specification allows us to assess whether varying levels of platform divergence cause heterogeneous effects. That is, platform divergence on issue  $k$  (immigration, tax, climate change), which enters the respective heterogeneity terms, is a chooser-specific variable that affects only the weight of issue  $k$  on choosing.

Table 2 reports the model estimates. The first column gives the log odds, followed by standard errors and t-values. The parameters related to the issue proximities in the location term all take positive values and are statistically different from zero at the 5% significance level. In line with proximity voting approaches, the estimates indicate that the closer voters perceive the parties to their own positions on the issues, the higher the weight they assign to them when voting, *ceteris paribus*. Inspecting the estimates on platform divergence in the heterogeneity term reveals interesting choice behavior. In all three models, the coefficients related to the concept of platform divergence are negative and statistically different from zero at the 5% significance level. The negative parameters indicate heterogeneity in choice behavior. The estimates imply that voters who perceive substantial divergence in party positions are more heterogeneous in choosing one party, *ceteris paribus*.

## 5.2 Fully-specified voter choice model

Next, we present the results of a fully-specified voter choice model in the sense that the location term includes both types of covariates (choice-specific and chooser-specific), which is in contrast to the models presented in Table 2, where we consider only a choice-specific variable in the location term. The chooser-specific variables  $s_{im}$  are socioeconomic voter characteristics. They account for the importance of voter's

**Table 2** Platform Divergence: GHMNL Model Estimates

Immigration Issue	coef.	s.e.	t-val.
<i>Location Term:</i>			
SPD	- 0.878	0.157	- 5.611
FDP	- 1.762	0.262	- 6.728
Greens	- 1.516	0.234	- 6.473
Left	- 1.693	0.258	- 6.571
AfD	- 1.961	0.282	- 6.952
Issue Proximity	0.473	0.072	6.552
<i>Heterogeneity Term:</i>			
Platform Divergence	- 0.064	0.018	- 3.476
Log-likelihood	- 2808.535		
N	1807		
Tax Issue	coef.	s.e.	t-val.
<i>Location Term:</i>			
SPD	- 0.829	0.153	- 5.405
FDP	- 1.722	0.237	- 7.272
Greens	- 1.632	0.223	- 7.324
Left	- 1.678	0.238	- 7.037
AfD	- 1.394	0.193	- 7.219
Issue Proximity	0.530	0.075	7.041
<i>Heterogeneity Term:</i>			
Platform Divergence	- 0.090	0.020	- 4.606
Log-likelihood	- 1986.497		
N	1251		
Climate Change Issue	coef.	s.e.	t-val.
<i>Location Term:</i>			
SPD	- 0.885	0.145	- 6.121
FDP	- 1.568	0.224	- 6.993
Greens	- 1.507	0.220	- 6.862
Left	- 1.646	0.229	- 7.181
AfD	- 1.119	0.181	- 6.167
Issue Proximity	0.470	0.070	6.684
<i>Heterogeneity Term:</i>			
Platform Divergence	- 0.052	0.020	- 2.583
Log-likelihood	- 1875.217		
N	1193		

*Source:* 2017 German election study (Roßteutscher et al. 2018). *Note:* The dependent variable is voter choice. The party abbreviations in the location term represent the party-specific constants. CDU is used as reference to identify them

**Table 3** Full Voter Choice Model: GHMNL Model Estimates

Location Term	coef.	s.e.	t-val.
Issue Proximity on Tax Issue	0.289	0.037	7.745
Constants			
SPD	- 0.133	0.113	- 1.173
FDP	- 0.584	0.141	- 4.134
Greens	- 0.759	0.169	- 4.482
Left	- 0.582	0.145	- 4.024
AfD	- 0.439	0.131	- 3.363
Religious Denomination (1 Catholic, 0 otherwise)			
SPD	- 0.816	0.217	- 3.765
FDP	- 0.525	0.237	- 2.211
Greens	- 0.257	0.208	- 1.238
Left	- 1.111	0.324	- 3.430
AfD	- 0.988	0.294	- 3.367
Gender (1 female, 0 male)			
SPD	- 0.359	0.217	- 1.653
FDP	- 0.859	0.357	- 2.406
Greens	- 0.569	0.307	- 1.856
Left	- 0.563	0.321	- 1.755
AfD	- 0.781	0.350	- 2.230
Union Membership (1 union member, 0 otherwise)			
SPD	0.166	0.257	0.645
FDP	- 0.439	0.387	- 1.135
Greens	0.193	0.294	0.656
Left	- 0.073	0.338	- 0.216
AfD	0.095	0.319	0.298
Heterogeneity Term			
Education	- 0.257	0.109	- 2.358
Gender (1 female, 0 male)	- 0.332	0.148	- 2.244
Region (1 former West Ger., 0 former East Ger.)	0.262	0.130	2.020
Log-likelihood	- 1960.986		

