Evaluation of Inequality Constrained Hypotheses Using an Akaike-Type Information Criterion

Yasin Altınışık
Utrecht University

EVALUATION OF INEQUALITY CONSTRAINED HYPOTHESES USING AN AKAIKE-TYPE INFORMATION CRITERION

EVALUATIE VAN ONGELIJKHEIDSRESTRICTIES DOOR MIDDEL VAN EEN AKAIKE-TYPE INFORMATIE CRITERIUM
(met een samenvatting in het Nederlands)

PROEFSCHRIFT

ter verkrijging van de graad van doctor aan de Universiteit Utrecht op gezag van de rector magnificus, prof.dr. G.J. van der Zwaan, ingevolge het besluit van het college voor promoties in het openbaar te verdedigen op vrijdag 2 februari 2018 des middags te 12.45 uur

door

Yasin Altınışık
geboren op 11 april 1986
te Trabzon, Turkije

Promotor:	Prof.dr. H.J.A. Hoijtink
Copromotoren:	Dr. R.M. Kuiper
	Dr. R.H. Klein Entink

This research was financially supported by the Republic of Turkey, Ministry of National Education and The Consortium on Individual Development (CID) is funded through the Gravitation program of the Dutch Ministry of Education, Culture, and Science and the Netherlands Organization for Scientific Research (NWO grant number 024.001.003).

Beoordelingscommissie:

Prof.dr. I.G. Klugkist
Prof.dr. J.K. Vermunt
Prof.dr. M.J. de Rooij
Prof.dr. M.J.C.R. Eijkemans
Prof.dr. W.A.M. Vollebergh

Altınışık, Yasin
Evaluation of inequality constrained hypotheses using an Akaike-type information criterion
Proefschrift Universiteit Utrecht, Utrecht. - Met lit. opg. - Met samenvatting in het Nederlands.
ISBN: 978-90-393-6918-0
Printed by Ridderprint, the Netherlands

Contents

Chapter 1

Introduction[1]

This dissertation focuses on how to evaluate equality and/or inequality constrained hypotheses using an Akaike-type information criterion (AIC) (Akaike, 1973, 1974) for a broad range of statistical models. (In)equality constrained hypotheses are formal representations of the directional relationships between model parameters. The AIC can be used to evaluate hypotheses containing equality constraints but not inequality constraints. Since the AIC cannot be used to examine the order of the parameters in hypotheses, Anraku (1999) modified the AIC such that it can be used to evaluate hypotheses containing simple order restrictions. This modification of the AIC is called order-restricted information criterion (ORIC) which is used to evaluate hypotheses of the form $\theta_1 < ... < \theta_K$, where "<" restrictions can be replaced by "=" restrictions and K represents the number of parameters in the model. However, the ORIC can only be used in the context of ANOVA models (Weiss, 2006; Rutherford, 2001, pp.18-24). Kuiper, Hoijtink, and Silvapulle (2011; 2012) propose a generalization of the ORIC which is called the generalized order-restricted information criterion (GORIC) (Kuiper et al., 2011, 2012). The GORIC can be utilized to evaluate more complicated (in)equality constrained hypotheses not only under ANOVA-type models but also under any univariate and multivariate normal linear models.

What we propose in this dissertation is a generalization of the GORIC which is called the generalized order-restricted information criterion approximation (GORICA). The GORICA can be utilized to evaluate the same type of (in)equality constrained hypotheses that the GORIC evaluates. However, the GORICA is applicable to not only normal linear models, but also applicable to generalized linear models (GLMs) (McCullagh & Nelder, 1989), generalized linear mixed models (GLMMs) (McCullogh & Searle, 2001), and structural equation models (SEMs) (Kline, 2010). In addition, the GORICA can be utilized in the context of contingency tables for which (in)equality constrained hypotheses do not necessarily contain linear restrictions on cell probabilities, but instead often contain non-linear restrictions on cell probabilities. Two software packages are presented that implement the GORICA in R.

[1] This chapter is written by Yasin Altınışık.

One of them is developed for the evaluation of (in)equality constrained hypotheses containing linear restrictions on model parameters in the context of GLMs, GLMMs, and SEMs. The other software package is developed to evaluate (in)equality constrained hypotheses containing linear and non-linear restrictions on cell probabilities.

1.1 (In)equality constrained hypotheses

Researchers often have theories and/or their own expectations when investigating their research questions. These theories/expectations can be precisely formulated using (in)equality constrained hypotheses. To illustrate how to create (in)equality constrained hypotheses, consider a high school in which the writing scores of students are measured within three academic programs: general, academic, and vocation. A researcher may have the following expectations regarding the association between the writing scores of students and the academic programs:

- The program involved does not impact the average writing scores of students.
- The students in the academic program have a higher average writing score than the students in the general and vocation programs and the average writing scores of the students in the general and vocation programs do not significantly differ from each other.
- No restrictions on the population means. Namely, there may be a relationship between the writing scores of students and the academic programs, but not necessarily the ones advised by hypotheses H_1 and H_2.

These expectations can be formulated by (in)equality constrained hypotheses:$H_1 : \theta_1 = \theta_2 = \theta_3$, $H_2 : \theta_1 > \theta_2 = \theta_3$ and the unconstrained hypothesis $H_u : \theta_1, \theta_2, \theta_3$, respectively, where $\theta = (\theta_1, \theta_2, \theta_3)$ denotes the vector of the population means of the writing scores of students in the three academic programs, respectively. The operator "," in the unconstrained hypothesis H_u denotes the absence of information regarding the directional relationships between the population means. The unconstrained hypothesis H_u is often evaluated together with the other hypotheses under evaluation to avoid choosing a hypothesis not supported by the data as the best hypothesis out of the set of hypotheses.

1.2 GORICA weights

The GORICA weights are calculated by taking into account the misfits and complexities of the hypotheses under evaluation. Analogous to the Akaike weights (Burnham & Anderson, 2002, p.75), these weights are used to quantify the support in the data for each hypothesis under evaluation. By looking at the pairwise ratios between the GORICA weights, one can determine the relative importance of one hypothesis over another hypothesis. To illustrate, suppose that the GORICA weights of the hypotheses H_1, H_2, and the unconstrained hypothesis H_u in the previous section are obtained as 0.05, 0.75, and 0.20, respectively. Based on the pairwise ratio between hypothesis H_1 and the unconstrained hypothesis H_u, hypothesis

H_1 is a weak hypothesis. Because the unconstrained hypothesis H_u is 0.20/0.05 = 4 times better supported by the data when compared to hypothesis H_1. Hypothesis H_2 is a much better hypothesis than hypothesis H_1. That is, hypothesis H_2 is 0.75/0.05 = 15 times better supported by the data when compared to hypothesis H_1. Moreover, hypothesis H_2 is a strong hypothesis. Because it is 0.75/0.20 = 3.75 times better supported by the data when compared to the unconstrained hypothesis H_u. Therefore, it is concluded that hypothesis H_2 is the best hypothesis in the set of hypotheses.

1.3 Outline of the dissertation

In this dissertation we developed a generalization of the AIC, named the GORICA, to extend the evaluation of (in)equality constrained hypotheses to a wider range of statistical models: GLMs, GLMMs, SEMs, and contingency tables.

Chapter 2 introduces the GORICA for GLMs, GLMMs, and SEMs. Evaluation of equality and/or inequality constrained hypotheses is illustrated for logistic regression model, multilevel regression model, and structural equation model as being representatives of GLMs, GLMMs, and SEMs, respectively. Two simulation studies are performed in Chapter 2. The first simulation study showed that the performance of the GORIC and GORICA are very similar for normal linear models. This implies that the GORICA can be utilized instead of the GORIC to evaluate (in)equality constrained hypotheses for these models. The second simulation study which mimics the model used in the logistic regression example showed that the performance of the GORICA in terms of selecting the true hypothesis in a set of hypotheses was convincing. The R script Gorica.R is provided to enable researchers to evaluate (in)equality constrained hypotheses for their own data.

Chapter 3 elaborates on the GORICA in the context of contingency tables. In this chapter we developed the GORICA for (sparse) high-dimensional contingency tables such that it can be used to evaluate (in)equality constrained hypotheses containing linear and non-linear restrictions on cell probabilities. The GORICA in the context of contingency tables is implemented in the R script GoricaCont.R.

Chapter 4 discusses the use of the GORICA in replication studies for contingency tables. We illustrated the use of the GORICA to evaluate hypotheses using multiple studies. We provided a consistent strategy for testing replication across studies using the GORICA.

Chapter 5 elaborates that the GORICA can be applied to evaluate (in)equality constrained hypotheses using three estimation methods: maximum likelihood estimation (Fisher, 1922), nonparametric bootstrapping (Efron & Tibshirani, 1993), and gibbs sampling (D. Spiegelhalter, 1994). We provided R functions ormle and gorica that are compatible with each of these estimation techniques. In the Supplementary material we provided the relevant R code that can be used to duplicate the results presented in this chapter.

Chapter 6 introduces a discussion about the position of the GORICA among other methods, its strengths and weaknesses, and its comparison with the other methods when evaluating (in)equality constrained hypotheses.

Chapter 2

Evaluation of inequality constrained hypotheses using a generalization of the AIC[1]

2.1 Introduction

The evaluation of the null hypothesis H_0 : "Nothing is going on" and the alternative hypothesis H_a : "Something is going on but I do not know what" by means of p-values is controversial. First of all, the null hypothesis is not often a reasonable representation of the population of interest (Cohen, 1994). For example, it is hard to come up with a population in which two means θ_1 and θ_2 are exactly equal, that is, $H_0 : \theta_1 = \theta_2$. In addition, Royall (1997, pp.79-81) elaborates that the null hypothesis can never be accepted, that is, failing to reject the null hypothesis $H_0 : \theta_1 = \theta_2$ against the alternative H_a essentially means that the alternative hypothesis is incorrect and not that the null hypothesis is correct. Secondly, p-values cannot measure the evidence in the data for H_0 or H_a (Wagenmakers, 2007). After rejecting the null hypothesis, for example, $p = .04$, it is still not quantifiable to what degree the alternative hypothesis is better than the null hypothesis.

Researchers are often not interested in evaluation of the null hypotheses by means of p-values. Researchers want to quantify the evidence in the data with respect to their own expectations. Rejecting ($p < .05$) or not rejecting ($p > .05$) the null hypothesis, does not address the interests of these researchers (Hoijtink, 2012, p.8). Evaluation of (in)equality

[1]This chapter will be submitted as Altınışık, Y., Nederhof, E., Hoijtink, H., Oldehinkel, A. J., & Kuiper, R. M. Evaluation of Inequality Constrained Hypotheses Using a Generalization of the AIC.
Author contributions: YA, HH, and RMK developed the study concept. All authors contributed to the study design. Data collection was performed by EN and AJO and the writing, programming, analysis, and interpretation was performed by YA under the supervision of RMK and HH. All authors read, commented, and approved the final manuscript.

constrained hypotheses representing the researchers' expectations will address the interests of these researchers.

Inequality constraints are denoted by the operators smaller than "<" and bigger than ">" while the operator "=" indicates equality constraints. Absence of a constraint is represented by ",". To illustrate, consider the logistic regression model:

$$\text{logit}(p_i) = \theta_0 + \theta_1 x_{1i} + \theta_2 x_{2i} + \theta_3 x_{3i}, \tag{2.1}$$

where $\text{logit}(p_i) = \ln(\frac{p_i}{1-p_i})$, p_i is the probability of success for the ith case, θ_0 is the intercept, and θ_1, θ_2, and θ_3 are the regression coefficients of the predictor variables x_1, x_2, and x_3, respectively. Hypotheses of interest are:

$$\begin{aligned} H_1 &: \theta_1 = \theta_2 = \theta_3 \qquad \text{(Equality constrained)}, \\ H_2 &: \theta_1 > \theta_2 > \theta_3 \qquad \text{(Inequality constrained)}, \\ H_3 &: \theta_1 = \theta_2 > \theta_3 \qquad \text{(Inequality constrained)}, \end{aligned} \tag{2.2}$$

and

$$H_4 : \theta_1, \theta_2, \theta_3 \qquad \text{(Unconstrained)},$$

where H_1 states that regression parameters θ_1, θ_2, and θ_3 are equal to each other, H_2 specifies that parameter θ_1 is bigger than parameter θ_2 which in turn is bigger than parameter θ_3, hypothesis H_3 states that θ_1 and θ_2 are equal to each other which are bigger than parameter θ_3, and hypothesis H_4 states that there is no constraint on any of the parameters.

In this article, an Akaike-type information criterion (AIC) (Akaike, 1973, 1974), is developed to evaluate (in)equality constrained hypotheses for a general class of models: generalized linear models (GLMs) (McCullagh & Nelder, 1989), generalized linear mixed models (GLMMs) (McCullogh & Searle, 2001), and structural equation models (SEMs) (Kline, 2010). The AIC selects the best of a set of hypotheses, that is, the hypothesis that has the shortest distance to the true hypothesis. However, the AIC cannot evaluate hypotheses containing inequality constraint(s). To evaluate inequality constrained hypotheses, Anraku (1999) proposed a modification of the AIC which is called the order-restricted information criterion (ORIC). However, the ORIC can only be applied to hypotheses that have simple order restrictions which are of the form $\theta_1 < ... < \theta_K$, where "<" may be replaced by "=", or "," and K is the number of groups in the context of analysis of variance (ANOVA) (Weiss, 2006; Rutherford, 2001, pp.18-24).

Kuiper et al. (2011, 2012) propose the GORIC which is a generalization of the ORIC that can be used for the evaluation of (in)equality constrained hypotheses going beyond the simple order constrained hypotheses (Kuiper et al., 2011, 2012). For example, the GORIC can be used to evaluate the hypothesis $H_5 : \theta_1 + \theta_2 > \theta_3$ for which the sum of parameters θ_1 and θ_2 is bigger than parameter θ_3. However, the GORIC can only be applied to univariate and multivariate normal linear models.

This paper introduces the generalized order-restricted information criterion approximation (GORICA) which is based on large-sample theory. The GORICA uses the property that the distribution of maximum likelihood estimates (MLEs) can be approximated by normal distribution (Fisher, 1922). Therefore, the performance of the GORICA approximates the performance of the GORIC and extends its use to a broader range of models: GLMs, GLMMs, and SEMs.

The outline of the paper is as follows. First of all, a general class of (in)equality constrained hypotheses will be discussed and an illustration of (in)equality constrained hypotheses will be given for a logistic regression model as being representative of the GLMs. Second, the GORICA will be introduced. Third, a simulation study will be performed to show that the performance of the GORIC and GORICA in the context of multiple linear regression models is comparable. Subsequently, the GORICA will be used to analyze the logistic regression example and a simulation study of the logistic regression model will be discussed. Next, illustrations of (in)equality constrained hypotheses will be given for multilevel and structural equation models as being representatives of the GLMMs and SEMs, respectively, and the GORICA will be used to analyze the two examples. The paper will be concluded with a short discussion.

2.2 A general class of restrictions

Let $\theta = (\theta_1, \theta_2, ..., \theta_K)^T \in \mathbb{R}^{K\times 1}$ denote the parameters of a GLM, GLMM, or a SEM, that will be used in the formulation of (in)equality constrained hypotheses. The GORICA, like GORIC, can be applied to (in)equality constrained hypotheses of the form:

$$H_m : S\theta = 0, R\theta > 0, \tag{2.3}$$

with $m = 1, ..., M$, and the unconstrained hypothesis $H_{M+1} : \theta_1, \theta_2, ..., \theta_K$. Note that S is a $q_s \times K$ matrix and R is a $q_r \times K$ matrix representing the equality and inequality constraints of hypothesis H_m, respectively. For example, for $H_1 : \theta_1 > \theta_2, \theta_3 = 0$, it holds that $\theta = (\theta_1, \theta_2, \theta_3)^T$, $K = 3$, $q_s = q_r = 1$, and

$$S = \begin{pmatrix} 0 & 0 & 1 \end{pmatrix},$$

$$R = \begin{pmatrix} 1 & -1 & 0 \end{pmatrix}.$$

One restriction in the use of equation (2.3) is that $\begin{bmatrix} S \\ R \end{bmatrix}$ has to be of full rank after discarding the redundant restrictions (Kuiper et al., 2012). Note that a redundant restriction is a restriction that is implied by one or more of the other restrictions. An important situation where $\begin{bmatrix} S \\ R \end{bmatrix}$ is not of full rank occurs with range constraints. For example, for hypothesis $H_2 : \theta_1, \theta_2, 3 > \theta_3 > 1$ which can be rewritten as $H_2 : \theta_1, \theta_2, \theta_3 > 1, -\theta_3 > -3$, it holds that

$$R = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & -1 \end{pmatrix}$$

and the right hand side values of hypothesis H_2 are not equal to zero. Note that R is not of full rank, because the two rows are linearly dependent on each other (the second row is minus the first row).

The form in equation (2.3) cannot always straightforwardly be applied to the statistical model at hand. For example, consider the regression model:

$$y_i = \theta_0 + \theta_1 x_{1i} + \theta_2 x_{2i} + \epsilon_i, \tag{2.4}$$

where θ_0 is the intercept of the model, θ_k for $k = 1, 2$ is the unstandardized regression coefficient that relates predictor x_k to response y, and ϵ_i are the residuals for $i = 1, 2, ..., N$. One might be interested in the evaluation of hypothesis $H_1 : \theta_1 > \theta_2$. However, unstandardized coefficients θ_1 and θ_2 not only consider the strength of the relationship between x_1 versus y and x_2 versus y, respectively, but also the scale of the predictors. This problem is solved if H_1 is formulated using standardized regression coefficients. How standardization is dealt with will be elaborated later in this paper.

2.3 A general class of models

In this section, GLMs are introduced and exemplified using logistic regression model. For the logistic regression example, (in)equality constrained hypotheses are formulated based on the expectations of researchers or the results of previous studies.

2.3.1 Generalized linear models (GLMs)

GLMs (McCullagh & Nelder, 1989) can be written as:

$$f(\hat{y}) = X\beta, \tag{2.5}$$

where $\hat{y} = E(y) = f^{-1}(X\beta)$ denotes the expected value of the dependent variable $y = (y_1, y_2, ..., y_N)^T \in \mathbb{R}^{N\times 1}$, $X\beta$ is the linear predictor with explanatory variables $X = (x_1, x_2, ..., x_p) \in \mathbb{R}^{N\times p}$, $x_j = (x_{j1}, x_{j2}, ..., x_{jN})^T \in \mathbb{R}^{N\times 1}$ for $j = 1, ..., p$, and model parameters $\beta = (\beta_1, \beta_2, ..., \beta_p)^T \in \mathbb{R}^{p\times 1}$. Here, $f(\hat{y})$ is the link function that relates $\hat{y}$ and $X\beta$. For example, if y is an outcome variable in a logistic regression model, then $f(\hat{y}) = log(\frac{\hat{y}}{1-\hat{y}}) = X\beta$, that is, $\hat{y} = \frac{exp(X\beta)}{1+exp(X\beta)}$.

In the sequel the parameter vector β will be separated into two parts as $\beta = \{\theta, \xi\}$, where $\theta \in \mathbb{R}^{K\times 1}$, with $K \leq p$, are the structural parameters, that is, the parameters that will be used in the formulation of (in)equality constrained hypotheses and ξ are the nuisance parameters, that is, the parameters that will not be used in the formulation of (in)equality constrained hypotheses. Note that although β has p parameters in total, only $K \leq p$ of them are used in the formulation of (in)equality constrained hypotheses.

2.3.2 Example 1: Logistic regression modeling

Logistic regression can be used to predict a binary dependent variable with continuous or categorical explanatory variable(s). In a study by Nederhof, Ormel, and Oldehinkel (2014), 11 years old participants ($N = 836$) are divided into $g = 1, 2, 3$ groups: 1 = Sustainers, 2 = Shifters, and 3 = Comparison group, based on their performance on a sustained-attention task and on a shifting-set task. The outcome depressive episode D $\in$ { 0 = "No depressive episode", 1= "Endorsed an episode" } is predicted by early life stress ES $\in$ { 0 = "High", 1 = "Low" }, recent stress RS which is a standardized continuous variable with values in the interval [-1.42, 2.56], and the interaction between both variables. The resulting logistic regression model is:

$$f(\hat{D}_{gi}) = \beta_{g0} + \beta_{g1}\mathrm{RS}_{gi} + \beta_{g2}\mathrm{ES}_{gi} + \beta_{g3}\mathrm{RS}_{gi}\mathrm{ES}_{gi}, \tag{2.6}$$

where $f(.)$ denotes the logit link function, β_{g0} denotes the intercept for group g, and β_{g1}, β_{g2}, and β_{g3} are the group dependent coefficients of the values of the three predictors for $g = 1, 2, 3$ and $i = 1, 2, ..., N_g$, where N_g denotes the number of persons in group g with $\sum_{g=1}^{3} N_g = N$. Based on the values of variable ES, the model in equation (2.6) can be converted to:

$$f(\hat{D}_{gi}) = \begin{cases} (\beta_{g0} + \beta_{g2}) + (\beta_{g1} + \beta_{g3})RS_{gi} & \text{if ES} = \text{"Low"} \\ \beta_{g0} + \beta_{g1}RS_{gi} & \text{if ES} = \text{"High"}. \end{cases} \tag{2.7}$$

In the study, there are two hypotheses: the mismatch and cumulative stress hypotheses. The mismatch hypothesis states that the risk of depression for adolescents with low levels of early life stress increases with high recent stress levels, while adolescents with high levels of early life stress are not affected by high recent stress levels. The cumulative stress hypothesis implies that there is no interaction and, therefore, only the main effect of recent stress predicts depression and that the relationship is positive (Nederhof et al., 2014).

Hypothesis H_1 is based on the theory in Nederhof and Schmidt (2012). It states that the mismatch hypothesis applies to the sustainers ($g = 1$): $\beta_{11} + \beta_{13} > 0$ if ES is low, $\beta_{11} = 0$ if ES is high, and the shifters ($g = 2$): $\beta_{21} + \beta_{23} > 0$ if ES is low, $\beta_{21} = 0$ if ES is high. In contrast, the cumulative stress hypothesis applies to the comparison groups ($g = 3$): $\beta_{33} = 0, \beta_{31} > 0$ if ES is low, $\beta_{31} > 0$ if ES is high. The hypothesis H_2 is based on the results presented in Nederhof et al. (2014, p.689). It expresses that the mismatch hypothesis applies to the sustainers, the cumulative stress hypothesis applies to the comparison groups and neither applies to the shifters: $\beta_{21} + \beta_{23} = 0$ if ES is low, $\beta_{21} = 0$ if ES is high:

$$\begin{aligned} & \beta_{11} = 0, \beta_{13} > 0 && \beta_{11} = 0, \beta_{13} > 0 && \text{(Sustainers)} \\ H_1: \ & \beta_{21} = 0, \beta_{23} > 0 && H_2: \ \beta_{21} = \beta_{23} = 0 && \text{(Shifters)} \\ & \beta_{33} = 0, \beta_{31} > 0 && \beta_{33} = 0, \beta_{31} > 0 && \text{(Comparison).} \end{aligned} \tag{2.8}$$

Note that the structural parameters addressed in hypotheses H_1 and H_2 are $\theta = (\beta_{11}, \beta_{13}, \beta_{21}, \beta_{23}, \beta_{31}, \beta_{33})$ and the nuisance parameters are $\xi = (\beta_{10}, \beta_{12}, \beta_{20}, \beta_{22}, \beta_{30}, \beta_{32})$.

2.4 The GORICA

This section elaborates on how to evaluate (in)equality constrained hypotheses using the GORICA for GLMs, GLMMs, and SEMs. As shown in the next section, the differences in the performance of the GORIC (Kuiper et al., 2011, 2012) and GORICA are quite small in the context of multiple regression models. This implies that, for the evaluation of (in)equality constrained hypotheses, the performance of the GORICA approximates the performance of the GORIC.

2.4.1 Preliminaries

The GORIC can only be implemented in normal linear models. The univariate normal linear model is defined as:

$$y = X\beta + \epsilon, \tag{2.9}$$

where $y = (y_1, y_2, ..., y_N)^T \in \mathbb{R}^{N\times 1}$ denotes the values of the dependent variable, $X = (x_1, x_2, ..., x_p) \in \mathbb{R}^{N\times p}$, $x_j = (x_{j1}, x_{j2}, ..., x_{jN})^T \in \mathbb{R}^{N\times 1}$ for $j = 1, 2, ..., p$ contains the explanatory variables, $\beta = (\beta_1, \beta_2, ..., \beta_p)^T \in \mathbb{R}^{p\times 1}$ are the model parameters, and $\epsilon_i \sim N(0, \sigma^2)$ represents residuals with mean 0 and variance σ^2 (Kuiper et al., 2011, 2012).

The GORIC is defined as:

$$\text{GORIC}_m = -2\, L(\tilde{\theta}^m, \tilde{\xi}^m|y) + 2\, [PT_m(\theta) + PT_m(\xi)], \tag{2.10}$$

where $\theta = \beta$ and $\xi = \sigma^2$. The GORIC consists of two parts: the fit and penalty terms. The log likelihood $L\,(\tilde{\theta}^m, \tilde{\xi}^m|y)$ represents the fit of hypothesis H_m to the data at hand. Note that $\tilde{\theta}^m$ and $\tilde{\xi}^m$ denote the values for which the log likelihood is maximized subject to the restrictions in hypothesis H_m. The penalty term consists of two parts: $PT_m(\theta)$ denotes the number of distinct regression parameters after taking into account the order constraints in H_m; and $PT_m(\xi) = 1$ because it concerns the unconstrained residual variance σ^2. A hypothesis that has large fit and small penalty terms will render a small GORIC value which means that this hypothesis has relatively small distance to the true hypothesis. For each statistical model of the family outside of normal linear models, the evaluation of (in)equality constrained hypotheses by means of the GORIC has to be redeveloped. The GORICA is being proposed to avoid these repetitions and to extend the GORIC's applicability to GLMs, GLMMs, and SEMs.

The GORICA can be obtained as follows: First of all, θ denotes the regression parameters for the GORIC, while θ denotes the structural parameters for the GORICA, that is, the parameters used in the specification of the (in)equality constrained hypotheses. For the GORIC ξ denotes the σ^2, for the GORICA ξ denotes all the parameters that are not used in the specification of the (in)equality constrained hypotheses. Secondly, in the GORICA the information with respect to θ is summarized in a normal approximation of the log likelihood:

$$L(\tilde{\theta}^m|\hat{\theta}, \hat{\Sigma}_{\hat{\theta}}) = -\frac{K}{2}\log(2\pi) - \frac{1}{2}\log|\hat{\Sigma}_{\hat{\theta}}| - \frac{1}{2}(\hat{\theta} - \tilde{\theta}^m)^T(\hat{\Sigma}_{\hat{\theta}})^{-1}(\hat{\theta} - \tilde{\theta}^m). \tag{2.11}$$

Here, $\hat{\theta} \in \mathbb{R}^{K\times 1}$ denotes the unconstrained MLEs and $\hat{\Sigma}_{\hat{\theta}} \in \mathbb{R}^{K\times K}$ is the covariance matrix of $\hat{\theta}$ (in the next subsection it will be elaborated how $\hat{\theta}$ and $\hat{\Sigma}_{\hat{\theta}}$ can be obtained). The restricted MLEs of the structural model parameters under H_m are denoted by $\tilde{\theta}^m \in \mathbb{R}^{K\times 1}$. After the next subsection, it will be elaborated how these estimates can be obtained. Thirdly, because the equality and inequality constraints are always on structural parameters and never on nuisance parameters, the part $PT_m(\xi)$ in equation (2.10) is removed keeping only $PT_m(\theta)$. Note that because there are no constraints on nuisance parameters, the number of ξ parameters is equal for all hypotheses $m = 1, 2, ..., M + 1$. This section will end with an elaboration on how $PT_m(\theta)$ can be computed.

The GORICA with respect to hypothesis H_m is defined as:

$$\text{GORICA}_m = -2\ L\ (\tilde{\theta}^m|\hat{\theta}, \hat{\Sigma}_{\hat{\theta}}) + 2PT_m(\theta), \tag{2.12}$$

for which the hypothesis with the smallest GORICA value has the smallest distance to the true hypothesis among the set of competing hypotheses under consideration which implies that it is the best from the set of hypotheses.

The extent to which a hypothesis is better than the other hypotheses can be quantified using GORICA weights (w_m), that is, numbers on a scale that ranges from 0 to 1, that are analogous to the Akaike weights (Burnham & Anderson, 2002, p.75) and the GORIC weights (Kuiper & Hoijtink, 2013):

$$w_m = \frac{\exp\{-\frac{1}{2}\ \text{GORICA}_m\}}{\sum_{m'=1}^{M+1}\exp\{-\frac{1}{2}\ \text{GORICA}_{m'}\}}. \tag{2.13}$$

For example, if hypotheses H_1, H_2, and the unconstrained hypothesis H_3 have GORICA weights 0.15, 0.03, and 0.82, respectively, H_1 has $0.15/0.03 = 5$ times more support than H_2. Note the value of including the unconstrained hypothesis H_3 in the model set: Although H_1 is a better hypothesis than H_2, it is not a good hypothesis, because it is outperformed by H_3, that is, H_3 has 0.82/0.15 $\approx$ 5.47 times more support than H_1. Stated otherwise, the constraints in H_2 are not supported by the data.

2.4.2 Estimation of $\hat{\theta}$ and $\hat{\Sigma}_{\hat{\theta}}$

To evaluate (in)equality constrained hypotheses with the GORICA, the structural model parameters and their covariance matrix need to be estimated. One common method to estimate parameters is the method of maximum likelihood (Fisher, 1922) which is often provided by standard software, e.g., Mplus, Amos, LISREL, SAS (Yuan & Hayashi, 2006). As elaborated earlier θ's are often standardized parameters to make them comparable. A disadvantage of maximum likelihood as implemented in standard software is that it does not always render $\hat{\Sigma}_{\hat{\theta}}$ if $\hat{\theta}$ is standardized. Therefore, nonparametric bootstrapping (Efron & Tibshirani, 1993, p.45) is used to compute $\hat{\theta}$ and $\hat{\Sigma}_{\hat{\theta}}$ since it is easy to obtain not only standardized parameter estimates but also their covariance matrix. It consists of the following steps:

1. a) **GLMs and SEMs:** Draw B independent bootstrap samples of size N with replacement from the observed data, where N is the number of data points.

   b) **GLMMs:** Roberts and Fan (2004) state that the empirical sampling distribution of an estimate is always related to a certain sample size. Therefore, in the framework of hierarchical linear modeling each bootstrap sample should have the same sample size N as the original data set. Similarly, each bootstrap sample should contain the same number of group J as the original data set. Based on Roberts and Fan (2004) the following approach is utilized to obtain bootstrap samples. Firstly, draw B bootstrap samples of size N_j with replacement from each group separately for $j = 1, 2, ..., J$. Secondly, combine the observations for the groups 1 to J into one bootstrap sample. The resulting bootstrap samples have $\sum_{j=1}^{J} N_j = N$ observations and the number of groups equals J.

2. **Optional:** Standardize the observed variables (and latent variables in the case of SEMs) in each bootstrap sample such that standardized estimates of θ's can be obtained. The β's in GLMs and GLMMs, Γ's, Π's, or Λ's in SEMs are often compared to each other. For example, consider the regression model:

$$y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \beta_3 x_{3i} + \epsilon_i, \tag{2.14}$$

   where β_0 is the intercept, β_1, β_2, and β_3 are the regression coefficients of the three predictors, and ϵ_i are the residuals for $i = 1, 2, ..., N$. Standardization is often used when quantifying the relative importance of independent variables on dependent variable(s) (Bring, 1994). For example, the hypothesis $H_1 : \beta_1 < \beta_2 < \beta_3$ only makes sense if these parameters are measured on the same scale. Another reason to standardize variables is to interpret main effects in the presence of interactions (Gelman, 2008).

3. Estimate the structural parameters using the "glm" (R Development Core Team, 2012) and "glmer" (Bates, Mächler, Bolker, & Walker, 2015) procedures for GLMs and GLMMs, and the "lavaan" package (Rosseel, 2012) for SEMs in R. This renders the values $\hat{\theta}_{bk}$ with $b = 1, 2, ..., B$ and $k = 1, 2, ..., K$, where B is the number of samples and K is the number of structural parameters.

4. Estimate the structural parameters and their covariance matrix using B bootstrap samples:

$$\hat{\theta}_k = \sum_{b=1}^{B} \frac{\hat{\theta}_{bk}}{B} \text{ with } k = 1, 2, ..., K, \tag{2.15}$$

$$\hat{\Sigma}_{\hat{\theta}}(k, k) = \frac{1}{B-1} \sum_{b=1}^{B} (\hat{\theta}_{bk} - \hat{\theta}_k)^2, \tag{2.16}$$

and

$$\hat{\Sigma}_{\hat{\theta}}(k,k') = \frac{1}{B-1}\sum_{b=1}^{B}(\hat{\theta}_{bk} - \hat{\theta}_k)(\hat{\theta}_{bk'} - \hat{\theta}_{k'}), \tag{2.17}$$

with $k \neq k'$.

2.4.3 The order-restricted MLEs

The order-restricted MLEs of the structural parameters $\tilde{\theta}^m \in \mathbb{R}^{K\times 1}$ are used to calculate the restricted maximum of the log likelihood for the hypothesis of interest. The values of $\tilde{\theta}^m$ are obtained by

$$\tilde{\theta}^m = \underset{\theta\in H_m}{\arg\min}\,[(\hat{\theta}-\theta)(\hat{\Sigma}_{\hat{\theta}})^{-1}(\hat{\theta}-\theta)^T], \tag{2.18}$$

which maximize the log likelihood in equation (2.11) subject to the restrictions in H_m. For example, if $\hat{\theta}_1 < 0$, $\hat{\theta}_2 > 0$, and hypothesis $H_1 : \theta_1 > 0, \theta_2 > 0$, then the corresponding order-restricted MLEs are $\tilde{\theta}_1^1 = 0$ and $\tilde{\theta}_2^1 = \hat{\theta}_2$. In this paper, to calculate the values of $\tilde{\theta}^m$, a quadratic programming algorithm the "solve.QP" subroutine of the "quadprog" package (Turlach, 2014, pp.2-4) in R is used.

2.4.4 Computing the penalty

Similar to the AIC and GORIC, the maximized log likelihood in the GORICA represents the fit of hypothesis H_m to the data at hand. The penalty part $PT_m(\theta)$ is used to penalize hypothesis H_m by taking into account the equality and/or inequality constraints imposed on the parameters in H_m.

First, a vector $z \in \mathbb{R}^{K\times 1}$ (see Figure 2.1, Step 1) is sampled from a normal distribution with mean vector 0 and covariance matrix $\hat{\Sigma}_{\hat{\theta}} \in \mathbb{R}^{K\times K}$. Second, the values of the vector of constrained estimates $\tilde{z}_m \in \mathbb{R}^{K\times 1}$ (see Figure 2.1, Step 2) are computed using the values of $z \in \mathbb{R}^{K\times 1}$, that is, $z \in \mathbb{R}^{K\times 1}$ in agreement with hypothesis H_m. Third, the number of the levels in $\tilde{z}_m$ is calculated, that is, $K - A_m$ where A_m is the number of active constraints in H_m (see Figure 2.1, Step 3). Note that a constraint becomes active when it is imposed on z. For example, if $z_1 < 0$ and $z_2 > 0$ for hypothesis $H_1 : \theta_1 > 0, \theta_2 > 0$, then the first constraint in H_1 is an active constraint but the second constraint is inactive because the constrained MLEs are $\tilde{z}_1^m = 0$ and $\tilde{z}_2^m = z_2$.

