



LUDWIG-
MAXIMILIANS-
UNIVERSITÄT
MÜNCHEN

INSTITUT FÜR STATISTIK



Gerhard Tutz

Modelling Heterogeneity: On the Problem of Group Comparisons with Logistic Regression and the Potential of the Heterogeneous Choice Model

Technical Report Number 220, 2019
Department of Statistics
University of Munich

<http://www.statistik.uni-muenchen.de>



Modelling Heterogeneity: On the Problem of Group Comparisons with Logistic Regression and the Potential of the Heterogeneous Choice Model

Gerhard Tutz

Ludwig-Maximilians-Universität München

Akademiestraße 1, 80799 München

January 3, 2019

Abstract

The comparison of coefficients of logit models obtained for different groups is widely considered as problematic because of possible heterogeneity of residual variances in latent variables. It is shown how the heterogeneous logit model can be used to account for this type of heterogeneity. A strategy to obtain models with identifiable parameters is proposed, which works in most applications. In contrast to the common understanding, the heterogeneous logit model is considered as a model that contains effect modifying terms, which are not necessarily linked to variances but can also represent other types of heterogeneity in the population. The alternative interpretation of the parameters in the heterogeneous logit model makes it a flexible tool that can account for various sources of heterogeneity. Although the model is typically derived from latent variables it is important that for the interpretation of parameters the reference to latent variables is not needed. Latent variables are considered as a motivation for binary models, but the effects in the models can be interpreted as effects on the binary response.

Keywords: Heterogeneous choice model; Location-scale model; Heterogeneity of variances; Logit model; Group comparisons

1 Introduction

Allison (1999) demonstrated that comparisons of binary model coefficients across groups can be misleading if one has underlying heterogeneity of residual variances. If one compares, the regression coefficients of a set of explanatory variables like age, income, social status on a binary response one might find different coefficients for the gender groups although the effects of the explanatory variables on

the response have equal strengths in the underlying model. The reason is that coefficients are confounded with differences in variation across gender groups. Since Allison's paper the issue has been investigated in various papers, see Williams (2009), Mood (2010), Rohwer (2015), Karlson et al. (2012), Breen et al. (2014).

More recently, Kuha and Mills (2017) (henceforth KM) tried to convince readers that the problem is much less serious. In their concluding remarks they say that if researchers make sure to be clear about their target quantities of their analysis "they will in most cases be able to conclude that comparisons of estimates from such models between different groups or between different models pose no fundamental problems or at least not the kinds of problems that have been raised in the literature on this question." They conclude "that the problem of group comparisons is largely chimerical and that any remaining difficulties arise from expecting these techniques to do things they were never designed to do in the first place" (p.4).

KM reasonably distinguish between two different versions of the problem, which have not been clearly separated in the literature. The first version is caused by heterogeneity of variances in the underlying latent regression model, the second refers to the comparison of effects between groups, when the individuals in different groups have different distributions of additional explanatory variables, which may also be seen as unobserved heterogeneity. The focus here is on the first problem and heterogeneity as used here refers to heterogeneity of variances. KM argue that the most important solutions to the problem are to conclude that for the current research question the problem does not exist or that one should improve the measurement of the response variable. Improving the response variable is certainly a good concept, however, no solution for researchers who have only binary data. The conclusion that the problem does not exist if one focuses on the binary response variables and ignores potentially existing latent variable is not without problems. It might have been inspired by the fact that when only one explanatory variable is available a comparison of effects between groups can indeed not account for heterogeneity because of identifiability issues. As will be shown in the case of just one explanatory variable one can definitely not identify the source of the differences between coefficients obtained for the groups. Since KM mostly consider the case with only one explanatory variable they ignore that the sources of differing coefficients can be modelled if one has more than one variable, which is the case in most studies. The modelling may be based on the heteroscedastic logit model, also known as heterogeneous choice model, which is strongly related to the modelling of interaction effects as has been noted by Rohwer (2015) and Tutz (2018).

The heterogeneous choice model has been used before to model unobserved heterogeneity by Williams (2009) and implicitly by Allison (1999), however, identifiability issues have not been investigated in detail. It will be shown that parameters are identifiable under weak conditions that are met in most modelling problems. The heterogeneous choice model uses a predictor that is composed of

two factors, one factor contains the linear effects of explanatory variables, the other modifies these effects as a function of further variables. The latter factor models heterogeneity in the population, it can be interpreted as a model for the heterogeneity of residual variances, but, as will be demonstrated, it can also be interpreted as a model for the response behaviour. The present article can be seen as an investigation of the potential of the heterogeneous choice model to account for different forms of heterogeneity.

In the first part of the article (Section 2 and 3) we consider the problem how to compare coefficients obtained from different groups. In Section 4 we investigate the effects of covariates on the binary response. In Section 5 an alternative derivation of the model is given that sheds some light on the problems of interpreting coefficients. The paper concludes with summarizing comments.

2 Comparing Coefficients Across Logit Models

We will first consider the technical side of the problem, which has been exposed in several papers before. It is needed to clarify issues and introduce the notation.

2.1 Binary Regression Models and Latent Variables

A common way to derive and motivate binary regression models is by assuming that a continuous latent variable is behind the response. Let the latent regression model have the form

$$Y_i^* = \alpha_0 + x_{i1}\alpha_1 + \cdots + x_{ip}\alpha_p + \sigma\varepsilon_i,$$

where ε_i has symmetric distribution function $F(\cdot)$. Although one may use any symmetric distribution function, for simplicity we restrict consideration to the logistic distribution $F(\eta) = \exp(\eta)/(1 + \exp(\eta))$. The essential concept is to consider Y_i as a dichotomized version of the latent variable Y_i^* with the link between the observable binary variable Y_i and the latent variable Y_i^* given by

$$Y_i = 1 \quad \text{if} \quad Y_i^* \geq \theta, \tag{1}$$

where θ is some unknown threshold. It is straightforward to derive that one obtains for the probability of $Y_i = 1$, $\pi_i = P(Y_i = 1|\mathbf{x}_i)$, the logit model

$$\text{logit}(\pi_i) = \beta_0 + x_{i1}\beta_1 + \cdots + x_{ip}\beta_p, \tag{2}$$

where the coefficients are given by

$$\beta_0 = \frac{\alpha_0 - \theta}{\sigma}, \beta_1 = \frac{\alpha_1}{\sigma}, \dots, \beta_p = \frac{\alpha_p}{\sigma}.$$

The basic assumption behind the derivation of the model is that the observable variable Y_i is a coarser, even binary version of the latent variable Y_i^* . In many

applications one can imagine that an underlying continuous variable steers the decision process that results in a categorical outcome.

It is obvious that the α -parameters are not identifiable if the variable Y^* is not observable, which is assumed here. Therefore, in common derivations of the model one typically postulates $\sigma = 1$ and $\theta = 0$. The latter postulate is equivalent to using $Y_i^* \geq 0$ in equ. (1), and only fixes the location on the latent scale. The more critical postulate is the former one. It fixes the scaling of the coefficients. Since σ is not known one could use any value for σ . The β -parameters of the binary model can be estimated but do not yield estimates of the α -parameters since the latter are only fixed up to a constant. This scaling problem raises problems when comparing coefficients obtained for different groups.