*Source:* 2017 German election study (Roßteutscher et al. 2018). N = 1251. *Note:* The dependent variable is voter choice. CDU is reference party

nonpolicy motivations in the voting calculus, which presents a central extension of the proximity voting model (e.g., Adams et al. 2005; Mauerer et al. 2015b). As nonpolicy factors  $s_i$ , we consider three dummy-coded voter attributes in the location term: religious denomination, gender, and a regional variable, indicating whether the respondent resides in former West or East Germany. In the heterogeneity term, we include gender and the regional variable to examine whether there are systematic gender or regional differences in choice behavior. We also examine whether the



voter's education causes heterogeneity in proximity voting. We note that one could maintain the variable platform divergence both in the location and heterogeneity terms (see Eq. 3) without causing identifiability problems. We opted to include three different chooser-specific variables to demonstrate how the model allows accounting for nonpolicy considerations in the voting calculus.

We focus on the tax issue and use the CDU again as the reference party. The voter choice model is based on 24 degrees of freedom: 1 issue proximities on taxes, 6 – 1 constants, and  $(6 - 1) \times 3$  parameters related to voter attributes in the location term and 3 coefficients in the heterogeneity term. Table 3 reports the estimation results. In the location term, the interpretation of the coefficients refers to the CDU as this party is used as the reference alternative to identify the model. For example, in line with central social cleavage structures in Germany, Catholics tend to prefer the Christian-Democratic Party CDU compared to the left parties SPD and the Left, *ceteris paribus*.

Regarding the heterogeneity term, the coefficients are not specific to a particular party. The corresponding effects are global and do not relate to a reference alternative. All three parameters in the heterogeneity term are statistically different from zero at the 5% significance level. The coefficient related to education is negative. This result indicates that voters with a higher level of education tend to react more heterogeneously to the tax issue. The coefficient associated with the variable gender is also negative. The negative value indicates that females show more heterogeneity in voter choice than males, *ceteris paribus*. By contrast, the coefficient related to the regional variable is positive. This result suggests that voters residing in former West Germany have more distinct party choice preferences than those in East Germany, *ceteris paribus*.

### 5.3 Empirical model comparisons

In this section, we compare our empirical GHMNL models and the competing models, and we do that as follows. First, we contrast the GHMNL models with the MNL models based on Likelihood Ratio tests. Then, we compare the performance of the GHMNL models with all alternative models.

Table 4 reports the results of the Likelihood Ratio tests across the four empirical models (immigration issue, tax issue, climate change issue, full model: tax issue). The test statistics indicate that the GHMNL models yield significantly better fits to the data than the standard MNL models in all applications.

Next, we compare the fit of the GHMNL models with the four alternative models (MNL, MXL, S-MNL, G-MNL). We use the AIC and BIC criteria to measure model

**Table 4** Likelihood Ratio Tests: Standard MNL vs. GHMNL

	Test Statistic	p-value
Table 2: Immigration Issue: MNL vs. GHMNL	11.035	0.001
Table 2: Tax Issue: MNL vs. GHMNL	20.278	0.000
Table 2: Climate Change Issue: MNL vs. GHMNL	6.154	0.013
Table 3: Full Model: MNL vs. GHMNL	15.216	0.002