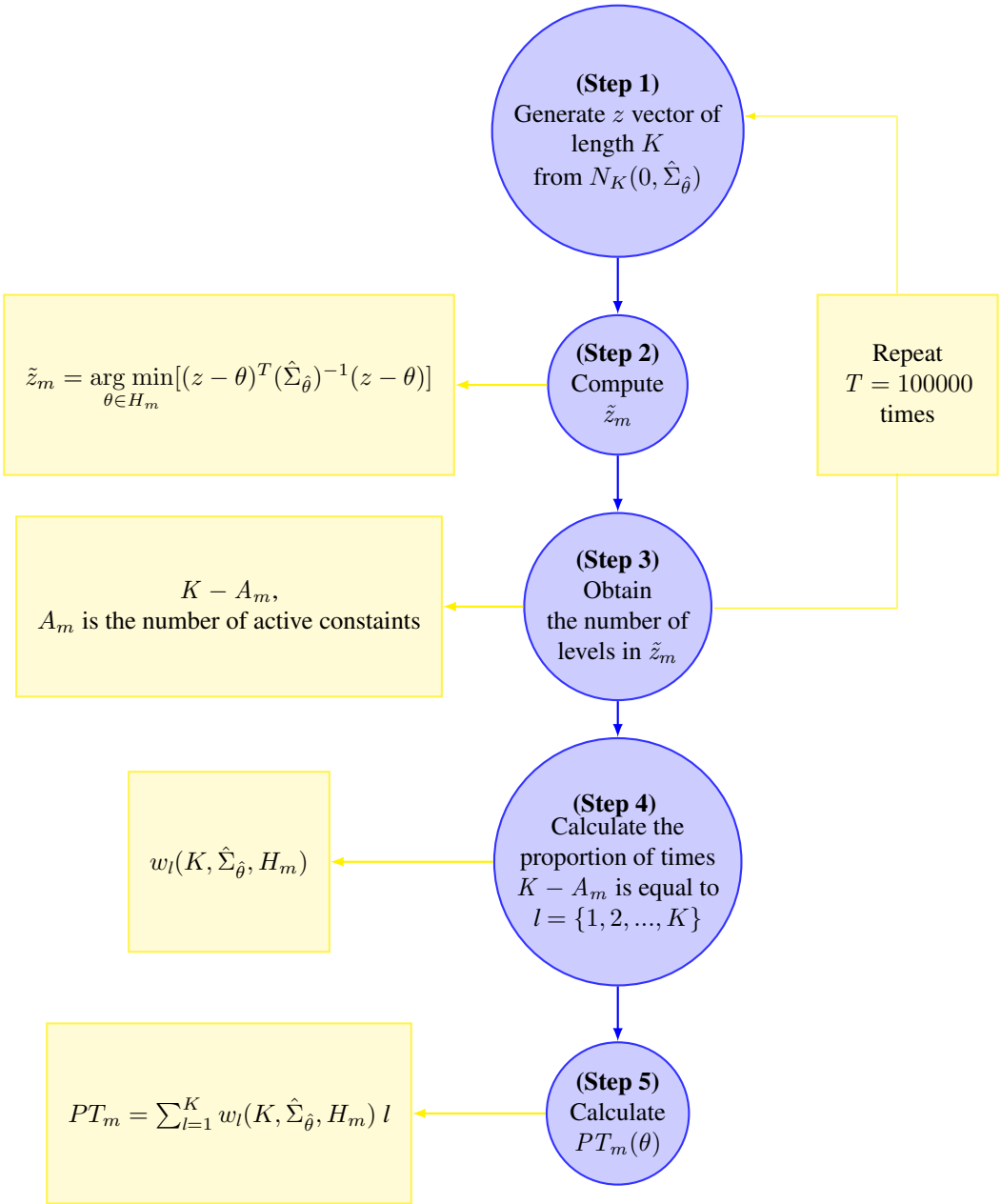

Figure 2.1: Working Scheme for Calculation of the Penalty Term

The first three steps are repeated, for example, for $T = 100000$ times. Afterwards, the level probabilities ($w_l(.)$) (Silvapulle & Sen, 2004, pp.78-81), that is, the probabilities that the vector of constrained estimates $\tilde{z}_m \in \mathbb{R}^{K\times 1}$ has l distinct values are calculated (see Figure 2.1, Step 4). Then, the penalty part $PT_m(\theta)$ (see Figure 2.1, Step 5) is defined as:

$$PT_m(\theta) = \sum_{l=1}^{K} w_l(K, \hat{\Sigma}_{\hat{\theta}}, H_m)\, l. \tag{2.19}$$

Note that the unconstrained hypothesis H_{M+1} has no restriction on parameters which is represented by $w_K = 1$ and $K - A_{M+1} = K$, consequently, $PT_m(\theta) = K$. For example, suppose $K = 3$ and $\hat{\Sigma}_{\hat{\theta}} = I_3$, where I_3 denotes a (3×3) identity matrix, then for the hypothesis $H_1 : \theta_1, \theta_2, \theta_3$, $PT_m = w_1 \times 1 + w_2 \times 2 + w_3 \times 3 = 3$, with $w_1 = w_2 = 0$ and $w_3 = 1$, while for the hypothesis $H_2 : \theta_1 > \theta_2 > \theta_3$, $PT_m = w_1 \times 1 + w_2 \times 2 + w_3 \times 3 = 1.834$, with $w_1 = 0.333$, $w_2 = 0.500$, and $w_3 = 0.167$.

2.5 A comparison of the GORICA with the GORIC

The performance of the GORIC and GORICA is compared in the context of a normal linear regression model by performing a simulation study. If the performance of the GORICA is similar to that of the GORIC, the GORICA is performing well. The model used is:

$$y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \beta_3 x_{3i} + \beta_4 x_{4i} + \epsilon_i, \tag{2.20}$$

where y_i is the response of the ith observation, β_0 is the intercept, $\beta_1, \beta_2, \beta_3$, and β_4 are the coefficients of the corresponding predictor variables, and $\epsilon_i \sim N(0, \sigma^2)$ is the residual of the ith observation.

Nonparametric bootstrapping is not needed here, because the four predictor variables and ϵ_i are generated from standard normal distributions and are therefore standardized. Based on the normal theory maximum likelihood, the "lm" function in R provides the covariance matrix of the parameter estimates for the model in equation (2.20). The correlation between each pair of predictors is $\rho_{kk'}$ and the correlation between each predictor and the residuals is zero.

The four hypotheses of interest are:

$$\begin{aligned} H_1 &: \beta_1 = \beta_2 = \beta_3 = \beta_4, \\ H_2 &: \beta_1 < \beta_2 < \beta_3 < \beta_4, \\ H_3 &: (\beta_1 - \beta_2) - (\beta_3 - \beta_4) > 0, \end{aligned} \tag{2.21}$$

and

$$H_4 : \beta_1, \beta_2, \beta_3, \beta_4,$$

where H_1 states that regression parameters $\beta_1, \beta_2, \beta_3$, and β_4 are equal to each other, H_2 states that parameter β_1 is the smallest parameter followed by parameters β_2, β_3, and β_4, respectively, H_3 specifies that the difference between parameters β_1 and β_2 is bigger than the difference between parameters β_3 and β_4, and hypothesis H_4 states that there is no constraint on any of the parameters.

One of the hypotheses above is selected as the true hypothesis. For these hypotheses the population values of the β's are obtained using selected values of $\rho_{kk'}$, and selected effect size values f^2 (Cohen, 1992, p.157). The resulting population values are displayed in Table 2.1. Subsequently, the response of the ith observation y_i is obtained by substituting β's, x_{1i}, x_{2i}, x_{3i}, x_{4i}, and ϵ_i into equation 2.20.

Table 2.1: Regression simulation: Population parameters for the regression coefficients

Case 1: $\mathrm{H}_1 : \beta_1 = \beta_2 = \beta_3 = \beta_4$ is true

	$\rho = 0$					$\rho = 0.3$					$\rho = 0.5$				
f^2	β_0	β_1	β_2	β_3	β_4	β_0	β_1	β_2	β_3	β_4	β_0	β_1	β_2	β_3	β_4
0.02	1.00	0.0700	0.0700	0.0700	0.0700	1.00	0.0508	0.0508	0.0508	0.0508	1.00	0.0443	0.0443	0.0443	0.0443
0.15	1.00	0.1806	0.1806	0.1806	0.1806	1.00	0.1310	0.1310	0.1310	0.1310	1.00	0.1142	0.1142	0.1142	0.1142
0.35	1.00	0.2546	0.2546	0.2546	0.2546	1.00	0.1847	0.1847	0.1847	0.1847	1.00	0.1610	0.1610	0.1610	0.1610

Case 2: $\mathrm{H}_2 : \beta_1 < \beta_2 < \beta_3 < \beta_4$ is true

	$\rho = 0$					$\rho = 0.3$					$\rho = 0.5$				
f^2	β_0	β_1	β_2	β_3	β_4	β_0	β_1	β_2	β_3	β_4	β_0	β_1	β_2	β_3	β_4
0.02	1.00	0.0000	0.0374	0.0748	0.1121	1.00	0.0000	0.0309	0.0618	0.0927	1.00	0.0000	0.0280	0.0560	0.0840
0.15	1.00	0.0000	0.0965	0.1930	0.2895	1.00	0.0000	0.0795	0.1590	0.2385	1.00	0.0000	0.0723	0.1446	0.2169
0.35	1.00	0.0000	0.1360	0.2720	0.4080	1.00	0.0000	0.1120	0.2240	0.3360	1.00	0.0000	0.1020	0.2040	0.3060

Case 3: $\mathrm{H}_3 : (\beta_1 - \beta_2) - (\beta_3 - \beta_4) > 0$ is true

	$\rho = 0$					$\rho = 0.3$					$\rho = 0.5$				
f^2	β_0	β_1	β_2	β_3	β_4	β_0	β_1	β_2	β_3	β_4	β_0	β_1	β_2	β_3	β_4
0.02	1.00	0.1190	0.0690	0.0250	0.0000	1.00	0.0990	0.0590	0.0250	0.0000	1.00	0.0890	0.0540	0.0250	0.0000
0.15	1.00	0.3550	0.0620	0.0250	0.0000	1.00	0.3200	0.0800	0.0250	0.0000	1.00	0.3000	0.0800	0.0250	0.0000
0.35	1.00	0.5000	0.0940	0.0250	0.0000	1.00	0.4590	0.1040	0.0250	0.0000	1.00	0.4390	0.0980	0.0250	0.0000

Case 4: $\mathrm{H}_4 : \beta_1, \beta_2, \beta_3, \beta_4$ is true

	$\rho = 0$					$\rho = 0.3$					$\rho = 0.5$				
f^2	β_0	β_1	β_2	β_3	β_4	β_0	β_1	β_2	β_3	β_4	β_0	β_1	β_2	β_3	β_4
0.02	1.00	0.0250	0.0000	0.1190	0.0690	1.00	0.0250	0.0000	0.0990	0.0590	1.00	0.0250	0.0000	0.0890	0.0540
0.15	1.00	0.0250	0.0000	0.3562	0.0542	1.00	0.0250	0.0000	0.3200	0.0800	1.00	0.0250	0.0000	0.3000	0.0800
0.35	1.00	0.0250	0.0000	0.5000	0.0940	1.00	0.0250	0.0000	0.4590	0.1040	1.00	0.0250	0.0000	0.4390	0.0980

Note. f^2 is Cohen's f^2 (Cohen, 1992, p.157), ρ is the correlation between each pair of predictors.

In Tables 2.2.1, 2.2.2, and 2.2.3, for nine scenarios, the proportion of times each of the hypotheses was selected by the GORIC and GORICA in 1000 independent samples is displayed. Table 2.2.1 shows that the GORICA performs as good as the GORIC no matter how the effect size measure differs for $\rho_{kk'} = 0.3, n = 200$. Table 2.2.2 shows that increasing sample size improves the performance of both methods and the performance of the methods are similar for any sample size for $\rho_{kk'} = 0.3, f^2 = 0.15$. Table 2.2.3 shows that increasing the correlations between the predictors reduces the performance of the methods for $f^2 = 0.15, n = 200$. However, again the performance of the methods are similar. Therefore, it can be concluded that the GORICA is performing equally good as the GORIC.

Table 2.2.1: Regression simulation: Percentage of times that the hypotheses H_1, H_2, H_3, and H_4 were selected by the GORIC and GORICA for $\rho_{kk'} = 0.3, n = 200$, and effect size values $0.02, 0.15$, and 0.35

		Case 1: H_1 is true				Case 2: H_2 is true				Case 3: H_3 is true				Case 4: H_4 is true			
f^2	Method	H_1	H_2	H_3	H_4	H_1	H_2	H_3	H_4	H_1	H_2	H_3	H_4	H_1	H_2	H_3	H_4
0.02	GORIC	74	14	10	2	54	40	5	1	73	3	21	3	60	28	10	2
	GORICA	75	13	10	2	55	39	5	1	73	3	21	3	60	28	10	2
0.15	GORIC	74	14	10	2	15	83	2	0	13	0	86	1	7	19	15	59
	GORICA	75	13	10	2	15	83	2	0	13	0	86	1	6	20	15	59
0.35	GORIC	74	14	10	2	3	96	1	0	1	0	99	0	0	12	12	76
	GORICA	75	13	10	2	3	96	1	0	1	0	99	0	0	12	12	76

Table 2.2.2: Regression simulation: Percentage of times that the hypotheses H_1, H_2, H_3, and H_4 were selected by the GORIC and GORICA for $\rho_{kk'} = 0.3, f^2 = 0.15$, and sample sizes $20, 100$, and 200

		Case 1: H_1 is true				Case 2: H_2 is true				Case 3: H_3 is true				Case 4: H_4 is true			
n	Method	H_1	H_2	H_3	H_4	H_1	H_2	H_3	H_4	H_1	H_2	H_3	H_4	H_1	H_2	H_3	H_4
20	GORIC	65	19	12	4	50	39	8	3	63	5	28	4	49	28	11	12
	GORICA	72	15	9	4	57	35	6	2	67	5	25	3	55	27	9	9
100	GORIC	72	16	9	3	32	65	2	1	36	0	62	2	24	28	16	32
	GORICA	73	15	9	3	33	64	2	1	37	0	61	2	24	29	16	31
200	GORIC	74	14	10	2	15	83	2	0	13	0	86	1	7	19	15	59
	GORICA	75	13	10	2	15	83	2	0	13	0	86	1	6	20	15	59

Table 2.2.3: Regression simulation: Percentage of times that the hypotheses H_1, H_2, H_3, and H_4 were selected by the GORIC and GORICA for $f^2 = 0.15$, $n = 200$, and the correlation values $0, 0.3$, and 0.5

		Case 1: H_1 is true				Case 2: H_2 is true				Case 3: H_3 is true				Case 4: H_4 is true			
$\rho_{kk'}$	Method	H_1	H_2	H_3	H_4	H_1	H_2	H_3	H_4	H_1	H_2	H_3	H_4	H_1	H_2	H_3	H_4
0	GORIC	72	16	9	3	4	95	1	0	2	0	98	0	2	11	12	75
	GORICA	73	15	9	3	4	95	1	0	2	0	98	0	2	11	12	75
0.3	GORIC	74	14	10	2	15	83	2	0	13	0	86	1	7	19	15	59
	GORICA	75	13	10	2	15	83	2	0	13	0	86	1	6	20	15	59
0.5	GORIC	74	16	8	2	30	68	2	0	28	0	71	1	22	36	16	26
	GORICA	74	16	8	2	30	68	2	0	28	0	71	1	23	36	15	26

2.6 The GORICA applied

In this section, (in)equality constrained hypotheses are evaluated for the logistic regression, multilevel regression, and structural equation models, respectively. In addition, the performance of the GORICA is examined by means of a simulation study in the context of logistic regression. The simulation study shows that the performance of the GORICA is convincing.

2.6.1 Logistic regression example (Continued)

Consider again the logistic regression model in equation (2.6). This model is utilized to evaluate hypotheses H_1 and H_2 in equation (2.8). As mentioned, hypothesis H_1 is based on the theory in Nederhof and Schmidt (2012) and hypothesis H_2 is based on the results presented in Nederhof et al. (2014, p.689). Now, a replication data set ($N = 310$) is analyzed to evaluate hypotheses H_1 and H_2. Each bootstrap sample resulting from the replication data set is standardized because of one main reason: to improve the interpretation of main effects when interactions exist (Gelman, 2008). The bootstrapped estimates and their covariance matrix are displayed in Table 2.3. Note that when bootstrapping is performed, a common problem in the framework of logistic regression modeling that is called "quasi complete separation of data points" is encountered. This problem occurs when the response variable separates an explanatory variable (or a combination of explanatory variables) to a certain level. In Appendix 2.A, it is elaborated how the separation problem is dealt with in this paper.

Table 2.3: Estimates and covariance matrix of the structural parameters for the logistic regression example

		$\hat{\Sigma}_{\hat{\theta}}$					
θ	$\hat{\theta}$	β_{11}	β_{13}	β_{21}	β_{23}	β_{31}	β_{33}
β_{11}	0.79	1.21					
β_{13}	-0.22	-1.21	1.52				
β_{21}	0.32	0.07	-0.06	1.24			
β_{23}	0.67	-0.06	0.04	-1.22	1.72		
β_{31}	0.63	-0.02	4.9e-3	-3.2e-3	3.3e-3	0.09	
β_{33}	0.20	0.06	-0.06	9.7e-3	-0.03	-0.09	0.32

Table 2.4: The order-restricted MLEs for the logistic regression example

H_m	$\tilde{\beta}^m_{11}$	$\tilde{\beta}^m_{13}$	$\tilde{\beta}^m_{21}$	$\tilde{\beta}^m_{23}$	$\tilde{\beta}^m_{31}$	$\tilde{\beta}^m_{33}$
H_1	0.00	0.57	0.00	0.99	0.69	0.00
H_2	0.00	0.61	0.00	0.00	0.70	0.00
H_3	0.79	-0.22	0.32	0.67	0.63	0.20

The order-restricted MLEs are given in Table 2.4 for the three hypotheses H_1, H_2, and the unconstrained hypothesis H_3 : $\beta_{11}, \beta_{13}, \beta_{21}, \beta_{23}, \beta_{31}, \beta_{33}$. The order-restricted MLEs are always in agreement with the constraints of the corresponding hypothesis. In Table 2.5, the values of the likelihood and penalty parts, GORICA and GORICA weights are displayed, respectively. Because the hypothesis with the smallest GORICA value is the best one among the three hypotheses, it is concluded that hypothesis H_1 is preferred over hypotheses H_2 and H_3. Based on the GORICA weights, hypothesis H_1 has 0.607/0.384 $\approx$ 1.58 times more support than hypothesis H_2. This is not a compeling evidence for H_1 but note that H_1 and H_2 only differ in one constraint, that is, $\beta_{23} > 0$ for hypothesis H_1 and $\beta_{23} = 0$ for hypothesis H_2. Both hypotheses are better than H_3. Therefore, it is concluded that both hypotheses H_1

Table 2.5: The likelihood and penalty parts, GORICA values, and GORICA weights of hypotheses H_1, H_2, and H_3 for the logistic regression example

H_m	$L(\tilde{\theta}^m \mid \hat{\theta}, \hat{\Sigma}_{\hat{\theta}})$	PT_m	$GORICA_m$	w_m
H_1	-3.188	1.447	9.270	0.607
H_2	-4.120	0.974	10.188	0.384
H_3	-2.862	6.000	17.724	0.009

and H_2 are supported by the data. However, in the replication data, the theory in Nederhof and Schmidt (2012) which is represented by hypothesis H_1 has more support than the results in Nederhof et al. (2014, p.689), that is, H_2.

2.6.2 Performance of the GORICA

A simulation study is performed based on the model and the study design presented in equation (2.6) to further illustrate the performance of the GORICA. The model is redefined as:

$$f(\hat{y}_{gi}) = \beta_{g0} + \beta_{g1}X_{gi} + \beta_{g2}I_{gi} + \beta_{g3}X_{gi}I_{gi}, \tag{2.22}$$

for $g = 1, 2, 3$ and $i = 1, 2, ..., N_g$, where $f(.)$ denotes the logit link function.

The parameters in Table 2.6 are chosen such that if we generate $y_{gi} \sim \text{Bernoulli}(\hat{y}_{gi})$ for $g = 1, 2, 3$ and $i = 1, 2, ..., N_g$, the expected correct classification rate (CCR) is equal to 60% or 75%, where the variable X_{gi} is generated from the standard normal distribution, the dummy variable I_{gi} is generated with both levels 0 and 1 having equal probability of being observed, and the X_{gi}'s and I_{gi}'s are independent.

Table 2.6: Logistic regression simulation: Population parameters for the regression coefficients

	Case 1: $H_1 : \beta_{11} = \beta_{21} = \beta_{33} = 0, \{\beta_{13}, \beta_{23}, \beta_{31}\} > 0$ is true											
	Group 1				Group 2				Group 3			
CCR	β_{10}	β_{11}	β_{12}	β_{13}	β_{20}	β_{21}	β_{22}	β_{23}	β_{30}	β_{31}	β_{32}	β_{33}
60 %	-0.15	0.00	-0.10	0.80	-0.15	0.00	-0.10	0.70	-0.15	0.61	-0.10	0.00
75 %	-0.15	0.00	-0.10	3.73	-0.15	0.00	-0.10	3.50	-0.15	3.40	-0.10	0.00
	Case 2: $H_2 : \beta_{11} = \beta_{21} = \beta_{23} = \beta_{33} = 0, \{\beta_{13}, \beta_{31}\} > 0$ is true											
	Group 1				Group 2				Group 3			
CCR	β_{10}	β_{11}	β_{12}	β_{13}	β_{20}	β_{21}	β_{22}	β_{23}	β_{30}	β_{31}	β_{32}	β_{33}
60 %	0.66	0.00	-0.55	0.11	0.66	0.00	-0.55	0.00	0.66	0.31	-0.55	0.00
75 %	1.28	0.00	-0.31	0.31	1.28	0.00	-0.31	0.00	1.28	0.45	-0.31	0.00
	Case 3: $H_3 : \beta_{11}, \beta_{13}, \beta_{21}, \beta_{23}, \beta_{31}, \beta_{33}$ is true											
	Group 1				Group 2				Group 3			
CCR	β_{10}	β_{11}	β_{12}	β_{13}	β_{20}	β_{21}	β_{22}	β_{23}	β_{30}	β_{31}	β_{32}	β_{33}
60 %	0.14	0.98	-0.02	-0.55	0.14	0.97	-0.02	-0.51	0.14	0.12	-0.02	-0.30
75 %	0.14	1.74	-0.02	-0.39	0.14	1.89	-0.02	-0.23	0.14	1.85	-0.02	-0.30

Note. CCR = Correct Classification Rate.

As can be seen in Table 2.7, the $\text{CCR} = (\text{A} + \text{D})/(\text{A} + \text{B} + \text{C} + \text{D})$ is defined based on the cut-off score 0.5.

Table 2.7: The expected correct classification rate (CCR) with the cut-off score 0.5

	$y_{gi} = 0$	$y_{gi} = 1$
$\hat{y}_{gi} \leq 0.5$	A	B
$\hat{y}_{gi} > 0.5$	C	D

Note. CCR = (A + D) / (A+B+C+D).

The hypotheses H_1 and H_2 in equation (2.8) and the unconstrained hypothesis H_3 : $\beta_{11}, \beta_{13}, \beta_{21}, \beta_{23}, \beta_{31}, \beta_{33}$ are evaluated. One of the three hypotheses is selected as the true hypothesis. The percentage of times each one of the hypothesis is selected in 1000 independent samples is calculated with different values of CCR and sample size (see Table 2.8). For each generated sample nonparametric bootstrapping with $B = 1000$ is utilized to obtain the MLEs and their covariance matrix. Table 2.8 shows that the performance of the GORICA is convincing. The GORICA choses the true hypothesis at least 88%, 82%, and 84% of times and at most 98%, 85%, and 100% of times when H_1, H_2, and H_3 is the true hypothesis, respectively.

Table 2.8: Percentage of times that the models H_1, H_2, and H_3 were selected by the GORICA for the logistic regression model

		H_1 is true			H_2 is true			H_3 is true		
CCR	N_g	H_1	H_2	H_3	H_1	H_2	H_3	H_1	H_2	H_3
60 %	100	88	11	1	16	82	2	10	6	84
	150	94	4	2	14	83	3	1	0	99
	200	97	1	2	15	82	3	0	0	100
75 %	100	98	0	2	15	85	0	1	0	99
	150	98	0	2	16	82	2	0	0	100
	200	98	0	2	15	82	3	0	0	100

2.6.3 Generalized linear mixed models (GLMMs)

GLMMs (McCullogh & Searle, 2001) are an extension of GLMs. The difference between GLMs and GLMMs is that, in the case of a GLM, one specifies a distribution of y and utilizes a linear predictor that consists of fixed β parameters, while in the case of GLMMs, one also incorporates normally distributed random effects τ into the linear predictor.

GLMMs can be defined as:

$$f(\hat{y}_j) = X_j\beta + Z_j\tau_j, \tag{2.23}$$

where $\hat{y}_j = E(y_j) = f^{-1}(X_j\beta + Z_j\tau_j)$ is an N_j vector that contains the expected values of the responses in group j, $X_j\beta + Z_j\tau_j$ is the linear predictor with $X_j \in \mathbb{R}^{N_j \times p}$, $\beta \in \mathbb{R}^{p\times 1}$, $Z_j \in \mathbb{R}^{N_j \times q}$ and $\tau_j \in \mathbb{R}^{q\times 1}$ for $j = 1, 2, ..., J$ groups each consisting of N_j observations. Similar to GLMs, $f(\hat{y}_j)$ is a link function that relates the fixed and random components to the linear predictor. The random effects τ_j are assumed to have normal distribution with mean vector 0 and covariance matrix $\Psi \in \mathbb{R}^{q\times q}$.

In the sequel, the parameter vectors β and τ_j, and covariance matrix Ψ will be described as $\{\beta, \tau_1, ..., \tau_J, \Psi\} = \{\theta, \xi\}$, where θ denotes the structural parameters and ξ denotes the nuisance parameters. Note that so far only (part of) the β's are used as the structural parameters. Note futhermore that the parameter vectors θ and ξ can be modified based on the interests of researchers, which will be illustrated in the next subsection.

2.6.4 Example 2: Multilevel regression modeling

Multilevel regression models are an extension of ordinary regression models that involve populations of interest with hierarchical data structures. The study in Hox (2010, p.59) is used to illustrate the specification of (in)equality constrained hypotheses in the context of multilevel regression models. In the study, the outcome variable PS represents the popularity scores of pupils that range from 0 (very unpopular) to 10 (very popular) for J = 100 classes with N_j pupils in each class. The popularity scores are predicted by a pupil level explanatory variable gender (G: 0 = boy, 1 = girl), a class level explanatory variable teacher experience (TE: ranging from 3 to 25 years), and the interaction between G and TE:

$$\mathrm{PS}_{ji} = \beta_{00} + \beta_{10}\mathrm{G}_{ji} + \beta_{01}\mathrm{TE}_j + \beta_{11}\mathrm{TE}_j\mathrm{G}_{ji} + \tau_{1j}\mathrm{G}_{ji} + \tau_{0j} + \epsilon_{ji}, \tag{2.24}$$

where β_{00} is the intercept, β_{10}, β_{01}, and β_{11} are the regression slopes for the pupil level variable gender, teacher experience, and their interaction, respectively, τ_{1j} and τ_{0j} are the random effects at the class level, and $\epsilon_{ji} \sim N(0, \psi_0^2)$ is the error at the pupil level with $j = 1, 2, ..., J$ and $i = 1, 2, ..., N_j$. The τ_j's are assumed to be normally distributed with mean vector 0 and covariance matrix:

$$\Psi = \begin{bmatrix} \psi_0^2 & \psi_{01} \\ \psi_{01} & \psi_1^2 \end{bmatrix}.$$

Note that in terms of the notation in equation (2.23),

$$X_j = \begin{pmatrix} 1 & \mathrm{G_{j1}} & \mathrm{TE_j} & \mathrm{TE_jG_{j1}} \\ 1 & \mathrm{G_{j2}} & \mathrm{TE_j} & \mathrm{TE_jG_{j2}} \\ . & . & . & . \\ . & . & . & . \\ . & . & . & . \\ 1 & \mathrm{G_{jN_j}} & \mathrm{TE_j} & \mathrm{TE_jG_{jN_j}} \end{pmatrix},$$

$$\beta = (\beta_{00}, \beta_{10}, \beta_{01}, \beta_{11})^T,$$

$$Z_j = \begin{pmatrix} 1 & \mathrm{G_{j1}} \\ 1 & \mathrm{G_{j2}} \\ . & . \\ . & . \\ . & . \\ 1 & \mathrm{G_{jN_j}} \end{pmatrix},$$

and

$$\tau_j = (\tau_{0j}, \tau_{1j})^T$$

for $i = 1, 2, ..., N_j$. The model in equation (2.24) can be converted to:

$$\mathrm{PS_{ji}} = \begin{cases} \beta_{00} + \beta_{01}\mathrm{TE_j} + \tau_{0j} + \epsilon_{ji} & \text{if G} = \text{“Boy”} \\ (\beta_{00} + \beta_{10}) + (\beta_{01} + \beta_{11})\mathrm{TE_j} + (\tau_{0j} + \tau_{1j}) + \epsilon_{ji} & \text{if G} = \text{“Girl”}. \end{cases} \tag{2.25}$$

Two hypotheses are formulated based on the figure in Hox (2010, p.61) to illustrate the applicability of the GORICA in the context of GLMMs. The first hypothesis H_1 states that TE has a positive effect on PS for each gender group (i.e., $H_1 : \beta_{01} > 0, \beta_{01} + \beta_{11} > 0$). The second hypothesis additionally specifies that girls have a higher PS than boys when TE = 0 (i.e., $\beta_{10} > 0$) and that the difference is smaller with more experienced teachers (i.e., $\beta_{11} < 0$), that is, $H_2 : \beta_{01} > 0, \beta_{01} + \beta_{11} > 0, \beta_{10} > 0, \beta_{11} < 0$.

The structural parameters of the model are $\theta = (\beta_{01}, \beta_{10}, \beta_{11})$ and the nuisance parameters are $\xi = (\beta_{00}, \tau_{0j}, \tau_{1j}, \psi_0^2, \psi_1^2, \psi_{01})$. As mentioned, components of the parameter vectors θ and ξ can differ based on the interests of researchers. For example, consider the correlation matrix of the random effects τ_{0j} and τ_{1j}:

$$\rho = \begin{bmatrix} 1 & \rho_{12} \\ \rho_{21} & 1 \end{bmatrix},$$

where ρ_{12} represents the correlation between τ_{0j} and τ_{1j}. If researchers only want to evaluate hypotheses, for example, $H_1 : \rho_{12} = 0$, that is, the correlation between the random effects is equal to zero and $H_2 : \rho_{12} > 0$, that is, the correlation between the random effects is bigger than zero, the parameter vectors θ and ξ need to be modified as $\theta = \rho_{12}$ and $\xi = (\beta_{00}, \beta_{01}, \beta_{10}, \beta_{11}, \tau_{0j}, \tau_{1j}, \psi_0^2, \psi_1^2, \psi_{01})$.

Consider again the multilevel regression model in equation (2.24). Note that the explanatory variables of the model are unstandardized (see the second column of Table 4.2 in Hox (2010, p.60)). The bootstrapped estimates and their covariance matrix are displayed in Table 2.9. Because the bootstrapped estimates are in agreement with the constraints in hypotheses H_1, H_2, and the unconstrained hypothesis $H_3 : \beta_{10}, \beta_{01}, \beta_{11}$, the order-restricted MLEs of these hypotheses are equal to the MLEs (see Table 2.10). Thus, the three hypotheses have the same log likelihood and, therefore, the GORICA distinguishes among these hypotheses only through the penalty term (see Table 2.11). Based on the GORICA weights, both hypotheses

Table 2.9: Estimates and covariance matrix of the parameters in the multilevel regression example

		$\hat{\Sigma}_{\hat{\theta}}$		
θ	$\hat{\theta}$	β_{10}	β_{01}	β_{11}
β_{10}	1.33	4.7e-3		
β_{01}	0.11	1.4e-4	1.0e-5	
β_{11}	-0.03	-2.8e-4	-1.0e-5	1.9e-5

Table 2.10: The order-restricted MLEs for the multilevel regression example for each hypothesis H_m

H_m	$\tilde{\beta}^m_{10}$	$\tilde{\beta}^m_{01}$	$\tilde{\beta}^m_{11}$
H_1, H_2, H_3	1.33	0.11	-0.03

H_1 and H_2 are good hypotheses, that is, hypothesis H_1 is 0.342 / 0.125 $\approx$ 2.74 and hypothesis H_2 is 0.533 / 0.125 $\approx$ 4.26 times better than the unconstrained hypothesis H_3. However, hypothesis H_2 is a better hypothesis than hypothesis H_1, because it is 0.533 / 0.342 $\approx$ 1.56 times better supported by the data compared to hypothesis H_1.

Table 2.11: The likelihood and penalty parts, GORICA values, and GORICA weights of hypotheses H_1, H_2, and H_3 for the multilevel regression example

H_m	$L(\tilde{\theta}^m \mid \hat{\theta}, \hat{\Sigma}_{\hat{\theta}})$	PT_m	$GORICA_m$	w_m
H_1	12.378	1.998	-20.760	0.342
H_2	12.378	1.555	-21.647	0.533
H_3	12.378	3.000	-18.756	0.125

2.6.5 Structural equation models (SEMs)

SEMs (Kline, 2010) typically contain two submodels: The measurement model and the structural model. The measurement model represents the connections between unobserved (latent) variables and observed (indicator) variables while the structural model describes the relations between dependent and independent variables.

Let $y_i = (y_{1i}, y_{2i}, ..., y_{pi}) \in \mathbb{R}^{p\times 1}$ and $x_i = (x_{1i}, x_{2i}, ..., x_{ti}) \in \mathbb{R}^{t\times 1}$ denote the vectors of indicator and latent variables, respectively, for person $i = 1, 2, ..., N$. The measurement model can be written as:

$$y_i = \Lambda x_i + \upsilon_i, \tag{2.26}$$

where $\Lambda \in \mathbb{R}^{p\times t}$ is the matrix of factor loadings relating observed and unobserved variables, $\upsilon \sim N(0, \Phi)$ represents measurement error with $\upsilon \in \mathbb{R}^{p\times 1}$ and covariance matrix $\Phi \in \mathbb{R}^{p\times p}$. The observed variables y_i have covariance matrix $\Sigma_y = \Lambda\Sigma_x\Lambda^T + \Phi$ in which $\Sigma_x \in \mathbb{R}^{t\times t}$ represents the covariance matrix of x.

Let $x_i = \{\eta_i, \alpha_i\}$ with $\eta = (\eta_1, \eta_2, ..., \eta_{t_1}) \in \mathbb{R}^{t_1\times 1}$ and $\alpha = (\alpha_1, \alpha_2, ..., \alpha_{t_2}) \in \mathbb{R}^{t_2\times 1}$ denote the vectors of latent dependent and independent variables, respectively. The structural model is then given by

$$\eta_i = \Pi\eta_i + \Gamma\alpha_i + \delta_i, \tag{2.27}$$

where zero-diagonal $\Pi \in \mathbb{R}^{t_1\times t_1}$ and $\Gamma \in \mathbb{R}^{t_1\times t_2}$ are the matrices of regression coefficients relating the latent variables and $\delta_i \sim N(0, \zeta)$ is the vector of residuals with $\delta \in \mathbb{R}^{t_1\times 1}$ and $\zeta \in \mathbb{R}^{t_1\times t_1}$. Note that the structural model can also be used to relate observed variables with latent variables. In that case, η and α are extended with latent variables that have only one indicator: the observed variable of interest. Note furthermore that the Π's, Γ's, and Λ's are often used as the structural parameters in (in)equality constrained hypotheses, while the entries of Φ, ζ, Σ_x, and Σ_y are generally considered as the nuisance parameters.