2.2 Coefficients from Different Groups

Let us now consider the comparison of coefficients of separately fitted logit models with a fixed set of covariates. Let the latent variable for the two groups, for example females and males, be given by

$$Y_i^* = \alpha_0^{(r)} + x_{i1}\alpha_1^{(r)} + \dots + x_{ip}\alpha_p^{(r)} + \sigma^{(r)}\varepsilon_i, \quad r = 0, 1,$$

with coefficients that may vary across groups. One obtains two logit models, one in each group,

$$\text{logit}(\pi_i) = \beta_0^{(r)} + x_{i1}\beta_1^{(r)} + \dots + x_{ip}\beta_p^{(r)} \quad r = 0, 1, \quad (3)$$

with the coefficients of the models given by

$$\beta_j^{(r)} = \frac{\alpha_j^{(r)}}{\sigma^{(r)}}, \quad j = 1, \dots, p \quad r = 0, 1.$$

The comparison of coefficients of the underlying regression model is determined by $\alpha_j^{(1)}/\alpha_j^{(0)}$, whereas the comparison of coefficients across groups obtained from the fitting of separate logit models is determined by

$$\frac{\beta_j^{(1)}}{\beta_j^{(0)}} = \frac{\alpha_j^{(1)} \sigma^{(0)}}{\alpha_j^{(0)} \sigma^{(1)}} = \frac{\alpha_j^{(1)}}{\alpha_j^{(0)}} \delta,$$

where $\delta = \sigma^{(0)}/\sigma^{(1)}$ is a scaling factor. For illustration, let us consider the simple case, in which the coefficients in the underlying regression model are the same in both groups, $\alpha_j^{(0)} = \alpha_j^{(1)}$. If the standard deviation in the second group is twice the standard deviation in the first group one obtains $\beta_j^{(1)}/\beta_j^{(0)} = 2$, although the effect strengths are the same in the underlying model.

Allison (1999) gave an illustration of this potentially misleading effect sizes. In his example the binary response was promotion to associate professor from

the assistant professor level and he found that the effect of the number of articles published by the end of each person year was about twice as large for men than for women. If coefficients are larger for men, "men get a greater payoff from their published work than do females" (Williams, 2009), which seems very unfair. However, the question is if the effect is substantial or due to unobserved heterogeneity of variances.

It is obvious that inference on the α -parameters is not possible because they are coefficients of a latent variable, which by definition can not be observed. In particular one cannot estimate the variances $\sigma^{(r)}$ of the latent variable, which are responsible for the different scaling in the groups. The consequences KM draw is to make some serious effort to measure the latent variable itself. This is indeed one way to go, however it does not solve the problems of all the researchers that have to work with binary data because they are the only data available.

KM argue that the derivation of the models by the latent variable mostly has no implications for the interpretation of the regression coefficients of the binary model. If the binary variable is the substantially interesting response variable the latent variable Y_i^* is "nothing but a hypothetical mathematical device that may be introduced if it is helpful for motivating the binary response." (p.9). Often one can imagine that a latent variable Y_i^* is behind the binary response, however, it is indeed hypothetical. We also see the latent variable as hypothetical, it primarily serves to motivate the binary response model. It is not needed when interpreting regression coefficients of the binary response model. However, as will be argued in the following, the consideration of the latent regression model is useful to obtain models, which allow to examine the impact of explanatory variables in an alternative way. It is important that one does not need to refer to the latent variable when interpreting parameters in the extended models.

2.3 Comparing Groups: A Closed Model Version of the Problem

Let us consider again the problem of comparing coefficients obtained from logit models that hold in subpopulations or groups. There is no need to fit logit models separately in groups. One can also fit a model for both groups simultaneously, that is, in the total population by including an indicator variable for the group. This has conceptual advantages and directly provides tests that compare coefficients within the framework of generalized linear models. Let the indicator variable be $x_{i0} = 1$ in group 1 and $x_{i0} = 0$ in group 0. The model for the total population that corresponds to the models for separate groups is the logit model with specific interactions

$$\text{logit}(\pi_i) = \beta_{00} + x_{i0}\beta_0 + x_{i1}\beta_1 + \cdots + x_{ip}\beta_p + x_{i0}x_{i1}\beta_{01} + \cdots + x_{i0}x_{ip}\beta_{0p}, \quad (4)$$

The model contains main effects of the indicator variable and the explanatory variables and all the interactions between the group indicator and the explanatory variables. It is just a reparameterized version of the logit models (3). It is

straightforward to derive the reparameterization, for example, one obtains that the main effects are given by $\beta_0 = \beta_0^{(1)} - \beta_0^{(0)}$, $\beta_j = \beta_j^{(0)}$ and the interaction effects by $\beta_{0j} = \beta_j^{(1)} - \beta_j^{(0)}$. Therefore, assuming that model (4) holds is equivalent to assuming that the models (3) hold in the corresponding subpopulations. Fitting of the model in the population is equivalent to fitting separate logit models in the subpopulations.

The model in the total population can be derived from an underlying regression model as before (see following section). The essential point here is that one has a model for the total population, which comprises both groups.

3 Models for Unobserved Heterogeneity

In this section we consider the modelling of unobserved heterogeneity. A widely used model that accounts for unobserved heterogeneity is the heterogeneous choice model. It has already been used by Allison (1999) in the original paper on heterogeneity but without reference to the model. Later, Williams (2009) showed that Allison’s method is a special case of the heterogeneous choice model. We will in particular investigate under which conditions it is a sensible approach to account for heterogeneity. We start with a brief derivation and description of the model.

3.1 The Heterogeneous Choice Model

The heterogeneous choice model can be derived from the assumption of an underlying latent variable in a similar way as the logit model. Let the latent variable be given by $Y_i^* = \beta_{00} + \mathbf{x}_i^T \boldsymbol{\alpha} + \sigma_i \varepsilon_i$, where the standard deviation of the noise, σ_i , is linked to the observation. One assumes that the standard deviation is determined by the so-called *variance equation*

$$\sigma_i = \exp(\mathbf{z}_i^T \boldsymbol{\gamma}),$$

where \mathbf{z}_i is a vector of covariates and $\boldsymbol{\gamma}$ a vector of coefficients. Simple derivation shows that the resulting *heterogeneous choice model* has the form

$$P(Y_i = 1 | \mathbf{x}_i, \mathbf{z}_i) = F\left(\frac{\alpha_{00} + \mathbf{x}_i^T \boldsymbol{\alpha}}{\exp(\mathbf{z}_i^T \boldsymbol{\gamma})}\right), \quad (5)$$

where \mathbf{x}_i is a vector of explanatory variables and $\boldsymbol{\alpha}$ collects the coefficients. The model explicitly accounts for heterogeneous variances, which are determined by the covariate vector \mathbf{z}_i . The term in the numerator, $\mathbf{x}_i^T \boldsymbol{\alpha}$, can be seen as the location term, referring to the location on the continuum of the underlying variable, and $\mathbf{z}_i^T \boldsymbol{\gamma}$ as the scaling term. The parameter $\boldsymbol{\gamma}$ is referred to as *scaling* or *dispersion parameter*. The model is also known as *heteroscedastic logit model* and in its ordinal version as *location-scale model* (McCullagh, 1980).