**Table 5** Empirical Model Comparisons

Immigration Issue	LogL	AIC	BIC
MNL	- 2814.052	5640.104	5673.101
MXL (without covariates)	- 2812.469	5638.964	5677.460
MXL	- 2811.731	5639.463	5683.458
S-MNL (without covariates)	- 2812.469	5638.937	5677.433
S-MNL	- 2811.719	5639.438	5683.433
G-MNL (without covariates)	- 2812.487	5642.973	5692.468
G-MNL	- 2811.603	5643.205	5698.199
GHMNL	- 2808.535	5631.069	5669.565
Tax Issue	LogL	AIC	BIC
MNL	- 1996.636	4005.272	4036.062
MXL (without covariates)	- 1993.144	4000.289	4036.211
MXL	- 1992.360	4000.720	4041.774
S-MNL (without covariates)	- 1993.818	4001.636	4037.558
S-MNL	- 1992.949	4001.897	4042.951
G-MNL (without covariates)	- 1991.781	4001.561	4047.747
G-MNL	- 1990.921	4001.843	4053.160
GHMNL	- 1986.497	3986.994	4022.916
Climate Change Issue	LogL	AIC	BIC
MNL	- 1878.293	3768.587	3801.584
MXL (without covariates)	- 1878.114	3770.227	3805.817
MXL	- 1877.758	3771.517	3812.191
S-MNL (without covariates)	- 1877.773	3769.547	3805.136
S-MNL	- 1877.479	3770.959	3811.632
G-MNL (without covariates)	- 1877.525	3773.051	3818.809
G-MNL	- 1877.267	3774.534	3825.376
GHMNL	- 1875.217	3764.433	3800.023
Full Model: Tax Issue	LogL	AIC	BIC
MNL	- 1968.593	3979.187	4086.953
MXL (without covariates)	- 1965.101	3974.202	4087.100
MXL	- 1958.191	3966.383	4094.675
S-MNL (without covariates)	- 1965.937	3975.873	4088.770
S-MNL	- 1959.647	3969.295	4097.587
G-MNL (without covariates)	- 1963.894	3975.789	4098.950
G-MNL	- 1956.316	3966.633	4105.189
GHMNL	- 1960.986	3935.971	3971.893

Section C in the Supplementary Information reports the parameter estimates and provides estimation details of the competing models

performance.<sup>6</sup> For each of the last three models, we estimated two variants. The first models only account for subject-specific heterogeneity according to their respective parameterizations (see Table 1). The second models additionally include covariates to explain the heterogeneity. The covariates are the same as in Tables 2 and 3. Table 5 summarizes the performance measures. The values indicate substantial improvements in both information criteria for the GHMNL models. In all settings, the AIC and BIC values for the alternative models are larger than for the GHMNL models, showing that the GHMNL models yield better model fits.

## 6 Conclusion

Categorical dependent variables are widespread in the social sciences. Applied social scientists studying nominal dependent variables as a choice among discrete alternatives frequently hypothesize heterogeneous effects. We presented, discussed, and applied a general multinomial logit model (GHMNL) to account for heterogeneity in choice behavior systematically. The statistical theory and empirical applications provided in this paper suggest that the GHMNL model offers exciting insights into heterogeneous effects and better captures differences in choice behavior than competing approaches.

The GHMNL model integrates a heterogeneity term into the systematic part of the utility function and accounts for behavioral choice tendencies without referring to latent variables. The heterogeneity term is linked to explanatory variables, indicates the degree of distinctiveness of choice or the impact of heterogeneity in choice behavior. As demonstrated, alternative approaches come with several drawbacks, such as a high number of parameters to be estimated, identifiability problems, or the need to specify a specific and appropriate distribution for the random effects. The GHMNL avoids these difficulties, is computationally straightforward, and has convenient properties.

We illustrated the approach by analyzing electoral choices, highlighting the important insights possible from systematically modeling heterogeneity in voting behavior. We also provided empirical comparisons with alternative models, demonstrating that the GHMNL models outperform the competing models in all applications. As many research questions in the social sciences involve theoretical expectations about heterogeneity, we see a wide range of applications. We hope this contribution fosters the application of this type of model in applied social science work.

---

<sup>6</sup> The Akaike Information Criterion (AIC) is defined by  $AIC = -2l(\hat{\delta}) + 2b$ , the Bayesian Information Criterion (BIC) by  $BIC = -2l(\hat{\delta}) + \log(n)b$ , where  $l(\hat{\delta})$  is the log-likelihood function computed at the maximum of the estimated parameter vector  $\hat{\delta}$  and  $b$  is the number of model parameters.

**Funding** Open Access funding provided thanks to the CRUE-CSIC agreement with Springer Nature. Funding for open access charge: Universidad de Málaga / CBUA.