2.6.6 Example 3: Structural equation modeling

The specification of (in)equality constrained hypotheses for structural equation models are exemplified based on the study in Stevens (1999, p.596) that evaluates the effects of the first year of the Sesame Street television series in a sample (N = 240) of 3-5 years old children in the USA. The sample contains two latent variables: The first latent variable is the prewatch (x_1) with prebody (y_1; pretest on knowledge of body parts), prelet (y_2; pretest on letters), preform (y_3; pretest on forms), prenumb (y_4; pretest on numbers), prerelat (y_5; pretest on relational terms), and preclas (y_6; pretest on classification skills) as indicators. The second latent variable is the postwatch (x_2) with the postbody (y_7; posttest on knowledge of body parts), postlet (y_8; posttest on letters), postform (y_9; posttest on forms), postnumb (y_{10}; posttest on numbers), postrel (y_{11}; posttest on relational terms), and postclas (y_{12}; posttest on classification skills) as indicators. Both latent variables have regression relations with four predictors, namely, the site, S $\in \{0 =$ "disadvantaged", $1 =$ "advantaged"$\}$ for which

0 represents spanish speaking children, disadvantaged rural children, and the children from inner city areas of the USA, while 1 represents advantaged rural and suburban children, the gender, G $\in \{0 = \text{male}, 1 = \text{female}\}$, the age in months (A), and the peabody score (P), that is, a mental age score which is obtained from the Peabody Picture Vocabulary test. The descriptives for each observed variable are displayed in Table 2.12. Moreover, Figure 2.2 shows how all variables are related to each other. The measurement model for the Sesame Street data can be defined for the ith person as in equation (2.26), where $y_i = (y_1, y_2, ..., y_{12})^T$ denotes a vector of indicators, $x_i = (x_1, x_2)^T$ denotes a vector of latent variables,

$$\Lambda = \begin{bmatrix} \lambda_1 & \lambda_2 & \lambda_3 & \lambda_4 & \lambda_5 & \lambda_6 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & \lambda_7 & \lambda_8 & \lambda_9 & \lambda_{10} & \lambda_{11} & \lambda_{12} \end{bmatrix}^T$$

is a 12×2 matrix of factor loadings, and υ_i is a vector of measurement errors with $\upsilon \sim N(0, \Phi)$ with Φ being its 12×12 covariance matrix. The covariance matrix of latent variables Σ_x is a 2×2 matrix:

$$\Sigma_x = \begin{bmatrix} \omega_{11} & \omega_{12} \\ \omega_{21} & \omega_{22} \end{bmatrix}.$$

As mentioned the structural model can be used to relate not only latent variables to each other but also latent variables with observed variables. Therefore, the structural model for the Sesame Street data can be defined for the ith person as:

$$x_i = \Gamma\alpha_i + \delta_i, \tag{2.28}$$

where $x_i = (x_1, x_2)^T$ denotes a vector of latent variables,

$$\Gamma = \begin{bmatrix} \gamma_1 & \gamma_2 & \gamma_3 & \gamma_4 \\ \gamma_5 & \gamma_6 & \gamma_7 & \gamma_8 \end{bmatrix}$$

Table 2.12: Descriptives for the variables in the SEM

	prebody	prelet	preform	prenumb	prerelat	preclas	postbody	postlet
Min:	6.000	1.000	2.000	1.000	2.000	0.000	11.000	0.000
Mean:	21.400	15.940	9.921	20.900	9.938	12.240	25.260	26.700
Max:	32.000	55.000	19.000	52.000	17.000	24.000	32.000	54.000
SD:	6.390	8.536	3.737	10.685	3.074	4.658	5.412	13.272
	postform	postnumb	postrel	postclas	site	gender	age	peabody
Min:	0.000	0.000	0.000	0.000	0.000	0.000	34.000	27.000
Mean:	13.740	30.050	11.650	15.740	0.496	0.521	51.520	46.770
Max:	20.000	54.000	17.000	24.000	1.000	1.000	69.000	99.000
SD:	4.001	12.846	2.832	5.151	0.501	0.500	6.281	15.987

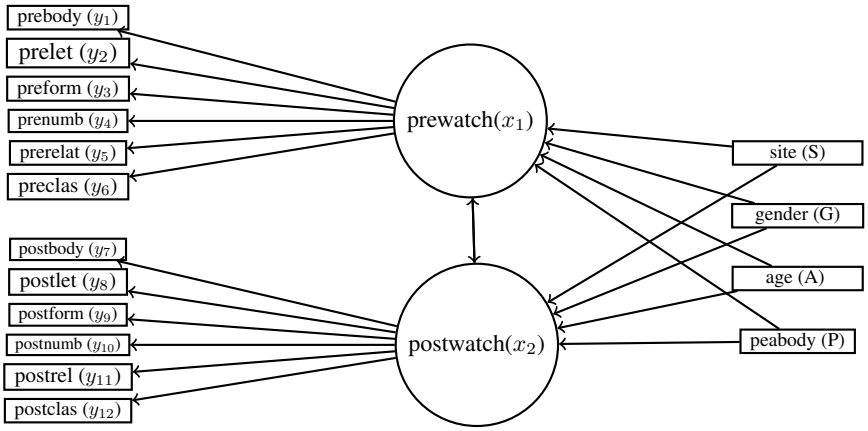

Figure 2.2: Structural Equation Model for the Sesame Street Data

is a 2×4 matrix that relates latent variables with predictors, $\alpha_i = (S_i, G_i, A_i, P_i)^T$ is a 4×1 vector of predictors, and $\delta_i \sim N(0, \zeta)$ is the vector of residuals with ζ is a 2×2 matrix. Note that η and α in equation (2.27) are extended with x which has only one indicator α, that is, the combination of the four predictors.

Six (in)equality constrained hypotheses are formulated based on expectations about the order of parameters. The first set of hypotheses which are about the regression relations between latent variables and predictors are:

$$\begin{aligned} H_1 &: \gamma_4 > \gamma_3 > \gamma_2 > \gamma_1,\ \gamma_8 > \gamma_7 > \gamma_6 > \gamma_5, \\ H_2 &: \gamma_8 > \gamma_4,\ \gamma_7 > \gamma_3,\ \gamma_6 > \gamma_2,\ \gamma_5 > \gamma_1, \end{aligned} \tag{2.29}$$

where H_1 states that the effect of site is the smallest effect for each latent variable followed by effects gender, age, and peabody, H_2 states that the effect of each predictor is bigger on postwatch when compared to prewatch. The second set of hypotheses are about the factor loadings:

$$\begin{aligned} H_3 &: \lambda_7 > \lambda_1,\ \lambda_8 > \lambda_2,\ \lambda_9 > \lambda_3,\ \lambda_{10} > \lambda_4,\ \lambda_{11} > \lambda_5,\ \lambda_{12} > \lambda_6, \\ H_4 &: \lambda_1 > \lambda_7,\ \lambda_2 > \lambda_8,\ \lambda_3 > \lambda_9,\ \lambda_4 > \lambda_{10},\ \lambda_5 > \lambda_{11},\ \lambda_6 > \lambda_{12}, \end{aligned} \tag{2.30}$$

where H_3 states that the factor loadings for postwatch are bigger than the factor loadings for prewatch, H_4 states that the factor loadings for prewatch are bigger than the factor loadings for postwatch. The third set of hypotheses are about the correlation between latent variables. Hypothesis H_5 specifies that the covariance between latent variables is zero, that is, $\omega_{12} = 0$, while hypothesis H_6 states that the covariance is positive, that is, $\omega_{12} > 0$.

The structural parameters used to formulate the six hypotheses are $\theta = (\gamma_1, ..., \gamma_8, \lambda_1, ..., \lambda_{12}, \omega_{12})$ and the nuisance parameters are $\xi = (\Phi, \zeta, \omega_{11}, \omega_{22})$.

Within each bootstrap sample both observed and latent variables are standardized to ensure that the structural parameters are comparable. To obtain the standardized parameter estimates within each bootstrap sample "lavaan" package (Rosseel, 2012) in R is used.

Consider again the structural equation model introduced earlier. To be able to compare the structural parameters to each other, both latent and observed variables are standardized within each bootstrap sample using the "lavaan" package (Rosseel, 2012) in R. Standardized bootstrapped estimates and their covariance matrix for the regression parameters are shown in Table 2.13.1.

Table 2.13.1: Estimates and covariance matrix of the regression parameters for the SEM example

		$\hat{\Sigma}_{\hat{\theta}}$							
θ	$\hat{\theta}$	γ_1	γ_2	γ_3	γ_4	γ_5	γ_6	γ_7	γ_8
γ_1	-0.08	2.5e-3							
γ_2	0.01	-4.1e-4	2.6e-3						
γ_3	0.29	-3.2e-4	9.5e-5	2.8e-3					
γ_4	0.60	-1.7e-4	4.8e-4	-1.6e-3	2.6e-3				
γ_5	0.04	1.6e-3	-7.2e-4	-2.2e-4	-7.5e-5	3.5e-3			
γ_6	0.03	-6.0e-4	1.6e-3	-1.2e-4	4.5e-4	-4.6e-4	3.0e-3		
γ_7	0.19	-3.7e-4	-1.2e-4	1.9e-3	-1.3e-3	1.7e-4	3.1e-4	3.2e-3	
γ_8	0.50	-1.1e-4	4.3e-4	-9.8e-4	1.9e-3	-8.0e-4	5.9e-5	-1.6e-3	3.1e-3

Table 2.13.2: Estimates and covariance matrix of the indicators for the SEM example

		$\hat{\Sigma}_{\hat{\theta}}$											
θ	$\hat{\theta}$	λ_1	λ_2	λ_3	λ_4	λ_5	λ_6	λ_7	λ_8	λ_9	λ_{10}	λ_{11}	λ_{12}
λ_1	0.79	7.8e-4											
λ_2	0.68	-2.8e-5	1.9e-3										
λ_3	0.79	1.4e-4	-1.2e-5	7.5e-4									
λ_4	0.91	-1.6e-5	4.5e-4	-1.7e-5	3.2e-4								
λ_5	0.78	9.3e-5	-2.8e-6	7.5e-5	3.3e-5	9.7e-4							
λ_6	0.80	-9.0e-6	4.1e-5	2.3e-4	4.0e-5	2.2e-4	9.0e-4						
λ_7	0.74	2.9e-4	5.7e-7	1.0e-4	4.3e-5	2.8e-6	4.1-5	9.5e-4					
λ_8	0.85	4.7e-5	1.8e-4	9.3e-5	7.7e-5	8.2e-5	4.6e-5	9.1e-5	3.5e-4				
λ_9	0.80	1.1e-4	2.7e-5	2.2e-4	2.4e-5	4.7e-7	2.1e-4	2.7e-4	-3.5e-5	8.2e-4			
λ_{10}	0.91	1.4e-5	1.0e-4	3.9e-5	5.7e-5	4.9e-5	8.6e-5	3.3e-5	1.3e-4	-4.1e-5	1.8e-4		
λ_{11}	0.73	2.8e-4	2.5e-4	1.8e-4	5.0e-5	1.4e-4	1.4e-4	2.7e-4	6.9e-5	3.5e-4	1.0e-4	1.1e-3	
λ_{12}	0.86	5.3e-6	-3.3e-6	1.7e-4	9.2e-6	-2.7e-5	2.2e-4	1.5e-4	-3.5e-5	4.9e-4	-2.3e-5	2.5e-4	5.6e-4

Table 2.13.3: Estimate and variance estimate of the correlation between the latent variables for the SEM example

θ	$\hat{\theta}$	$\hat{\Sigma}_{\hat{\theta}}$
ω_{12}	0.64	4.0e-3

The order-restricted MLEs for hypotheses H_1, H_2, with the unconstrained hypothesis H_u : $\gamma_1, ..., \gamma_8$ are given in Table 2.14.1. Based on the GORICA weights in Table 2.15.1, hypothesis H_1 is a very strong hypothesis, that is, $0.969/0.031 \approx 31.258$ times better than the unconstrained hypothesis, while hypothesis H_2 is a weak hypothesis which has the GORICA weight zero. These results imply that when assessing social and mental intelligence of children before and after watching the Sesame Street television series, peabody score and site are the most and least important predictors, respectively, and their importance for the assessment of social and mental intelligence, that is measured by prewatch and postwatch did not increase after watching the television series.

Table 2.14.1: The order-restricted MLEs of the regression parameters for the SEM example

H_m	$\tilde{\gamma}_1^m$	$\tilde{\gamma}_2^m$	$\tilde{\gamma}_3^m$	$\tilde{\gamma}_4^m$	$\tilde{\gamma}_5^m$	$\tilde{\gamma}_6^m$	$\tilde{\gamma}_7^m$	$\tilde{\gamma}_8^m$
H_1	-0.09	0.01	0.29	0.60	0.04	0.04	0.19	0.50
H_2	-0.08	-0.01	0.27	0.56	0.007	0.03	0.27	0.56
H_u	-0.08	0.01	0.29	0.60	0.04	0.03	0.19	0.50

Table 2.14.2: The order-restricted MLEs of the indicators for the SEM example

H_m	$\tilde{\lambda}_1^m$	$\tilde{\lambda}_2^m$	$\tilde{\lambda}_3^m$	$\tilde{\lambda}_4^m$	$\tilde{\lambda}_5^m$	$\tilde{\lambda}_6^m$	$\tilde{\lambda}_7^m$	$\tilde{\lambda}_8^m$	$\tilde{\lambda}_9^m$	$\tilde{\lambda}_{10}^m$	$\tilde{\lambda}_{11}^m$	$\tilde{\lambda}_{12}^m$
H_3	0.77	0.68	0.79	0.91	0.75	0.80	0.77	0.85	0.82	0.92	0.75	0.87
H_4	0.78	0.84	0.79	0.94	0.79	0.84	0.73	0.84	0.79	0.92	0.74	0.84
H_u	0.79	0.68	0.79	0.91	0.78	0.80	0.74	0.85	0.80	0.91	0.73	0.86

Table 2.14.3: The order-restricted MLEs of the correlation between the latent variables for the SEM example

H_m	$\tilde{\omega}_{12}^m$
H_5	0.00
H_6	0.64
H_u	0.64

Table 2.15.1: The likelihood and penalty parts, GORICA values, and GORICA weights of hypotheses H_1, H_2, and H_u for the SEM example

H_m	$L(\tilde{\theta}^m \mid \hat{\theta}, \hat{\Sigma}_{\hat{\theta}})$	PT_m	$GORICA_m$	w_m
H_1	17.514	4.558	-25.912	0.969
H_2	8.347	5.872	-4.950	0.000
H_u	17.522	8.000	-19.045	0.031

Table 2.15.2: The likelihood and penalty parts, GORICA values, and GORICA weights of hypotheses H_3, H_4, and H_u for the SEM example

H_m	$L(\tilde{\theta}^m \mid \hat{\theta}, \hat{\Sigma}_{\hat{\theta}})$	PT_m	$GORICA_m$	w_m
H_3	32.743	9.339	-46.808	0.784
H_4	23.742	9.336	-28.812	0.001
H_u	34.112	12.000	-44.224	0.215

Table 2.15.3: The likelihood and penalty parts, GORICA values, and GORICA weights of hypotheses H_5, H_6, and H_u for the SEM example

H_m	$L(\tilde{\theta}^m \mid \hat{\theta}, \hat{\Sigma}_{\hat{\theta}})$	PT_m	$GORICA_m$	w_m
H_5	-48.692	0.000	97.384	0.000
H_6	1.834	0.499	-2.670	0.623
H_u	1.834	1.000	-1.668	0.377

Standardized bootstrapped estimates and their covariance matrix for the indicators are displayed in Table 2.13.2. The order-restricted MLEs for hypotheses H_3, H_4, with the unconstrained hypothesis $H_u : \lambda_1, ..., \lambda_{12}$ are given in Table 2.14.2. As can be seen in Table 2.15.2, hypothesis H_3 is preferred over hypothesis H_u, that is, $0.784/0.215 \approx 3.647$ times better than hypothesis H_u. This result means that postwatch is measured more reliably than prewatch and consequently hypothesis H_4 is a bad hypothesis.

Standardized bootstrap estimate and its variance estimate for the covariance between latent variables are displayed in Table 2.13.3 and the order-restricted MLEs for the hypotheses with the unconstrained hypothesis $H_u : \omega_{12}$ are given in Table 2.14.3. Note in Table 2.15.3 that hypothesis H_6 (a positive covariance between factors) receives support from the data, that is, $0.623/0.377 \approx 1.652$ times better than the unconstrained hypothesis, while hypothesis H_5 (zero covariance between factors) is not supported by the data at all.

2.7 Discussion

In this paper, an Akaike-type information criterion (AIC) (Akaike, 1973, 1974), the GORICA, has been proposed to extend the applicability of the GORIC to a large class of statistical models, that is, GLMs, GLMMs, and SEMs. To illustrate the approach, (in)equality constrained hypotheses were evaluated for logistic regression, multilevel regression, and structural equation models as being representatives of GLMs, GLMMs, and SEMs, respectively. The investigation of the examples rendered GORICA weights that quantified the distance of the current hypothesis to the true hypothesis.

A simulation study was performed to compare the performance of the GORIC and GORICA in a normal linear regression model. The comparison showed that the performance of the GORICA is similar to that of the GORIC which implies that the loss of information is negligible when the GORICA is utilized instead the GORIC. Another simulation study was performed to investigate the performance of the GORICA in the context of a logistic regression model. Again, the performance of the GORICA was convincing.

Note that the GORICA is implemented in the R script Gorica.R using the MLEs of structural parameters and their covariance matrix. The R script Gorica.R is discussed in Appendix 2.B.

2.A Separation problem

In the framework of logistic regression modeling, when a binary outcome separates the values of an explanatory variable or a combination of explanatory variables completely, that is called "complete separation", or to a certain level, that is called "quasi complete separation". The probability of the outcome being 0 or 1 is perfectly predicted or nearly perfectly predicted in the cases of complete and quasi complete separation, respectively (Zorn, 2005). Two small data sets are given to illustrate the complete and quasi complete separation, respectively:

Data 1		**Data 2**	
R	E	R	E
0	3	0	3
0	8	0	8
0	1	0	1
0	6	0	6
1	10	1	8
1	15	1	10
1	9	1	15
1	11	1	9

where R is a binary response and E is a continuous explanatory variable.

For the first data set, the values of R are 0 for $\mathrm{E} \leq 8$, and 1 for $\mathrm{E} > 8$ which implies that $\mathrm{P}(\mathrm{R} = 0|\mathrm{E} \leq 8) = 1$ and $\mathrm{P}(\mathrm{R} = 1|\mathrm{E} > 8) = 1$ with respect to the expected probabilities. There is no value of E for which the values of R differ which implies that the probability of R being 0 or 1 is perfectly predicted.

For the second data set, the values of R are 0 for $\mathrm{E} < 8$, 1 for $\mathrm{E} > 8$, and 0 or 1 for $\mathrm{E} = 8$ which implies that $\mathrm{P}(\mathrm{R} = 0|\mathrm{E} < 8) = 1$ and $\mathrm{P}(\mathrm{R} = 1|\mathrm{E} > 8) = 1$ with respect to the expected probabilities. The only value of E for which the values of R differ is 8 which implies that the probability of R being 0 or 1 is nearly perfectly predicted.

When complete or quasi complete separation occurs, the maximum likelihood method in the "glm" (R Development Core Team, 2012) procedure, produces large parameter estimates and standard errors (Webb, Wilson, & Chong, 2004, pp.282-283). Consequently, R gives the warning message: "probabilities numerically 0 or 1 occured". Note that this warning message is encountered when fitting the logistic regression model in equation (2.6) to bootstrap samples. We used the approach that (Kamakura, 2011) suggests, that is, the samples with the separation problem are removed from the bootstrap samples. The separation problem is encountered with only 30 out of $B = 1000$ bootstrap samples for the example data, that is, $B = 970$ bootstrap samples are used to obtain the bootstrapped estimates and their covariance matrix.

2.B User manual for Gorica.R

In this appendix, it will be discussed how to evaluate (in)equality constrained hypotheses using the R script Gorica.R. Note that the evaluation process can be executed as long as structural parameters and their covariance matrix are obtained. Users only need an input file called input.txt to be able to execute the code in Gorica.R. The name of the Input.txt file has been fixed in the R script Gorica.R and therefore cannot be altered. The script Gorica.R and the file Input.txt are available on the website `http://informative-hypotheses.sites.uu.nl/software` under folder GORICA. The entries of the input file for the three examples in this paper are also provided in folder GORICA. The input file for hypothesis H_1 in the logistic regression example is displayed below and elaborated on next:

K	Eq	Ineq	Seed	Iterations
6	3	3	111	100000

Estimates of structural parameters

0.7862	-0.2179	0.3154	0.6676	0.6300	0.2004

Covariance matrix of structural parameters

1.20998	-1.20834	0.06543	-0.06445	-0.01702	0.06137
-1.20834	1.52253	-0.06385	0.04148	0.00496	-0.05989
0.06543	-0.06385	1.23877	-1.21600	-0.00321	0.00975
-0.06445	0.04148	-1.21600	1.71649	0.00335	-0.02673
-0.01702	0.00496	-0.00321	0.00335	0.09153	-0.09155
0.06137	-0.05989	0.00975	-0.02673	-0.09155	0.31680

Restriction matrix

1	0	0	0	0	0
0	0	1	0	0	0
0	0	0	0	0	1
0	1	0	0	0	0
0	0	0	1	0	0
0	0	0	0	1	0

Right hand side values

0	0	0	0	0	0

The first line consists of five labels: label K for the number of structural parameters in H_m, label Eq for the number of the equality constraints in H_m, label $Ineq$ for the number of the inequality constraints in H_m, label Seed for the index number used in the random number generator, and label Iterations for the number of the iterations when calculating the penalty term. The third and fifth lines label structural parameters and their covariance matrix, respectively. Although the input file is for one hypothesis, structural parameters are based on the whole set of hypotheses. Both matrices S and R in equation (2.3), representing the equality and inequality constraints of hypothesis H_m, are reflected by label Restriction matrix, while the zeros of the equality and inequality constraints in H_m are reflected by label Right hand side values. The script Gorica.R always considers equality constraints first. Therefore, equality constraints must be specified before any of the inequality constraints. Both the number of the rows below label Restriction matrix and the length of the vector below label Right hand side values are equal to the total number of constraints, that is, $Eq + Ineq$. The first Eq lines below label Restriction matrix and the first Eq elements below label Right hand side values represent the equality constraints in H_m together, while the last $Ineq$ lines below label Restriction matrix and the last $Ineq$ elements below label Right hand side values represent the inequality constraints in H_m together. To illustrate how to specify the restriction matrix and right hand side values, consider the following example with $K = 5$ structural parameters, $Eq = 2$ equality constraints, and $Ineq = 3$ inequality constraints:

Hypotheses	Restriction matrix	Right hand side values
• $\theta_1 - \theta_2 + \theta_4 = 0$	1 -1 0 1 0	0,
• $\theta_4 - \theta_5 = 0$	0 0 0 1 -1	0,
• $\theta_1 - \theta_2 + \theta_3 - \theta_5 > 0$	1 -1 1 0 -1	0,
• $\theta_2 - \theta_4 > 0$	0 1 0 -1 0	0,

• $\theta_3 - \theta_5 > 0$ 0 0 1 0 -1 0.

In general, for equality constraints,

• $a\theta_1 + b\theta_2 + c\theta_3 + d\theta_4 + e\theta_5 = 0$ $a\ b\ c\ d\ e$ 0,

and for inequality constraints

• $a\theta_1 + b\theta_2 + c\theta_3 + d\theta_4 + e\theta_5 > 0$ $a\ b\ c\ d\ e$ 0.

Each row corresponds to only one equality or inequality constraint. Therefore, for example, $H_1 : \theta_1 = \theta_2 > \theta_3$ has two rows. The first two elements for label Right hand side values correspond to the right hand side values for equality constraints, while the last three elements correspond to the right hand side values for inequality constraints.

The script Gorica.R for the input file can easily be executed in R. The execution procedure consists of the following basic steps:

• Install the packages "matrixcalc", "MASS", "quadprog", and "FRACTION" in R.

• Change the working directory to folder GORICA with the "setwd()" function in R. For example, setwd("C:/Users/(Username)/Desktop/GORICA"), if the GORICA file is saved on the desktop of computer.

• Specify the entries in the input file.

• Save the changes (CTRL+S).

•Apply the "source()" function in R as shown below:

source("Gorica.R").

Executing script Gorica.R based on the input file renders the following results in R:

$MLEs

0.7862 -0.2179 0.3154 0.6676 0.6300 0.2004

$Covariance_matrix

1.20998	-1.20834	0.06543	-0.06445	-0.01702	0.06137
-1.20834	1.52253	-0.06385	0.04148	0.00496	-0.05989
0.06543	-0.06385	1.23877	-1.21600	-0.00321	0.00975
-0.06445	0.04148	-1.21600	1.71649	0.00335	-0.02673
-0.01702	0.00496	-0.00321	0.00335	0.09153	-0.09155
0.06137	-0.05989	0.00975	-0.02673	-0.09155	0.31680

$Restricted_MLEs

0.0000	0.5663	0.0000	0.9869	0.6875	0.0000

	LogLikelihood	Penalty	GORICA_value	GORICA_weight
H_m	-3.1879	1.4464	9.2689	0.9856
H_u	-2.8621	6.0000	17.7243	0.0144

When using the script Gorica.R, only one hypothesis of interest H_m can be evaluated simultaneously with the unconstrained hypothesis H_u.

Chapter 3

An AIC-based information criterion evaluating (in)equality constrained hypotheses for contingency tables[1]

3.1 Introduction

Contingency tables, sometimes called cross-classification tables, can be utilized to investigate the relationships between two or more nominal or ordinal variables. In contingency tables, often the hypothesis of no association is tested, while researchers generally expect some association between the variables. Evaluation of researchers' expectations is naturally handled using equality and/or inequality constrained hypotheses with respect to model parameters. To illustrate (in)equality constrained hypotheses in the context of contingency tables, consider the study of (Moore, McCabe, & Craig, 2009, p.305) investigating the degrees earned in higher education in terms of gender in the USA during the academic year 2010-2011. The nominal variable gender is defined as G $\in$ {1 = Female, 2 = Male}. Similarly, the ordinal variable academic degree is defined as AD $\in$ {1 = Bachelor's, 2 = Master's, 3 = Professional, 4 = Doctorate}. The population probability of a person having the ith level of gender and the jth level of academic degree is represented by π_{ij} with $i = 1, 2$ and $j = 1, 2, 3, 4$, which are displayed (together with the observed cell frequencies) in Table 3.1. In this study, a hypothesis of interest could be: H_1 : After correction for the prevalence of each degree, a bachelor's degree being obtained by males is more likely than any other degrees being obtained by males. This hypothesis can formally be represented by the (in)equality constrained

[1]This chapter will be submitted as Altınışık, Y., Hessels, R. S., & Kuiper, R. M. An AIC-based Information Criterion Evaluating (In)equality Constrained Hypotheses for Contingency Tables.
Author contributions: YA and RMK developed the study concept. All authors contributed to the study design. Data collection was performed by RHS and the writing, programming analysis, and interpretation was performed by YA under the supervision of RMK. All authors read, commented, and approved the final manuscript.

Table 3.1: Population probabilities π_{ij} and observed cell frequencies between brackets for the combinations of gender and academic degree

π_{ij}	1 = Bachelor's	2 = Master's	3 = Professional	4 = Doctorate
1 = Female	π_{11} (933)	π_{12} (402)	π_{13} (51)	π_{14} (26)
2 = Male	π_{21} (661)	π_{22} (260)	π_{23} (44)	π_{24} (26)

hypothesis:

$$H_1 : \frac{\pi_{21}}{\pi_{+1}} > \{\frac{\pi_{22}}{\pi_{+2}}, \frac{\pi_{23}}{\pi_{+3}}, \frac{\pi_{24}}{\pi_{+4}}\}, \tag{3.1}$$

where the operator "+" denotes a sum of cell probabilities over one level of a variable. For example, $\pi_{+1} = \pi_{11} + \pi_{21}$ is the sum of proportions of the bachelor's degrees of females and males.

Analysis of contingency tables can be rather difficult. An important model used to analyze contingency tables is the log-linear model (Agresti, 2007; Azen & Walker, 2010). However, when the contingency table is high-dimensional it might be virtually impossible to interpret the log-linear parameters (Hendrickx, 2004, p.603). For example, consider a $5\times5\times5$ contingency table modeled using a full log-linear model. This model contains 124 parameters in total, that is, 12 main effects, 48 two-way interaction effects, and 64 three-way interaction effects. There are too many parameters in this model, and it is obvious that the three-way interaction effects are hard to interpret by researchers. In addition, high-dimensional contingency tables often have empty cells which complicate parameter estimation. In this paper, an alternative for log-linear analysis to evaluate the hypotheses of interest will be proposed. First, a set of (in)equality constrained hypotheses should be formulated. Second, the evidence in the data for these hypotheses will be quantified by evaluating them using an AIC-based information criterion (Akaike, 1973, 1974). This approach avoids interpretation problems and can be used for analyzing high-dimensional sparse contingency tables containing small counts and/or zeros.

Evaluation of (in)equality constrained hypotheses using an AIC-based information criterion has not yet been investigated for contingency tables.[2] The generalized order-restricted information criterion (GORIC) (Kuiper et al., 2011, 2012) is an AIC-based information criterion that can be used to evaluate (in)equality constrained hypotheses. However, the GORIC can only be used in the context of normal linear models. Altınışık, Nederhof, Hoijtink, Oldehinkel, and Kuiper (unpublished) propose an approximation to the GORIC, the generalized order restricted information criterion approximation (GORICA), that evaluates (in)equality constrained hypotheses for generalized linear models (GLMs) (McCullagh & Nelder, 1989),

[2]Readers who are more in favor of the Bayesian framework instead of the classical inferential one are referred to Klugkist, Laudy, and Hoijtink (2010), who evaluate (in)equality constrained hypotheses in the context of contingency tables using a model selection criterion called Bayes factor (Kass, 1993; Kass & Raftery, 1995).

generalized linear mixed models (GLMMs) (McCullogh & Searle, 2001), and structural equation models (SEMs) (Bollen, 1989). However, the GORICA can only be used to evaluate (in)equality constrained hypotheses specifying linear restrictions on model parameters. As is shown by hypothesis H_1 in equation (3.1), hypotheses in the context of contingency tables are often expressed through non-linear restrictions on cell probabilities. In this paper, we elaborate how the GORICA can be applied to hypotheses containing linear or non-linear restrictions on cell probabilities. It will be shown that the GORICA can easily be implemented for high-dimensional contingency tables with and without empty cells.

The outline of the paper is as follows. First, five classes of restrictions often used in hypotheses regarding contingency tables are introduced. Second, we introduce the GORICA that can be used to evaluate hypotheses specified using these restrictions. Third, we elaborate three major problems that may be encountered when evaluating (in)equality constrained hypotheses and give solutions we provide in our software. Fourth, the GORICA is applied to two examples: one with hypotheses containing linear restrictions on cell probabilities for two high-dimensional contingency tables with empty cells and the other with hypotheses containing non-linear restrictions on cell probabilities. The paper will end with a discussion.

3.2 Classes of restrictions common in contingency tables

Let $\pi = (\pi_{111}, ..., \pi_{ijv}, ..., \pi_{IJV})$ denote the cell probabilities in a three-way contingency table, without loss of generalization to higher-way contingency tables. The parameter π_{ijv} is the cell probability for the ith level of the first variable, the jth level of the second variable, and the vth level of the third variable in the contingency table with $i = 1, 2, ..., I$, $j = 1, 2, ..., J$, and $v = 1, 2, ..., V$. The GORICA can straightforwardly be utilized to evaluate hypotheses containing linear restrictions on (functions of) these cell probabilities. These hypotheses are of the type:

$$H_m : S_m\eta = s_m, R_m\eta > r_m, \tag{3.2}$$

where $S_m \in \mathbb{R}^{h_s \times K}$ and $R_m \in \mathbb{R}^{h_r \times K}$ are the restriction matrices representing the equality and inequality constraints in H_m respectively, $\eta = g(\pi) > 0$, denotes (functions of) cell probabilities in contingency tables, and s_m is a h_s-vector and r_m is a h_r-vector consisting of the constants in H_m. To illustrate, consider the hypothesis $H_1 : \frac{\pi_{21}}{\pi_{+1}} > \{\frac{\pi_{22}}{\pi_{+2}}, \frac{\pi_{23}}{\pi_{+3}}, \frac{\pi_{24}}{\pi_{+4}}\}$ in equation (3.1) for the gender by earned degrees example, which is non-linear in π's. However, hypothesis H_1 can be rewritten as $H_1 : \eta_1 > \{\eta_2, \eta_3, \eta_4\}$ with $\eta = (\eta_1, \eta_2, \eta_3, \eta_4)^T = (\frac{\pi_{21}}{\pi_{+1}}, \frac{\pi_{22}}{\pi_{+2}}, \frac{\pi_{23}}{\pi_{+3}}, \frac{\pi_{24}}{\pi_{+4}})^T$ and

$$R_1 = \begin{pmatrix} 1 & -1 & 0 & 0 \\ 1 & 0 & -1 & 0 \\ 1 & 0 & 0 & -1 \end{pmatrix}, r_1 = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}.$$

A second hypothesis containing only linear restrictions on the cell probabilities could be:

$$H_2 : \pi_{11} = \pi_{12}, \pi_{23} > \pi_{24}, \tag{3.3}$$

which states that obtaining a bachelor's degree and a master's degree is equally likely for females, and obtaining a professional degree is more likely than obtaining a doctorate degree for males. Since hypothesis H_2 contains only cell probabilities and not functions of them and these cell probabilities are linearly dependent on each other (i.e., they sum to one), it holds that $\eta = (\eta_1, \eta_2, ..., \eta_8)^T = (\pi_{11}, \pi_{12}, ..., \pi_{24})^T$,

$$S_2 = \begin{pmatrix} 1 & -1 & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}, s_2 = 0,$$

and

$$R_2 = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 1 & -1 \end{pmatrix}, r_2 = 0.$$

As shown by Tables 3.2 and 3.3, five classes of restrictions commonly used in the context of contingency table analysis will be considered in this paper: Linear restrictions of the type $H_m : S_m\pi = s_m$, $R_m\pi > r_m$, restrictions on conditional cell probabilities, restrictions on local odds ratios, restrictions on marginal odds ratios, and restrictions on conditional odds ratios. Note that the class of marginal cell probabilities is a subclass of linear restrictions, which will be elaborated upon later in this paper by means of an example, involving two high-dimensional sparse contingency tables. The last four classes of restrictions above contain non-linear restrictions on cell probabilities.

In Tables 3.2 and 3.3, each line represents a subclass of the corresponding class of restrictions. The GORICA evaluation of (in)equality constrained hypotheses is only investigated in cases when the restrictions of the hypotheses under consideration all belong to the same subclass. For example, the hypotheses $H_1 : \frac{\pi_{21}}{\pi_{+1}} > \{\frac{\pi_{22}}{\pi_{+2}}, \frac{\pi_{23}}{\pi_{+3}}, \frac{\pi_{24}}{\pi_{+4}}\}$ and $H_2 : \frac{\pi_{12}}{\pi_{1+}} > \frac{\pi_{22}}{\pi_{2+}}$ in Table 2 are not comparable. Although their restrictions belong to the same class, namely the class of restrictions on conditional cell probabilities, they do not belong to the same subclass. That is, the functions of the cell probabilities in hypothesis H_1 are row conditional cell probabilities, while the functions of the cell probabilities in hypothesis H_2 are column conditional cell probabilities. The GORICA evaluation of hypotheses containing restrictions that are associated with the same subclass will be elaborated in the next section.

Table 3.2: Three classes of restrictions for (at least) two-dimensional contingency tables, and the corresponding reparameterization such that they are in accordance with $H_m : S_m\eta = s_m, R_m\eta > r_m$, which are exemplified using a 2×4 contingency table

	Classes of Restrictions	Reparameterization into η's	Examples
Linear Restrictions	Restrictions of Type	$\eta = \pi = (\pi_{11}, ..., \pi_{IJ})^T$	$H_1 : \pi_{11} - \pi_{21} > \pi_{14} - \pi_{24}$
	$H_m : S_m\pi = s_m, R_m\pi > r_m$		H_2: $\pi_{21} > \pi_{22} > \pi_{23} > \pi_{24}$
	Restrictions on Marginal	$\eta_1 = (\pi_{+1}, ..., \pi_{+J})^T$	$H_1 : \pi_{+1} > \{\pi_{+2}, \pi_{+3}, \pi_{+4}\}$
	Cell Probabilities	$\eta_2 = (\pi_{1+}, ..., \pi_{I+})^T$	$H_2 : \pi_{1+} > \pi_{2+}$
Non-linear Restrict.	Restrictions on Conditional	$\eta_1 = (\frac{\pi_{i1}}{\pi_{+1}}, ..., \frac{\pi_{iJ}}{\pi_{+J}})^T$	$H_1 : \frac{\pi_{21}}{\pi_{+1}} > \{\frac{\pi_{22}}{\pi_{+2}}, \frac{\pi_{23}}{\pi_{+3}}, \frac{\pi_{24}}{\pi_{+4}}\}$
	Cell Probabilities	$\eta_2 = (\frac{\pi_{1j}}{\pi_{1+}}, ..., \frac{\pi_{Ij}}{\pi_{I+}})^T$	$H_2 : \frac{\pi_{12}}{\pi_{1+}} > \frac{\pi_{22}}{\pi_{2+}}$
	Restrictions on Local Odds Ratios	$\eta = (\frac{\pi_{11}\pi_{22}}{\pi_{12}\pi_{21}}, ..., \frac{\pi_{I-1,J-1}\pi_{I,J}}{\pi_{I-1,J}\pi_{I,J-1}})^T$	$H_1 : \frac{\pi_{11}\pi_{22}}{\pi_{12}\pi_{21}} = 1, \frac{\pi_{13}\pi_{24}}{\pi_{14}\pi_{23}} > 1$
			$H_2 : \frac{\pi_{12}\pi_{23}}{\pi_{13}\pi_{22}} > \frac{\pi_{13}\pi_{24}}{\pi_{14}\pi_{23}} > 1$

Note. η : The vector of cell probabilities for linear restrictions of type $H_m : S_m\eta = s_m, R_m\eta > r_m$ for $i = 1, ..., I$ and $j = 1, ..., J$ or the vector of local odds ratios for $i = 1, ..., I-1$ and $j = 1, ..., J-1$.
η_1, η_2 : The vectors of row and column marginal cell probabilities or row and column conditional cell probabilities for $i = 1, ..., I$ and $j = 1, ..., J$, representing the parameters of four different subclasses.
π_{i+}, π_{+j} : The cell probabilities summed over the ith row and the jth column, respectively. For example, $\pi_{1+} = \pi_{11} + \pi_{12} + \pi_{13} + \pi_{14}$ and $\pi_{+1} = \pi_{11} + \pi_{21}$ for a 2×4 contingency table.