3.2 The Heterogeneous Choice Model and the Comparison of Groups

The heterogeneous choice model seems to offer the possibility to model underlying heterogeneity of variances. It certainly is a sensible model if the vectors \mathbf{x}_i and \mathbf{z}_i , which determine location and variance, are distinct. However, some care is needed if the two vectors contain the same variables. This is exactly the situation that occurs if one considers the heterogeneous choice model that corresponds to the logit models in two groups.

Let again x_{i0} denote the group indicator and let the vector \mathbf{x}_i and the corresponding coefficient vector be given by

$$\mathbf{x}_i^T = (x_{i0}, x_{i1}, \dots, x_{ip}, x_{i0}x_{i1}, \dots, x_{i0}x_{ip}), \quad \boldsymbol{\alpha}^T = (\alpha_0, \alpha_1, \dots, \alpha_p, \alpha_{01}, \dots, \alpha_{0p}).$$

The parameters $\alpha_0, \alpha_1, \dots, \alpha_p$ represent main effects, the parameters $\alpha_{01}, \dots, \alpha_{0p}$ represent interaction effects between the group indicator and the variables. Heterogeneity of variances over groups is obtained by specifying $\sigma_i = \exp(x_{i0}\gamma)$, which means $\sigma_i = \exp(\gamma)$ if $x_{i0} = 1$ and $\sigma_i = 1$ if $x_{i0} = 0$. Implicitly group 0 with a standard deviation $\sigma_i = 1$ is used as the reference group.

It is straightforward to derive the logit models in the two groups, which are given by

$$\begin{aligned} \text{logit}(\pi_i) &= \alpha_{00} + x_{i1}\alpha_1 + \dots + x_{ip}\alpha_p && \text{in group 0,} \\ \text{logit}(\pi_i) &= \{\alpha_{00} + \alpha_0 + x_{i1}(\alpha_1 + \alpha_{01}) + \dots + x_{ip}(\alpha_p + \alpha_{0p})\} / \exp(\gamma) && \text{in group 1.} \end{aligned}$$

Thus the parameters in group 0 are given by $\beta_0^{(0)} = \alpha_{00}$, $\beta_j^{(0)} = \alpha_j$, $j = 1, \dots, p$, in group 1 by $\beta_0^{(1)} = (\alpha_{00} + \alpha_0) / \exp(\gamma)$, $\beta_j^{(1)} = (\alpha_j + \alpha_{0j}) / \exp(\gamma)$, $j = 1, \dots, p$. However, in the groups fully flexible models, which means with potentially different coefficients for all variables, are already obtained if one sets $\gamma = 0$. Consequently, γ is not identifiable in the corresponding heterogeneous choice model. That means, potential heterogeneity of variances cannot be captured by the model.

The derivation of the heterogeneous choice model given above uses the term $\mathbf{x}_i^T \boldsymbol{\alpha}$ in the latent model. The obtained parameterization is very general but it hides the link to the parameters in the latent models for the separate groups. For completeness we give the derivation in a parameterization that provides this link. An alternative parameterization of the latent model (with just one explanatory variable) is given by

$$Y_i^* = \alpha_0^{(0)} + x_{i0}(\alpha_0^{(1)} - \alpha_0^{(0)}) + x_{i1}\alpha_1^{(0)} + x_{i0}x_{i1}(\alpha_1^{(1)} - \alpha_1^{(0)}) + \{\sigma^{(0)} + x_{i0}(\sigma^{(1)} - \sigma^{(0)})\}\varepsilon_i, \quad (6)$$

where $\sigma_i = \sigma^{(0)} + x_{i0}(\sigma^{(1)} - \sigma^{(0)})$ represents the heterogeneity of variances. It is immediately seen that the latent traits are $Y_i^* = \alpha_0^{(0)} + x_{i1}\alpha_1^{(0)} + \sigma^{(0)}\varepsilon_i$ if $x_{i0} = 0$ and $Y_i^* = \alpha_1^{(1)} + x_{i1}\alpha_1^{(1)} + \sigma^{(1)}\varepsilon_i$ if $x_{i0} = 1$.

Thus, the parameters of the logit models in the separate groups, $\text{logit}(\pi_i) = \beta_0^{(j)} + \mathbf{x}_i^T \boldsymbol{\beta}^{(j)}$, $j = 1, 2$, are given by

$$\beta_0^{(1)} = \frac{\alpha_0^{(1)}}{\sigma^{(1)}}, \beta_1^{(1)} = \frac{\alpha_1^{(1)}}{\sigma^{(1)}}, \quad \beta_0^{(0)} = \frac{\alpha_0^{(0)}}{\sigma^{(0)}}, \beta_1^{(0)} = \frac{\alpha_1^{(0)}}{\sigma^{(0)}},$$

Again the comparison of α - parameters yields different results than the comparison of β - parameters because

$$\frac{\alpha_j^{(1)}}{\alpha_j^{(0)}} = \frac{\beta_j^{(1)} \sigma^{(1)}}{\beta_j^{(0)} \sigma^{(0)}} = \frac{\beta_j^{(1)}}{\beta_j^{(0)}} \gamma \text{ for } j = 0, \dots, p.$$

Consequently, comparison is distorted by the scaling factor $\gamma = \sigma^{(1)}/\sigma^{(0)}$.

3.3 Obtaining Models with Identifiable Parameters

It seems that the modelling of latent variances in the two group problem fails because γ is not identified in the heterogeneous choice model. This is indeed true if one has only one predictor, however, in the case of more than one predictors there are strategies to obtain estimates of γ that typically are successful.

Models with One Explanatory Variable

Let us first consider the heterogeneous choice model with just one explanatory variable x_{i1} ,

$$\text{logit}(\pi_i) = \frac{\alpha_{00} + x_{i0}\alpha_0 + x_{i1}\alpha_1 + x_{i0}x_{i1}\alpha_{01}}{\exp(x_{i0}^T \boldsymbol{\gamma})}. \quad (7)$$

If one has only *one* explanatory variable in the model the situation is certainly hopeless. The model is not identified and variances cannot be modelled. This is the case that is considered in detail in KM. They consider among other examples the case in which the group represent experimental conditions with $x_{i0} = 1$ indicating treatment and $x_{i0} = 0$ the control group in a randomized experiment. That means all the conditions are fulfilled to examine the effect of treatment on the binary response. One is in a situation that reminds of the paper of Mood (2010), which says in the header ‘‘Logistic regression: Why we cannot do what we think we can do...’’. Indeed one is in a situation where one cannot obtain what one might want. This should be accepted and admitted but may be seen under a wider view. When we fit logit models we make several assumptions without being able to examine all of them. We assume that random variables that generate observations are independent, that they have the same distribution given covariates, for continuous variables often a linear effect is assumed. In particular the assumption that all individuals follow the same distribution (given covariates) is simplistic but necessary. In repeated measurements one could at least include a random intercept that allows each person an own level, however,

with just one observation per individual this does not work because random effects are not identifiable. The problem here is of a similar nature, heterogeneity of variances cannot be identified in a simple setting with just one predictor. It might be present but cannot be examined. Thus, when comparing parameters from different groups one should be aware of the problem that one implicitly assumes that no heterogeneity is present, which might be seen as one further assumption when using the logit model.