**Supplementary Information** The online version contains supplementary material available at <https://doi.org/10.1007/s10260-022-00642-5>.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

## References

- Adams J, Merrill S III, Grofman B (2005) A unified theory of party competition: a cross-national analysis integrating spatial and behavioral factors. Cambridge University Press, New York, NY
- Agresti A (2007) An introduction to categorical data analysis, 2nd edn. Wiley, Hoboken, NJ
- Alvarez RM, Nagler J (2004) Party system compactness: measurement and consequences. *Polit Anal* 12(1):46–62
- Converse PE (1964) The nature of belief systems in the mass public. In: Apter DE (ed) *Ideology and Discontent*. Free Press, New York, NY
- Davis OA, Hinich MJ, Ordeshook PC (1970) An expository development of a mathematical model of the electoral process. *Am Polit Sci Rev* 64(2):426–448
- DeShazo J, Fermo G (2002) Designing choice sets for stated preference methods: the effects of complexity on choice consistency. *J Environ Econ Manag* 44(1):123–143
- Downs A (1957) An economic theory of democracy. Harper & Row, New York, NY
- Edwards GC III, Mitchell W, Welch R (1995) Explaining presidential approval: The significance of issue salience. *Am J Polit Sci* 39(1):108–134
- Fan J, Zhang W (1999) Statistical estimation in varying coefficient models. *Ann Stat* 27(5):1491–1518
- Fiebig DG, Keane MP, Louviere J et al. (2010) The generalized multinomial logit model: accounting for scale and coefficient heterogeneity. *Mark Sci* 29(3):393–421
- Gerber D, Nicolet S, Sciarini P (2015) Voters are not fools, or are they? Party profile, individual sophistication and party choice. *Eur Polit Sci Rev* 7(1):145–165
- Greene WH, Hensher DA, Rose J (2006) Accounting for heterogeneity in the variance of unobserved effects in mixed logit models. *Trans Res Part B: Methodol* 40(1):75–92
- Hastie T, Tibshirani R (1993) Varying-coefficient models. *J Royal Stat Soc B* 55(4):757–796
- Hensher D, Louviere J, Swait J (1998) Combining sources of preference data. *J Econ* 89(1–2):197–221
- Hensher DA, Greene WH (2003) The mixed logit model: the state of practice. *Transportation* 30(2):133–176
- Hensher DA, Rose JM, Greene WH (2015) *Applied choice analysis*, 2nd edn. Cambridge University Press, Cambridge, England
- Long JS (1997) *Regression models for categorical and limited dependent variables*. Sage, Thousand Oaks, CA
- Luskin RC (1987) Measuring political sophistication. *Am J Political Sci* 31(4):856–899
- MacDonald SE, Rabinowitz G, Lijsthaug O (1995) Political sophistication and models of issue voting. *British J Political Sci* 25(4):453–483
- Mauerer I (2016) A party-varying model of issue voting. A cross-national study. PhD thesis, University of Munich (LMU), Germany
- Mauerer I, Pöbnecker W, Thumer PW et al. (2015) Modeling electoral choices in multiparty systems with high-dimensional data: a regularized selection of parameters using the lasso approach. *J Choice Model* 16(3):23–42

- Mauerer I, Thurner PW, Debus M (2015) Under which conditions do parties attract voters' reactions to issues? Party-varying issue voting in German elections 1987–2009. *West Eur Polit* 38(6):1251–1273
- McFadden DL (1974) Conditional logit analysis of qualitative choice behaviour. In: Zarembka P (ed) *Frontiers in Econometrics*. Academic Press, New York, NY, pp 105–142
- McFadden DL, Train KE (2000) Mixed MNL models for discrete response. *J Appl Economet* 15(5):447–470
- Park BU, Mammen E, Lee YK et al. (2015) Varying coefficient regression models: a review and new developments. *Int Stat Rev* 83(1):36–64
- RePass DE (1971) Issue salience and party choice. *Am Political Sci Rev* 65(2):389–400
- Rivers D (1988) Heterogeneity in models of electoral choice. *Am J Political Sci* 32(3):737–757
- Roßteutscher S, Schmitt-Beck R, Schoen H, et al. (2018) Pre- and post-election cross section (cumulation) (GLES 2017). ZA6802, Data file Version 3.0.0. Cologne, Germany: GESIS Data Archive. <https://doi.org/10.4232/1.13139>
- Thurner PW (2000) The empirical application of the spatial theory of voting in multiparty systems with random utility models. *Elect Stud* 19(4):493–517
- Train KE (2009) *Discrete choice methods with simulation*, 2nd edn. Cambridge University Press, New York
- Tutz G (2012) *Regression for categorical data* Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press, New York
- Tutz G (2021) Uncertain choices: the heterogeneous multinomial logit model. *Sociol Methodol* 51(1):86–111
- Weßels B, Schmitt H (2008) Meaningful choices, political supply, and institutional effectiveness. *Elect Stud* 27(1):19–30

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.