Table 3.3: Two other classes of restrictions for (at least) three-dimensional contingency tables, and the corresponding reparameterization such that they are in accordance with $H_m : S_m\eta = s_m, R_m\eta > r_m$, which are exemplified using a $4 \times 4 \times 4$ contingency table

Classes of Restrictions	Reparameterization into η's	Examples
Non-linear Restrictions		
Restrictions on Marginal	$\eta_1 = (\frac{\pi_{+11}\pi_{+22}}{\pi_{+12}\pi_{+21}}, ..., \frac{\pi_{+J-1,V-1}\pi_{+JV}}{\pi_{+J-1,V}\pi_{+J,V-1}})^T$	$H_1 : \frac{\pi_{+11}\pi_{+22}}{\pi_{+12}\pi_{+21}} = \frac{\pi_{+22}\pi_{+33}}{\pi_{+23}\pi_{+32}} = \frac{\pi_{+33}\pi_{+44}}{\pi_{+34}\pi_{+43}}$
Odds Ratios	$\eta_2 = (\frac{\pi_{1+1}\pi_{2+2}}{\pi_{1+2}\pi_{2+1}}, ..., \frac{\pi_{I-1,+,V-1}\pi_{I+V}}{\pi_{I-1,+V}\pi_{I+,V-1}})^T$	$H_2 : \frac{\pi_{1+1}\pi_{2+2}}{\pi_{1+2}\pi_{2+1}} > \frac{\pi_{2+2}\pi_{3+3}}{\pi_{2+3}\pi_{3+2}} > \frac{\pi_{3+3}\pi_{4+4}}{\pi_{3+4}\pi_{4+3}}$
	$\eta_3 = (\frac{\pi_{11+}\pi_{22+}}{\pi_{12+}\pi_{21+}}, ..., \frac{\pi_{I-1,J-1,+}\pi_{IJ+}}{\pi_{I-1,J+}\pi_{I,J-1,+}})^T$	$H_3 : \frac{\pi_{11+}\pi_{22+}}{\pi_{12+}\pi_{21+}} > \{\frac{\pi_{22+}\pi_{33+}}{\pi_{23+}\pi_{32+}}, \frac{\pi_{33+}\pi_{44+}}{\pi_{34+}\pi_{43+}}\}$
Restrictions on Conditional	$\eta_1 = (\frac{\pi_{(1)11}\pi_{(1)22}}{\pi_{(1)12}\pi_{(1)21}}, ..., \frac{\pi_{(I-1)J-1,V-1}\pi_{(I-1)JV}}{\pi_{(I-1)J-1,V}\pi_{(I-1)J,V-1}})^T$	$H_1 : \frac{\pi_{(1)11}\pi_{(1)22}}{\pi_{(1)12}\pi_{(1)21}} = \frac{\pi_{(1)22}\pi_{(1)33}}{\pi_{(1)23}\pi_{(1)32}} = \frac{\pi_{(1)33}\pi_{(1)44}}{\pi_{(1)34}\pi_{(1)43}}$
Odds Ratios	$\eta_2 = (\frac{\pi_{1(1)1}\pi_{2(1)2}}{\pi_{1(1)2}\pi_{2(1)1}}, ..., \frac{\pi_{I-1(J-1)V-1}\pi_{I(J-1)V}}{\pi_{I-1(J-1)V}\pi_{I(J-1)V-1}})^T$	$H_2 : \frac{\pi_{1(1)1}\pi_{2(1)2}}{\pi_{1(1)2}\pi_{2(1)1}} > \frac{\pi_{2(1)2}\pi_{3(1)3}}{\pi_{2(1)3}\pi_{3(1)2}} > \frac{\pi_{3(1)3}\pi_{4(1)4}}{\pi_{3(1)4}\pi_{4(1)3}}$
	$\eta_3 = (\frac{\pi_{11(1)}\pi_{22(1)}}{\pi_{12(1)}\pi_{21(1)}}, ..., \frac{\pi_{I-1,J-1(V-1)}\pi_{IJ(V-1)}}{\pi_{I-1,J(V-1)}\pi_{I,J-1(V-1)}})^T$	$H_3 : \frac{\pi_{11(1)}\pi_{22(1)}}{\pi_{12(1)}\pi_{21(1)}} > \{\frac{\pi_{22(1)}\pi_{33(1)}}{\pi_{23(1)}\pi_{32(1)}}, \frac{\pi_{33(1)}\pi_{44(1)}}{\pi_{34(1)}\pi_{43(1)}}\}$

Note. η_1, η_2, η_3 : The vectors of marginal or conditional odds ratios summed over the first, second, and third variables for $i = 1, ..., I-1$, $j = 1, ..., J-1$, and $v = 1, ..., V-1$, respectively, representing the parameters of six different subclasses.

$\pi_{+jv}, \pi_{i+v}, \pi_{ij+}$: The cell probabilities summed over one variable. For example, $\pi_{+11} = \pi_{111} + \pi_{211} + \pi_{311} + \pi_{411}$; $\frac{\pi_{1(1)1}\pi_{2(1)2}}{\pi_{1(1)2}\pi_{2(1)1}}$ states that our focus is on the first level of the second variable, and therefore, this level is fixed.

3.3 The GORICA

The GORICA (Altınışık, Nederhof, et al., unpublished) is an AIC-based information criterion that can be used to evaluate (in)equality constrained hypotheses in the context of GLMs, GLMMs, and SEMs. The GORICA evaluates hypotheses based on their fit and complexity. The fit is quantified by the maximum of the log likelihood under the restriction(s) of hypothesis H_m based on the data at hand. The complexity is a penalty term for the number of distinct parameters used in the specification of hypothesis H_m, after taking into consideration its restrictions.

In this section, we first introduce the GORICA evaluation of hypotheses specifying linear and non-linear restrictions on cell probabilities in the context of contingency tables, which are of the type in equation (3.2). Second, we elaborate how to calculate the log likelihood and penalty parts of the GORICA, respectively. Third, a measure that can be used in conjunction with the GORICA to determine the extent of support in the data for each hypothesis under evaluation, GORICA weight, will be discussed.

3.3.1 The GORICA for the evaluation of hypotheses regarding cell probabilities

The GORICA selects the best hypothesis from a set of competing hypotheses, that is, the hypothesis that has the smallest distance to the true hypothesis relative to the others. The GORICA value of hypothesis H_m is:

$$\text{GORICA}_m = -2\, L(\tilde{\eta}_m|\hat{\eta}, \hat{\Sigma}_{\hat{\eta}}) + 2PT_m(\eta), \tag{3.4}$$

with $L(\tilde{\eta}_m|\hat{\eta}, \hat{\Sigma}_{\hat{\eta}})$ denotes the log likelihood part and $PT_m(\eta)$ denotes the penalty part for $m = 1, 2, ..., M$, where M is the number of hypotheses under evaluation. The hypothesis with the smallest GORICA value is selected as the best hypothesis out of the set of M hypotheses. The following two subsections elaborate on how to calculate the log likelihood and penalty parts of the GORICA.

3.3.2 Using a normal approximation of the log likelihood

Based on large sample theory (Fisher, 1922), the log likelihood of the data is approximated using a normal distribution:

$$L(\tilde{\eta}_m|\hat{\eta}, \hat{\Sigma}_{\hat{\eta}}) = -\frac{K}{2}\log(2\pi) - \frac{1}{2}\log|\hat{\Sigma}_{\hat{\eta}}| - \frac{1}{2}(\hat{\eta} - \tilde{\eta}_m)^T(\hat{\Sigma}_{\hat{\eta}})^{-1}(\hat{\eta} - \tilde{\eta}_m), \tag{3.5}$$

where $\hat{\eta} \in \mathbb{R}^{K\times 1}$ and $\hat{\Sigma}_{\hat{\eta}} \in \mathbb{R}^{K\times K}$ are the maximum likelihood estimates (MLEs) of the η parameters and their covariance matrix, respectively, with K representing the number of η parameters. The order-restricted MLEs $\tilde{\eta}_m \in \mathbb{R}^{K\times 1}$ represent the MLEs after the (in)equality constraints in H_m are imposed on the η parameters, that is, the MLEs that are in agreement with the constraints in H_m.

The GORICA utilizes nonparametric bootstrapping (Efron & Tibshirani, 1993, p.45) to obtain the MLEs of (functions of) cell probabilities, the $\hat{\eta}$'s, and their covariance matrix $\hat{\Sigma}_{\hat{\eta}}$. First of all, the MLEs of the cell probabilities, the $\hat{\pi}$'s, need to be obtained from the observed contingency tables using nonparametric bootstrapping. We provide the general estimation procedure below.

For simplicity of notation, the observed cell frequencies in a multidimensional contingency table are stacked together in a $D \times 1$ vector. For example, in case of the three-way contingency table introduced in the previous section, the vector of the observed cell frequencies is defined as $y = (y_{111}, ..., y_{ijv}, ..., y_{IJV}) = (y_1, ..., y_d, ..., y_D) \in \mathbb{R}^{D\times 1}$, where $N = \sum_{d=1}^{D} y_d$ is the sum of the cell frequencies. Then, nonparametric bootstrapping consists of the following steps:

1. Create y_d data points with the value d for $d = 1, 2, ..., D$. Note that the data contain numbers from 1 to D.

2. Draw B independent bootstrap samples of size N with replacement from the data such that each bootstrap sample represents one contingency table for which the entries are y_{bd}, cell frequency $d = 1, 2, ..., D$ in bootstrap sample $b = 1, 2, ..., B$.

3. Create a matrix $P \in \mathbb{R}^{B\times D}$

$$P = \begin{pmatrix} \hat{\pi}_{11} & \hat{\pi}_{12} & . & . & . & . & . & \hat{\pi}_{1D-1} & \hat{\pi}_{1D} \\ \hat{\pi}_{21} & \hat{\pi}_{22} & . & . & . & . & . & \hat{\pi}_{2D-1} & \hat{\pi}_{2D} \\ . & . & . & . & . & . & . & . & . \\ \hat{\pi}_{b1} & \hat{\pi}_{b2} & . & . & \hat{\pi}_{bd} & . & . & \hat{\pi}_{bD-1} & \hat{\pi}_{bD} \\ . & . & . & . & . & . & . & . & . \\ \hat{\pi}_{B1} & \hat{\pi}_{B2} & . & . & . & . & . & \hat{\pi}_{BD-1} & \hat{\pi}_{BD} \end{pmatrix},$$

   where $\hat{\pi}_{bd} = \frac{y_{bd}}{N}$ is the MLE of the dth cell probability in the bth bootstrap sample for $b = 1, 2, ..., B$ and $d = 1, 2, ..., D$.

4. Obtain the MLEs of the η's for each bootstrap sample using the corresponding rows of matrix P, that is, $\hat{\eta}_{bk} = g(\hat{\pi}_{b1}, ..., \hat{\pi}_{bD})$, reparameterized cell probability $k = 1, 2, ..., K$ in bootstrap sample $b = 1, 2, ..., B$. For example, consider the hypothesis $H_1 : \frac{\pi_{21}}{\pi_{+1}} > \{\frac{\pi_{22}}{\pi_{+2}}, \frac{\pi_{23}}{\pi_{+3}}, \frac{\pi_{24}}{\pi_{+4}}\}$ in equation (3.1) for which $D = 2 \times 4 = 8$ and $K = 4$, and $(\eta_1, \eta_2, \eta_3, \eta_4) = (\frac{\pi_{21}}{\pi_{+1}}, \frac{\pi_{22}}{\pi_{+2}}, \frac{\pi_{23}}{\pi_{+3}}, \frac{\pi_{24}}{\pi_{+4}})$. Then, since maximum likelihood estimates are invariant to transformations, the MLE of parameter η_1 in the bth bootstrap sample is $\hat{\eta}_{b1} = \frac{\hat{\pi}_{b21}}{\hat{\pi}_{b11}+\hat{\pi}_{b21}}$.

5. Estimate the η's and their covariance matrix:

$$\hat{\eta}_k = \sum_{b=1}^{B} \frac{\hat{\eta}_{bk}}{B} \text{ for } k = 1, 2, ..., K, \tag{3.6}$$

$$\hat{\Sigma}_{\hat{\eta}}(k,k') = \frac{1}{B-1}\sum_{b=1}^{B}(\hat{\eta}_{bk} - \hat{\eta}_k)(\hat{\eta}_{bk'} - \hat{\eta}_{k'}). \tag{3.7}$$

Once the MLEs of the η parameters and their covariance matrix are obtained, the values of $\tilde{\eta}$ for $m = 1, 2, ..., M$ are estimated by

$$\tilde{\eta}_m = \underset{\eta \in H_m}{\arg\min}\ (\hat{\eta} - \eta)^T(\hat{\Sigma}_{\hat{\eta}})^{-1}(\hat{\eta} - \eta). \tag{3.8}$$

The right hand side of equation (3.8) represents quadratic function which needs to be minimized accounting for the (in)equality constraints in hypothesis H_m. In our software, this is done using a quadratic programming algorithm the "solve.QP" subroutine of the "quadprog" R-package (Turlach, 2014, pp.2-4). In Appendix 3.A, the use of our software will be elaborated.

3.3.3 Computing the penalty part

This subsection elaborates on how to calculate the penalty part of the GORICA. For ease of calculation, we first rewrite hypotheses when they contain non-zero constants. Hypotheses of the form $H_m : S_m\eta = s_m, R_m\eta > r_m$, can be rewritten in the form $H_m^* : S_m\eta^* = 0, R_m\eta^* > 0$, where $\eta^* = \eta$ when $s_m = r_m = 0$ and $\eta^* = \eta - q$ when $s_m \neq 0$ and/or $r_m \neq 0$, with $\begin{bmatrix} S_m \\ R_m \end{bmatrix} q = \begin{bmatrix} s_m \\ r_m \end{bmatrix}$, where $q \in \mathbb{R}^{K\times 1}$ represents a vector on the boundary of the parameter space defined by the restrictions in H_m and (Kuiper & Hoijtink, 2013, p.7). Note that, since $\eta^* = \eta$ (i.e., no shift on the parameter space) or $\eta^* = \eta - q$ (i.e., linear shift on the parameter space) it holds that $\hat{\Sigma}_{\hat{\eta}} = \hat{\Sigma}_{\hat{\eta}^*}$.

The penalty part remains the same when we convert hypothesis H_m to hypothesis H_m^*. However, its calculation involves simulations of the η's for hypothesis H_m and the η^*'s for hypothesis H_m^*. We use the latter when calculating the penalty part, since the simulation of the η^*'s is easier than the simulation of the η's (Kuiper & Hoijtink, 2013). In our software, the penalty part is computed as follows:

1. Sample a vector $z \in \mathbb{R}^{K\times 1}$ from a normal distribution with mean vector $\mu_{\eta^*} = 0$ and covariance matrix $\Sigma_{\eta^*} = c\hat{\Sigma}_{\hat{\eta}^*} \in \mathbb{R}^{K\times K}$, where c is a positive constant.

   The choice of the mean vector μ_{η^*} for hypothesis H_m^* does not influence the penalty part for hypotheses of the type $H_m : S_m\eta = s_m, R_m\eta > r_m$ (closed convex cone; (Silvapulle & Sen, 2004, p.82) and relocated closed convex cone; (Kuiper et al., 2012, p.2460). However, there are exceptions (see Kuiper et al., 2012, p.2460), for which μ_{η^*} has to be zero, which is therefore used as a convenient choice when sampling the vector z.

   Calculation of the penalty part with the mean vector $\mu_{\eta^*} = 0$ and covariance matrix $\Sigma_{\eta^*} = c\hat{\Sigma}_{\hat{\eta}^*}$ is independent of the choice of the scale factor c as soon as hypotheses

involve (relocated) closed convex cone (Kuiper & Hoijtink, 2013). But calculation of the penalty part for hypotheses containing non-linear restrictions on cell probabilities is often dependent on the choice of the scale factor c. This issue is solved if hypotheses containing non-linear restrictions on cell probabilities are specified by means of functions of cell probabilities, the η's, a topic which will be revisited in the next section.

2. Calculate the vector of the order-restricted maximum likelihood estimates

$$\tilde{z}_m = \underset{\eta^* \in H_m^*}{\arg\min}\ (z-\eta^*)^T(\hat{\Sigma}_{\hat{\eta}^*})^{-1}(z-\eta^*), \tag{3.9}$$

such that $\tilde{z}_m \in \mathbb{R}^{K\times 1}$ is in agreement with H_m^*.

3. Obtain the number of levels in $\tilde{z}_m$, that is, $K - A_m$ where A_m is the number of active constraints in hypothesis H_m^*.

The calculation of the number of active constraints is illustrated by means of two examples. Suppose $K = 3$ and $\hat{\Sigma}_{\hat{\eta}} = I_3$, where I_3 denotes a 3×3 identity matrix and $H_m : \eta_1 > \eta_2, \eta_3 > 0.5$. To calculate the order-restricted MLEs, we determine $\tilde{z}_m$ which is the argument that minimizes the objective function under the restrictions $\eta_1^* > \eta_2^*$ and $\eta_3^* > 0$. Suppose we sampled $z_1 > z_2$ and $z_3 < 0$. Then, z_1 and z_2 are in accordance with $\eta_1^* > \eta_2^*$. Therefore, the constraint $z_1 > z_2$ is inactive, namely, it is not on the boundary of $\eta_1^* > \eta_2^*$, and consequently, $\tilde{z}_{m1} = z_1$ and $\tilde{z}_{m2} = z_2$. However, z_3 is not in accordance with $\eta_3^* > 0$. To let $\tilde{z}_{m3}$ be in accordance with the constraint, $\tilde{z}_{m3}$ should be chosen to be on the boundary of the parameter space such that it minimizes the objective function in equation (3.9). In this case, $\tilde{z}_{m3} = 0$ and the corresponding constraint is called active. Then, the number of levels in $\tilde{z}_m$ is $K - A_m = 3 - 1 = 2$. Now, consider the same hypothesis $H_m : \eta_1 > \eta_2, \eta_3 > 0$ for different sampled values: $z_1 < z_2$ and $z_3 > 0$. In this case, the first restriction is an active restriction and the second restriction is not an active restriction. The resulting order-restricted MLEs[3] are $\tilde{z}_{m1} = \tilde{z}_{m2} = \frac{z_1+z_2}{2}$ and $\tilde{z}_{m3} = z_3$, and therefore, the number of levels in $\tilde{z}_m$ is $K - A_m = 3 - 1 = 2$.

Repeat steps 1-3 above for $t = 1, 2, ..., T$ times.[4]

4. Calculate the level probabilities, the $w_l(.)$'s, that is, the proportion of times the number of levels in $\tilde{z}_m$ is equal to l for $l = 1, 2, ..., K$.

5. Calculate the penalty part of the GORICA which is defined as:

$$PT_m(\eta) = \sum_{l=1}^{K} w_l(K, c\hat{\Sigma}_{\hat{\eta}^*}, H_m^*)\ l = \sum_{l=1}^{K} w_l(K, c\hat{\Sigma}_{\hat{\eta}}, H_m)\ l, \tag{3.10}$$

[3] In case of other matrices for which the off-diagonals are not equal to zero, the order-restricted MLEs will be different than the ones in the text, because of the correlations between the sampled values. Consequently, the resulting penalty part will be different.

[4] Note that, we used $T = 100,000$ in this paper.

where $w_l(.)$ represents the lth level probability for $l = 1, 2, ..., K$. The penalty part denotes the number of η parameters used in the specification of the hypotheses under evalution minus the expected number of active constraints in H_m, that is, $K - E(A_m)$. In case of simple order restrictions the penalty part equals the expected number of distinct parameters used to specify the hypothesis under evaluation.

Once the GORICA values defined in equation (3.4) are calculated, they can be transformed into GORICA weights, which quantify the support in the data for each hypothesis under evaluation compared to the others in the set. This is elaborated in the next subsection.

3.3.4 GORICA weights and unconstrained hypothesis

After computing the GORICA values for each hypothesis separately, these values can be used to calculate the GORICA weights. Comparable to the Akaike weights (Burnham & Anderson, 2002, p.75), the GORICA weight of hypothesis H_m are defined as:

$$\psi_m = \frac{\exp\{-\frac{1}{2}\,\mathrm{GORICA}_m\}}{\sum_{m'=1}^{M}\exp\{-\frac{1}{2}\,\mathrm{GORICA}_{m'}\}}, \tag{3.11}$$

where ψ_m can take a value between the numbers 0 and 1 and the ψ_m's sum to one. The GORICA weight is utilized to quantify the support in the data set at hand for hypothesis H_m compared to the set of M hypotheses. We advise the use of the unconstrained hypothesis H_u, sometimes called the traditional alternative hypothesis, as the last hypothesis in the set of M hypotheses under evaluation. This hypothesis can be included in the set of hypotheses to avoid that the best of a set of informative hypotheses is nevertheless a weak hypothesis.

Consider hypotheses H_1, H_2 and the unconstrained hypothesis H_u with the GORICA weights 0.84, 0.04, and 0.12, respectively. Then, hypothesis H_2 is $0.04/0.12 \approx 0.33$ times more supported by the data compared to the unconstrained hypothesis H_u, that is, H_u is $0.12/0.04 = 3$ times more supported by the data compared to H_2. Therefore, hypothesis H_2 is a weak hypothesis, that is, a hypothesis which is not supported by the data. Hypothesis H_1 is better than hypothesis H_2, because it is $0.84/0.04 = 21$ times more supported by the data compared to H_2 and it is not a weak hypothesis, since H_1 is $0.84/0.12 = 7$ times more supported by the data compared to H_u.

3.4 Problems and solutions for the GORICA in contingency tables

The general framework presented in the previous section to evaluate (in)equality constrained hypotheses in the context of contingency tables is not always applicable. Three main problems may be encountered in evaluating these hypotheses using the GORICA. First, when (in)equality constrained hypotheses contain non-linear restrictions on cell probabilities, the calculation of the penalty part (see Step 1) is often dependent on the choice of a scale factor. Second, maximum likelihood estimates for the η parameters and their covariance matrix

need to be obtained in calculating the GORICA. This covariance matrix has to be a positive definite matrix to evaluate hypotheses that are (relocated) convex cones. However, this is not the case for $\eta = \pi$. Because cell probabilities are linearly dependent on each other since they sum to one. Third, empty cells in contingency tables complicate parameter estimation when evaluating (in)equality constrained hypotheses containing linear or non-linear restrictions on cell probabilities. We elaborate on these three problems and how these problems are solved in our software.

3.4.1 Problem 1: Evaluating non-linear restrictions

The GORICA (Altınışık, Nederhof, et al., unpublished) cannot be used to evaluate hypotheses specifying non-linear restrictions on cell probabilities, because these hypotheses are not of the type (relocated) closed convex cone: $H_m : S_m\eta = s_m, R_m\eta > r_m$. When (in)equality constrained hypotheses are of the type (relocated) closed convex cone, the calculation of the penalty part does not depend on the choice of c, see equation (3.10). In contrast, for (in)equality constrained hypotheses containing non-linear restrictions on cell probabilities, the calculation of the penaty part often depends on the choice of c, when these hypotheses are not rewritten in terms of the η parameters. To illustrate, consider the gender by earned degrees example for which the data set is given in Table 3.1. For this example, the hypothesis $H_1 : \frac{\pi_{21}}{\pi_{+1}} > \{\frac{\pi_{22}}{\pi_{+2}}, \frac{\pi_{23}}{\pi_{+3}}, \frac{\pi_{24}}{\pi_{+4}}\}$ in equation (3.1) contains non-linear restrictions on the cell probabilities and the hypothesis $H_2 : \pi_{11} = \pi_{12}, \pi_{23} > \pi_{24}$ in equation (3.3) consists of linear restrictions on the cell probabilities. The penalty parts for hypotheses H_1 and H_2, and the unconstrained hypothesis H_u are given in Table 3.4.1.

As can be seen, the penalty part for hypothesis H_2, as being representative of hypotheses containing only linear restrictions on cell probabilities is invariant with c, that is, $PT_2(\eta)$ = 5.499 for $c = 1$ and $c = 100000$. The penalty part for hypothesis H_1, as being representative of hypotheses containing non-linear restrictions on cell probabilities, does depend on c, that is, $PT_1(\eta)$ = 6.275 for $c = 1$, while $PT_1(\eta)$ = 5.366 for $c = 100000$. The penalty part for the unconstrained hypothesis $H_u : \pi_{11}, \pi_{12}, ..., \pi_{24}$ represents the number of distinct cell probabilities, and therefore, $PT_u(\eta) = D - 1 = 7.000$ for $c = 1$ and $c = 100000$, since the last cell probability π_{24} is one minus the sum of the other cell probabilities.

The penalty part is invariant for a positive constant c for hypotheses containing linear restrictions on cell probabilities, that is, $H_m : S_m\eta = s_m$, $R_m\eta > r_m$, with η representing the cell probabilities for contingency tables, i.e., $\eta = \pi$. This suggests a solution to the problem of the penalty part not being invariant with c for hypotheses containing non-linear restrictions on cell probabilities, that is, reparameterizing the cell probabilities such that the restrictions become linear in the new parameters (see Tables 3.2 & 3.3).

The penalty parts for hypothesis H_1 and the unconstrained hypothesis H_u after reparameterizing the cell probabilities are given in Table 3.4.2. As is to be expected for hypotheses of the form in equation (3.2), the penalty part for hypothesis

$$H_1 : \eta_1 > \{\eta_2, \eta_3, \eta_4\} \tag{3.12}$$

Table 3.4.1: Calculation of the penalty part for hypotheses in terms of π's based on the covariance matrix $c\hat{\Sigma}_{\hat{\pi}}$

	Penalty term	
Hypothesis	$c = 1$	$c = 100000$
H_1: $\frac{\pi_{21}}{\pi_{+1}} > \{\frac{\pi_{22}}{\pi_{+2}}, \frac{\pi_{23}}{\pi_{+3}}, \frac{\pi_{24}}{\pi_{+4}}\}$	6.275	5.366
H_2: $\pi_{11} = \pi_{12}, \pi_{23} > \pi_{24}$	5.499	5.499
H_u: $\pi_{11}, \pi_{12}, ..., \pi_{24}$	7.000	7.000

Note. The penalty part for H_1 is calculated using the non-linear optimization algorithm the "auglag" subroutine of "alabama" R-package (Varadhan, 2015, pp.2-5).

Table 3.4.2: Calculation of the penalty part based on the covariance matrix $c\hat{\Sigma}_{\hat{\eta}}$, where the restrictions are linear in the new parameter η

	Penalty term	
Hypothesis	$c = 1$	$c = 100000$
H_1: $\eta_1 > \{\eta_1, \eta_2, \eta_3\}$	2.582	2.582
H_u: $\eta_1, \eta_2, \eta_3, \eta_4$	4.000	4.000

Note. $\eta = (\eta_1, \eta_2, \eta_3, \eta_4)^T = (\frac{\pi_{21}}{\pi_{+1}}, \frac{\pi_{22}}{\pi_{+2}}, \frac{\pi_{23}}{\pi_{+3}}, \frac{\pi_{24}}{\pi_{+4}})^T$.

with $\eta = (\eta_1, \eta_2, \eta_3, \eta_4)^T = (\frac{\pi_{21}}{\pi_{+1}}, \frac{\pi_{22}}{\pi_{+2}}, \frac{\pi_{23}}{\pi_{+3}}, \frac{\pi_{24}}{\pi_{+4}})^T$, is independent of c, that is, $PT_1(\eta) =$ 2.582 for $c = 1$ and $c = 100000$. The penalty part for the unconstained hypothesis H_u is now based on the number of distinct η parameters, that is, $H_u : \eta_1, \eta_2, \eta_3, \eta_4$, and therefore, $PT_u(\eta) = 4.000$ for the four η's in H_u.

3.4.2 Problem 2: Linear dependency

In this subsection, we elaborate on two cases of linear dependency. The first case of linear dependency occurs between the η parameters when $\eta = \pi$, because cell probabilities sum to one. As a result, the covariance matrix of the MLEs of the η parameters is not a positive definite matrix, and therefore, the log likelihood and penalty parts of the GORICA cannot be obtained. In this case, the vector of the cell probabilities should be defined after discarding the last[5] cell probability as $\eta_{-1} = (\pi_1, \pi_2, ..., \pi_K) \in \mathbb{R}^{K \times 1}$, where the subscript "-1" in η_{-1} represents removing the last cell probability from the π's and $K = D - 1$ is the number of the cell probabilities after discarding the last cell probability. The vector η_{-1} represents all the information in the π's, since all the π's sum to one and the last cell probability can be replaced by one minus the sum of the other cell probabilities, $\pi_D = 1 - \sum_{d=1}^{D-1} \pi_d$.

[5]Because of possible empty cells, the last non-zero estimated cell probability is discarded in the software.

Two adjustments are required in evaluating (in)equality constrained hypotheses containing linear restrictions on cell probabilities. First, the covariance matrix of the cell probabilities in P is not a positive definite matrix as needed in evaluating hypotheses using the GORICA, since they are linearly dependent on each other. Therefore, the matrix P is made to correspond to the parameter vector η_{-1}. This is done by discarding the last column of the matrix P, and consequently, using this submatrix to evaluate the hypotheses. Second, because the last column of matrix P is discarded, the last cell probability should be taken into account when specifying restriction matrices and constants in hypotheses. For example, consider a 2×2 contingency table to evaluate hypothesis $H_1 : \eta_1 > \eta_2 > \eta_3 > \eta_4$, where $\eta = (\eta_1, \eta_2, \eta_3, \eta_4)^T = (\pi_{11}, \pi_{12}, \pi_{21}, \pi_{22})^T$ denotes the four cell probabilities in the contingency table. Because the fourth column in P is discarded, this hypothesis should be reformulated as $H_1 : \eta_1 > \eta_2 > \eta_3 > 1 - \eta_1 - \eta_2 - \eta_3$. The restriction matrix R_1 and constants r_1 for this hypothesis are:

$$R_1 = \begin{pmatrix} 1 & -1 & 0 \\ 0 & 1 & -1 \\ 1 & 1 & 2 \end{pmatrix}, r_1 = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}.$$

Note that, after rewriting the last cell probability as one minus the sum of the others, we still have linear restrictions on cell probabilities. Therefore, the GORICA evaluation of hypotheses as described in the previous section is still applicable.

The second case of linear dependency may occur between the η's when $\eta = g(\pi)$ for hypotheses containing restrictions on marginal cell probabilities and conditional cell probabilities. For example, consider a 2×2 contingency table with $\eta = (\eta_1, \eta_2)^T = (\frac{\pi_{11}}{\pi_{11}+\pi_{21}}, \frac{\pi_{21}}{\pi_{11}+\pi_{21}})^T$, when evaluating hypothesis $H_1 : \eta_1 > \eta_2$. Because $\eta_1 + \eta_2 = 1$, there is a linear dependency between these η's. When the linear dependency between the η's occurs for hypotheses containing restrictions on marginal or conditional cell probabilities, our software cannot calculate the log likelihood and penalty parts of the GORICA, since the covariance matrix of the MLEs of η's is not a positive definite matrix. However, rewriting renders $H_1 : \eta_1 > 1 - \eta_1$, that is, $H_1 : \eta_1 > 0.5$ which can easily be evaluated with our software.

In the next subsection, we will discuss another situation where the covariance matrix is not positive definite which may occur in the presence of empty cells in contingency tables.

3.4.3 Problem 3: Hypothesis evaluation in the presence of empty cells

Empty cells in the context of contingency tables are the cells for which no observation is obtained. This means that, for the vector of the frequencies observed in a contingency table, $y = (y_1, y_2, ..., y_D) \in \mathbb{R}^{D \times 1}$, one or more of the y_d's are zero. Then, reparametrizing cell probabilities (π's) into the η's may cause some of the η's to be estimated as (1) zero, (2) one, and (3) infinite. These three cases are dealt with by rewriting the hypotheses under evaluation such that empty cells do not cause problems for evaluating them, which will be elaborated in the next subsections.

Rewriting hypotheses in case of $\hat{\eta} = 0$

Empty cells in contingency tables may cause some of the η's to be zero, which can occur in all classes mentioned in Tables 3.2 and 3.3. This results in η's without variation, which leads to a covariance matrix that is not positive definite, as needed in evaluating hypotheses using the GORICA. To illustrate, consider the parameter vector $\eta = (\eta_1, \eta_2, \eta_3)^T$, with the second parameter estimated as zero because of empty cell(s), that is, $\hat{\eta}_2 = 0$. Then, the covariance matrix of $\hat{\eta} = (\hat{\eta}_1, 0, \hat{\eta}_3)^T$ is:

$$\hat{\Sigma}_{\hat{\eta}} = \begin{bmatrix} \mathrm{Var}(\hat{\eta}_1) & 0 & \mathrm{Cov}(\hat{\eta}_1, \hat{\eta}_3) \\ 0 & 0 & 0 \\ \mathrm{Cov}(\hat{\eta}_1, \hat{\eta}_3) & 0 & \mathrm{Var}(\hat{\eta}_3) \end{bmatrix}.$$

This covariance matrix is not a positive definite matrix, because it contains only zeros in the second row and column.

The solution is to rewrite the hypothesis of interest by defining $\tilde{\eta}_m = \{\tilde{\eta}_E, \tilde{\eta}_F\}$, where $\tilde{\eta}_E$ and $\tilde{\eta}_F$ represent the order-restricted MLEs corresponding to the η's that are estimated as zero and not zero, respectively. The solution requires adjusted restriction matrices S_m^{adj} and/or R_m^{adj} and may require adjusted constants s_m^{adj} and/or r_m^{adj}, as will be described in the following procedure. To keep the exposition simple, below we provide this procedure in terms of hypotheses containing only inequality restrictions, which can be extended for hypotheses containing equality restrictions in the same manner. Subsequently, we will give three examples for the three different scenarios discussed in the procedure.

1. Discard the rows and thus columns from $\hat{\Sigma}_{\hat{\eta}}$ for which all elements are zero. Then, discard the corresponding η's which leads to η^{adj} and the corresponding columns from the restriction matrix R_m, which leads to R_m^{adj}.

2. Check the rows of R_m^{adj} to determine whether r_m has to be adjusted as well.

   a) If there are no row(s) with only zeros, r_m does not change and $\tilde{\eta}_E = \hat{\eta}_E = 0$.

   b) If there are row(s) with only zeros and the corresponding constants are ...

      i. ... non-positive, r_m does not change, because $\tilde{\eta}_E = \hat{\eta}_E = 0$ is in agreement with the restrictions in H_m.

      ii. ... positive, r_m has to be adjusted, since $\tilde{\eta}_E = \hat{\eta}_E = 0$ is not in agreement with the restrictions in H_m. Therefore, we first need to calculate the $\tilde{\eta}_E$ that are in accordance with the restrictions in H_m, and then adjust r_m accordingly. This is done as follows:

         - We calculate the values $q = \{q_E, q_F\}$ for which $R_m q = r_m$.
           The q_E and q_F represent values on the boundary of the parameter space defined by the restrictions in H_m. These q values are obtained using a least distance programming algorithm the "ldei" subroutine of the "limSolve" package (Soetart, Van den Meersche, & Van Oevelen, 2014,

pp.11-13) in R. The subroutine ldei finds the "least distance" in the sense that the sum of squared q's is minimal:

$$\begin{aligned} &\min(\sum_{k=1}^{K} q_k^2) \text{ subject to,} \\ &R_m q = r_m, \\ &q_E \geq 0. \end{aligned} \tag{3.13}$$

- Based on the resulting q values, we can calculate the adjusted r_m values by $r_m^{adj} = R_m^{adj} q_F$.

Note that this solution does not imply that $\tilde{\eta}_E = 0$, instead it implies that $\tilde{\eta}_E = q_E$.

This procedure leads to rewritten hypothesis $H_m^{adj} : R_m^{adj}\eta^{adj} > r_m^{adj}$ and $\tilde{\eta}_F$ will be calculated based on this rewritten hypothesis.

In Step 2(a), r_m does not need to be adjusted, while R_m and η have to be adjusted. For example, for hypothesis $H_m : \eta_1 > \eta_2, \eta_1 > \eta_3, 2\eta_1 + \eta_2 + \eta_3 > 1$ with $\hat{\eta}_2 = 0$, after discarding the second parameter η_2 and thus the second column in R_m, r_m does not change:

$$H_m : R_m\eta > r_m \qquad\qquad H_m^{adj} : R_m^{adj}\eta^{adj} > r_m^{adj}$$

$$\begin{pmatrix} 1 & -1 & 0 \\ 1 & 0 & -1 \\ 2 & 1 & 1 \end{pmatrix} \begin{pmatrix} \eta_1 \\ \eta_2 \\ \eta_3 \end{pmatrix} > \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 & 0 \\ 1 & -1 \\ 2 & 1 \end{pmatrix} \begin{pmatrix} \eta_1 \\ \eta_3 \end{pmatrix} > \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}.$$

In Step 2(b)i, similarly, it is necessary to adjust R_m and η but not r_m. For example, for hypothesis $H_m : \eta_1 > \{\eta_2, \eta_3, \eta_4\}, \eta_2 < 0.3$ with $\hat{\eta}_2 = 0$, there is a row with only zeros, after discarding the second column in R_m:

$$H_m : R_m\eta > r_m \qquad\qquad H_m^{adj} : R_m^{adj}\eta^{adj} > r_m^{adj}$$

$$\begin{pmatrix} 1 & -1 & 0 & 0 \\ 1 & 0 & -1 & 0 \\ 1 & 0 & 0 & -1 \\ 0 & -1 & 0 & 0 \end{pmatrix} \begin{pmatrix} \eta_1 \\ \eta_2 \\ \eta_3 \\ \eta_4 \end{pmatrix} > \begin{pmatrix} 0 \\ 0 \\ 0 \\ -0.3 \end{pmatrix}, \begin{pmatrix} 1 & 0 & 0 \\ 1 & -1 & 0 \\ 1 & 0 & -1 \\ 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} \eta_1 \\ \eta_3 \\ \eta_4 \end{pmatrix} > \begin{pmatrix} 0 \\ 0 \\ 0 \\ -0.3 \end{pmatrix}.$$

Because $\tilde{\eta}_2 = \hat{\eta}_2 = 0$ is in accordance with the restrictions in H_m, r_m does not change. Note that the last row in R_m^{adj} above does not impose any restriction on the η parameters and it is in agreement with the last element in r_m (i.e., $0 \times \eta_1 + 0 \times \eta_3 + 0 \times \eta_4 = 0 > 0.3$).