Models with a Vector of Explanatory Variables

The situation is different when more predictors are available. Let us consider the heterogeneous choice model with a vector of explanatory variables. Since we still consider the two group problem all interactions between the indicator and the covariates are included in the predictor, yielding the model

$$\text{logit}(\pi_i) = \frac{\alpha_{00} + x_{i0}\alpha_0 + x_{i1}\alpha_1 + \cdots + x_{ip}\alpha_p + x_{i0}x_{i1}\alpha_{01} + \cdots + x_{i0}x_{ip}\alpha_{0p}}{\exp(x_{i0}^T\gamma)}. \quad (8)$$

The linear term is kept very general such that the resulting logit models in the separate groups $\text{logit}(\pi_i) = \alpha_0^{(j)} + \mathbf{x}_i^T \boldsymbol{\alpha}^{(j)}$, $j = 0, 1$, can have any parameters $\alpha_0^{(j)}$, $\boldsymbol{\alpha}^{(j)}$. Consequently, in this general model parameters are again not identified. However, when several predictors are available not all parameters are unequal zero, typically at least some interaction effects can be omitted. This can be used to obtain identifiable sub models.

The testing strategy is simple. First, one exploits that the relevance of interaction terms can be tested within the framework of generalized linear models. Since γ is not identifiable model (8) is equivalent to the model

$$\text{logit}(\pi_i) = \tilde{\alpha}_{00} + x_{i0}\tilde{\alpha}_0 + x_{i1}\tilde{\alpha}_1 + \cdots + x_{ip}\tilde{\alpha}_p + x_{i0}x_{i1}\tilde{\alpha}_{01} + \cdots + x_{i0}x_{ip}\tilde{\alpha}_{0p}, \quad (9)$$

in which parameters are identified (given sufficient diversity of observations). One can consider the parameters $\tilde{\alpha}_{index}$ as being given by $\alpha_{index}/\exp(x_{i0}^T\gamma)$.

Given model (9) one can test the hypotheses $H_0 : \tilde{\alpha}_{0j} = 0$, $j = 1, \dots, p$. If one of the hypotheses, say $H_0 : \tilde{\alpha}_{0p} = 0$, is not rejected one can leave out the corresponding interaction term $x_{i0}x_{ip}$ and assume that the reduced model

$$\text{logit}(\pi_i) = \tilde{\alpha}_{00} + x_{i0}\tilde{\alpha}_0 + x_{i1}\tilde{\alpha}_1 + \cdots + x_{ip}\tilde{\alpha}_p + x_{i0}x_{i1}\tilde{\alpha}_{01} + \cdots + x_{i0}x_{i,p-1}\tilde{\alpha}_{0,p-1} \quad (10)$$

holds. Then one can fit the corresponding heterogeneous logit model

$$\text{logit}(\pi_i) = \frac{\alpha_{00} + x_{i0}\alpha_0 + x_{i1}\alpha_1 + \cdots + x_{ip}\alpha_p + x_{i0}x_{i1}\alpha_{01} + \cdots + x_{i0}x_{i,p-1}\alpha_{0,p-1}}{\exp(x_{i0}^T\gamma)}, \quad (11)$$

which can be estimated because parameters are identifiable (for a proof see Appendix). The model is actually more general than model (10) but as a heterogeneous logit model it allows to investigate which variables and interactions are

actually needed if one allows for the term $\exp(x_{i0}^T \gamma)$ and therefore heterogeneity of variances. Typically the model can be further simplified by omitting irrelevant terms.

The basic procedure is simple. Since models (8) and (9) are equivalent one investigates if interaction terms can be excluded. If $\tilde{\alpha}_{0p} = 0$ is zero also $\alpha_{0p} = 0$. But model (10) with $\alpha_{0p} = 0$ has identifiable parameters and therefore one can estimate parameters and further simplify the obtained model. Of course, when simplifying model (9) one can also leave out all the interaction terms that are not needed simultaneously. The modelling strategy is illustrated in the next section.

TABLE 1: *Interaction model for importance of sports data*

	Estimate	Std. Error	z value	Pr(> z)
Gender	-1.4485	0.5216	-2.78	0.0055
MaritalStatus2	-0.6944	0.2690	-2.58	0.0098
MaritalStatus3	-1.4005	0.4627	-3.03	0.0025
MaritalStatus4	-1.7674	0.8829	-2.00	0.0453
age	-0.0288	0.0081	-3.55	0.0004
Family	-0.2062	0.0708	-2.91	0.0036
Gender:MaritalStatus2	0.6285	0.3347	1.88	0.0605
Gender:MaritalStatus3	1.2774	0.5831	2.19	0.0285
Gender:MaritalStatus4	1.2976	0.9951	1.30	0.1922
Gender:age	-0.0014	0.0104	-0.14	0.8922
Gender:Family	0.1168	0.0897	1.30	0.1925

Illustrative Examples

For illustration we use a data set that is available from <https://cran.r-project.org/web/packages/CUB/index.html> (dataset relgoods). The original study aimed at measuring the evaluation of various relational goods and leisure time, and was collected in December 2014. In our use of the data the binary response refers to the importance of sports with category 1 representing that sport is important in life. As explanatory variables we include gender (0: men, 1: women), age in years, the number of members of the family (variable Family) Marital Status (1=Unmarried, 2=Married/Cohabitant, 3= Separated/Divorced, 4=Widower). We investigate if gender generates scaling effects.

Table 1 shows the fit of a logit model with all main effects and interactions with gender. Among the interaction effects only the interaction with marital status is significant (likelihood ratio test is 9.833 on 3 df). Therefore, we keep only the interaction effect between gender and marital status and fit the corresponding heterogeneous logit model. Since interactions were omitted the model parameters are identifiable. In the following tables we give fits of the reduced model in the parameterization (11), which includes a dispersion parameter. It is seen from

Table 2 that the interaction with marital status becomes irrelevant if a dispersion parameter is included. In a further step it is excluded yielding Table 3. In the reduced model the dispersion parameter is highly significant and also the main effects of the other variables, with the exception of gender, cannot be omitted. Nevertheless, compared to the other models the reduced model given in Table 3 is rather sparse. It is also the best model in terms of AIC, AIC is 2394.00 for the interaction model, 2393.77 for the heterogeneous choice model with interaction, and 2389.99 for the main effects heterogeneous choice model.