In Step 2(b)ii, not only R_m and η but also r_m needs to be adjusted. For example, for hypothesis $H_m : \eta_1 > \{\eta_2, \eta_3, \eta_4\}, \eta_2 > 0.3$ with $\hat{\eta}_2 = 0$, there is a row with only zeros, after discarding the second column from R_m:

$$H_m : R_m\eta > r_m \qquad\qquad H_m^{adj} : R_m^{adj}\eta^{adj} > r_m^{adj}$$

$$\begin{pmatrix} 1 & -1 & 0 & 0 \\ 1 & 0 & -1 & 0 \\ 1 & 0 & 0 & -1 \\ 0 & 1 & 0 & 0 \end{pmatrix} \begin{pmatrix} \eta_1 \\ \eta_2 \\ \eta_3 \\ \eta_4 \end{pmatrix} > \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0.3 \end{pmatrix}, \begin{pmatrix} 1 & 0 & 0 \\ 1 & -1 & 0 \\ 1 & 0 & -1 \\ 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} \eta_1 \\ \eta_3 \\ \eta_4 \end{pmatrix} > \begin{pmatrix} 0.3 \\ 0 \\ 0 \\ 0 \end{pmatrix}.$$

In this case, $\hat{\eta}_2 = 0$ is not in accordance with the restrictions in H_m. This conflict is denoted by the last line in R_m, where $0 \times \eta_1 + 0 \times \eta_3 + 0 \times \eta_4 = 0$ is not bigger than 0.3. Using equation (3.13), we obtained $\tilde{\eta}_2 = q_E = 0.3$ and $q_F = (0.3, 0.3, 0.3)$, which leads to:

$$r_m^{adj} = R_m^{adj} q_F = \begin{pmatrix} 1 & 0 & 0 \\ 1 & -1 & 0 \\ 1 & 0 & -1 \\ 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} 0.3 \\ 0.3 \\ 0.3 \end{pmatrix} = \begin{pmatrix} 0.3 \\ 0 \\ 0 \\ 0 \end{pmatrix}.$$

Note that the procedure Step 2(b)ii gives one possible solution for the $\tilde{\eta}_E$'s. However, sometimes there might be multiple solutions for the order-restricted MLEs, when two or more η's are estimated as zero. We highlight three such cases. For the first situation, the actual solution is irrelevant for the computation of the GORICA. For example, for hypothesis $H_m : \eta_1 + \eta_2 + \eta_3 = 0.4, \eta_1 + \eta_2 > 0.3$ with $\hat{\eta}_1 = \hat{\eta}_2 = 0$, equation (3.13) gives $\tilde{\eta}_E = (\tilde{\eta}_{m1}, \tilde{\eta}_{m2}) = (0.15, 0.15)$ and based on this solution we calculate $\tilde{\eta}_F = \tilde{\eta}_{m3} = 0.1$. There is actually an infinite number of solutions for the $\tilde{\eta}_{m1}$ and $\tilde{\eta}_{m2}$. Stated otherwise, both the $\tilde{\eta}_{m1}$ and $\tilde{\eta}_{m2}$ can take any values ranging from 0 to 0.3 such that their sum is 0.3. All these solutions, irrelevant of the values, give $\tilde{\eta}_F = \tilde{\eta}_{m3} = 0.1$. Therefore, our solution for the $\tilde{\eta}_E$'s is optimal, namely the sum of squared $\tilde{\eta}_E$'s is minimized. For the second situation, the choice of the values for $\tilde{\eta}_E$ may influence the results, but the optimal solution is still obtained. For example, for hypothesis $H_m : \eta_1 + \eta_2 + \eta_3 = 0.4, \eta_1 - \eta_2 > 0.3$, there are multiple solutions for the $\tilde{\eta}_{m1}$ and $\tilde{\eta}_{m2}$, and these solutions influence the value of the $\tilde{\eta}_{m3}$, and consequently, the results. For example, one possible solution could be $\tilde{\eta}_E = (\tilde{\eta}_{m1}, \tilde{\eta}_{m2}) = (0.35, 0.05)$, and consequently, $\tilde{\eta}_F = \tilde{\eta}_{m3} = 0$ and another possible solution could be $\tilde{\eta}_E = (\tilde{\eta}_{m1}, \tilde{\eta}_{m2}) = (0.3, 0)$, and consequently, $\tilde{\eta}_F = \tilde{\eta}_{m3} = 0.1$. Note that the latter solution is given by equation (3.13) and it is optimal, that is, it gives the minimal sum of squared $\tilde{\eta}_E$'s. For the third situation, equation (3.13) does not give the optimal solution, but it gives the sub-optimal solution, namely the sum of squared $\tilde{\eta}_E$'s is not minimized, but the sum of all squared $\tilde{\eta}$'s is minimized. For example, for hypothesis $H_m : \eta_1 + \eta_2 + \eta_3 = 0.6, \eta_1 - \eta_2 > 0.2$, the optimal solution is $\tilde{\eta}_E = (\tilde{\eta}_{m1}, \tilde{\eta}_{m2}) = (0.2, 0)$, and thus, $\tilde{\eta}_F = \tilde{\eta}_{m3} = 0.4$ for which the sum of squared the $\tilde{\eta}_E$'s is minimal (i.e., $0.2^2 + 0^2 = 0.04$ is the minumum value among all the other solutions). However, equation (3.13) does not give this solution, instead it gives the solution $\tilde{\eta}_E = (\tilde{\eta}_{m1}, \tilde{\eta}_{m2}) = (0.3, 0.1)$, and therefore, $\tilde{\eta}_F = \tilde{\eta}_{m3} = 0.2$, for which the sum of all squared η's are minimized. Note, however, that this last situation is rarely encountered even if these three conditions hold: (1) two or more η's are estimated as zero, (2) the hypothesis under evaluation contains a restriction on only these η's, and (3) $\tilde{\eta}_E = 0$ is not in agreement with this restriction.

Rewriting hypotheses in case of $\hat{\eta} = 1$ and $\hat{\eta} = \infty$

Sometimes some of the η parameters can be estimated as one or cannot be estimated because of empty cells. Our software gives a warning if any of these situations occur. In these cases, users have to rewrite the hypotheses. In this section, we will give three examples. Note that the situations can be solved analogously, however, the procedure is not fully automatic and needs intervention by users.

For the first situation, some of the η parameters are estimated as one because of empty cells. This happens for the class of restrictions on conditional cell probabilities in Table 3.2. Consider hypothesis $H_m : \eta_1 > \eta_2$ with $\eta = (\eta_1, \eta_2)^T = (\frac{\pi_{11}}{\pi_{11}+\pi_{21}}, \frac{\pi_{12}}{\pi_{12}+\pi_{22}})$ and $\hat{\pi}_{22} = 0$ for a 2×2 contingency table. This results in $\hat{\eta}_2 = \frac{\hat{\pi}_{12}}{\hat{\pi}_{12}+\hat{\pi}_{22}} = 1$ which has no variation. Because of the lack of variation, we cannot calculate the log likelihood and penalty parts when inspecting this hypothesis. Thus, the procedure in our software stops and gives a warning message stating that users should rewrite themselves the hypotheses under evaluation. For example, since $\hat{\eta}_2 = \frac{\hat{\pi}_{12}}{\hat{\pi}_{12}+\hat{\pi}_{22}} = 1$ with $\hat{\pi}_{22} = 0$, $\eta_1 > \eta_2$ can easily be rewritten as $\eta_1 > 1$, which can be evaluated by our software.

For the second situation, some of the η parameters cannot be estimated because of empty cells. This occurs for the classes of restrictions on conditional cell probabilities in Table 3.2 and the odds ratios in Tables 3.2 and 3.3. For example, it is not possible to estimate the odds ratio $\eta_1 = \frac{\pi_{11}\pi_{22}}{\pi_{12}\pi_{21}}$, if the cell probability corresponding to an empty cell appears in the denominator, namely when $\hat{\pi}_{12} = 0$ and/or $\hat{\pi}_{21} = 0$. When inspecting $\eta_1 > 1$, for example, we cannot calculate the log likelihood and penalty parts. In this case, our software will give a warning and a suggestion to rewrite the hypotheses. For example, $H_m : \eta_1 > 1$ can be inspected equivalently by means of $\eta_1^\diamond = \pi_{11}\pi_{22} - \pi_{12}\pi_{21}$ and $\eta_1^\diamond > 0$. Note that $\hat{\eta}_1^\diamond = \hat{\pi}_{11}\hat{\pi}_{22}$ when $\hat{\pi}_{12} = 0$ and/or $\hat{\pi}_{21} = 0$, which is still estimable and has variation. The parameter $\eta_1^\diamond$ is not contained in Table 3.2, but is another reparameterization of the same hypothesis in this class. For example, consider hypothesis $H_m : \eta_1 > \eta_2 > \eta_3$, which is formulated in terms of three local odds ratios with $\eta = (\eta_1, \eta_2, \eta_3)^T = (\frac{\pi_1\pi_6}{\pi_2\pi_5}, \frac{\pi_2\pi_7}{\pi_3\pi_6}, \frac{\pi_3\pi_8}{\pi_4\pi_7})^T$ and $\hat{\pi}_2 = \hat{\pi}_7 = 0$ because of two empty cells. In this case, it is not possible to estimate the parameters $\eta_1 = \frac{\pi_1\pi_6}{\pi_2\pi_5}$ and $\eta_3 = \frac{\pi_3\pi_8}{\pi_4\pi_7}$ and the parameter $\eta_2 = \frac{\pi_2\pi_7}{\pi_3\pi_6}$ is estimated as zero. To be able to evaluate this hypothesis, it should be rewritten as $H_m : \eta_1^\diamond > \eta_2^\diamond, \eta_3^\diamond > \eta_4^\diamond$ with $\eta^\diamond = (\eta_1^\diamond, \eta_2^\diamond, \eta_3^\diamond, \eta_4^\diamond)^T = (\pi_1\pi_3\pi_6^2, \pi_2^2\pi_5\pi_7, \pi_2\pi_4\pi_7^2, \pi_3^2\pi_6\pi_8)^T$. Note that we have one parameter more, but we use the same number of restrictions, since $\eta_1^\diamond > \eta_2^\diamond$ and $\eta_3^\diamond > \eta_4^\diamond$ correspond to the restrictions $\eta_1 > \eta_2$ and $\eta_2 > \eta_3$ in hypothesis H_m, respectively.

3.5 The GORICA illustrated

In this section, we apply the GORICA to two examples. The first example is used to illustrate the evaluation of (in)equality constrained hypotheses containing linear restrictions on cell probabilities, in case of two high-dimensional contingency tables with empty cells. The second example is used to exemplify the evaluation of (in)equality constrained hypotheses containing non-linear restrictions on cell probabilities.

3.5.1 The eye-tracking example

In this example, we illustrate the use of the GORICA for restrictions on marginal cell probabilities with high-dimensional contingency tables containing empty cells. Examining marginal cell probabilities, that is, the cell probabilities marginalized (summed) over one or more variable(s) in a contingency table results in linear restrictions. Consider a study carried out to determine the relation between the eye movements of two participants (i.e., person A and person B). The eye movements of person A and person B were tracked and recorded at a series of 9051 time frames each representing 33.3 msec. The eye-movement data are assigned to the categorical variable Eye-Tracking Screen, defined as $\text{ETS} \in \{1 = \text{Nose}, 2 = \text{Mouth}, 3 = \text{Right Eye}, 4 = \text{Left Eye}, 5 = \text{Non}\}$. The levels denote the five areas of interest presented in Hessels, Cornelissen, Hooge, and Kemner (2017) where the last level, the Non, is used to indicate that the eye-movement data was collected but gaze was not on the other areas of interest. The categorical variable ETS is investigated using the cell probabilities in Table 3.5. The cell probability π_{ijv} is the true probability of person A looking at the ith level of ETS at time point t, person B looking at the jth level of ETS at time point $t + 20$, and person B looking at the vth level of ETS at time point $t + 200$ for $i, j, v = 1, 2, ..., 5$ and $t = 1, 2, ..., N = 8851$. As can be seen in Figure 3.1, the data set contain $N + 200$ time frames. However, only N of them are used to investigate the eye movements of the two persons at different time points.

In the study, there are two types of hypotheses: the hypotheses of interaction and no interaction. The hypotheses of interaction state a specific interaction, where the eye movements of one participant guide the eye movements of the other participant, such that the second participant mimics where the first participant looks (i.e., the gaze location). This influence is more explicit at time frames that are relatively close to each other. Similarly, the hypotheses of no interaction state that the participants do not influence the gaze location of each other for a given time point. Hypotheses H_1 and H_2 are based on the hypotheses of interaction and hypotheses H_3 and H_4 are based on the hypotheses of no interaction. Hypothesis H_1 states that the probability of person A at time point t and person B at time point $t + 20$ looking at the same object with all possible levels at time point $t + 200$ (i.e., π_{11+} for nose, π_{22+} for mouth, π_{33+} for right eye, π_{44+} for left eye, respectively) is bigger than the probability of person A at time point t and person B at time point $t + 200$ looking at the same object with all possible levels at time point $t + 20$ (i.e., π_{1+1} for nose, π_{2+2} for mouth, π_{3+3} for right eye, π_{4+4} for left eye, respectively):

$$H_1 : \pi_{11+} > \pi_{1+1},\ \pi_{22+} > \pi_{2+2},\ \pi_{33+} > \pi_{3+3},\ \pi_{44+} > \pi_{4+4}. \tag{3.14}$$

The marginal cell probabilities in equation (3.14) represent the cell probabilities that are summed over the levels of one variable, for example, $\pi_{11+} = \pi_{111}+\pi_{112}+\pi_{113}+\pi_{114}+\pi_{115}$, that is, the probability of person A at time point t looking at the nose and person B at time point $t + 20$ looking at the nose with all possible levels at time point $t + 200$. We provide hypothesis H_1 with the other three hypotheses, namely, hypotheses H_2, H_3, and H_4, and the unconstrained hypothesis H_u, which are described below in a similar manner as hypothesis

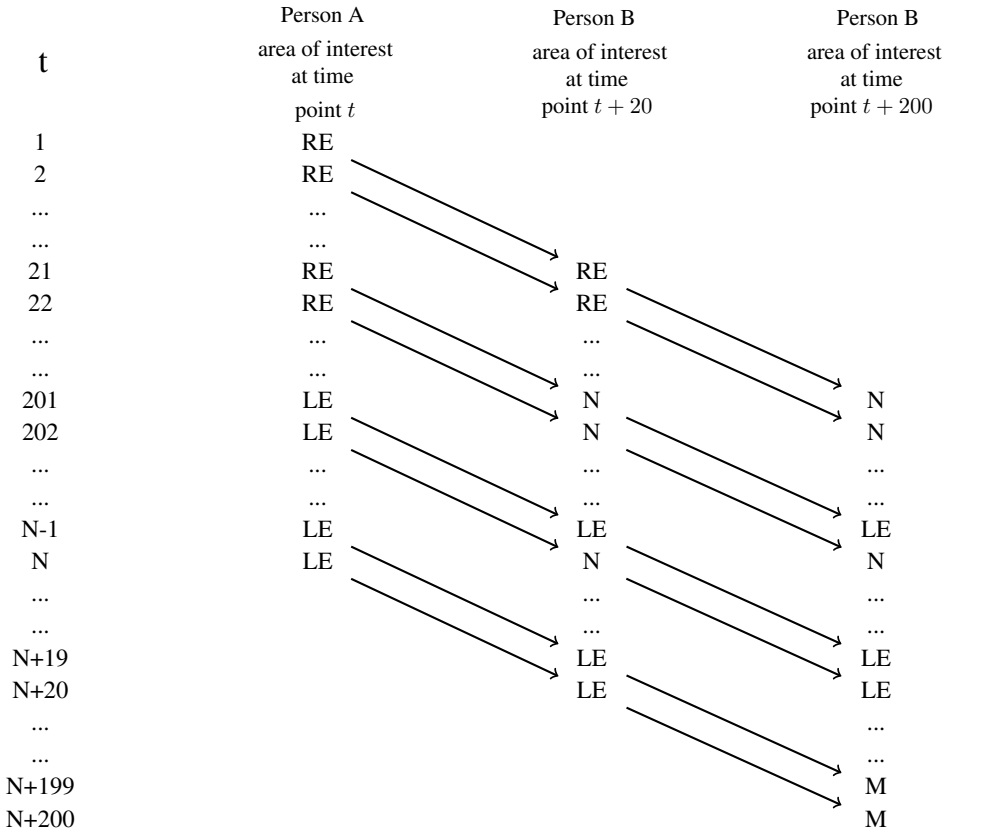

Figure 3.1: The Eye-Tracking Dataset under the model that the gaze location of person A influences the gaze location of person B

H_1, but in less detail:

$$\begin{aligned}
&H_1 : \pi_{11+} > \pi_{1+1},\ \pi_{22+} > \pi_{2+2},\ \pi_{33+} > \pi_{3+3},\ \pi_{44+} > \pi_{4+4},\\
&H_2 : \pi_{11+} > \pi_{1+1},\ \pi_{22+} > \pi_{2+2},\ \pi_{34+} > \pi_{3+4},\ \pi_{43+} > \pi_{4+3},\\
&H_3 : \pi_{11+} = \pi_{1+1},\ \pi_{22+} = \pi_{2+2},\ \pi_{33+} = \pi_{3+3},\ \pi_{44+} = \pi_{4+4},\\
&H_4 : \pi_{11+} = \pi_{1+1},\ \pi_{22+} = \pi_{2+2},\ \pi_{34+} = \pi_{3+4},\ \pi_{43+} = \pi_{4+3},\\
&H_u : \pi_{11+},\ \pi_{1+1}, ..., \pi_{4+4}.
\end{aligned} \tag{3.15}$$

Note that we reduce the number of the parameters from 124 to 12 (without loss of information) by estimating the twelve η's to evaluate the hypotheses in equation (3.15). Hypothesis H_2 states that the situation in H_1 remains valid for only the objects nose and mouth, and because persons A and B sit across from each other the left eye and the right eye should be shifted, that is, the probability of person A at time point t looking at the left eye and person B at time point $t + 20$ looking at the right eye with all possible levels at time point $t + 200$ (i.e., π_{34+}) is bigger than the probability of person A at time point t looking at the left eye and person B at time point $t + 200$ looking at the right eye with all possible levels at time point $t + 20$ (i.e., π_{3+4}) and vice versa (i.e., π_{43+} compared to π_{4+3}). Compared to hypothesis H_1, hypothesis H_3 specifies that the gaze location of person A does not influence the gaze location of person B, and therefore, the corresponding marginal cell probabilities in hypothesis H_1 are equal to each other. Similarly, hypothesis H_4 specifies that the gaze location of person A does not influence the gaze location of person B, and therefore, the marginal cell probabilities in hypothesis H_2 are equal to each other. The unconstrained hypothesis H_u covers the relationships between marginal cell probabilities that are represented by hypotheses H_1, H_2, H_3, and H_4, and all the other situations which is used to avoid choosing a weak hypothesis out of the four hypotheses.

During the experiment, both persons may influence the gaze location of each other. Therefore, a data set ($N = 8851$ video frames) is analyzed under two possible scenarios. For both scenarios, the same hypotheses, namely those in equation (3.15), are evaluated. In the first scenario, as mentioned, the gaze location of person A is considered to influence the gaze location of person B, such that person B follows person A. In the second scenario, the gaze location of person B is considered to influence the gaze location of person A, and therefore, the roles of person A and person B are turned around. In this case, for example, $\pi_{11+} = \pi_{111} + \pi_{112} + \pi_{113} + \pi_{114} + \pi_{115}$ means that the probability of person B at time point t looking at the nose and person A at time point $t + 20$ looking at the nose with all possible levels at time point $t + 200$. In Tables 3.5 and 3.6, the true cell probabilities and observed cell frequencies are displayed for the two scenarios, respectively.

Table 3.5: Population probabilities π_{ijv} and observed cell frequencies between brackets for the eye-tracking example under the model that the gaze location of person A influences the gaze location of person B

π_{ijv}		Person B at time point $t+200$				
Person A at time point t	Person B at time point $t+20$	Nose	Mouth	Right Eye	Left Eye	Non
1 = Nose	Nose	π_{111} (51)	π_{112} (2)	π_{113} (42)	π_{114} (46)	π_{115} (7)
	Mouth	π_{121} (4)	π_{122} (0)	π_{123} (5)	π_{124} (3)	π_{125} (0)
	Right Eye	π_{131} (44)	π_{132} (4)	π_{133} (174)	π_{134} (39)	π_{135} (2)
	Left Eye	π_{141} (14)	π_{142} (3)	π_{143} (34)	π_{144} (36)	π_{145} (4)
	Non	π_{151} (4)	π_{152} (1)	π_{153} (26)	π_{154} (11)	π_{155} (0)
2 = Mouth	Nose	π_{211} (1)	π_{212} (0)	π_{213} (1)	π_{214} (2)	π_{215} (0)
	Mouth	π_{221} (0)	π_{222} (0)	π_{223} (0)	π_{224} (0)	π_{225} (0)
	Right Eye	π_{231} (5)	π_{232} (0)	π_{233} (4)	π_{234} (2)	π_{235} (0)
	Left Eye	π_{241} (1)	π_{242} (0)	π_{243} (3)	π_{244} (0)	π_{245} (0)
	Non	π_{251} (0)	π_{252} (0)	π_{253} (0)	π_{254} (0)	π_{255} (0)
3 = Right Eye	Nose	π_{311} (1006)	π_{312} (112)	π_{313} (921)	π_{314} (495)	π_{315} (91)
	Mouth	π_{321} (81)	π_{322} (1)	π_{323} (79)	π_{324} (22)	π_{325} (2)
	Right Eye	π_{331} (1084)	π_{332} (109)	π_{333} (1909)	π_{334} (397)	π_{335} (134)
	Left Eye	π_{341} (486)	π_{342} (42)	π_{343} (419)	π_{344} (220)	π_{345} (33)
	Non	π_{351} (63)	π_{352} (3)	π_{353} (151)	π_{354} (26)	π_{355} (1)
4 = Left Eye	Nose	π_{411} (34)	π_{412} (3)	π_{413} (7)	π_{414} (21)	π_{415} (1)
	Mouth	π_{421} (8)	π_{422} (0)	π_{423} (2)	π_{424} (1)	π_{425} (0)
	Right Eye	π_{431} (23)	π_{432} (0)	π_{433} (20)	π_{434} (13)	π_{435} (3)
	Left Eye	π_{441} (6)	π_{442} (11)	π_{443} (5)	π_{444} (0)	π_{445} (0)
	Non	π_{451} (0)	π_{452} (0)	π_{453} (3)	π_{454} (0)	π_{455} (0)
5 = Non	Nose	π_{511} (28)	π_{512} (0)	π_{513} (18)	π_{514} (4)	π_{515} (7)
	Mouth	π_{521} (0)	π_{522} (0)	π_{523} (2)	π_{524} (0)	π_{525} (0)
	Right Eye	π_{531} (22)	π_{532} (2)	π_{533} (98)	π_{534} (14)	π_{535} (1)
	Left Eye	π_{541} (5)	π_{542} (0)	π_{543} (25)	π_{544} (0)	π_{545} (0)
	Non	π_{551} (2)	π_{552} (0)	π_{553} (0)	π_{554} (0)	π_{555} (0)

Table 3.6: Population probabilities π_{ijv} and observed cell frequencies between brackets for the eye-tracking example under the model that the gaze location of person B influences the gaze location of person A

π_{ijv}		Person A at time point $t + 200$				
Person B at time point t	Person A at time point $t + 20$	Nose	Mouth	Right Eye	Left Eye	Non
1 = Nose	Nose	π_{111} (5)	π_{112} (0)	π_{113} (155)	π_{114} (3)	π_{115} (9)
	Mouth	π_{121} (0)	π_{122} (0)	π_{123} (10)	π_{124} (0)	π_{125} (0)
	Right Eye	π_{131} (124)	π_{132} (5)	π_{133} (2350)	π_{134} (38)	π_{135} (82)
	Left Eye	π_{141} (5)	π_{142} (3)	π_{143} (47)	π_{144} (0)	π_{145} (1)
	Non	π_{151} (5)	π_{152} (0)	π_{153} (61)	π_{154} (0)	π_{155} (2)
2 = Mouth	Nose	π_{211} (1)	π_{212} (0)	π_{213} (8)	π_{214} (0)	π_{215} (0)
	Mouth	π_{221} (0)	π_{222} (0)	π_{223} (0)	π_{224} (0)	π_{225} (0)
	Right Eye	π_{231} (14)	π_{232} (0)	π_{233} (162)	π_{234} (15)	π_{235} (6)
	Left Eye	π_{241} (0)	π_{242} (0)	π_{243} (2)	π_{244} (1)	π_{245} (0)
	Non	π_{251} (1)	π_{252} (0)	π_{253} (1)	π_{254} (0)	π_{255} (0)
3 = Right Eye	Nose	π_{311} (14)	π_{312} (2)	π_{313} (264)	π_{314} (4)	π_{315} (9)
	Mouth	π_{321} (1)	π_{322} (0)	π_{323} (8)	π_{324} (0)	π_{325} (0)
	Right Eye	π_{331} (365)	π_{332} (47)	π_{333} (3009)	π_{334} (78)	π_{335} (86)
	Left Eye	π_{341} (4)	π_{342} (4)	π_{343} (69)	π_{344} (8)	π_{345} (1)
	Non	π_{351} (10)	π_{352} (0)	π_{353} (122)	π_{354} (4)	π_{355} (0)
4 = Left Eye	Nose	π_{411} (3)	π_{412} (0)	π_{413} (46)	π_{414} (0)	π_{415} (1)
	Mouth	π_{421} (0)	π_{422} (0)	π_{423} (0)	π_{424} (0)	π_{425} (0)
	Right Eye	π_{431} (45)	π_{432} (2)	π_{433} (1154)	π_{434} (13)	π_{435} (15)
	Left Eye	π_{441} (6)	π_{442} (14)	π_{443} (9)	π_{444} (0)	π_{445} (1)
	Non	π_{451} (1)	π_{452} (0)	π_{453} (17)	π_{454} (1)	π_{455} (0)
5 = Non	Nose	π_{511} (5)	π_{512} (0)	π_{513} (27)	π_{514} (0)	π_{515} (0)
	Mouth	π_{521} (0)	π_{522} (0)	π_{523} (0)	π_{524} (0)	π_{525} (0)
	Right Eye	π_{531} (15)	π_{532} (0)	π_{533} (214)	π_{534} (20)	π_{535} (8)
	Left Eye	π_{541} (0)	π_{542} (0)	π_{543} (0)	π_{544} (5)	π_{545} (1)
	Non	π_{551} (3)	π_{552} (0)	π_{553} (0)	π_{554} (0)	π_{555} (0)

Table 3.7: The estimates $\hat{\eta}$ and the covariance matrix $\hat{\Sigma}_{\hat{\eta}}$ for the eye-tracking example

		Under the model that the gaze location of person A influences the gaze location person B											
		$\hat{\Sigma}_{\hat{\eta}}$											
η	$\hat{\eta}$	η_1	η_2	η_3	η_4	η_5	η_6	η_7	η_8	η_9	η_{10}	η_{11}	η_{12}
η_1	0.017	1.9e-6											
η_2	0.013	6.5e-7	1.5e-6										
η_3	0.000	0.000	0.000	0.000									
η_4	0.000	0.000	0.000	0.000	0.000								
η_5	0.410	-1.1e-6	-6.5e-7	0.000	0.000	2.7e-5							
η_6	0.393	-8.9e-7	-6.7e-7	0.000	0.000	7.0e-6	2.7e-5						
η_7	0.136	-2.3e-7	-2.8e-7	0.000	0.000	-5.3e-6	-8.2e-7	1.2e-5					
η_8	0.131	-2.2e-7	-8.7e-8	0.000	0.000	-6.7e-7	-5.2e-6	1.2e-6	1.3e-5				
η_9	0.007	1.7e-8	-1.2e-8	0.000	0.000	-1.1e-7	-3.6e-7	-1.7e-7	2.9e-9	7.2e-7			
η_{10}	0.004	7.8e-9	-1.2e-8	0.000	0.000	-1.6e-7	-3.5e-7	-1.2e-7	-9.3e-8	2.7e-7	4.9e-7		
η_{11}	0.003	-4.3e-8	-2.8e-8	0.000	0.000	-9.7e-8	-2.5e-7	-3.7e-8	4.3e-8	-1.7e-8	5.0e-8	2.8e-7	
η_{12}	0.004	-2.1e-8	4.8e-9	0.000	0.000	-1.4e-7	-1.9e-7	-4.1e-8	-2.8e-8	1.3e-7	-1.9e-8	-5.7e-10	4.3e-7

		Under the model that the gaze location of person B influences the gaze location person A											
		$\hat{\Sigma}_{\hat{\eta}}$											
η	$\hat{\eta}$	η_1	η_2	η_3	η_4	η_5	η_6	η_7	η_8	η_9	η_{10}	η_{11}	η_{12}
η_1	0.020	2.3e-6											
η_2	0.016	4.9e-8	1.6e-6										
η_3	0.000	0.000	0.000	0.000									
η_4	0.000	0.000	0.000	0.000	0.000								
η_5	0.405	-1.2e-6	-5.6e-7	0.000	0.000	2.7e-5							
η_6	0.392	-1.0e-6	-6.2e-7	0.000	0.000	2.1e-5	2.8e-5						
η_7	0.010	-8.4e-8	-2.6e-8	0.000	0.000	-6.0e-7	3.2e-7	1.2e-6					
η_8	0.011	-4.8e-8	6.1e-8	0.000	0.000	5.5e-7	-5.0e-7	1.4e-7	1.2e-6				
η_9	0.139	-2.6e-7	-1.3e-7	0.000	0.000	-5.4e-6	-4.8e-6	-1.2e-7	-2.7e-7	1.2e-5			
η_{10}	0.139	-1.8e-7	-1.2e-7	0.000	0.000	-5.8e-6	-5.0e-6	-9.5e-8	-3.1e-7	1.1e-5	1.2e-5		
η_{11}	0.003	-8.9e-9	-4.0e-8	0.000	0.000	-2.3e-7	-2.4e-7	-3.8e-9	-1.1e-8	-6.9e-10	1.7e-7	3.9e-7	
η_{12}	0.002	-2.2e-8	-1.3e-8	0.000	0.000	-2.0e-8	-8.7e-8	1.4e-8	-4.8e-9	1.6e-7	-4.5e-8	-1.0e-8	1.9e-7

Note. $\eta = (\eta_1, \eta_2, \eta_3, \eta_4, \eta_5, \eta_6, \eta_7, \eta_8, \eta_9, \eta_{10}, \eta_{11}, \eta_{12})^T$ =$(\pi_{11+}, \pi_{1+1}, \pi_{22+}, \pi_{2+2}, \pi_{33+}, \pi_{3+3}, \pi_{34+}, \pi_{3+4}, \pi_{43+}, \pi_{4+3}, \pi_{44+}, \pi_{4+4})^T$.

The corresponding MLEs and their covariance matrix in terms of the reparameterized cell probabilities, that is, the marginal cell probabilities in equation (3.15), are given in Table 3.7 for both scenarios, respectively. Similarly, the order-restricted MLEs are given in Table 3.8 for the five hypotheses H_1, H_2, H_3, H_4, and the unconstrained hypothesis H_u. In Tables 3.9.1 and 3.9.2, the values of the likelihood and penalty parts, GORICA and GORICA weights are displayed for the two scenarios, respectively. Based on the GORICA weights in Table 3.9.1, hypotheses H_3 and H_4 do not receive any support from the data set for the first scenario. Hypothesis H_2 receives quite considerable amount of support from the data, that is, hypothesis H_2 is $0.684/0.155 \approx 4.41$ times better than the unconstrained hypothesis H_u.

Table 3.8: The order-restricted MLEs for the eye-tracking example

Under the model that the gaze location of person A influences the gaze location person B												
H_m	$\tilde{\eta}_{m1}$	$\tilde{\eta}_{m2}$	$\tilde{\eta}_{m3}$	$\tilde{\eta}_{m4}$	$\tilde{\eta}_{m5}$	$\tilde{\eta}_{m6}$	$\tilde{\eta}_{m7}$	$\tilde{\eta}_{m8}$	$\tilde{\eta}_{m9}$	$\tilde{\eta}_{m10}$	$\tilde{\eta}_{m11}$	$\tilde{\eta}_{m12}$
H_1	0.017	0.013	0.000	0.000	0.410	0.393	0.136	0.131	0.006	0.004	0.003	0.003
H_2	0.017	0.013	0.000	0.000	0.410	0.393	0.136	0.131	0.007	0.004	0.003	0.004
H_3	0.015	0.015	0.000	0.000	0.402	0.402	0.138	0.129	0.006	0.004	0.003	0.003
H_4	0.015	0.015	0.000	0.000	0.412	0.392	0.133	0.133	0.005	0.005	0.003	0.003
H_u	0.017	0.013	0.000	0.000	0.410	0.393	0.136	0.131	0.007	0.004	0.003	0.004
Under the model that the gaze location of person B influences the gaze location person A												
H_m	$\tilde{\eta}_{m1}$	$\tilde{\eta}_{m2}$	$\tilde{\eta}_{m3}$	$\tilde{\eta}_{m4}$	$\tilde{\eta}_{m5}$	$\tilde{\eta}_{m6}$	$\tilde{\eta}_{m7}$	$\tilde{\eta}_{m8}$	$\tilde{\eta}_{m9}$	$\tilde{\eta}_{m10}$	$\tilde{\eta}_{m11}$	$\tilde{\eta}_{m12}$
H_1	0.020	0.016	0.000	0.000	0.405	0.392	0.010	0.011	0.139	0.139	0.003	0.002
H_2	0.020	0.016	0.000	0.000	0.404	0.392	0.010	0.010	0.139	0.139	0.003	0.002
H_3	0.017	0.017	0.000	0.000	0.400	0.400	0.011	0.010	0.140	0.139	0.002	0.002
H_4	0.017	0.017	0.000	0.000	0.405	0.393	0.010	0.010	0.139	0.139	0.003	0.002
H_u	0.020	0.016	0.000	0.000	0.405	0.392	0.010	0.011	0.139	0.139	0.003	0.002

Note. The vector $\tilde{\eta}_m = (\tilde{\eta}_{m1}, \tilde{\eta}_{m2}, \tilde{\eta}_{m3}, \tilde{\eta}_{m4}, \tilde{\eta}_{m5}, \tilde{\eta}_{m6}, \tilde{\eta}_{m7}, \tilde{\eta}_{m8}, \tilde{\eta}_{m9}, \tilde{\eta}_{m10}, \tilde{\eta}_{m11}, \tilde{\eta}_{m12})^T$ denotes the reparameterized cell probabilities that are in agreement with the restrictions of the hypotheses in equation (3.15).

Table 3.9.1: The likelihood parts $L(\tilde{\eta}_m|\hat{\eta}, \hat{\Sigma}_{\hat{\eta}})$, the penalty parts $PT_m(\eta)$, the GORICA values $GORICA_m$, and the GORICA weights ψ_m for hypothesis H_m with $m = 1, 2, 3, 4$ and H_u under the model that the gaze location of person A influences the gaze location person B

H_m	$L(\tilde{\eta}_m\|\hat{\eta}, \hat{\Sigma}_{\hat{\eta}})$	$PT_m(\eta)$	$GORICA_m$	ψ_m
H_1	52.238	8.493	-87.489	0.159
H_2	53.714	8.513	-90.402	0.684
H_3	45.344	7.000	-76.688	< 0.001
H_4	45.693	7.000	-77.386	0.001
H_u	53.714	10.000	-87.428	0.155

Hypothesis H_2 is $0.684/0.159 \approx 4.30$ times better than hypothesis H_1. Therefore, hypothesis H_2 is selected as the best hypothesis among the four hypotheses for the first scenario. However, as can be seen in Table 3.9.2, in the second scenario, hypothesis H_2 is not the only hypothesis supported by the data. Hypotheses H_1 and H_4 also receive some support from the data. Hypothesis H_1 has $0.379/0.085 \approx 4.46$, hypothesis H_2 has $0.324/0.085 \approx 3.81$, and hypothesis H_4 has $0.212/0.085 \approx 2.49$ times more support than the unconstrained hypothesis H_u. Since the support relative to each other is close to 1, for example, hypothesis

Table 3.9.2: The likelihood parts $L(\tilde{\eta}_m|\hat{\eta}, \hat{\Sigma}_{\hat{\eta}})$, the penalty parts $PT_m(\eta)$, the GORICA values $GORICA_m$, and the GORICA weights ψ_m for hypothesis H_m with $m = 1, 2, 3, 4$ and H_u under the model that the gaze location of person B influences the gaze location person A

H_m	$L(\tilde{\eta}_m\|\hat{\eta}, \hat{\Sigma}_{\hat{\eta}})$	$PT_m(\eta)$	$GORICA_m$	ψ_m
H_1	54.316	8.502	-91.628	0.379
H_2	54.132	8.477	-91.311	0.324
H_3	43.361	7.000	-72.722	< 0.001
H_4	52.230	7.000	-90.461	0.212
H_u	54.316	10.000	-88.632	0.085

H_1 has only $0.379/0.324 \approx 1.17$ times more support than hypothesis H_2, we conclude that although hypothesis H_1 is the best hypothesis among the four hypotheses, the evidence for this conclusion is not quite obvious in the second scenario.

The use of the log-linear model in evaluating hypotheses for a $5 \times 5 \times 5$ contingency table involves 124 parameters in total, including two-way and three-way interaction terms, which are difficult to interpret by researchers. However, by using the GORICA, only twelve η parameters and their covariance matrix need to be estimated to evaluate the hypotheses in equation (3.15).

3.5.2 The gender by earned degrees example (Continued)

Consider again the gender by earned degrees example from the introduction for which the true cell probabilities $(D = 8)$ and the data set $(N = 2403)$ were introduced earlier in Table 3.1. It was mentioned in the introduction that the evaluation of the hypothesis $H_1 : \frac{\pi_{21}}{\pi_{+1}} > \{\frac{\pi_{22}}{\pi_{+2}}, \frac{\pi_{23}}{\pi_{+3}}, \frac{\pi_{24}}{\pi_{+4}}\}$ in equation (3.1) might be of interest to researchers.

Now, one could also be interested in the evaluation of two more competing hypotheses together with the unconstrained hypothesis H_u:

$$
\begin{aligned}
H_1 &: \frac{\pi_{21}}{\pi_{+1}} > \{\frac{\pi_{22}}{\pi_{+2}}, \frac{\pi_{23}}{\pi_{+3}}, \frac{\pi_{24}}{\pi_{+4}}\},\\
H_2 &: \{\frac{\pi_{21}}{\pi_{+1}} = \frac{\pi_{22}}{\pi_{+2}}\}, \{\frac{\pi_{23}}{\pi_{+3}} > \frac{\pi_{24}}{\pi_{+4}}\},\\
H_3 &: \frac{\pi_{21}}{\pi_{+1}} > \frac{\pi_{22}}{\pi_{+2}} > \frac{\pi_{23}}{\pi_{+3}} > \frac{\pi_{24}}{\pi_{+4}},\\
H_u &: \frac{\pi_{21}}{\pi_{+1}}, \frac{\pi_{22}}{\pi_{+2}}, \frac{\pi_{23}}{\pi_{+3}}, \frac{\pi_{24}}{\pi_{+4}}.
\end{aligned}
\tag{3.16}
$$

The hypothesis H_2 states that, after taking into account the prevalence of each degree across males and females, a bachelor's degree and a master's degree being obtained by males are equally likely and a professional degree being obtained by males is more likely than a doctorate degree being obtained by males. Similarly, hypothesis H_3 specifies that the probability of a degree being obtained by males given that the degree is a doctorate degree is the smallest probability followed by those for who the degree is a professional, a master's, and a bachelor's degree, respectively. The unconstrained hypothesis H_u states that there may not be any relationship between the four conditional cell probabilities or there may be relationships between them but not necessarily the ones specified in hypotheses H_1, H_2, and H_3.

The MLEs of the conditional cell probabilities in equation (3.16) and their covariance matrix are given in Table 3.10. The order-restricted MLEs with respect to hypotheses H_1, H_2, H_3, and the unconstrained hypothesis H_u are displayed in Table 3.11. The corresponding values of the likelihood and penalty parts, GORICA and GORICA weights for the four hypotheses are given in Table 3.12. Based on the GORICA weights, hypothesis H_3 is a weak hypothesis, that is, hypothesis H_3 $0.156/0.174 \approx 0.90$ times better than the unconstrained hypothesis H_u. Hypothesis H_1 is not a strong hypothesis, because it is only $0.211/0.174 \approx 1.21$ times better than the unconstrained hypothesis H_u. Hypothesis H_2 receives the highest support

Table 3.10: The estimates $\hat{\eta}$ and the covariance matrix $\hat{\Sigma}_{\hat{\eta}}$ of the estimates of the conditional cell probabilities for the gender by earned degrees example

		$\hat{\Sigma}_{\hat{\eta}}$			
η	$\hat{\eta}$	η_1	η_2	η_3	η_4
η_1	0.415	1.6e-4			
η_2	0.393	-3.6e-6	3.5e-4		
η_3	0.465	-2.3e-5	1.7e-5	2.5e-3	
η_4	0.502	-1.7e-6	-5.7e-5	-8.4e-5	4.8e-3

Note. $\eta = (\eta_1, \eta_2, \eta_3, \eta_4)^T = (\frac{\pi_{21}}{\pi_{+1}}, \frac{\pi_{22}}{\pi_{+2}}, \frac{\pi_{23}}{\pi_{+3}}, \frac{\pi_{24}}{\pi_{+4}})^T$.

Table 3.11: The order-restricted MLEs of the η's for the gender by earned degrees example

H_m	$\tilde{\eta}_{m1}$	$\tilde{\eta}_{m2}$	$\tilde{\eta}_{m3}$	$\tilde{\eta}_{m4}$
H_1	0.421	0.393	0.421	0.421
H_2	0.408	0.408	0.478	0.478
H_3	0.415	0.409	0.409	0.409
H_u	0.415	0.393	0.465	0.502

Note. The vector $\tilde{\eta}_m = (\tilde{\eta}_{m1}, \tilde{\eta}_{m2}, \tilde{\eta}_{m3}, \tilde{\eta}_{m4})^T$ denotes the reparameterized cell probability estimates that are in agreement with the restrictions of the hypotheses H_1, H_2 and H_3 in equation (3.16).

Table 3.12: The likelihood parts $L(\tilde{\eta}_m|\hat{\eta}, \hat{\Sigma}_{\hat{\eta}})$, the penalty parts $PT_m(\eta)$, the GORICA values $GORICA_m$, and the GORICA weights ψ_m for hypothesis H_m with $m = 1, 2, 3, 4$ and H_u for the gender by earned degrees example

H_m	$L(\tilde{\eta}_m\|\hat{\eta}, \hat{\Sigma}_{\hat{\eta}})$	$PT_m(\eta)$	$GORICA_m$	ψ_m
H_1	9.116	2.590	-13.052	0.211
H_2	9.806	2.500	-14.612	0.459
H_3	8.407	2.183	-12.448	0.156
H_u	10.334	4.000	-12.668	0.174

in the set of the three hypotheses, and because hypothesis H_2 is $0.459/0.174 \approx 2.64$ times better than the unconstrained hypothesis H_u, it is not a weak hypothesis. Therefore, it is concluded that hypothesis H_2 better explains the relationship between the cell probabilities compared to the other three hypotheses based on the data set at hand.

3.6 Discussion

Contingency tables can be analyzed using the log-linear model. However, analysis and interpretation of the data for high-dimensional contingency tables and/or tables with empty cells is problematic by means of the log-linear model. We implemented the GORICA (Altınışık, Nederhof, et al., unpublished) in the context of contingency tables to remedy this problem. We showed that the GORICA can be used to evaluate hypotheses containing linear or nonlinear restrictions on cell probabilities in the context of (sparse) high-dimensional contingency tables. Two examples are analyzed to illustrate the applicability of the GORICA. The method is implemented in the R script GoricaCont.R which is discussed in Appendix 3.A and available in the research archive sent with this dissertation.

3.A User manual for GoricaCont.R

This appendix discusses how to implement the R script GoricaCont.R which is available as supplementary material. To be able to execute the R code in GoricaCont.R, the user should adjust two files called data.txt and input.txt in their working directory based on their data and hypotheses. The names of these files has been fixed in the R code and therefore cannot be changed. The data in data.txt must be provided in terms of observed cell frequencies and without any missing value. The data in data.txt contain $J + 1$ columns with J representing the dimension of the contingency table, for example, $J = 3$ for a $2 \times 2 \times 2$ contingency table. The values in the first J columns represent the levels of the variables in the contingency table and the values of the last column are the corresponding observed cell frequencies. The use of $J + 1$ columns is obligatory. For example, consider the gender by earned degrees example with $J = 2$ dimensions for which data.txt is given as:

1	1	933
1	2	402
1	3	51
1	4	26
2	1	661
2	2	260
2	3	44
2	4	26

The first two columns contain the levels of the variable gender, G $\in$ {1 = Female, 2 = Male}, and the variable academic degree, AD $\in$ {1 = Bachelor's, 2 = Master's, 3 = Professional, 4 = Doctorate}, respectively, and the last column consists of the corresponding cell frequencies. The numbers in the third, that is, $(J + 1)$th, column are stacked into a vector of observed cell frequencies based on this order. That is, the number in the third column in row d denotes y_d with $d = 1, 2, ..., D$. Therefore, users must enter their data into the file data.txt as shown above. The input file for this example is displayed below and elaborated on next:

#D	K	Seed	B	Iterations
8	4	113	1000	10000

#Parameters used in evaluation

x[5] / (x[1] + x[5])

x[6] / (x[2] + x[6])

x[7] / (x[3] + x[7])

x[8] / (x[4] + x[8])

```
#Number of models to be compared
  4
#Number of equality and inequality constraints per model
  0        3
  1        1
  0        3
  0        0
#Model 1
#Restriction matrix
  1     -1     0     0
  1      0    -1     0
  1      0     0    -1
#Constants
  0     0     0
#Model 2
#Restriction matrix
  1     -1     0     0
  0      0     1    -1
#Constants
  0     0
#Model 3
#Restriction matrix
  1     -1     0     0
  0      1    -1     0
  0      0     1    -1
#Constants
  0     0     0
#Model 4 (unconstrained hypothesis)
#Restriction matrix
  0     0     0     0
#Constants
  0
```

The use of the "#" is compulsory for the lines of comments. The first line with "#" contains five labels: label D for the number of cells in the contingency table, label K for the number of parameters used in evaluation, label Seed for the index number used in the random number generator, label B for the number of bootstrap samples when obtaining the MLEs and their covariance matrix, label Iterations for the number of the iterations when calculating the penalty part. The second line contains the corresponding numbers for which $D = 8$ and $K = 4$ are the compulsory numbers to evaluate the hypotheses in equation (3.16). One needs to use the recommended values seed = 113, $B = 1000$, and $T = 10000$ in order to duplicate the results presented in Tables 3.10, 3.11, and 3.12. After the line "#Parameters used in evaluation (eta)", $K = 4$ conditional cell probabilities are expressed as functions of the $D = 8$ cell probabilities. Note, for example, that x[5] / (x[1] + x[5]) represents the first parameter in $\eta = (\eta_1, \eta_2, \eta_3, \eta_4)^T = (\frac{\pi_{21}}{\pi_{+1}}, \frac{\pi_{22}}{\pi_{+2}}, \frac{\pi_{23}}{\pi_{+3}}, \frac{\pi_{24}}{\pi_{+4}})^T$ used for the gender by earned degrees example, that is, $\eta_1 = \frac{\pi_{21}}{\pi_{+1}} = \frac{\pi_5}{\pi_1 + \pi_5}$. Note that the four operators "+", "-",

"$*$", and "/" can be used to specify the η's for the classes of restrictions in Tables 3.2 and 3.3. The third line with "#" is followed by the number of hypotheses in evaluation, here $M = 4$, that is, the hypotheses H_1, H_2, H_3, and the unconstrained hypothesis H_u in equation (3.16), which is represented by hypothesis H_4 above. The fourth line with "#" is followed by the numbers of the equality and inequality constraints in each of hypotheses. Hence, the two numbers in each of the four lines represent the number of equality and inequality constraints in the corresponding hypothesis, respectively. For example, the lines with "0 3" imply that hypotheses H_1 and H_3 have no equality constraints and have three inequality constraints. Labels Model 1, 2, 3, and 4 denote hypotheses H_1, H_2, H_3, and H_4, each of which must be followed by a restriction matrix and a vector of constants. Label Restriction matrix reflects both matrices S_m and R_m in equation (3.2), while label Constants reflects both vectors s_m and r_m in equation (3.2) for the corresponding hypothesis. The script GoricaCont.R always considers equality constraints first. It is therefore compulsory for users to first introduce equality constraints of hypothesis H_m in both restriction matrices and constants. Each row in a restriction matrix corresponds to only one equality or inequality constraint. To illustrate how to specify the restriction matrix and constants, consider the hypotheses $H_1 : \frac{\pi_{21}}{\pi_{+1}} > \{\frac{\pi_{22}}{\pi_{+2}}, \frac{\pi_{23}}{\pi_{+3}}, \frac{\pi_{24}}{\pi_{+4}}\}$, $H_2 : \{\frac{\pi_{21}}{\pi_{+1}} = \frac{\pi_{22}}{\pi_{+2}}\}, \{\frac{\pi_{23}}{\pi_{+3}} > \frac{\pi_{24}}{\pi_{+4}}\}$, and $H_3 : \frac{\pi_{21}}{\pi_{+1}} > \frac{\pi_{22}}{\pi_{+2}} > \frac{\pi_{23}}{\pi_{+3}} > \frac{\pi_{24}}{\pi_{+4}}$ and the unconstrained hypothesis $H_u : \frac{\pi_{21}}{\pi_{+1}}, \frac{\pi_{22}}{\pi_{+2}}, \frac{\pi_{23}}{\pi_{+3}}, \frac{\pi_{24}}{\pi_{+4}}$ in equation (3.16) with $\eta = (\eta_1, \eta_2, \eta_3, \eta_4)^T = (\frac{\pi_{21}}{\pi_{+1}}, \frac{\pi_{22}}{\pi_{+2}}, \frac{\pi_{23}}{\pi_{+3}}, \frac{\pi_{24}}{\pi_{+4}})^T$:

Hypotheses	Restriction matrix	Right hand side values
H_1: $\eta_1 - \eta_2 > 0$	1 -1 0 0	0,
$\eta_1 - \eta_3 > 0$	1 0 -1 0	0,
$\eta_1 - \eta_4 > 0$	1 0 0 -1	0,
H_2: $\eta_1 - \eta_2 = 0$	1 -1 0 0	0,
$\eta_3 - \eta_4 > 0$	0 0 1 -1	0,
H_3: $\eta_1 - \eta_2 > 0$	1 -1 0 0	0,
$\eta_2 - \eta_3 > 0$	0 1 -1 0	0,
$\eta_3 - \eta_4 > 0$	0 0 1 -1	0,
H_u: $\eta_1, \eta_2, \eta_3, \eta_4$	0 0 0 0	0.

Note that the unconstrained hypothesis H_u does not contain any restriction on the η's.

The execution procedure of the GoricaCont.R comprises the following basic steps:

• Change the working directory to the folder you are working in using the "setwd()" function in R. For example, setwd("C:/Users/(Username)/Desktop/GoricaContingency"), if the GoricaContingency file is saved on the desktop of the computer.

• Enter the data into the file data.txt, specify the entries in the input file input.txt, and save the changes (CTRL+S).

• Copy and paste the following code into R console:

source("GoricaCont.R").

Upon the first execution in R, script GoricaCont.R checks whether the packages "matrixcalc", "MASS", "quadprog", "limSolve", and "FRACTION" are installed and if they are not installed, it installs them automatically. The program may produce regular warnings while installing these packages, which are not a problem for the execution of the R code in GoricaCont.R. Executing script GoricaCont.R based on the data and input files renders the following results in R:

$MLEs

p1	p2	p3	p4
0.4146	0.3928	0.4654	0.5022

$Covariance_matrix

	p1	p2	p3	p4
p1	1.6e-04	-3.6e-06	-2.3e-05	-1.7e-06
p2	-3.6e-06	3.5e-04	1.7e-05	-5.7e-05
p3	-2.3e-05	1.7e-05	2.5e-03	-8.4e-05
p4	-1.7e-06	-5.7e-05	-8.4e-05	4.8e-03

$Restricted_MLEs

	p1	p2	p3	p4
H_1	0.4209	0.3933	0.4209	0.4209
H_2	0.4077	0.4077	0.4783	0.4783
H_3	0.4151	0.4087	0.4087	0.4087
H_4	0.4146	0.3928	0.4654	0.5022

	LogLikelihood	Penalty	GORICA_value	GORICA_weight
H_1	9.1160	2.5901	-13.0519	0.2107
H_2	9.8063	2.5003	-14.6120	0.4596
H_3	8.4072	2.1831	-12.4482	0.1558
H_4	10.3339	4.0000	-12.6677	0.1739

For hypotheses containing linear restrictions on cell probabilities where $\eta = \pi$, replacing the last cell probability (or replacing the last non-zero estimated cell probability when $\hat{\pi}_D = 0$ because of an empty cell), by one minus the sum of the other cell probabilities is embedded in the software. To illustrate, consider a 2×2 contingency table which contains four cell probabilities, with $\eta = (\eta_1, \eta_2, \eta_3, \eta_4) = (\pi_{11}, \pi_{12}, \pi_{21}, \pi_{22})^T$. Consider evaluation of hypotheses $H_1 : \pi_{11} = \pi_{12}, \pi_{21} > \pi_{22}$, $H_2 : \pi_{11} > \pi_{12}, \pi_{21} > \pi_{22}$ and the unconstained hypothesis $H_u : \pi_{11}, \pi_{12}, \pi_{21}, \pi_{22}$.

#D	K	Seed	B	Iterations
4	4	113	1000	10000

#Number of models to be compared

3

#Number of equality and inequality constraints per model

```
  1          1
  0          2
  0          0
#Model 1
#Restriction matrix
  1    -1    0    0
  0     0    1   -1
#Constants
  0    0
#Model 2
#Restriction matrix
  1    -1    0    0
  0     0    1   -1
#Constants
  0    0
#Model 3 (unconstrained hypothesis)
#Restriction matrix
  0    0    0    0
#Constants
  0
```

The difference is now that $\eta = \pi$ and hence $K = D = 4$. Then, the software knows that the hypotheses should be transformed by replacing the last cell probability (or the last non-zero estimated cell probability) with one minus the sum of the other cell probabilities. To prevent the user from typing out all the cell probabilities, we adjust the format for input.txt when $D = K$. Namely, in this case, one has to remove the line "#Parameters used in evaluation (eta)" and can right away start with "#Number of models to be compared".

3.A.1 Error and warning messages

In script GoricaCont.R, when file input.txt is not correctly specified, the procedure is stopped and error messages are given. We discuss some of these wrongly specified input.txt files and the corresponding error messages below.

- In case of $\eta = g(\pi)$, functions of cell probablities may be linearly dependent on each other. In such cases, our software cannot calculate the GORICA values, because the η parameters, for example, sum to one, and therefore, the covariance matrix of the estimates for these parameters is not positive definite. For example, below the four η parameters used in evaluation are linearly dependent on each other, because they sum to one.

```
#D          K          Seed       B          Iterations
  8          4          113        1000       10000
#Parameters used in evaluation
  x[ 1 ] / ( x[ 1 ] + x[ 2 ] + x[ 3 ] + x[ 4 ] )
```

x[2] / (x[1] + x[2] + x[3] + x[4])

x[3] / (x[1] + x[2] + x[3] + x[4])

x[4] / (x[1] + x[2] + x[3] + x[4])

#Number of models to be compared

...

Therefore, the program stops and gives the error message: "The functions of cell probabilities (eta) are linearly dependent on each other. The covariance matrix of the eta's is not positive definite."

- When $\eta = g(\pi)$, sometimes some of the η's are estimated as one or cannot be estimated because of empty cells. For the first situation, consider the comparison of the hypothesis $H_1 : \frac{\pi_{21}}{\pi_{+1}} > \{\frac{\pi_{22}}{\pi_{+2}}, \frac{\pi_{23}}{\pi_{+3}}, \frac{\pi_{24}}{\pi_{+4}}\}$ and the unconstrained hypothesis $H_u : \frac{\pi_{21}}{\pi_{+1}}, \frac{\pi_{22}}{\pi_{+2}}, \frac{\pi_{23}}{\pi_{+3}}, \frac{\pi_{24}}{\pi_{+4}}$ in equation (3.16), where, for example, $\hat{\pi}_{11} = 0$ because of an empty cell.

#D	K	Seed	B	Iterations
8	4	113	1000	10000

#Parameters used in evaluation

x[5] / (x[1] + x[5])

x[6] / (x[2] + x[6])

x[7] / (x[3] + x[7])

x[8] / (x[4] + x[8])

#Number of models to be compared

2

#Number of equality and inequality constraints per model

0 3

0 0

#Model 1

#Restriction matrix

1 -1 0 0
1 0 -1 0
1 0 0 -1

#Constants

0 0 0

#Model 2 (unconstrained hypothesis)

#Restriction matrix

0 0 0 0

#Constants

0

Below $\hat{\eta}_1 = \frac{\hat{\pi}_{21}}{\hat{\pi}_{11}+\hat{\pi}_{21}} = 1$ because of the empty cell, and therefore, has no variation.

$MLEs

p1	p2	p3	p4
1.0000	0.3920	0.4632	0.5035

Thus, our software stops and gives the warning message: "Some of the eta parameters are estimated as one because of the empty cell(s). The user needs to rewrite the hypotheses under evaluation."

For the second situation, consider the comparison of hypothesis $H_1 : \frac{\pi_1\pi_6}{\pi_2\pi_5} > \frac{\pi_2\pi_7}{\pi_3\pi_6} > \frac{\pi_3\pi_8}{\pi_4\pi_7}$ and the unconstrained hypothesis $H_u : \frac{\pi_1\pi_6}{\pi_2\pi_5}, \frac{\pi_2\pi_7}{\pi_3\pi_6}, \frac{\pi_3\pi_8}{\pi_4\pi_7}$, where $\hat{\pi}_3 = 0$ because of an empty cell. In this case, $\hat{\eta}_2 = \frac{\hat{\pi}_2\hat{\pi}_7}{\hat{\pi}_3\hat{\pi}_6} = \infty$ with $\hat{\pi}_3 = 0$.

#D	K	Seed	B	Iterations
8	3	113	1000	10000

#Parameters used in evaluation

(x[1] * x[6]) / (x[2] * x[5])

(x[2] * x[7]) / (x[3] * x[6])

(x[3] * x[8]) / (x[4] * x[7])

#Number of models to be compared

2

#Number of equality and inequality constraints per model

0 3

0 0

#Model 1

#Restriction matrix

1 -1 0
0 1 -1

#Constants

0 0

#Model 2 (unconstrained hypothesis)

#Restriction matrix

0 0 0

#Constants

0

Below $\hat{\eta}_2 = \infty$ because of the empty cell.

$MLEs

p1	p2	p3
0.9164	∞	0.0000

Thus, our software stops and gives the warning message: "Some of the eta parameters cannot be estimated because of empty cell(s). The user needs to rewrite the hypotheses under evaluation."

- In case of $\eta = \pi$, and therefore, $K = D$, since cell probabilities themselves are used to evaluate hypotheses containing linear restrictions on cell probabilities, specification of the η's in file input.txt has to be discarded. For example, the five lines which are crossed out below have to be removed from the file input.txt:

#D	K	Seed	B	Iterations
4	4	113	1000	10000

~~#Parameters used in evaluation~~

~~x[1]~~

~~x[2]~~

~~x[3]~~

~~x[4]~~

#Number of models to be compared

....

Otherwise, the program stops and gives the error message: "The hypotheses under evaluation contain linear restrictions on cell probabilities, which is implied by $K = D$. Specification of "#Parameters used in evaluation (eta)" has to be discarded."

- When the number of columns in restriction matrices does not agree with the number of η parameters, the program is stopped and another error message is given. For example, below the number of columns in the restriction matrix for Model 1 is 4, while it has to be 8.

#D K Seed B Iterations

8 8 113 1000 10000

#Number of models to be compared

2

#Number of equality and inequality constraints per model

0 3

1 1

#Model 1

#Restriction matrix

1 -1 0 0
1 0 -1 0
1 0 0 -1

#Constants

0 0 0

#Model 2

#Restriction matrix

....

Because this number is not equal to $K = 8$, the program stops and gives the error message: "The number of eta parameters (K) and the number of columns in restriction matrices are not equal to each other." Thus, although the hypotheses under evaluation have restrictions on less than 8 π's, one needs to specify the hypotheses in terms of all 8 π's.

- When $D = K$, the order-restricted MLEs are between zero and one and they sum to one for each hypothesis under evaluation. However, if the restrictions of hypotheses are specified with unreasonable constants, the order-restricted MLEs can be negative. For example, consider the comparison of hypotheses $H_1 : \pi_5 > 0.3, \pi_5 > \{\pi_6, \pi_7, \pi_8\}$ and the unconstrained hypothesis $H_u : \pi_1, \pi_2, ..., \pi_8$, where $\hat{\pi}_5 = 0$ because of an empty cell.

#D K Seed B Iterations

8 8 113 1000 10000

#Number of models to be compared

2

#Number of equality and inequality constraints per model

0 4

0 0

#Model 1

#Restriction matrix

0	0	0	0	1	-1	0	0
0	0	0	0	1	0	-1	0
0	0	0	0	1	0	0	-1
0	0	0	0	1	0	0	0

#Constants

0	0	0	0.3

#Model 2 (unconstrained hypothesis)

#Restriction matrix

0	0	0	0	0	0	0	0

#Constants

0

As can be seen, the order-restricted MLE of the last cell probability for hypothesis H_1 is negative.

....

$Restricted_MLEs

	p1	p2	p3	p4	p5	p6	p7	p8
H_1	0.5355	0.2308	0.0292	0.0147	0.3000	0.1491	0.0254	-0.2849
H_2	0.5355	0.2308	0.0292	0.0147	0.0000	0.1491	0.0254	0.0150

....

In such a case, the program does not stop but gives the following warning message: "WARNING: A negative value is encountered for the order-restricted MLEs. Please check the order-restricted MLEs and inspect the hypothesis for which the order-restricted MLE(s) are negative. The restrictions of the corresponding hypothesis might not be reasonable."

Chapter 4

Evaluation of directional hypotheses using contingency tables for replication in mice studies[1]

4.1 Introduction

The replication crisis has become a main concern in psychological science and neuroscience. Replication studies are used to contribute to an original study with the aim of confirming its results. However, many replication studies in psychology and animal models in neurology do not confirm the original findings. This lack of reproducibility has given rise to speculation that the results of most research in these areas are not reliable (loannidis, 2005; Button et al., 2013). In response to this growing concern, researchers started paying more attention to the reasons for non-replication (Pashler & Wagenmakers, 2012; Maxwell, Lau, & Howard, 2015) and appropriate analysis procedures to assess original findings (Nosek & Lakens, 2014; Anderson & Maxwell, 2016). Contextual differences between the original and replication studies, sometimes called hidden moderators, appear to be one of the major reasons for non-replication in psychology (Stroebe & Strack, 2014; Lindsay, 2015; Maxwell et al., 2015). For example, consider the study of van Bavel, Mende-Siedlecki, Brady, and Reinero (2016, p.6455) investigating 100 replication studies in psychology. This study concludes that replication success heavily depends on contextual differences like time, culture, or location. In addition, van Bavel et al. (2016, p.6455) state that this issue is applicable not only to psychology, but also rodent studies in neuroscience.

[1]This chapter will be submitted as Altınışık, Y., Joëls, M., Kanatsou, S., Naninck, E. F. G., Krugers, H., & Klein Entink, R. H. Evaluation of Directional Hypotheses Using Contingency Tables for Replication in Mice Studies.
Author contributions: YA and RHKE developed the study concept. All authors contributed to the study design. Data collection was performed by MJ, SK, EFGN, and HK and the writing, programming analysis, and interpretation was performed by YA under the supervision of RHKE. All authors read, commented, and approved the final manuscript.

Many rodent studies have been and are being conducted in different environmental settings such as food, bedding, and light and these settings play a vital role in replicating the initial results. Contextual differences make it difficult to fairly compare variables across studies, and consequently, cause non-replication. Establishing similar settings between studies can assist to replicate the original findings. However, this requires a considerable amount of effort in each study and can be very difficult if not impossible to achieve. Furthermore, many studies involve dependent variables that are measured on different scales for different groups of participants. Therefore, the problem of scale dependency has to be addressed when replicating studies.

In response to this challenge, one could consider to utilize correlation coefficients to investigate the magnitude and direction of the association between variables across studies, since correlation coefficients are scale independent. However, comparison of relations between the original study and replication studies by means of the correlation coefficients is highly controversial. Correlation coefficients should not be relied on either to assess replication success (Hubbard, 2015) or to quantify relationships between variables across studies (König, 2009). When testing correlation coefficients, statistical power is highly dependent on sample size. Therefore, contradictory results with regard to the original study and replication studies are often unavoidable, even when the only differences between the studies are sample size (Hubbard, 2015). In addition, inferences are focused on rejecting or not rejecting the null hypothesis H_0 : There is no relationship between variables. However, the results of the original study in most research are described by means of an alternative hypothesis H_a : There is a relationship between variables. Porte (2012, p.101) states that the method used for replication should be capable of giving reliable information with regard to not only the null hypothesis H_0 but also an alternative hypothesis H_a, which cannot be done using null hypothesis significance testing. Another problem arises if there is more than one correlation coefficient. Instead of testing whether one correlation replicates, now multiple tests have to be executed. Unfortunately, all these tests are not straightforward to summarize and interpret by researchers, because often there is not one conclusive summary.

In this paper we elaborate how to solve these problems when replicating studies. In the presence of continuous variables, we ensure that they are comparable across studies by discretizing them into categories. We compute odds ratios within these categories to ascertain the relationships between variables. Subsequently, hypotheses specifying directional relationships between these odds ratios are tested across studies to assess the agreement between the original and replication studies.

The outline of the paper is as follows. First, we introduce three data sets to illustrate the problem of scale dependency in evaluating hypotheses across studies. Second, we elaborate how to discretize variables into categories by means of selecting a cut-off value. Third, we demonstrate how to formulate directional hypotheses, containing the relationships between variables, using odds ratios. Fourth, we elaborate evaluation of directional hypotheses containing odds ratios. Fifth, two examples are given. The paper will be completed with a discussion.

4.2 The problem of scale dependency across replication studies

Replication studies are designed to investigate the same research question as the original study. However, this investigation is often carried out by means of observed variables that are measured on different scales, because of different environmental settings and/or groups of participants. These scale differences do not allow fair comparisons of relationships between variables across studies. To illustrate, consider studies investigating how early life stress influences the contextual memory of mice. The following three data sets are analyzed for which contextual memory of mice is measured in different environmental settings. The first data set ($N_1 = 14$) is based on a study performed at University of Amsterdam to test contextual memory for only male mice in terms of an object in context task. The second data set ($N_2 = 100$) is based on the study in Kanatsou et al. (2017, pp.2-3) for male mice and the study in Kanatsou et al. (2016, p.3) for female mice testing spatial memory in terms of an object location task. The third data set ($N_3 = 26$) is based on the study in Naninck et al. (2015, p.320) testing contextual memory for both male and female mice in terms of an object location task. In each study, the outcome, cognitive impairement, is measured in terms of the discrimination ratio (DR), representing the percentage of time spent on a novel object divided by the sum of time spent on the novel object and a familiar object. The performance of the mice in the tasks are measured for two groups, that is, the control and early life stress (ELS) groups, for each of the three data sets. The descriptives for the discrimination ratio with respect to the stress exposure groups and gender are given for each data set in Tables 4.1.1, 4.1.2, and 4.1.3, respectively. Note that DR is measured by means of different tasks and/or genders for the control and ELS groups. This may prevent a valid comparison of discrimination ratio across the three studies.

Four hypotheses might be of interest to researchers. The first hypothesis is based on a well known theory in neuroscience and states that early life stress causes cognitive impairment for mice in adulthood (Naninck et al., 2015, p.309). The second hypothesis states that the early life stress exposure impairs the cognitive ability for male mice but the impairment is not apparent for female mice (Naninck et al., 2015, p.325). One could also be interested in the evaluation of a third competing hypothesis which states that the early life stress exposure improves the cognitive ability for both male and female mice in adulthood. We describe these three hypotheses below with the unconstrained hypothesis, which covers the situations not specified by hypotheses H_1, H_2, H_3.

$$
\begin{aligned}
&H_1 : \text{Exposure to ELS impairs the cognitive ability for both male and female mice,}\\
&H_2 : \text{Exposure to ELS impairs the cognitive ability for male mice, but the impairement is not apparent for female mice,}\\
&H_3 : \text{Exposure to ELS improves the cognitive ability for both male and female mice,}\\
&H_u : \text{Something is going on but not necessarily hypotheses } H_1, H_2, \text{ and } H_3.
\end{aligned}
\tag{4.1}
$$

In this section, we introduced three data sets each containing the variable discrimination ratio, which is measured on different scales in each of these studies. In the following section,

Table 4.1.1: Descriptives for the discrimination ratio for the first data set

	Control	ELS
Min:	0.550	0.470
Mean:	0.618	0.520
Max:	0.680	0.650
SD:	0.051	0.063

Table 4.1.2: Descriptives for the discrimination ratio for the second data set

	Male		Female	
	Control	ELS	Control	ELS
Min:	0.270	0.250	0.300	0.150
Mean:	0.719	0.498	0.588	0.553
Max:	0.840	0.740	0.870	0.730
SD:	0.132	0.108	0.155	0.135

Table 4.1.3: Descriptives for the discrimination ratio for the third data set

	Male		Female	
	Control	ELS	Control	ELS
Min:	0.620	0.450	0.400	0.450
Mean:	0.692	0.517	0.644	0.574
Max:	0.810	0.560	0.750	0.700
SD:	0.076	0.038	0.119	0.087

we elaborate on how the scale dependency can be addressed by discretizing DR into two categories for each study.

4.3 Discretization

Many studies have been conducted to investigate discretization of continuous variables (Muhlenbach & Rakotomalala, 2005; Tsai, Lee, & Yang, 2008; Qureshi & Zighed, 2009). One of the main aims of these studies is to determine the most effective cut-off value, that is, the value that minimizes the loss of information that was captured in continuous variables before discretization. Our purpose is to obtain a discretized variable by means of using a predefined cut-off value supplied by literature and/or expert knowledge, which is scale independent and therefore, comparable across studies.

In some situations the choice of cut-off value is logical and therefore, easily justified. To illustrate, consider the outcome discrimination ratio, which is described in the previous

section. The discrimination ratio represents the ratio of the time the mouse spends on the novel object divided by the sum of time spent on the novel and familiar objects. If one wants to discretize the discrimination ratio as low and high to make it scale independent and comparable across studies, the cut-off value 0.5 would be a natural choice. Because if both familiar and novel objects are equally explored by a mouse, the percentage of time the mouse spends with the novel object becomes 50% (Akkerman et al., 2012, p.10). This implies that there is no discrimination between familiar and novel objects, which is known as the chance level of performance (Mumby, Gaskin, Glenn, Schramek, & Lehmann, 2002, p.52). Therefore, supported by literature (see Mumby et al., 2002, p.52; Kanatsou et al., 2015, p.4; Akkerman et al., 2012, p.10), the cut-off value 0.5 is chosen for the variable discrimination ratio. Then, mice with values of the discrimination ratio below and above the chance level 0.5 are considered to have low and high performance levels in the task. This discretized variable is independent of the task and the group and gender of the mice under consideration. Therefore, it is scale independent and comparable across studies. In Tables 4.2.1, 4.2.2, and 4.2.3, descriptives for the discretized DR are displayed.

Note that there may be situations where the choice of cut-off value is not straightforward. In such cases, it can be achieved by utilizing statistical methods or algorithms (Zighed, Rabaseda, & Rakotomalala, 1998; Boulle, 2003; Catlett, 2005). However, because the choice of cut-off value highly depends on the context of the study, none of these methods and algorithms guarantees that the chosen cut-off value is the most appropriate one to achieve scale independence. To avoid this, the choice of cut-off value should be justified carefully on the basis of an expert knowledge. In many cases expert knowledge is considered to be the best available source of information in discretizing variables into categories, when the cut-off value can be determined by the literature. An expert determines the most appropriate discretization based on the context of the study, and consequently, interprets the converted attributes (Muhlenbach & Rakotomalala, 2005).

Once the continuous variable discrimination ratio is dichotomized, the hypotheses H_1, H_2, H_3, and the unconstrained hypothesis H_u (see equation (4.1)) can be formulated in the context of 3-way (DR $\times$ group $\times$ gender) contingency table. This will be elaborated in the next section.