TABLE 2: *Heterogeneous choice model for importance of sports data*

	Estimate	Std. Error	z value	Pr(> z)
Gender	-0.8813	0.4419	-1.994	0.0461
MaritalStatus2	-0.6119	0.2555	-2.395	0.0166
MaritalStatus3	-1.2667	0.4468	-2.835	0.0045
MaritalStatus4	-1.5906	0.8644	-1.840	0.0657
age	-0.0315	0.0076	-4.151	0.0000
Family	-0.1435	0.0576	-2.492	0.0127
Gender:MaritalStatus2	0.4769	0.3679	1.296	0.1949
Gender:MaritalStatus3	1.0308	0.6137	1.680	0.0930
Gender:MaritalStatus4	0.9333	1.0721	0.871	0.3840
disp.Gender	0.1010	0.3334	0.303	0.762

TABLE 3: *Reduced heterogeneous choice model for importance of sports data*

	Estimate	Std. Error	z value	Pr(> z)
Gender	-0.2394	0.2426	-0.987	0.3236
MaritalStatus2	-0.4682	0.2147	-2.180	0.0292
MaritalStatus3	-0.9394	0.3866	-2.430	0.0151
MaritalStatus4	-1.3643	0.6005	-2.272	0.0230
age	-0.0360	0.0065	-5.492	0.0000
Family	-0.1791	0.0565	-3.170	0.0015
disp.Gender	0.4795	0.1728	2.774	0.00553

As a second example we consider data from the general social survey of social science, in short ALLBUS, a study by the German institute GESIS. The data is available from <http://www.gesis.org/allbus>. Our analysis is based on a subset containing 2935 respondents of the ALLBUS in 2012. The response is the confidence in the federal government (1: strong confidence, 0: no confidence). As explanatory variables we consider the gender (0: male, 1: female), the income in thousands of Euros with a linear and a quadratic term, the age in decades (centered at 50) with a linear and a quadratic term and the self reported interest

in politics from 1 (very strong interest) to 5 (no interest at all). Table 4 shows the fit of the interaction model. Since the interaction between gender and income is not significant it is omitted, yielding the heterogeneous logit model given in Table 5. After deleting the interaction effects, which are not needed, one obtains the reduced main effects model in Table 6.

TABLE 4: *Fit of interaction model for confidence data*

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	0.5250	0.2012	2.61	0.0091
gender	0.5680	0.3092	1.84	0.0662
interest	-0.1703	0.0495	-3.44	0.0006
income	0.1526	0.0504	3.03	0.0024
age	-0.0364	0.0320	-1.14	0.2548
agesqrt	0.0726	0.0183	3.96	0.0001
gender:interest	-0.1413	0.0758	-1.86	0.0624
gender:income	0.0315	0.0932	0.34	0.7357
gender:age	-0.1360	0.0475	-2.86	0.0042
gender:agesqrt	0.0288	0.0259	1.11	0.2659

TABLE 5: *Fit of heterogeneous model for confidence data*

	Estimate	Std. Error	z value	Pr(> z)
gender	0.0215	0.1210	0.178	0.8586
interest	-0.1856	0.0418	-4.440	0.0000
income	0.1327	0.0395	3.359	0.0007
age	-0.0367	0.0319	-1.148	0.2508
agesqrt	0.0708	0.0181	3.907	0.0009
gender:age	-0.0692	0.0437	-1.583	0.1135
gender:agesqrt	-0.0073	0.024503	-0.299	0.7647
disp:gender	-0.4747	0.2565	-1.851	0.0642

TABLE 6: *Fit of reduced heterogeneous model for confidence data*

	Estimate	Std. Error	z value	Pr(> z)
gender	-0.0399	0.1133	-0.352	0.7247
interest	-0.1748	0.0373	-4.678	0.0000
income	0.1281	0.0382	3.350	0.0008
age	-0.0810	0.0195	-4.153	0.0000
agesqrt	0.0610	0.0137	4.454	0.0000
disp:gender	-0.5567	0.2168	-2.568	0.0102

In both applications the linear predictor in the heterogeneous model is reduced to a main effects model after including the dispersion parameter. Thus, the interactions between gender and the other covariates, which will show in different effects in groups when separate models are fitted, seem to be generated by variance heterogeneity and one might be tempted to see them as artifacts. However, then interpretation has to refer to the latent model, which should and can be avoided. Instead of sticking to the latent trait model one should take the binary response seriously and interpret effect of covariates instead of single coefficients, which is outlined in the next section.

4 Beyond Unobserved Heterogeneity: The Effects of Variables in the Heterogeneous Choice Model

Following the construction of the model, the parameters in the heterogeneous choice model are traditionally interpreted as location and dispersion effects, which means that implicitly one uses the latent variable when interpreting parameters. However, as in the derivation of the simple logit model it can be seen as a mere motivation and effects on the variables can be interpreted without reference to a latent variable.

Let us first consider the case where the location term contains the binary variable x_{i0} , which is also present in the variance term, but no interactions between x_{i0} and explanatory variables are included. Models of this form were obtained in the applications. In general, the model has the form

$$\text{logit}(\pi_i) = \{\alpha_{00} + x_{i0}\alpha_0 + x_{i1}\alpha_1 + \dots + x_{ip}\alpha_p\} / \exp(x_{i0}\gamma).$$

The predictor term on the right hand side is composed of two components, the usual linear term, which contains the explanatory variables, and the factor $\exp(-x_{i0}^T\gamma)$. The latter can be seen as an *effect modifying term* since it modifies the effects of the explanatory variables. Since the seminal paper of Hastie and Tibshirani (1993) varying coefficients models are in common use, see also Cai et al. (2000), Fan and Zhang (1999), Park et al. (2015). They allow that all or parts of the parameters in a regression model are functions of other variables, the so-called effect modifiers. As shown in the following x_{i0} is a effect modifier which simultaneously modifies the effects of x_{i1}, \dots, x_{ip} .

We use in particular that in logit models the impact of a covariate effect can be measured by the increase of the logits if covariate x_{ij} increases to $x_{ij} + 1$. Let $\pi(x_{ij})$ denote the probability of observing $Y_i = 1$ when the j -th covariate has value x_{ij} and $\pi(x_{ij} + 1)$ denote the probability if the j -th covariate has value $x_{ij} + 1$; all other variables are kept fixed. One obtains for the effect strength of the j -th covariate ($j \geq 1$)

$$\text{logit}(\pi(x_{ij} + 1)) - \text{logit}(\pi(x_{ij})) = \alpha_j / \exp(x_{i0}\gamma).$$

That means the effect is α_j if $x_{i0} = 0$ and α_j/e^γ if $x_{i0} = 1$. The effect modifying factor changes the effects of all the explanatory variables, which may be seen as a specific interaction between x_{i0} and the explanatory variables. As in general varying coefficients models one allows for a modification of effects but there is no need to postulate a latent variable.

Before looking at examples let us consider the effect of the group variable, that is, we consider only the term $(\alpha_{00} + x_{i0}\alpha_0)/\exp(x_{i0}^T\gamma)$. It is immediately seen that the term is α_{00} in group 0, and $(\alpha_{00} + \alpha_0)/\exp(\gamma)$ in group 1. That means the effects, and therefore the basic response levels, are typically different in the two groups, provided $\gamma \neq 0$, even if $\alpha_0 = 0$.

For illustration let us consider the importance of sports example. The effect modifying factor generated by gender is $e^\gamma = e^{0.4795} = 1.6152$. Since $x_{i0} = 1$ codes women, the effects of variables are weaker in the female population. This is illustrated in Figure 1 where the odds for the two metrically scaled variables age and family are shown; drawn line refers to men, dashed line refers to women. The level for women is generally lower than for men, which is an effect of the modification of the term $\alpha_{00} + x_{i0}\alpha_0$, and means that women have a weaker tendency to consider sports important. The coefficient of gender in the linear predictor is very small, it could be even considered as being equal to zero. Nevertheless, the probabilities for men and women are distinctly different due to the effect modifying term. The modification of the effect strengths of the explanatory variables age and family is seen from the different slopes of the curves. For men the decrease is much stronger than for women.

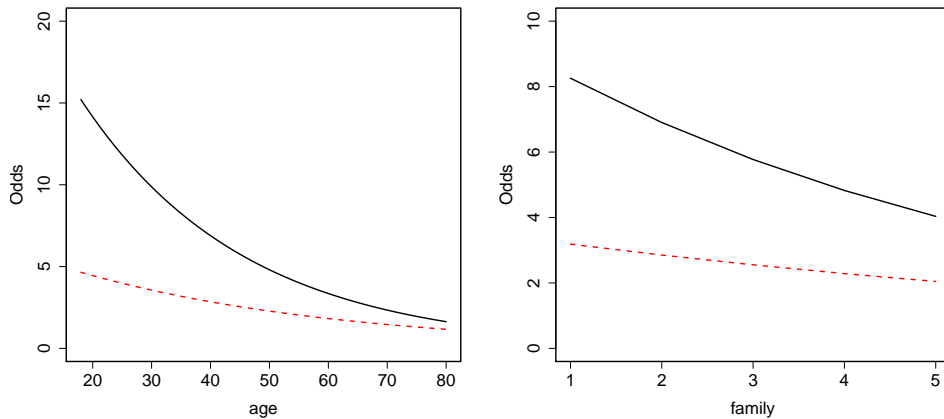


FIGURE 1: Odds for men (drawn line) and women (dashed line) for sports example plotted against age (Marital Status=0, Family=20) and family size (Marital Status=0, Age=40).

The effect modifying term seems rather restrictive since it modifies all the

effects of explanatory variables in the same way. However, this restriction is only strong if no further interactions are included in the linear predictor. If interaction effects are present the effects of variables are modified by the effect modifying term but they can nevertheless have stronger or weaker effects. Let us consider the general case in which some interactions are included in the linear part of the model. Then, $\eta_i = \alpha_{00} + x_{i0}\alpha_0 + x_{i1}\alpha_1 + \dots + x_{ip}\alpha_p$ is replaced by $\eta_i = \alpha_{00} + x_{i0}\alpha_0 + x_{i1}\alpha_1 + \dots + x_{ip}\alpha_p + \dots + x_{i0}x_{ij}\alpha_{ij} + \dots$, which contains some but, for reasons of identifiability, not all interaction terms between x_{i0} and explanatory variables. If an interaction term with the j -th covariate is included, the effect has to be modified to

$$\text{logit}(\pi(x_{ij} + 1)) - \text{logit}(\pi(x_{ij})) = (\alpha_j + x_{i0}\alpha_{0j})/\exp(x_{i0}\gamma).$$

This means, in addition to the effect modifying term one has the interaction effect α_{0j} in the linear predictor, which can be positive or negative and therefore change the effect. In particular, the interaction term can take values that allow to have the same effect in both groups. If $\alpha_{0j} = \alpha_j \exp(\gamma) - \alpha_j$ one obtains in both groups the coefficient α_j . Therefore, the general model does not postulate that all effects have to be different in the groups. The flexibility is obtained since the interaction effect α_{0j} is variable-specific in contrast to the effect modifying term. Although interactions in the linear effects are allowed, in our examples main effects turned out to be sufficient.

5 Models with Individual-Level Heterogeneity

As has already been shown in the previous section for the interpretation of effects latent variables are not needed. The case is strengthened in the following by giving a quite different motivation for the effect modifying term in the heterogeneous logit model. Let us consider the model

$$\text{logit}(\pi_i) = e^{\delta_i}(\alpha_0 + \mathbf{x}_i^T \boldsymbol{\alpha}), \quad (12)$$

where \mathbf{x}_i is a vector of covariates. The essential modification is that the factor e^{δ_i} is subject-specific and very general. Thus, each individual is allowed to have its own parameter e^{δ_i} .

As considered above, in logit models the impact of a covariate effect can be measured by the increase of the logits if covariate x_{ij} increases to $x_{ij} + 1$. One obtains for the effect of the j -th covariate

$$\text{logit}(\pi(x_{ij} + 1)) - \text{logit}(\pi(x_{ij})) = e^{\delta_i} \alpha_j.$$

That means the effect strength increases with δ_i . If δ_i is small the effect strength is weak, in the extreme case, $\delta_i \rightarrow -\infty$, the covariate has no effect at all. Thus, e^{δ_i} represents the subject-specific modification of the impact of covariates. One gets for extreme values of δ_i the following.

If $\delta_i \rightarrow \infty$ one has

$$P(Y_i = 1) = 1 \text{ if } \alpha_0 + \mathbf{x}_i^T \boldsymbol{\alpha} > 0, \quad P(Y_i = 0) = 1 \text{ if } \alpha_0 + \mathbf{x}_i^T \boldsymbol{\alpha} < 0.$$

Thus, individuals with a large δ_i show a strong preference for categories. They are strong discriminators, in a questionnaire, they definitely know which category, 0 or 1, they prefer.

If $\delta_i \rightarrow -\infty$ one obtains $P(Y_i = 0) = P(Y_i = 1) = 0.5$. Thus, individuals with a small δ_i choose rather randomly between the response categories. In a questionnaire this may be a sign that the person is undecided or lacks concentration.

In summary, the parameter δ_i can be seen as modelling the subject-specific decisiveness or discriminatory power. For large δ_i the person has strong preferences, for small δ_i the person tends to choose one of the response categories at random.

Although each person may have its own parameter, it can not be estimated without having repeated measurements. However, one can link the discriminatory power to covariates. A model of this type is obtained if one replaces the parameter δ_i by $x_{i0}\delta$. It can be seen as a model in which discriminatory power depends on the group. More general one can use the model

$$\text{logit}(\pi_i) = e^{\mathbf{z}_i^T \boldsymbol{\delta}} (\alpha_0 + \mathbf{x}_i^T \boldsymbol{\alpha}), \quad (13)$$

where \mathbf{z}_i is a vector of covariates and $\boldsymbol{\delta}$ is a parameter vector that determines how the distinctness of the response is determined by the covariates in \mathbf{z}_i . It is immediately seen that the model is equivalent to the heterogeneous choice model (5) if one sets $\boldsymbol{\delta} = -\boldsymbol{\gamma}$.

Let us first consider the case of a binary indicator variable in the factor and consider the effects of variables and their interpretation as distinctness when the linear predictor includes interactions. Then the effect strength of a variable has to be modified. One obtains for the effect strength of the j -th covariate, which is assumed to have an interaction effect,

$$\text{logit}(\pi(x_{ij} + 1)) - \text{logit}(\pi(x_{ij} - 1)) = e^{x_{i0}\delta} (\alpha_j + x_{i0}\alpha_{0j}).$$

Thus, the effect strength is determined by two components

$e^{x_{i0}\delta}$ represents the distinctness of the response, which depends on the group. It is e^δ in group 1 and one in the reference group 0. It accounts for the possibility that, for example, women might have more distinct preferences than men. The term $e^{x_{i0}\delta}$ is the same for all variables and modifies the effect strengths of all variables in the same way.

In contrast, $\alpha_j + x_{i0}\alpha_{0j}$ contains the *variable-specific* effect strength. The first parameter, α_j , can be seen as the effect strength in group 0, the interaction parameter α_{0j} represents the change if one switches to group 1. These are variable-specific effects beyond the impact of possible group-specific distinctiveness. If groups refer to gender, α_{0j} and $\alpha_j + \alpha_{0j}$ might represent the group-specific effects of age or gender on the outcome beyond the strength of preference found in gender groups.