4.4 Hypotheses in terms of odds ratios

Evaluation of directional hypotheses offers three main advantages for replication studies compared to the classical null hypothesis testing. First, replication should not be considered as a black and white dichotomy such as interpreting statistically significant findings as success and nonsignificant findings as failure of replication (Anderson & Maxwell, 2016). Evaluation of directional hypotheses provides a straightforward interpretation of the results of replication studies without simply concluding that replication was successful or failed. This is achieved by means of using the GORICA elaborated in the next section, which provides the degree of support in the data for each hypothesis under evaluation. Second, null hypoth-

Table 4.2.1: The observed numbers of mice in each category of discretized DR for the first data set

		Group	
Gender	DR	Control	ELS
Male	Low	0	5
	High	6	3

Table 4.2.2: The observed numbers of mice in each category of discretized DR for the second data set

		Group	
Gender	DR	Control	ELS
Male	Low	1	10
	High	21	11
Female	Low	11	7
	High	18	21

Table 4.2.3: The observed numbers of mice in each category of discretized DR for the third data set

		Group	
Gender	DR	Control	ELS
Male	Low	0	1
	High	5	5
Female	Low	1	1
	High	6	7

esis testing is limited to two hypotheses (i.e., H_0 : There is no relationship between variables versus H_a: There is a relationship between variables). But evaluation of directional hypotheses provides the best hypothesis from a set of competing hypotheses (see equation (4.1)). Third, the null hypothesis often fails to address the expectations of researchers (Anderson & Maxwell, 2016). However, evaluation of directional hypotheses (see equation (4.1)) overcomes this problem in the sense that researchers can specify their expectations in terms of directional relationships.

Hypotheses H_1, H_2, and H_3 explicitly state how exposure to ELS and the performance of mice in the tasks are associated with each other. In that sense, they give information on the directional relationships between variables. These directional relationships can be formalized by means of odds ratios (OR). An odds ratio denotes the association between the levels of the outcome (i.e., low and high task performance) and early life stress exposure (i.e,

the control and ELS groups) for either male or female mice. As can be seen in Table 4.3, the direction of this association is interpreted relative to the number 1, that is, odds ratios smaller than 1 imply negative associations, odds ratios equal to 1 imply no associations, and odds ratios larger than 1 imply positive associations. For example, consider the hypothesis H_1 in equation (4.1) for which there is a negative association between the levels of the variables. Hypothesis H_1 implies that the mice in the early life stress group have lower performance scores compared to the mice in the control group. This is implied by the restrictions $a < b$ and $d < c$ in Table 4.3(A), for male mice. Note that the restrictions $\mathrm{a} < \mathrm{b}$ and $\mathrm{d} < \mathrm{c}$ together imply the restriction $\mathrm{ad} < \mathrm{bc}$, which leads to the odds ratio $\mathrm{OR}_{\mathrm{Male}} = \mathrm{ad}/\mathrm{bc} < 1$. In terms

Table 4.3: Three possible associations between the levels of the variables discrimination ratio and early life stress exposure in terms of an odds ratio

(A) $\mathrm{OR}_{\mathrm{Male}} < 1$				(B) $\mathrm{OR}_{\mathrm{Male}} = 1$				(C) $\mathrm{OR}_{\mathrm{Male}} > 1$			
	Control		ELS		Control		ELS		Control		ELS
Low	a	<	b	Low	a	=	b	Low	a	>	b
High	c	>	d	High	c	=	d	High	c	<	d

Note. $\mathrm{OR}_{\mathrm{Male}} = \mathrm{ad}/\mathrm{bc}$. For simplicity, only the odds ratio for male mice is displayed. The odds ratio for female mice can obtained analogously.
a: Number of male mice in control group with low DR.
b: Number of male mice in early life stress group with low DR.
c: Number of male mice in control group with high DR.
d: Number of male mice in early life stress group with high DR.
(A) $\mathrm{OR}_{\mathrm{Male}} < 1$: Exposure to stress associated with lower performance for male mice.
(B) $\mathrm{OR}_{\mathrm{Male}} = 1$: Exposure to stress does not affect the performance for male mice.
(C) $\mathrm{OR}_{\mathrm{Male}} > 1$: Exposure to stress associated with higher performance for male mice.

of odds ratios for male and female mice the hypotheses H_1, H_2, H_3, and the unconstrained hypothesis H_u in equation (4.1) are described below as:

$$\begin{aligned} H_1 &: \mathrm{OR}_{\mathrm{Male}} < 1, \mathrm{OR}_{\mathrm{Female}} < 1, \\ H_2 &: \mathrm{OR}_{\mathrm{Male}} < 1, \mathrm{OR}_{\mathrm{Female}} = 1, \\ H_3 &: \mathrm{OR}_{\mathrm{Male}} > 1, \mathrm{OR}_{\mathrm{Female}} > 1, \\ H_u &: \mathrm{OR}_{\mathrm{Male}}, \mathrm{OR}_{\mathrm{Female}}, \end{aligned} \tag{4.2}$$

where the operator "," in the unconstrained hypothesis H_u denotes the absence of information on the relationship between the odds ratios.

In this section, we illustrated how to specify directional hypotheses using odds ratios. Once directional hypotheses are created, they should be evaluated across studies to choose

the best hypothesis among them. To achieve this, we need a measure of the support in the data for each hypothesis under evaluation. This will be discussed in the following section.

4.5 Evaluation of directional hypotheses

In this paper, an AIC-based information criterion (AIC) (Akaike, 1973, 1974) the GORICA (Altınışık, Hessels, & Kuiper, unpublished), is utilized to evaluate directional hypotheses in the context of contingency tables. The information criterion GORICA measures the support in the data for each hypothesis under evaluation using the distance of hypothesis H_m to the true hypothesis. The GORICA evaluates hypotheses based on their misfit and complexity. The GORICA value for hypothesis H_m is defined as:

$$\text{GORICA}_\text{m} = \text{misfit}_\text{m} + \text{complexity}_\text{m}. \tag{4.3}$$

Misfit is a measure indicating to what extent the restrictions in hypothesis H_m disagree with the data. To illustrate, consider comparision of hypotheses $H_1 : \text{OR}_1 = \text{OR}_2$ and $H_2 : \text{OR}_1 < \text{OR}_2$, and suppose that population values of these odds ratios are estimated as $\widehat{\text{OR}}_1 = 1.5$ and $\widehat{\text{OR}}_2 = 2.0$ based on the data set at hand. Then, the equality restriction in hypothesis H_1 lacks the underlying relationship between the odds ratios, because it is not in agreement with the estimated values of them. The inequality restriction in hypothesis H_2 better identifies the relationship between the estimated values of the odds ratios. Therefore, the misfit for hypothesis H_2 is smaller than the misfit for hypothesis H_1. Note, however, that hypothesis H_2 is less specific than hypothesis H_1, because it gives less information on the relationship between the odds ratios than hypothesis H_1. Complexity penalizes less specific hypotheses by considering the number of distinct parameters used in the specification of the hypotheses under evaluation. The hypothesis with the smallest GORICA value has the smallest distance to the true hypothesis, and therefore, is selected to be the best one in a set of competing hypotheses. Once theories, the expectations of researchers, or the results of the original study are translated into directional hypotheses, these directional hypotheses should be evaluated for each replication study to assess replication success.

Evaluation of replication success is achieved using GORICA weights (Altınışık, Nederhof, et al., unpublished). The GORICA weigths denoted by w_m for hypothesis H_m can be used to quantify the relative importance of one hypothesis compared with the other hypotheses under evaluation. GORICA weights can take on values between the numbers 0 and 1 and they sum to 1. For example, consider hypotheses H_1, H_2, and the unconstrained hypothesis H_u, that is, something is going on but not necessarily hypotheses H_1 and H_2 with GORICA weights 0.75, 0.05, and 0.20. Then, hypothesis H_1 is much better than hypothesis H_2, that is, $0.75/0.05 = 15$ times more supported by the data compared to H_2 and hypothesis H_1 is a strong hypothesis, that is, H_1 is $0.75/0.20 = 3.75$ times better than H_u. However, hypothesis H_2 is a weak hypothesis, because the unconstrained hypothesis H_u is $0.20/0.05 = 4$ times better than H_2. Therefore, it is concluded that hypothesis H_1 is the best hypothesis among the three hypotheses.

We elaborate on two possible scenarios that researchers often encounter when executing replication studies. These two scenarios are based on whether researchers aim to verify information obtained from a theory, or to confirm the results of original study. Evaluating directional hypotheses is suitable for this framework, because theories, the expectations of researchers, and the results of the original study can easily be translated into directional hypotheses. These two replication scenarios are displayed below:

Scenario 1

1. Translate theories or the expectations of researchers into directional hypotheses (see equations (4.1) and (4.2)) and discretize continuous variables into categories.

Repeat steps 2-3 below for each replication data set.

2. Calculate the values of GORICA and GORICA weights for the directional hypotheses specified in Step 1.
3. Interpret the results by comparing the values of GORICA and GORICA weights for each hypothesis under evaluation, that is, the hypothesis with the smallest GORICA value and largest GORICA weight is the best hypothesis in the set of competing hypotheses.
4. Summarize the findings by taking the average GORICA weight across replication studies for each hypothesis under evaluation.

Scenario 2

1. Collect the original data and summarize the resulting odds ratios.
2. Create one main directional hypothesis based on the results obtained in the original study and one or more competing hypotheses and categorize continuous variables.

Repeat steps 3-4 below for each replication data set.

3. Calculate the values of GORICA and GORICA weights for the directional hypotheses specified in Step 2.
4. Interpret the results by comparing the values of GORICA and GORICA weights for each hypothesis under evaluation, that is, the hypothesis with the smallest GORICA value and largest GORICA weight is the best hypothesis in the set of competing hypotheses.
5. Summarize the findings by taking the average GORICA weight across replication studies for each hypothesis under evaluation.

In the next section, both replication scenarios will be illustrated in the context of 2-way (DR $\times$ group) and 3-way (DR $\times$ group $\times$ gender) contingency tables, respectively.

4.6 Examples

In this section, evaluation of directional hypotheses for replication studies are illustrated in terms of two types of replication scenarios. The first example is used to illustrate the first replication scenario in which researchers aim to verify information obtained from a theory. The second example involves the second replication scenario in which the goal is to confirm the results of the original study.

4.6.1 Example 1

Researchers should directly evaluate directional hypotheses if there is an available theory (or a strong belief) about the relationships between variables (van de Schoot & Strohmeier, 2011). To illustrate, consider again the theory in (Naninck et al., 2015, p.309), which claims that exposure to early life stress reduces the performance of the mice in the task, see the hypothesis H_1 in equation (4.1). As is shown in Table 4.3, this theory can be translated into a directional hypothesis for male mice as:

$$H_1 : \mathrm{OR}_{\mathrm{Male}} < 1.$$

One of the advantages of evaluating directional hypotheses is that researchers are not limited to evaluate only two hypotheses (i.e., H_0 versus H_a). For example, although not stated in the theory, one could also be interested in the evaluation of two competing hypotheses H_2 and H_3, which are described below in a similar manner as hypothesis H_1:

$$\begin{aligned} H_1 &: \mathrm{OR}_{\mathrm{Male}} < 1, \\ H_2 &: \mathrm{OR}_{\mathrm{Male}} = 1, \\ H_3 &: \mathrm{OR}_{\mathrm{Male}} > 1. \end{aligned} \tag{4.4}$$

Hypothesis H_2 states that exposure to early life stress does not affect the performance of male mice in the task. Hypothesis H_3 specifies that exposure to early life stress increases the performance of male mice in the task. Then, hypotheses H_1, H_2, and H_3 can be evaluated from one study to another to strenghten (or weaken) our confidence in hypothesis H_1 or to build up a new belief in hypothesis H_2 or hypothesis H_3. Note that the unconstrained hypothesis H_u is not needed for this set of hypotheses. Because already all possible values for the odds ratio $\mathrm{OR}_{\mathrm{Male}}$ are taken into account by hypotheses H_1, H_2, and H_3, that is smaller than one, equal to one, and bigger than one.

Hypotheses H_1, H_2, and H_3 are evaluated for male mice by means of the three data sets introduced earlier. In Tables 4.4.1, 4.4.2, and 4.4.3, the values of the misfit and complexity parts, GORICA, and GORICA weights are displayed for the three hypotheses H_1, H_2, and H_3. The information obtained from the theory is validated by the first two data sets. Because the GORICA value for hypothesis H_1 is the smallest one among the three hypotheses for these data sets. This means that hypothesis H_1 has the smallest distance to the true hypothesis relative to the other hypotheses under evaluation based on these data sets. However, the

Table 4.4.1: The misfit and complexity parts, $GORICA_m$, and the weights w_m for hypotheses H_1, H_2, and H_3 for the first data set

H_m	misfit_m	$\text{complexity}_\text{m}$	$GORICA_m$	w_m
H_1	-4.312	1.003	-3.309	0.973
H_2	4.809	0.000	4.809	0.017
H_3	4.809	1.006	5.815	0.010

Table 4.4.2: The misfit and complexity parts, $GORICA_m$, and the weights w_m for hypotheses H_1, H_2, and H_3 for the second data set

H_m	misfit_m	$\text{complexity}_\text{m}$	$GORICA_m$	w_m
H_1	-5.167	0.998	-4.169	0.993
H_2	6.740	0.000	6.740	0.004
H_3	6.740	0.997	7.737	0.003

Table 4.4.3: The misfit and complexity parts, $GORICA_m$, and the weights w_m for hypotheses H_1, H_2, and H_3 for the third data set

H_m	misfit_m	$\text{complexity}_\text{m}$	$GORICA_m$	w_m
H_1	-4.846	0.998	-3.848	0.380
H_2	-3.881	0.000	-3.881	0.386
H_3	-3.881	1.001	-2.880	0.234

support for this theory is not quite valid for the third data set. Because, although they are close to each other, not hypothesis H_1 but instead hypothesis H_2 has the smallest GORICA value among the three hypotheses for this data set. This implies that the effect of early life stress exposure on the task performance of male mice is not clear-cut for the third replication study.

Looking at the GORICA weights for the first two data sets, it can be seen that hypothesis H_1 is a very strong hypothesis. That is, for the first data set hypothesis H_1 has $0.973/0.017 \approx 57.2$ times more support than hypothesis H_2 and $0.973/0.010 \approx 97.3$ times more support than hypothesis H_3. Similarly, for the second data set, hypothesis H_1 is $0.993/0.004 \approx 248.3$ times better than hypothesis H_2 and $0.993/0.003 \approx 331$ times better than hypothesis H_3. Note that hypotheses H_2 and H_3 are not supported by these data sets at all, because the GORICA weights for these hypotheses are much smaller than the GORICA weights for hypothesis H_1. However, there is no compeling evidence for hypothesis H_1 in the third data set. In this case, hypothesis H_1 is only $0.380/0.234 \approx 1.62$ times better than hypothesis H_3 and it is not better than hypothesis H_2, that is, hypothesis H_2 $0.386/0.380 \approx 1.02$ times better than hypothesis H_1.

Two out of the three data sets are in favor of hypothesis H_1 when compared to hypotheses H_2 and H_3. The average GORICA weights for hypotheses H_1, H_2, and H_3 across the three data sets are 0.782, 0.136, and 0.082, respectively. Therefore, the overall conclusion is that the joined evidence in the three studies strongly supports hypothesis H_1.

4.6.2 Example 2

Researchers sometimes use replication studies not to investigate the information provided by a theory, but to assess the validity of their initial results that are obtained by analyzing an original data set. In this case, instead of evaluating directional hypotheses directly based on the theory, researchers should first create hypotheses based on the results of their original study. Once hypotheses are created, the same hypotheses should be evaluated by means of one or more replication studies. To illustrate, consider again the results obtained in (Naninck et al., 2015, p.325), which states that early life stress impairs the cognitive ability for male mice, but it does not influence this ability for female mice. We evaluated the hypotheses H_1, H_2, H_3, and the unconstrained hypothesis H_u in equation (4.2) to assess the validity of these results across studies, where hypothesis H_2 represents the results obtained in the original study.

These four hypotheses are evaluated for the second ($N_2 = 100$) and third ($N_3 = 26$) replication data sets introduced earlier. In Tables 4.5.1 and 4.5.2, the values of the misfit and complexity parts, GORICA values, and GORICA weights are displayed for the four hypotheses H_1, H_2, and H_3 and the unconstrained hypothesis H_u. Hypothesis H_2 has the smallest GORICA value among the three hypotheses and the unconstrained hypothesis H_u for both data sets. This implies that hypothesis H_2 is better than the other hypotheses under evaluation based on the two replication data sets.

Based on the GORICA weights in Table 4.5.1, hypothesis H_1 has $0.296/0.187 \approx 1.583$ and hypothesis H_2 has $0.506/0.187 \approx 2.706$ times more support than the unconstrained hypothesis H_u. Therefore, it is concluded that these hypotheses are supported by the data. Note that hypothesis H_3 is considered to be too weak, that is, it does not receive any support from the data set. Hypothesis H_2 is better than hypothesis H_1, that is, hypothesis H_2 has $0.506/0.296 \approx 1.709$ times more support than hypothesis H_1. Therefore, it is concluded that although H_2 is the best among the three hypotheses, hypothesis H_1 should be considered along with hypothesis H_1. Based on the GORICA weights in Table 4.5.2, all the three hypotheses receive considerable amount of support from the data. Hypothesis H_1 has $0.268/0.010 \approx 26.800$, hypothesis H_2 has $0.448/0.010 \approx 44.800$, and hypothesis H_3 has $0.185/0.010 \approx 18.500$ times more support than the unconstrained hypothesis H_u. Hypothesis H_2 has $0.448/0.268 \approx 1.672$ and $0.448/0.185 \approx 2.422$ times more support than hypotheses H_1 and H_3, respectively. Therefore, we conclude that hypothesis H_2 is again the best hypothesis in the set of the three hypotheses, but that hypothesis H_1 cannot be discarded.

Both data sets are in favor of hypothesis H_2 when compared to hypotheses H_1, H_2, H_3, and the unconstrained hypothesis H_u. The average GORICA weights for hypotheses H_1, H_2, H_3, and the unconstrained hypothesis H_u are 0.282, 0.477, 0.098, and 0.143, respectively.

Table 4.5.1: The misfit and complexity parts, $GORICA_m$, and the weights w_m for hypotheses H_1, H_2, H_3, and H_u for the second data set

H_m	misfit_m	$\text{complexity}_\text{m}$	$GORICA_m$	w_m
H_1	-14.442	1.037	-12.370	0.296
H_2	-14.442	0.499	-13.443	0.506
H_3	-7.810	1.026	-5.759	0.011
H_u	-15.450	2.000	-11.450	0.187

Table 4.5.2: The misfit and complexity parts, $GORICA_m$, and the weights w_m for hypotheses H_1, H_2, H_3, and H_u for the third data set

H_m	misfit_m	$\text{complexity}_\text{m}$	$GORICA_m$	w_m
H_1	-14.514	1.009	-12.496	0.268
H_2	-14.514	0.496	-13.523	0.448
H_3	-13.762	1.006	-11.750	0.185
H_u	-14.524	2.000	-10.523	0.099

Therefore, we conclude that the joint evidence in both studies supports hypothesis H_2, but hypothesis H_1 cannot yet be completely excluded.

4.7 Discussion

Researchers use replication studies to evaluate the information obtained from a theory or to confirm the results of their original study. However, contextual differences often complicate the comparison of the variables between the original and replication studies, and consequently, cause non-replication. To remedy this, we provided a coherent strategy for testing replication. In cases where variables are measured on a continuous scale, these variables should be discretized into categories on the basis of expert knowledge and/or literature search. In addition, we consider that the traditional null hypothesis testing for replication studies in the context of contingency tables is not flexible enough to address the directional relationships between the odds ratios upon which hypotheses are based. Hypothesis evaluation with the GORICA not only appropriately addresses the interests of researchers but also allows them to investigate replication success of their studies. Two concrete examples were given to illustrate the applicability of the strategy. The GORICA is implemented in the R script GoricaCont.R, which is available in research archive sent with this dissertation. The interested reader is referred to Altınışık, Hessels, and Kuiper (unpublished), who explain how to execute the R code in GoricaCont.R to evaluate hypotheses in the context of contingency tables.

Chapter 5

The GORICA applied: An AIC-based information criterion for evaluating informative hypotheses[1]

5.1 Introduction

Informative hypotheses are formal representations of theories, the expectations of researchers, and the results of previous studies. Translation of theories, expectations, and results into informative hypotheses is done by using equality and/or inequality restrictions. To illustrate the specification of informative hypotheses, consider the academic awards data from UCLA Academic Technology Services (data available on the Web at `https://stats.idre.ucla.edu/sas/dae/poisson-regression`). In the study, the academic performance of students at a high school ($N = 200$) is measured in an academic year. The dependent variable the number of awards earned by students $\mathrm{A} \in \{0, 1, 2, 3, 4, 5, 6\}$ is predicted by academic program $\mathrm{P} \in \{1 = \mathrm{General},\ 2 = \mathrm{Academic},\ 3 = \mathrm{Vocational}\}$ which is a factor variable with three levels and math scores MS which is a standardized continuous variable with the values in the interval [-2.096, 2.386], and the interaction between these variables. The poisson regression model indicating this relationship is:

$$\begin{aligned} f(\hat{A}_i) = \beta_0 + \beta_1 \mathrm{P}_{2i} + \beta_2 \mathrm{P}_{3i} + \beta_3 \mathrm{MS}_i + \\ \beta_4 \mathrm{P}_{2i}\mathrm{MS}_i + \beta_5 \mathrm{P}_{3i}\mathrm{MS}_i, \end{aligned} \tag{5.1}$$

[1] This chapter will be submitted as Altınışık, Y. & Kuiper, R.M. The GORICA applied: An AIC-based information criterion for evaluating informative hypotheses.
Author contributions: YA and RMK developed the study concept. Both authors contributed to the study design. The writing, programming analysis, and interpretation was performed by YA under the supervision of RMK. Both authors read,, commented, and approved the final manuscript.

where $f(\hat{A}) = log(\mathrm{E}(\mathrm{A}|\mathrm{P}_2, \mathrm{P}_3, \mathrm{MS}))$ is the log link function for which $\hat{\mathrm{A}} = \mathrm{E}(\mathrm{A}|\ \mathrm{P}_2, \mathrm{P}_3, \mathrm{MS})$ denotes the expected number of awards earned by the students given the academic programs and math scores of the students, β_0 denotes the intercept, and $\beta_1, \beta_2, \beta_3, \beta_4$, and β_5 are the standardized regression parameters for the five predictors for $i = 1, 2, ..., 200$. These standardized parameters are used to quantify the strength of the relationship between the response variable and predictor variables on the same scale. The variables $\mathrm{P}_{2\mathrm{i}} \in \{0, 1\}$ and $\mathrm{P}_{3\mathrm{i}} \in \{0, 1\}$ are two dummies, where $\mathrm{P}_{2\mathrm{i}}$ = 0 and $\mathrm{P}_{3\mathrm{i}}$ = 0; $\mathrm{P}_{2\mathrm{i}}$ = 1 and $\mathrm{P}_{3\mathrm{i}}$ = 0; and $\mathrm{P}_{2\mathrm{i}}$ = 0 and $\mathrm{P}_{3\mathrm{i}}$ = 1 indicate that the ith student is in the first, second, and third academic program (i.e., $\mathrm{P} = 1, 2$, and 3), respectively. Based on the values of the two dummies, the model in equation (5.1) can be converted to:

$$f(\hat{A}_i) = \begin{cases} \beta_0 + \beta_3 MS_i & \text{if } P = 1 \\ (\beta_0 + \beta_1) + (\beta_3 + \beta_4) MS_i & \text{if } P = 2 \\ (\beta_0 + \beta_2) + (\beta_3 + \beta_5) MS_i & \text{if } P = 3. \end{cases} \tag{5.2}$$

Researchers might be interested in the evaluation of the following three hypotheses together with the unconstrained hypothesis H_u:

H_1 : Math scores of students do not influence the number of earned awards in any educational program,

H_2 : Math scores of students in the general program ($\mathrm{P} = 1$) have the smallest effect on the number of earned awards followed by the effect of math scores of students in the vocational program ($\mathrm{P} = 3$), which in turn is followed by the effect in the academic program ($\mathrm{P} = 2$),

H_3 : The difference between the academic and vocational programs ($\mathrm{P} = 2$ vs. $\mathrm{P} = 3$) is more apparent than the difference between the academic and general programs ($\mathrm{P} = 2$ vs. $\mathrm{P} = 1$) in terms of the effect of math scores on the number of earned awards,

H_u : There are no restrictions on the association between the math scores of students and the number of earned awards. That is, there may be an association between them but not necessarily the ones suggested by hypotheses H_1, H_2, and H_3.

Using the model parameters in equation (5.2), these hypotheses can be translated into the following informative hypotheses:

$$\begin{aligned} H_1 &: \beta_3 = (\beta_3 + \beta_4) = (\beta_3 + \beta_5) = 0, \\ H_2 &: (\beta_3 + \beta_4) > (\beta_3 + \beta_5) > \beta_3, \\ H_3 &: (\beta_3 + \beta_4) - (\beta_3 + \beta_5) > (\beta_3 + \beta_4) - \beta_3, \\ H_u &: \beta_3, \beta_4, \beta_5, \end{aligned} \tag{5.3}$$

where the operator "=" indicates equality restriction, the operators larger than ">" and smaller than "<" indicate inequality restrictions, and the operator "," denotes the absence of a restriction. Note that the unconstrained hypothesis H_u is included in the set of hypotheses as a safeguard against choosing a weak hypothesis as the best hypothesis among the others. This will be elaborated upon later in this paper.

The AIC (Akaike, 1973, 1974) can be used to evaluate informative hypotheses containing only equality restrictions on (standardized) model parameters. For example, the AIC can be utilized to evaluate hypothesis H_1 and the unconstrained hypothesis H_u, which contain only equality restrictions and no restriction on the standardized regression parameters, respectively. The generalized order-restricted information criterion (GORIC) (Kuiper et al., 2011, 2012) is an AIC-based information criterion that enables researchers to evaluate hypotheses containing equality and/or inequality restrictions. For example, the GORIC can also be used to evaluate hypotheses H_2 and H_3. However, because the GORIC can only be used for (multivariate) normal linear models, it cannot be used to evaluate hypotheses for the model in equation (5.1). Evaluation of the informative hypotheses with respect to this model can be achieved using the generalized version of the GORIC (GORICA) (Altınışık, Nederhof, et al., unpublished; Altınışık, Hessels, & Kuiper, unpublished). The GORICA is applicable to generalized linear models (GLMs) (McCullagh & Nelder, 1989), generalized linear mixed models (GLMMs) (McCullogh & Searle, 2001), structural equation models (SEMs) (Bollen, 1989), and contingency tables (Agresti, 2007). The GORICA evaluation of informative hypotheses can be performed in case the hypotheses under evaluation contain linear restrictions on model parameters for GLMs, GLMMs, and SEMs and linear and non-linear restrictions on cell probabilities for contingency tables.

To evaluate hypotheses using the GORICA, one needs to estimate (standardized) model parameters and their covariance matrix. Three methods that can be used to estimate these parameters and their covariance matrix are maximum likelihood estimation (Fisher, 1922), nonparametric bootstrapping (Efron & Tibshirani, 1993), and gibbs sampling (D. Spiegelhalter, 1994). In this chapter, we apply the GORICA for evaluating informative hypotheses in the context of GLMs, GLMMs, SEMs, and contingency tables and illustrate the use of these three estimation methods.

The outlline of the paper is as follows. First, the general class of informative hypotheses will be elaborated. Second, the GORICA that can be applied to this class of informative hypotheses will be presented. Third, we elaborate on the three methods for estimating (standardized) model parameters and their covariance matrix. Fourth, four different examples are used to illustrate the GORICA for GLMs, GLMMs, SEMs, and contingency tables. The paper will end with a brief discussion. The Supplementary material contains R code (and additional explanation) to duplicate our results in the paper for each of the four examples.

5.2 General class of informative hypotheses

Consider the poisson regression model in equation (5.1). Some of the parameters of this model are used to formulate the four hypotheses in equation (5.3), namely β_3, β_4, and β_5. These parameters are called structural parameters, which are primary interests of researchers. The other parameters that are not used to specify informative hypotheses, that is, β_0, β_1, and β_2, are called nuisance parameters. These parameters are important for model fit, but they do not represent researchers' primary expectations. Hence, given the hypotheses in

equation (5.3), the parameter vector for the model in equation (5.1) can be described as $\{\theta, \xi\} = \{\beta_0, \beta_1, \beta_2, \beta_3, \beta_4, \beta_5\}$, where $\theta = (\beta_3, \beta_4, \beta_5)$ denotes the structural parameters that are used to formulate the four hypotheses and $\xi = (\beta_0, \beta_1, \beta_2)$ denotes the nuisance parameters, which are not used to formulate the hypotheses.

Since the hypotheses under evaluation do not contain nuisance parameters, the ξ's, the class of informative hypotheses is described only by using the structural parameters, the θ's. The general class of informative hypotheses consists of linear restrictions on the structural parameters and is of the form:

$$H_m : S_m\theta = s_m, R_m\theta > r_m, \tag{5.4}$$

where $S_m \in \mathbb{R}^{q_s \times K}$ and $R_m \in \mathbb{R}^{q_r \times K}$ indicate the q_s equality and q_r inequality restrictions in H_m respectively, θ denotes the K-vector of (standardized) structural model parameters, and s_m and r_m are the q_s-vector and q_r-vector of the constants in hypothesis H_m. For example, the hypothesis $H_1 : \beta_3 = (\beta_3 + \beta_4) = (\beta_3 + \beta_5) = 0$ in equation (5.3) containing only equality restrictions on the standardized model parameters can be described using $\theta = (\beta_3, \beta_4, \beta_5)^T$ and

$$S_1 = \begin{pmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 1 & 0 & 1 \end{pmatrix}, s_1 = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix},$$

with $K = q_s = 3$. As another example, the hypothesis $H_2 : (\beta_3 + \beta_4) > (\beta_3 + \beta_5) > \beta_3$ containing only inequality restrictions on these parameters can be specified using

$$R_2 = \begin{pmatrix} 0 & 1 & -1 \\ 0 & 0 & 1 \end{pmatrix}, r_2 = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix},$$

with $K = 3$ and $q_r = 2$. Note that the GORICA also enables researchers to evaluate hypotheses containing both equality and inequality restrictions, as can be seen in the examples section.

For contingency tables, specification of informative hypotheses consisting of linear restrictions on cell probabilities is analogous to the specification in equation (5.4), where the only difference is that the θ's are replaced by the cell probabilities π's, namely $\theta = \pi$. Note that the cell probabilities are linearly dependent on each other, namely they sum to one. This means that any cell probability in contingeny tables can be described by one minus the sum of the other cell probabilities. This implies that all the cell probabilities are used to formulate the hypotheses under evaluation. Therefore, the vector of the nuisance parameters consisting of the parameters that are not used in formulating hypotheses is empty for contingency tables. However, informative hypotheses in the context of contingency tables do not necessarily specify linear restrictions on cell probabilities. Readers are referred to Altınışık, Hessels, and Kuiper (unpublished), in which informative hypotheses are evaluated for five classes (and their subclasses) of restrictions commonly used in the context of contingency tables. These hypotheses often contain non-linear restrictions on cell probabilities, but they are

linear in terms of functions of cell probabilities, like odds ratios. For example, consider the 2×2 contingency table for which the hypothesis of interest is $H_1 : \frac{\pi_{11}\pi_{22}}{\pi_{12}\pi_{21}} > 1$, specifying a positive association between the levels of two variables by means of using an odds ratio. This hypothesis contains a non-linear restriction on the four cell probabilities, but a linear restriction on the odds ratio which is a function of the four cell probabilities. If we define θ as $\frac{\pi_{11}\pi_{22}}{\pi_{12}\pi_{21}}$, the hypothesis $H_1 : \theta > 1$ is of the linear form as specified in equation (5.4). Thus, the GORICA can also be used to evaluate hypotheses containing non-linear restrictions on cell probabilities (i.e., π's), where the hypotheses are specified in terms of functions of cell probabilities (i.e., $\theta = g(\pi)$) such that they are linear in terms of θ as in equation (5.4). The GORICA evaluation of informative hypotheses containing linear restrictions on (functions of) cell probabilities can be performed with and without empty cell(s). In the examples, we will apply the GORICA to a $2 \times 2 \times 2$ contingency table containing an empty cell.

Once the hypotheses of interest are specified like in equation (5.4), these hypotheses can be evaluated using the GORICA, which will be elaborated in the next section.

5.3 The GORICA

Like the AIC (Akaike, 1973, 1974), the GORICA comprises misfit and complexity parts. For hypothesis H_m, the GORICA is defined as:

$$\text{GORICA}_m = \text{misfit}_\text{m} + \text{complexity}_\text{m}, \tag{5.5}$$

for $m = 1, 2, ..., M$, where M is the number of hypotheses in the set of hypotheses under evaluation. Misfit measures the disagreement with the restrictions of the hypothesis under evaluation and the data set at hand. To illustrate, suppose that $\hat{\theta} = \{\hat{\theta}_1, \hat{\theta}_2, \hat{\theta}_3\} = \{0.40, 0.30, 0.20\}$ denotes the maximum likelihood estimates (MLEs) of three standardized structural parameters, which are used for the comparison of hypotheses $H_1 : \theta_1 = \theta_2, \theta_3$ and $H_2 : \theta_1 > \theta_2, \theta_3$. In this case, the restriction $\theta_1 > \theta_2$ in hypothesis H_2 is in accordance with the MLEs. But the restriction $\theta_1 = \theta_2$ in hypothesis H_1 is not supported by the data set, namely $\hat{\theta}_1 = 0.40$ is not equal to $\hat{\theta}_2 = 0.30$. Since hypothesis H_2 better fits the data set, it has smaller misfit when compared to hypothesis H_1.[1] However, hypothesis H_1 is a more specific hypothesis when compared to hypothesis H_2. Complexity takes this into account by penalizing less specific hypotheses in the set of hypotheses under evaluation. Complexity can be seen as two times the number of distinct parameters used in the specification of the hypothesis.[2] For example, the complexity for hypothesis H_1 is 4, because the number of distinct parameters after taking into account the equality restriction in hypothesis H_1 is 2. That

[1] Hypothesis H_2 fits the data perfectly. Therefore, one could expect that the misfit is zero in such a case. However, because the misfit contains a negative constant, it is negative (and not zero) in case of a perfect fit. This is thus also the case for the unconstrained hypothesis H_u, which always has a perfect fit. Thus in the examples, one will see a negative misfit for the unconstrained hypothesis H_u and the other hypotheses that have perfect fit to the data.

[2] Complexity is calculated via simulation, see Kuiper et al. (2011, 2012) for more details. We use $\texttt{iter} = 100,000$ iterations in the simulation.

is, the parameters θ_1 and θ_2 are equal to each other (resulting into one distinct parameter) and there is not any restriction on the parameter θ_3 (resulting into the other distinct parameter). Similarly, since the number of distinct parameters after taking into account the inequality restriction in hypothesis H_2 is 2.5, the complexity for this hypothesis is 5. Namely, there is not any restriction on parameter θ_3 (resulting into one distinct parameter) and the number of θ parameters in the restriction $\theta_1 > \theta_2$ is 1.5. This is because in case the restriction in $\theta_1 > \theta_2$ is in agreement with the data, we have two distinct parameters; and when it is not, θ_1 and θ_2 will be set equal to each other leading to one distinct parameter. Both situations happen 50% of the time, when sampling from a null hypothesis where the θ's are equal to each other. Then, the expected number of distinct parameters in the restriction $\theta_1 > \theta_2$ is $0.5 \times 2 + 0.5 \times 1 = 1.5$. Thus, hypothesis H_1 has smaller complexity, stated otherwise, penalization, when compared to hypothesis H_2. Note that misfit and complexity comprise (standardized) structural parameters/estimates directly and the nuisance ones indirectly, because these parameters/estimates depend on each other.

The hypothesis with the smallest GORICA value in a set of hypotheses is considered to have the smallest distance to the true hypothesis, when compared to the other hypotheses under evaluation. Therefore, this hypothesis is selected as the best hypothesis among the others based on the data set at hand. Note that the unconstrained hypothesis H_u is the hypothesis that fits the data set perfectly, and therefore, has the smallest misfit. But also the unconstrained hypothesis H_u has the largest complexity (penalization) in the set of hypotheses, because it does not give any information in terms of the relations between model parameters. Therefore, the unconstrained hypothesis H_u is included in the set of hypotheses as a safeguard for not choosing a weak hypothesis as the best hypothesis among the others.

For information criteria, and thus also for the GORICA, decision in terms of which hypothesis is better supported by the data than another one is made based on the differences between the values. Unfortunately, the GORICA values themselves do not quantify the support in the data for the hypotheses under evaluation. For doing this, it is necessary to translate the GORICA values into the GORICA weights (w_m), comparable to the AIC weights (Burnham & Anderson, 2002, p.75). The relative support in the data for each hypothesis under evaluation can be determined by looking at pairwise ratios between these weights. For example, consider hypotheses H_1, H_2, and the unconstrained hypothesis H_u with GORICA weights of 0.60, 0.10, and 0.30, respectively. Then, hypothesis H_2 is a weak hypothesis, because the unconstrained hypothesis H_u is $0.30/0.10 = 3$ times more supported by the data when compared to hypothesis H_2 (i.e, $w_2/w_u = 0.10/0.30 < 1$). Hypothesis H_1 is not a weak hypothesis, because it is supported by the data $w_1/w_u = 0.60/0.30 = 2$ (>1) times better than the unconstrained hypothesis H_u. Since hypothesis H_1 is $w_1/w_2 = 0.60/0.10 = 6$ times more supported than hypothesis H_2 and it is not a weak hypothesis, hypothesis H_1 is the best hypothesis in the set of the three hypotheses.