The important point is that the derivation of the model does not use latent variables. The model can be seen as linking the response directly to the covariates with a clear interpretation of the coefficients. The factor $e^{\mathbf{z}_i^T \boldsymbol{\delta}}$ accounts for the heterogeneity or the distinctness in the preference for response categories, which is not available in the simple model. The same factor is considered to represent latent heterogeneity of variances in the traditional derivation of the heterogeneous logit model. Since the models are the same there is no way to decide which interpretation is to be preferred from the formulae. If one considers, for example, the importance of sports as response, both interpretations might apply. In other cases the interpretation as distinctiveness might not be the best choice. However, even in Allison's example it is not so clear what the correct interpretation of effects is. In his example the response was promotion to associate professor and effects of variables like number of articles was twice the size for gender. This might have been generated by heterogeneity of variances in the hypothetical latent trait model with assumed standard deviations being twice the size for women than for men. However, it is also conceivable that the committee that decides on promotion handles men and women in a different way. If the committee, given the values of variables like number of articles, tends to have a more distinct opinion for male applicants than for females the effect is the same as the assumed heterogeneity of variances.

Essentially one has the multiplicative structure of the predictor

$$\eta_i = e^{\mathbf{z}_i^T \boldsymbol{\delta}} (\alpha_0 + \mathbf{x}_i^T \boldsymbol{\alpha}),$$

which contains two factors,

the linear predictor $\alpha_0 + \mathbf{x}_i^T \boldsymbol{\alpha}$, which is always found in logit models, and which can contain variables from \mathbf{z}_i and interactions between variables,

the factor $e^{\mathbf{z}_i^T \boldsymbol{\delta}}$, which may represent heterogeneity of variances or distinctiveness of the response.

In both cases the descriptive effect of covariates is the same.

6 More General Heterogeneous Choice Models

So far the heterogeneous choice model has been used to model the coefficients in two groups. This case has been considered in most of the literature on heterogene-

TABLE 7: *General heterogeneous model for the importance of sports*

	Estimate	Std. Error	z value	Pr(> z)
Gender	-0.2428	0.0783	-3.098	0.0019
age	-0.0278	0.0042	-6.499	0.0000
Family	-0.0730	0.0255	-2.864	0.0041
disp.Gender	0.3249	0.1204	2.698	0.0069
disp.MaritalStatus2	0.5351	0.1584	3.377	0.0007
disp.MaritalStatus3	1.2722	0.7031	1.809	0.0703
disp.MaritalStatus4	-0.7168	0.3941	-1.819	0.0689
disp.age	-0.0246	0.0050	-4.906	0.0000

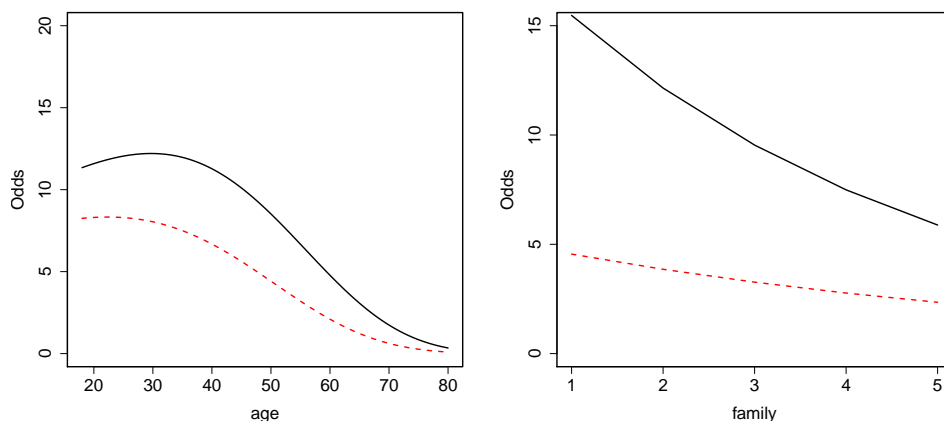


FIGURE 2: *Odds for men (drawn line) and women (dashed line) for sports example plotted against age (Marital Status=0, Family=20) and family size (Marital Status=0, Age=40) with multiple variables in the effectmodifying term.*

ity of variances. It is definitely interesting if the main question is the comparison of groups but ignores that the effect modification by variance heterogeneity or distinctness can be generated by more variables. The general heterogeneous model, however, explicitly allows to include a vector of covariates in the effect modifying term. Therefore we briefly consider an application of the general model.

For illustration we consider again the sports example and include all the variables in the effect modifying term. Wald tests show that not all of them are needed. In Table 7 the parameter estimates of the already reduced model is given. It is seen that in addition to gender at least age should be considered as an effect modifier. The presence of age in the effect modifier changes the effect strength of the other covariables. Figure 2 shows the corresponding effects of age and gender, which differ from the effects found in Figure 1, which shows the effects if only

gender is considered an effect modifier. It is seen that the inclusion of the other variables modifies the curves, for income the difference between gender groups becomes definitely stronger.

For the interpretation of parameter it is important that the parameters in Table 7 are the γ -parameters from the representation $\text{logit}(\pi_i) = (\alpha_0 + \mathbf{x}_i^T \boldsymbol{\alpha}) / e^{z_i^T \boldsymbol{\gamma}}$, since the software that was used fits the heterogeneous logit model. To obtain the δ -parameters one simply has to change the sign because $\boldsymbol{\delta} = -\boldsymbol{\gamma}$. The gender effect in the liner term is -0.242, which means that for women (coded as 1) sports tends to be less interesting than for men. The dispersion effect of gender is $e^{\gamma_{\text{Gender}}} = e^{0.379} = 1.46$, which means stronger dispersion in the female population. Alternatively, one can consider the factor $e^{\delta_{\text{Gender}}} = e^{-0.379} = 0.68$, which means a less distinct response for females than for males. In both cases the effect is that the response probabilities for females tend to be closer to 0.5 than the response probabilities for males.

It should be noted that the model given in Table 7 is rather sparse. The response probabilities are described by few parameters. A logit model (without dispersion) that includes all the interactions contained in the model would have to include almost all two factor interactions between explanatory variables and the effects would be hard to interpret.

7 Concluding Remarks

It has been shown that the heterogeneous choice model can be used to model underlying heterogeneity in most cases because reduced models are identifiable. However, when one has only one observation per individual one should be cautious when interpreting single coefficients in the heterogeneous model. Since the interpretation of effects is not unique in many cases it might be preferable to see the factor in the model as an effect modifier without trying to specify its substantial meaning. One should focus on the effects of explanatory variables on the binary response rather than trying to estimate effects of hypothetical latent variables and their variances. This is in agreement with the arguments of KM, however, they did not offer a modelling strategy beyond saying that the problem does not exist if the research questions are about the binary responses.

In the present study the potential of the heterogeneous model is investigated. It is a useful tool to investigate the possibly complex effects of explanatory variables and account for interactions in a specific sparse way. Specific advantages of the model are:

- it can account for effect modifiers, which might represent heterogeneity *or* distinctiveness,

- several variables can be included in the effect modifying term,

it allows for sparse models, most often one obtains a main effects model in the linear predictor.

The second point is important since the investigation of the effect structure by using simple logit models with interaction terms is restricted to very few, typically just one, potential effect modifier. If more than one variable is involved in interactions, which is quite possible, the inclusion of all the possible interactions yields such large models that they are very hard to interpret, and often estimates will not exist. In contrast, the heterogeneous model can include several variables and it can be examined which ones are needed.

It should be noted that in our use of the model, we do not have to assume that one knows at least for one of the coefficients of the underlying model that it is the same in both groups, as was assumed by Allison (1999). Constraints of this type have rightfully criticized by KM since they might shift assumptions from one part of the model to another, they produce different estimates but are equivalent in terms of fit.

We considered the binary logit model because it is the most important categorical regression model, but the same modelling strategies can be used in ordered categorical regression. In particular the family of cumulative models, which contains the proportional odds model as its most prominent member, can be extended to include heterogeneity terms. Ordered categorical regression models are extensively discussed in Tutz (2012), Agresti (2013), heterogeneous choice models with ordinal responses were considered by McCullagh (1980), Williams (2009), Williams (2010).

All the computations in the present paper were done with the statistical software R (R Core Team, 2016). We used the function `clm()` to estimate the heterogeneous choice model, it is available in the package `ordinal` (Christensen, 2015). Alternatively, one can also use STATA programs, see Williams (2010).

Appendix

Let us consider the model (4) and assume that one of the interaction parameters is zero. Without loss of generality we assume $\alpha_{0p} = 0$. Then one has the model

$$\text{logit}(\pi_i) = \frac{\alpha_{00} + x_{i0}\alpha_0 + x_{i1}\alpha_1 + \cdots + x_{ip}\alpha_p + x_{i0}x_{i1}\alpha_{01} + \cdots + x_{i0}x_{i,p-1}\alpha_{0,p-1}}{\exp(x_{i0}^T\delta)}.$$

Let $\alpha_{00}, \dots, \alpha_{0,p-1}, \delta$ and $\tilde{\alpha}_{00}, \dots, \tilde{\alpha}_{0,p-1}, \tilde{\delta}$ be two parameterizations of the model. It has to be shown that the two parameterizations are identical.

Let again $\pi(x_{ij})$ denote the probability of observing $Y_i = 1$ when the j -th covariate has value x_{ij} and $\pi(x_{ij} + 1)$ denote the probability if the j -th covariate has value $x_{ij} + 1$; all other variables are kept fixed. In addition we let $\pi(x_{ij}, x_{i0} = g)$ denote the probability of observing $Y_i = 1$ when the j -th covariate

has value x_{ij} and $x_{i0} = g$, correspondingly $\pi(x_{ij} + 1)$ denotes the probability if the j -the covariate has value $x_{ij} + 1$ and $x_{i0} = g$; all other variables are kept fixed.

(1) One obtains immediately

$$\text{logit}(\pi(x_{ip} + 1)) - \text{logit}(\pi(x_{ip})) = e^{x_{i0}\delta} \alpha_p$$

and therefore, provided $\alpha_p \neq 0$,

$$\frac{\text{logit}(\pi(x_{ip} + 1, x_{i0} = 1)) - \text{logit}(\pi(x_{ip}, x_{i0} = 1))}{\text{logit}(\pi(x_{ip} + 1, x_{i0} = 0)) - \text{logit}(\pi(x_{ip}, x_{i0} = 0))} = e^\delta.$$

Since the equations hold for both parameterizations one obtains $e^\delta = e^{\tilde{\delta}}$ and therefore $\delta = \tilde{\delta}$.

(2) For all variables $j \neq p$ one has

$$\text{logit}(\pi(x_{ij} + 1)) - \text{logit}(\pi(x_{ij})) = e^{x_{i0}\delta} (\alpha_j + x_{i0}\alpha_{0j}).$$

This yields for $x_{i0} = 0$ that $\alpha_j = \tilde{\alpha}_j$ holds, and for $x_{i0} = 1$ that $\alpha_{0j} = \tilde{\alpha}_{0j}$ holds.

(3) The only left parameters, which have to be identified, are α_{00} and α_0 . By using for $x_{i0} = 0$

$$\text{logit}(\pi_i) = \alpha_{00} + x_{i1}\alpha_1 + \cdots + x_{ip}\alpha_p$$

and for $x_{i0} = 1$

$$\text{logit}(\pi_i) = \frac{\alpha_{00} + x_{i0}\alpha_0 + x_{i1}\alpha_1 + \cdots + x_{ip}\alpha_p + x_{i0}x_{i1}\alpha_{01} + \cdots + x_{i0}x_{i,p-1}\alpha_{0,p-1}}{\exp(x_{i0}^T \delta)}$$

one obtains $\alpha_{00} = \tilde{\alpha}_{00}$ and $\alpha_0 = \tilde{\alpha}_0$, which concludes the proof.

References

- Agresti, A. (2013). *Categorical Data Analysis, 3d Edition*. New York: Wiley.
- Allison, P. D. (1999). Comparing logit and probit coefficients across groups. *Sociological Methods & Research* 28(2), 186–208.
- Breen, R., A. Holm, and K. B. Karlson (2014). Correlations and nonlinear probability models. *Sociological Methods & Research* 43(4), 571–605.
- Cai, Z., J. Fan, and R. Li (2000). Efficient estimation and inferences for varying-coefficient models. *Journal of the American Statistical Association* 95(451), 888–902.
- Christensen, R. H. B. (2015). ordinal—regression models for ordinal data. R package version 2015.6-28. <http://www.cran.r-project.org/package=ordinal/>.

- Fan, J. and W. Zhang (1999). Statistical estimation in varying coefficient models. *Annals of Statistics*, 1491–1518.
- Hastie, T. and R. Tibshirani (1993). Varying-coefficient models. *Journal of the Royal Statistical Society B* 55, 757–796.
- Karolson, K. B., A. Holm, and R. Breen (2012). Comparing regression coefficients between same-sample nested models using logit and probit: a new method. *Sociological Methodology* 42(1), 286–313.
- Kuha, J. and C. Mills (2017). On group comparisons with logistic regression models. *Sociological Methods & Research*, DOI:10.1177/0049124117747306.
- McCullagh, P. (1980). Regression model for ordinal data (with discussion). *Journal of the Royal Statistical Society B* 42(2), 109–127.
- Mood, C. (2010). Logistic regression: Why we cannot do what we think we can do, and what we can do about it. *European sociological review* 26(1), 67–82.
- Park, B. U., E. Mammen, Y. K. Lee, and E. R. Lee (2015). Varying coefficient regression models: a review and new developments. *International Statistical Review* 83(1), 36–64.
- R Core Team (2016). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing.
- Rohwer, G. (2015). A note on the heterogeneous choice model. *Sociological Methods & Research* 44(1), 145–148.
- Tutz, G. (2012). *Regression for Categorical Data*. Cambridge University Press.
- Tutz, G. (2018). Binary response models with underlying heterogeneity: Identification and interpretation of effects. *European Sociological Review*, to appear.
- Williams, R. (2009). Using heterogeneous choice models to compare logit and probit coefficients across groups. *Sociological Methods & Research* 37(4), 531–559.
- Williams, R. (2010). Fitting heterogeneous choice models with oglm. *Stata Journal* 10(4), 540–567.