5.3.1 The R function gorica

Misfit depends on the order-restricted estimates of (standardized) structural parameters (i.e., $\tilde{\theta}_m$), which are the estimates that are in accordance with the restrictions of the hypothesis under evaluation. These estimates are utilized to achieve the restricted minumum of the misfit (by maximizing a log likelihood function) for hypothesis H_m. The order-restricted estimates for the hypothesis under consideration are obtained using *ormle* function as follows:

Hm$restricted <- ormle(est = strest, covmtrx = strcovmtrx, const = constr, nec = nec, rhs = rhs).

Here, "str" in *str*est and *str*covmtrx denotes the estimates of (standardized) structural parameters (i.e., $\hat{\theta}$) and their covariance matrix (denoted by $\hat{\Sigma}_{\hat{\theta}}$), respectively. Constr and rhs represent the restriction matrices S_m and/or R_m and the constants s_m and/or r_m in equation (5.4), respectively. *ormle* considers equality constraints first, where nec is the number of these equality restrictions in hypothesis H_m (i.e., q_s). An overview of these relations are given in Table 5.1. To illustrate how the arguments in *ormle* and the order-restricted estimates look like, consider the poisson regression example given in the introduction. The MLEs of the three standardized structural parameters are $\hat{\theta} = (\hat{\theta}_1, \hat{\theta}_2, \hat{\theta}_3) = (\hat{\beta}_3, \hat{\beta}_4, \hat{\beta}_5) = (0.412, 0.266, 0.215)$ and their covariance matrix is:

$$\hat{\Sigma}_{\hat{\theta}} = \begin{pmatrix} \mathrm{Var}(\hat{\theta}_1) & & \\ \mathrm{Cov}(\hat{\theta}_1, \hat{\theta}_2) & \mathrm{Var}(\hat{\theta}_2) & \\ \mathrm{Cov}(\hat{\theta}_1, \hat{\theta}_3) & \mathrm{Cov}(\hat{\theta}_2, \hat{\theta}_3) & \mathrm{Var}(\hat{\theta}_3) \end{pmatrix}$$

$$= \begin{pmatrix} 0.196 & & \\ -0.196 & 0.208 & \\ -0.196 & 0.196 & 0.258 \end{pmatrix}.$$

Consider the hypotheses $H_1 : \beta_3 = (\beta_3 + \beta_4) = (\beta_3 + \beta_5) = 0$ and $H_2 : (\beta_3 + \beta_4) > (\beta_3 + \beta_5) > \beta_3$ in (3) for which the restriction matrices S_1 and R_2 (constr) and the vectors of constants s_1 and r_2 (rhs) are given in the previous section. The numbers of equality restrictions in these hypotheses are $q_s = 3$ and $q_s = 0$, respectively (i.e, nec). These are used as inputs in the *ormle* function, which then renders the following order-restricted MLEs, $\tilde{\theta}_m = (\tilde{\theta}_{m1}, \tilde{\theta}_{m2}, \tilde{\theta}_{m3})$, that are evidently in agreement with the restrictions of the hypotheses: H1$restricted $= (0, 0, 0)$ and H2$restricted $= (0.412, 0.266, 0.215)$. Note that because the restrictions in hypothesis H_2 are in accordance with the MLEs, the order-restricted MLEs for this hypothesis are equal to the MLEs (i.e., H2$restrictedest = H2$est). This situation always occurs for the unconstrained hypothesis, here, $H_u : \beta_3, \beta_4, \beta_5$ for which the restriction matrix, the constant, and the order-restricted MLEs are $R_u = (0, 0, 0)$, $r_u = 0$, and Hu$restricted = Hu$est = (0.412, 0.266, 0.215), respectively.

Table 5.1: An overview of the relations between the arguments of *ormle* and the order-restricted estimates of (standardized) structural parameters.

Inputs in *ormle*	Notations in the article	Description
strest	$\hat{\theta}$	The estimates of structural parameters.
strcovmtrx	$\hat{\Sigma}_{\hat{\theta}}$	The covariance matrix of the estimates of structural parameters.
constr	$\begin{bmatrix} S_m \\ R_m \end{bmatrix}$	The restriction matrices S_m and/or R_m in hypothesis H_m.
nec	q_s	The number of equality restrictions in hypothesis H_m.
rhs	$\begin{bmatrix} s_m \\ r_m \end{bmatrix}$	Right hand side values (rhs), that is, constants s_m and/or r_m in hypothesis H_m.
Output		
Hm$est	$\hat{\theta}_m$	The estimates of (standardized) structural parameters for hypothesis H_m.
Hm$restrictedest	$\tilde{\theta}_m$	The order-restricted estimates of (standardized) structural parameters for hypothesis H_m.

Once the order-restricted estimates are obtained, the values of misfit, complexity, GORICA, and GORICA weights for the set of hypotheses in equation (5.3) can easily be obtained using the *gorica* function as follows:

$$\mathsf{gorica}(\mathsf{H1}, \mathsf{H2}, \mathsf{H3}, \mathsf{Hu}, \mathsf{iter} = 100000),$$

where $\mathsf{iter} = 100000$ represents the number of iterations when calculating complexity.

The last three inputs in the function *ormle* (i.e., constr, nec, and rhs) result from the hypotheses under consideration, but the first two inputs (i.e., strest and strcovmtrx) need to be calculated beforehand based on the data (independently from the hypotheses). The next section elaborates on three estimation methods that can be used to obtain strest and strcovmtrx.

5.4 Estimation methods

Model parameters and their covariance matrix can be estimated using maximum likelihood estimation (Fisher, 1922), nonparametric bootstrapping (Efron & Tibshirani, 1993), and gibbs sampling (D. Spiegelhalter, 1994). In this section, we discuss these three estimation methods in detail.

5.4.1 Maximum likelihood estimation

Maximum likelihood estimation (Fisher, 1922) concerns estimating model parameters by maximizing a log likelihood function. Maximum likelihood estimation is more commonly applied in estimating model parameters when compared to nonparametric bootstrapping and gibbs sampling. One reason for the use of maximum likelihood estimation more frequently than the other two methods is that it is easy to obtain MLEs in almost any standard software. This is because many standard softwares (e.g., MPlus, R, SAS, Amos, EQS, Lisrel) provide maximum likelihood estimation as the default option when estimating model parameters (Yuan & Hayashi, 2006). Another reason is that maximum likelihood estimation is often faster than the other two methods. Although maximum likelihood estimation requires an iterative process when estimating model parameters, it usually takes only a few seconds to obtain the results of estimation.

In the function *ormle* (see previous section), the MLEs of (standardized) structural parameters and their covariance matrix are denoted by strest and strcovmtrx, respectively. These arguments for the model in equation (5.3), that is, the estimates of standardized structural parameters $\hat{\theta} = (\hat{\theta}_1, \hat{\theta}_2, \hat{\theta}_3) = (\hat{\beta}_3, \hat{\beta}_4, \hat{\beta}_5)$ and their covariance matrix $\hat{\Sigma}_{\hat{\theta}}$, can be obtained using the *glm* function in R as follows:

```
model <- glm(num_awards ~ prog + zmath + prog * zmath, family = "poisson",
 data = academic_awards)
strest <- model$coefficients[c(4, 5, 6)]
strcovmtrx <- vcov(model)[c(4, 5, 6), c(4, 5, 6)]
```

As mentioned before, the outcome the number of awards earned by the students (defined here as A; and in the data set as num_awards) is predicted by the independent variables the standardized math scores of the students (MS; zmath) and the academic programs (P; prog) and their interaction (P * MS; prog * zmath). Note that numbers 4, 5, and 6 are used to select the estimates of three standardized structural parameters and their covariance matrix (for more details see Supplementary material, p.123).

Maximum likelihood estimation has two main drawbacks when evaluating informative hypotheses using the GORICA. First, model parameters should often be standardized to make them comparable, but standard softwares do not always produce the covariance matrix of these standardized parameter estimates. Second, when the response variable is not

normally distributed, the standard errors that are produced by maximum likelihood estimation are incorrect (Eliason, 1993) and this problem does not always disappear when sample size increases (Hox, 2010). In this case, a better alternative that can be used in conjunction with the GORICA would be nonparametric bootstrapping, which will be elaborated in the following subsection.

5.4.2 Nonparametric bootstrapping

Nonparametric bootstrapping (Efron & Tibshirani, 1993) is a procedure in which multiple samples are drawn from the original sample with replacement. When model assumptions are not satisfied, but researchers have moderate or large sample sizes, nonparametric bootstrapping may be considered to be a better method than maximum likelihood estimation. Standard errors that are obtained using nonparametric bootstrapping take into account the violations of distributional assumptions (Yuan & Hayashi, 2006), which makes these standard errors to be viable alternatives to their counterparts with maximum likelihood estimation (Hox, 2010).

We provide a part of the code to obtain strest and strcovmtrx for the poisson regression example using the *boot* function in R:

$$\text{boot\_sim} <\text{- boot}(\text{academic\_awards}, \text{boot.fn}, \text{R} = 1000, \text{sim} = \text{“ordinary”})$$
$$\text{strest} <\text{- apply}(\text{boot\_sim\$t}, 2, \text{mean})[\text{c}(4, 5, 6)]$$
$$\text{strcovmtrx} <\text{- cov}(\text{boot\_sim\$t})[\text{c}(4, 5, 6), \text{c}(4, 5, 6)]$$

In *boot*, the function boot.fn is utilized to estimate the standardized structural parameters in each bootstrap sample, which incorporates the model and data set into the bootstrapping process. This process uses the ordinary nonparametric bootstrapping technique (by stating $\text{sim} = \text{“ordinary”}$) to create $\text{R} = 1000$ bootstrap samples. The estimates of the model parameters for these bootstrap samples can be displayed by simply typing boot_sim$t in the R console. Subsequently, the mean and cov commands are used to obtain the overall bootstrap estimates and their covariance matrix, respectively. For more details in estimating the standardized structural parameters and their covariance matrix using nonparametric bootstrapping technique see Supplementary material, pp.124-125.

Nonparametric bootstrapping assumes that the original sample is a good representative of the population of interest. Thus, it often needs an adequate sample size to perform well (Hox, 2010, p.264). On the other hand, as mentioned before, the standard errors that are produced by maximum likelihood estimation are not correct when the response variable is not normally distributed, which is often the case for small samples. Therefore, maximum likelihood estimation and nonparametric bootstrapping may not be viable estimation methods for small samples. In such cases, gibbs sampling is a good alternative in estimating model parameters and their covariance matrix, which will be elaborated in the next subsection.

5.4.3 Gibbs sampling

Gibbs sampling (D. Spiegelhalter, 1994) is an iterative sampling process that generates samples by incorporating the observed data and the prior knowledge into the process. Gibbs sampling can also be used in case of small samples, whereas maximum likelihood and nonparametric bootstrapping cannot (Hox, 2010).

We provide a part of the code to obtain strest and strcovmtrx for the poisson regression example using the *bugs* and *read.bugs* functions in R. Notably, the software OpenBUGS needs to be downloaded (from `http://www.openbugs.net/w/Downloads`) to estimate (standardized) structural parameters and their covariance matrix using gibbs sampling for GLMs and GLMMs. This software is implemented to estimate the standardized model parameters and their covariance matrix, using the function bugs in the package R2OpenBUGS. The codaPkg = TRUE in *bugs* makes it easy to estimate (standardized) model parameters and their covariance matrix. To attain this goal, the function *read.bugs* is utilized as follows:

codaobject <- as.matrix(read.bugs(gibbs.sim))

strest <- apply(codaobject, 2, mean)[c(3, 4, 5)]

strcovmtrx <- cov(codaobject)[c(3, 4, 5), c(3, 4, 5)],

where gibbs.sim represents the output of *bugs* function. Similar to the function boot.fn in nonparametric bootstrapping, the function bugs incorporates the model and the data set into the sampling process. Note that the column index numbers need to be selected as 3, 4, and 5 to obtain the estimates of the three standardized structural parameters and their covariance matrix for gibbs sampling. For more details in obtaining these estimates and their covariance matrix using gibbs sampling see Supplementary material, pp.125-127.

There are two main drawbacks of using gibbs sampling in conjunction with the GORICA. First, the GORICA evaluation of informative hypotheses uses the MLEs of (standardized) structural parameters and their covariance matrix. However, gibbs sampling does not always produce these MLEs when the data are not normally distributed (Scheines et al., 1999). Second, gibbs sampling is often very slow when compared to maximum likelihood estimation and nonparametric bootstrapping in estimating model parameters and their covariance matrix.

In this section, we elaborated on three methods to estimate (standardized) structural parameters and their covariance matrix for which each technique has some advantages over the others in certain situations. In the next section, we give four different examples to illustrate the evaluation of informative hypotheses using the GORICA together with these techniques for GLMs, GLMMs, SEMs, and contingency tables, respectively.

5.5 Examples

We illustrate evaluating informative hypotheses using the GORICA by means of four examples. The illustrations will be given for a poisson regression model, two-level multilevel

regression model, structural equation model, and a $2 \times 2 \times 2$ contingency table as being representatives of GLMs, GLMMs, SEMs, and contingency tables, respectively. For the first three examples, the model parameters and their covariance matrix are estimated using each of the three estimation methods, namely maximum likelihood estimation, nonparametric bootstrapping, and gibbs sampling. For simplicity, the MLEs and their covariance matrix for contingency tables are obtained using only nonparametric bootstrapping, which is implemented in our R script GoricaCont.R (see Altınışık, Hessels, and Kuiper, unpublished, pp.46-57).

5.5.1 Example 1: Poisson regression modeling

Consider again the poisson regression model in equation (5.1). As elaborated earlier, the structural parameters used to formulate the four hypotheses in equation (5.3) are $\theta = (\beta_3, \beta_4, \beta_5)$ and the nuisance parameters are $\xi = (\beta_0, \beta_1, \beta_2)$. The structural parameters are standardized to make them comparable. The estimates of the standardized structural parameters and their covariance matrix can be obtained using maximum likelihood estimation, nonparametric bootstrapping, and gibbs sampling. These estimates and, for simplicity, only their standard errors are displayed for each estimation method in Table 5.2.

Table 5.2: Estimates and standard errors of the structural parameters for the poisson regression example.

	Maximum likelihood estimation		Nonparametric bootstrapping		Gibbs sampling	
θ	$\hat{\theta}$	SE($\hat{\theta}$)	$\hat{\theta}$	SE($\hat{\theta}$)	$\hat{\theta}$	SE($\hat{\theta}$)
β_3	0.412	0.442	0.467	0.478	0.448	0.448
β_4	0.266	0.456	0.220	0.491	0.234	0.461
β_5	0.215	0.508	0.067	0.625	0.153	0.518

In Table 5.3, the misfits and complexities, the values of the GORICA and GORICA weights for the four hypotheses H_1, H_2, H_3, and the unconstrained hypothesis H_u are displayed. The misfits for hypothesis H_2 and the unconstrained hypothesis H_u are the same within each estimation method. Hence, hypothesis H_2 fits the data set perfectly well regardless of the method used for estimation. Note that in the complexities, because the unconstrained hypothesis H_u is always the least specific hypothesis in the set of hypotheses, it is penalized more when compared to the others. In the same manner, hypothesis H_3 is less specific than hypotheses H_1 and H_2 and provides less information about the relationships between the model parameters. Therefore, hypothesis H_3 is more penalized than hypotheses H_1 and H_2. By inspecting the GORICA, we combine the misfit and complexity of the hypotheses. Hypothesis H_2 has the smallest GORICA value among the other competing hypotheses for each estimation method followed by the corresponding GORICA values for hypotheses H_3, H_u, and H_1.

Table 5.3: The misfits and complexities, GORICA values, and GORICA weights of hypotheses H_1, H_2, H_3, and H_u for the poisson regression example based on the three estimation methods.

Maximum likelihood estimation				
H_m	misfit	complexity	GORICA_m	w_m
H_1	40.548	0.000	40.548	0.000
H_2	-3.272	3.703	0.432	0.557
H_3	-3.093	4.997	1.904	0.267
H_u	-3.272	6.000	2.728	0.176
Nonparametric bootstrapping				
H_m	misfit	complexity	GORICA_m	w_m
H_1	36.073	0.000	36.073	0.000
H_2	-1.974	3.569	1.595	0.561
H_3	-1.963	4.997	3.034	0.273
H_u	-1.974	6.000	4.026	0.166
Gibbs sampling				
H_m	misfit	complexity	GORICA_m	w_m
H_1	39.948	0.000	39.948	0.000
H_2	-3.185	3.693	0.509	0.551
H_3	-3.098	4.997	1.900	0.275
H_u	-3.185	6.000	2.815	0.174

The GORICA weight for hypothesis H_1 is zero for each estimation method, which means that hypothesis H_1 is not supported by the data set at all. From the pairwise ratios between the GORICA weights, one can see that the relative support in the data for hypothesis H_2 is the largest one when compared to the other hypotheses. Hence, hypothesis H_2 is not a weak hypothesis, that is, it is $0.557/0.176 \approx 3.16$, $0.561/0.166 \approx 3.38$, and $0.551/0.174 \approx 3.17$ times better than the unconstrained hypothesis H_u when the model parameters are estimated using maximum likelihood estimation, nonparametric bootstrapping and gibbs sampling, respectively. Hypothesis H_2 is $0.557/0.266 \approx 2.09$, $0.561/0.273 \approx 2.05$, and $0.551/0.275 \approx 2.00$ times better than hypothesis H_3 for the three estimation methods, respectively. Therefore, hypothesis H_2 is chosen as the best hypothesis out of the set of four hypotheses under evaluation.

5.5.2 Example 2: Multilevel regression modeling

Many studies involve multilevel data structures in which lower level observations are clustered at higher level units, e.g., students are clustered at classrooms or schools. Multilevel regression is a convenient method for analyzing such data structures. The evaluation of infor-

mative hypotheses for multilevel regression models is illustrated based on the study in Finch, Bolin, and Kelley (2014, p.32). The outcome variable RA is a standardized continuous variable with the values in the interval [-1.861, 3.284], representing the reading achievement test scores of $N = 10320$ students in $J = 160$ schools. This outcome is predicted by standardized continuous variables vocabulary achievement test score V and age A, with the values within the intervals [-1.898, 2.832] and [-5.046, 5.429] respectively, the nominal variable gender which is defined as $\mathrm{G} \in \{0 = \mathrm{Female}, 1 = \mathrm{Male}\}$, and two-way interactions between these variables. Note that all the variables are measured at student level and the continuous variables are standardized to ensure the comparability of the structural parameters. We use the grand mean centering (i.e., the mean centering which involves all groups together) to standardized the continuous variables. Thus, the overall means of the continuous variables (i.e., $\overline{RA}$, $\overline{V}$, $\bar{A}$) are subtracted from the score of each student (i.e., R_{ji}-$\overline{RA}$, V_{ji}-$\overline{V}$, A_{ji}-$\bar{A}$).[3] The variable vocabulary test scores is considered to have both fixed and random effects on the outcome, namely its effect on the outcome is investigated on average, but this effect is allowed to vary across the schools. The effects of the predictors age and gender are considered to be fixed effects, namely their impacts on the outcome are on average and they do not change across the schools. Therefore, we describe a mixed-effect model as:

$$\begin{aligned} RA_{ji} =& \beta_0 + \beta_1 \mathrm{A}_{ji} + \beta_2 \mathrm{V}_{ji} + \beta_3 \mathrm{G}_{ji} + \beta_4 \mathrm{A}_{ji}\mathrm{V}_{ji} + \\ & \beta_5 \mathrm{A}_{ji}\mathrm{G}_{ji} + \beta_6 \mathrm{V}_{ji}\mathrm{G}_{ji} + \tau_{1j}\mathrm{V}_{ji} + \tau_{0j} + \epsilon_{ji}, \end{aligned} \tag{5.6}$$

where β_0 is the intercept, β_1, β_2, β_3 are the regression slopes for the variables age, vocabulary test score, and gender respectively, β_4, β_5, and β_6 are the coefficients for the two-way interactions between these variables, τ_{1j} and τ_{0j} are the random effects at the school level, and $\epsilon_{ji} \sim N(0, \sigma^2)$ is the error at the student level for student $i = 1, 2, ..., N_j$ at school $j = 1, 2, ..., J$. The random effects are assumed to have a normal distribution with mean vector 0 and covariance matrix:

$$\Psi = \begin{bmatrix} \psi_0^2 & \psi_{01} \\ \psi_{01} & \psi_1^2 \end{bmatrix}.$$

This model can be converted to:

$$R_{ji} = \begin{cases} \beta_0 + \beta_1 \mathrm{A}_{ji} + \beta_2 \mathrm{V}_{ji} + \beta_4 \mathrm{A}_{ji}\mathrm{V}_{ji} + \nu_{ji}, & \text{if G} = \text{“Female”}, \\ (\beta_0 + \beta_3) + (\beta_1 + \beta_5) A_{ji} + & \\ (\beta_2 + \beta_6) V_{ji} + \beta_4 A_{ji} V_{ji} + \nu_{ji}, & \text{if G} = \text{“Male”}, \end{cases}$$

where $\nu_{ji} = \tau_{1j}\mathrm{V}_{ji} + \tau_{0j} + \epsilon_{ji}$ represents the random part of the model in equation (5.6).

[3] Another standardization technique in multilevel regression context is group mean centering. In this centering, the group means of continuous variables (e.g., $\overline{RA}_j$, $\overline{V}_j$, $\bar{A}_j$) are substracted from the scores of individuals (e.g., R_{ji}-$\overline{RA}_j$, V_{ji}-$\overline{V}_j$, A_{ji}-$\bar{A}_j$). Readers are referred to Algina and Swaminathan (2011, pp.285-312), who elaborate on the grand mean and group mean centering for two-level multilevel models.

Three informative hypotheses and the unconstrained hypothesis H_u are evaluated to investigate the effects of these variables on the outcome. Hypothesis H_1 states that age has no impact on reading achievement for females (i.e., $\beta_1 = 0$) and for males (i.e., $\beta_1 + \beta_5 = 0$). However, vocabulary test score has a positive impact on reading achievement for females (i.e., $\beta_2 > 0$) and for males (i.e., $\beta_2 + \beta_6 > 0$), and their interaction has a positive effect on reading achievement for both males and females (i.e., $\beta_4 > 0$). Hypotheses H_2 and H_3 and the unconstrained hypothesis H_u are specified in a similar manner as hypothesis H_1:

$$\begin{aligned}
H_1 &: \beta_1 = 0, \quad\;\; \beta_2 > 0, \\
&\quad\; \beta_1 + \beta_5 = 0, \; \beta_2 + \beta_6 > 0, \; \beta_4 > 0, \\
H_2 &: \beta_1 < 0, \quad\;\; \beta_2 > 0, \\
&\quad\; \beta_1 + \beta_5 < 0, \; \beta_2 + \beta_6 > 0, \; \beta_4 > 0, \\
H_3 &: \beta_1 > 0, \quad\;\; \beta_2 > 0, \\
&\quad\; \beta_1 + \beta_5 < 0, \; \beta_2 + \beta_6 < 0, \; \beta_4 = 0, \\
H_u &: \beta_1, \beta_2, \beta_4, \beta_5, \beta_6.
\end{aligned} \tag{5.7}$$

Hypothesis H_2 states that age has a negative effect and vocabulary test score has a positive effect on reading achievement for both males and females, while their interaction has a positive effect on reading achievement. Hypothesis H_3 specifies that age and vocabulary test score have positive impacts on reading achievement for females, but negative impacts for males, while their interaction has no effect on reading achievement for both males and females. These hypotheses are nested within the unconstrained hypothesis H_u, which also represents the other possible relations between the structural parameters.

The structural parameters that are used in the specification of the three hypotheses above are $\theta = (\beta_1, \beta_2, \beta_4, \beta_5, \beta_6)$ and the nuisance parameters that are used for model fit but not used in the hypotheses are $\xi = (\beta_0, \beta_3, \sigma^2, \psi_0^2, \psi_1^2, \psi_{01})$. Like in GLMs represented by the poisson regression model in equation (5.1), in GLMMs, the structural parameters and their covariance matrix can be estimated by maximum likelihood estimation, nonparametric bootstrapping, and gibbs sampling. These estimates and their standard errors are displayed in Table 5.4.[4]

Table 5.5 displays the misfits and complexities, the values of the GORICA and GORICA weights for the four hypotheses H_1, H_2, H_3, and the unconstrained hypothesis H_u. Based on the GORICA weights, hypothesis H_3 is not supported by the data set at all, because the corresponding value of the GORICA weight is zero using each estimation method. Hypotheses H_1 and H_2 are strong hypotheses. For example, hypothesis H_1 is $0.327/0.054 \approx 6.06$ and hypothesis H_2 is $0.619/0.054 \approx 11.46$ times better than the unconstrained hypothesis H_u

[4]Note that there is not a function in R that implements nonparametric bootstrapping to estimate model parameters and their covariance matrix for GLMMs in our framework, namely, for two-level regression models. Thus, we introduce a function named *covglmm*, which can be used to estimate (un)standardized model parameters and their covariance matrix using nonparametric bootstrapping for two-level regression models (see Supplementary material, pp.128-130).

Table 5.4: Estimates and standard errors of the structural parameters for the multilevel regression example.

	Maximum likelihood estimation		Nonparametric bootstrapping		Gibbs sampling	
θ	$\hat{\theta}$	SE($\hat{\theta}$)	$\hat{\theta}$	SE($\hat{\theta}$)	$\hat{\theta}$	SE($\hat{\theta}$)
β_1	-0.008	0.012	-0.006	0.011	-0.010	0.012
β_2	0.519	0.017	0.525	0.014	0.512	0.012
β_4	0.031	0.009	0.031	0.011	0.025	0.009
β_5	-0.011	0.016	-0.011	0.014	-0.011	0.016
β_6	0.016	0.017	0.018	0.021	0.018	0.017

using maximum likelihood estimation. Hypothesis H_2 is $0.619/0.327 \approx 1.89$, $0.609/0.357 \approx 1.71$, and $0.682/0.268 \approx 2.54$ times better than hypothesis H_1 using maximum likelihood estimation, nonparametric bootstrapping, and gibbs sampling, respectively. Since both H_1 and H_2 are very strong hypotheses when compared to the unconstrained hypothesis H_u and there is no compeling evidence that hypothesis H_2 is better than hypothesis H_1, it is concluded that both hypotheses H_1 and H_2 are the best hypothesis in the set of hypotheses (for each of the estimation methods).

5.5.3 Example 3: Structural equation modeling

Structural equation modeling is a useful analysis technique that can be conveniently used to ascertain the relationship between observed and unobserved/latent variables and/or between dependent and independent variables. The evaluation of informative hypotheses for structural equation modeling often involves two types of parameters, which are factor loadings relating observed and unobserved variables and regression coefficients relating dependent and independent variables. The study in McArdle and Prescott (1992, p.90) is used to illustrate evaluating informative hypotheses using the GORICA for SEMs. This study evaluates intelligence and cognitive ability in a sample of individuals above 18 years old ($N = 1,680$) using the IQ test Wechsler Adult Intelligence Scale-Revised (WAIS-R) (Wechsler, 1981). This sample involves two latent variables: crystallized intelligence factor (Cry) and fluid intelligence factor (Fld). The variable Cry is related to the following four indicators, here four verbal tests: information (y_1; general knowledge of participants), comprehension (y_2; ability of abstract reasoning or judgement), similarities (y_3; unifying a theme), and vocabulary (y_4; verbal definition). Similarly, in addition to the verbal tests comprehension (y_2) and similarities (y_3), the variable Fld is related to the following four indicators, here four performance subtests: picture completion (y_5; perceiving visual images with missing features), block design (y_6; arranging blocks to match a design), picture arrangement (y_7; ordering cards with true story lines), and object assembly (y_8; reassembling puzzles). Both unobserved/latent variables Cry and Fld are regressed on the following two observed independent variables:

Table 5.5: The misfits and complexities, the values of the GORICA and GORICA weights of hypotheses H_1, H_2, H_3, and H_u for the multilevel regression example based on the three estimation methods.

Maximum likelihood estimation				
H_m	misfit	complexity	$\mathrm{GORICA_m}$	w_m
H_1	-31.597	3.340	-28.257	0.327
H_2	-34.650	5.120	-29.530	0.619
H_3	985.960	3.738	989.698	0.000
H_u	-34.650	10.000	-24.650	0.054
Nonparametric bootstrapping				
H_m	misfit	complexity	$\mathrm{GORICA_m}$	w_m
H_1	-33.066	2.940	-30.126	0.357
H_2	-35.429	4.232	-31.196	0.609
H_3	1412.827	4.233	1417.060	0.000
H_u	-35.429	10.000	-25.429	0.034
Gibbs sampling				
H_m	misfit	complexity	$\mathrm{GORICA_m}$	w_m
H_1	-32.085	3.035	-29.050	0.268
H_2	-35.729	4.810	-30.919	0.682
H_3	1902.683	4.048	1906.731	0.000
H_u	-35.729	10.000	-25.729	0.050

edc $\in \{0 =$ "non high school graduate", $1 =$ "high school graduate"$\}$ and age with mean 42.2 and standard deviation 18. The descriptives for the eight indicators and two independent variables are displayed in Table 5.6. In addition, Figure 5.1 displays how the variables are related to each other.

In structural equation modeling, the relations between latent variables and indicators are represented by a measurement model. Let $\eta_i = (Cry_i, Fld_i)^T$ be the 2×1 vector of the latent variables and $y_i = (y_{1i}, y_{2i}, ..., y_{8i})^T$ is the 8×1 vector of the indicators for individual $i = 1, 2, ..., 1680$. Then, the measurement model can be defined as:

$$y_i = \Lambda\eta_i + \upsilon_i, \tag{5.8}$$

where

$$\Lambda = \begin{bmatrix} \lambda_1 & \lambda_2 & \lambda_3 & \lambda_4 & 0 & 0 & 0 & 0 \\ 0 & \lambda_5 & \lambda_6 & 0 & \lambda_7 & \lambda_8 & \lambda_9 & \lambda_{10} \end{bmatrix}^T$$

is an 8×2 matrix of factor loadings and υ_i is an 8×1 vector of measurement errors for person i, with $\upsilon_i \sim N(0, \Phi)$ and Φ is an 8×8 diagonal covariance matrix of these measurement

Table 5.6: Descriptives for the observed variables in the SEM example.

	y_1	y_2	y_3	y_4	y_5
Min:	2.000	0.000	0.000	3.000	0.000
Mean:	18.226	19.871	17.477	44.310	14.090
Max:	29.000	32.000	28.000	70.000	20.000
SD:	5.953	6.174	6.215	14.779	4.134
	y_6	y_7	y_8	edc	age
Min:	0.000	0.000	2.000	0.000	18.500
Mean:	26.607	11.272	28.369	0.661	42.238
Max:	51.000	20.000	41.000	1.000	72.000
SD:	11.471	5.242	7.338	0.473	18.031

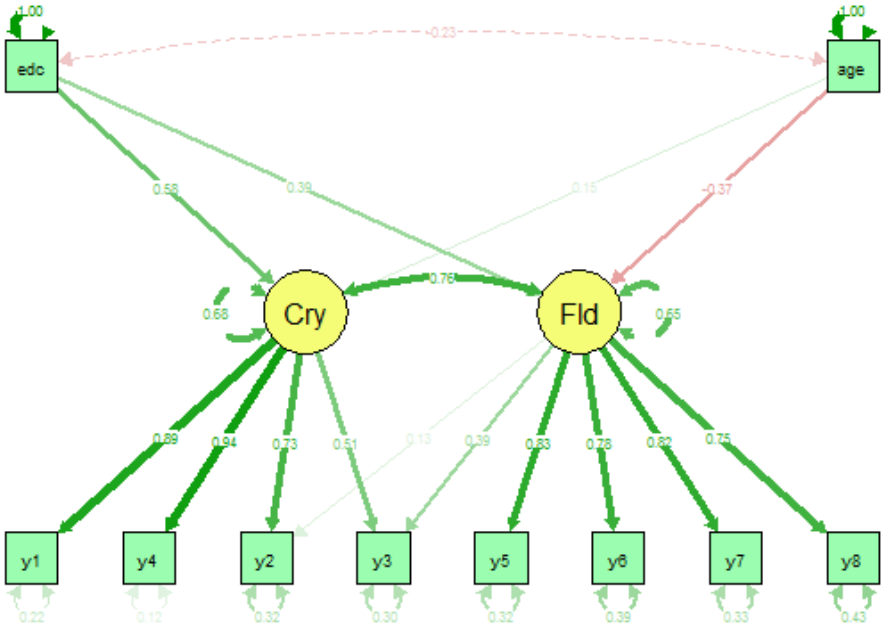

Figure 5.1: Relations between the variables for the SEM example. The figure including the standardized parameter estimates is obtained using the package *lavaan* (using maximum likelihood estimation).

errors. Similarly, the relations between the unobserved/latent variables, $\eta_i = (Cry_i, Fld_i)^T$, and the independent variables, $x_i = (edc_i, age_i)^T$, are given by a structural equation model as:

$$\eta_i = \Gamma x_i + \delta_i, \tag{5.9}$$

where

$$\Gamma = \begin{bmatrix} \gamma_1 & \gamma_2 \\ \gamma_3 & \gamma_4 \end{bmatrix}$$

is a 2×2 matrix of regression coefficients relating the two latent variables, $\eta_i = (Cry_i, Fld_i)^T$, with the two independent variables, $x_i = (edc_i, age_i)^T$, and δ_i is a 2×1 vector of residuals for person i, with $\delta_i \sim N(0, \zeta)$ and ζ is a 2×2 covariance matrix of these residuals. With this SEM notation it is not explicitly clear, but the covariance matrix of the latent variables is also modeled and is denoted by:

$$\Sigma_\eta = \begin{bmatrix} \omega_{11} & \omega_{21} \\ \omega_{21} & \omega_{22} \end{bmatrix}.$$

For illustration purposes, we will inspect three sets of fictional hypotheses. The first set of informative hypotheses involving factor loadings are:

$$\begin{aligned} H_1 &: \lambda_2 > \lambda_5,\ \lambda_3 > \lambda_6, \lambda_1 > 0, \lambda_4 > 0, \\ H_2 &: \lambda_7 > \lambda_8 > \lambda_9 > \lambda_{10}, \\ H_3 &: H_1\ \&\ H_2, \\ H_{u1} &: \lambda_1, \lambda_2, \lambda_3, \lambda_4, \lambda_5, \lambda_6, \lambda_7, \lambda_8, \lambda_9, \lambda_{10}. \end{aligned} \tag{5.10}$$

Hypothesis H_1 specifies that the factor loadings for the verbal tests comprehension and similarities related to the crystallized intelligence factor are bigger than the corresponding factor loadings related to the fluid intelligence factor (i.e., $\lambda_2 > \lambda_5$ for comprehension and $\lambda_3 > \lambda_6$ for similarities) and the factor loadings for the verbal tests information and vocabulary related to the cystallized intelligence factor are positive (i.e., $\lambda_1 > 0$ for information and $\lambda_4 > 0$ for vocabulary). Hypothesis H_2 states that the factor loading for the performance test picture completion related to the fluid inteligence factor is the biggest one among the other factor loadings for the performance tests and is followed by the factor loadings for the tests block design, picture arrangement, and object assembly, respectively (i.e., $\lambda_7 > \lambda_8 > \lambda_9 > \lambda_{10}$). Hypothesis H_3 is the combination of the two hypotheses, namely, it contains all the restrictions specified in hypotheses H_1 and H_2. Note that the structural parameters used to formulate the four hypotheses above are $\theta = (\lambda_1, \lambda_2, ..., \lambda_{10})$ and the nuisance parameters are $\xi = (\gamma_1, \gamma_2, ..., \gamma_4, \omega_{11}, \omega_{21}, \omega_{22}, \zeta, \Phi)$.

The second set contains two hypotheses and the corresponding unconstrained hypothesis H_{u2}, which are formulated using regression coefficients. These hypotheses are:

$$\begin{aligned} H_4 &: \gamma_1 > \gamma_3 > 0, \\ H_5 &: \gamma_1 > \gamma_3 > 0,\ \gamma_2 > 0,\ \gamma_4 < 0, \\ H_{u2} &: \gamma_1, \gamma_2, \gamma_3, \gamma_4. \end{aligned} \tag{5.11}$$

Hypothesis H_4 states that the impact of education for both latent variables is positive (i.e., $\gamma_1 > 0$ and $\gamma_3 > 0$), but its effect for crystallized intellingence factor is bigger than its effect for fluid intelligence factor (i.e., $\gamma_1 > \gamma_3$). Hypothesis H_5 additionally states that the effect of age is positive for the crystallized intelligence factor (i.e., $\gamma_2 > 0$), but its effect is negative for the fluid intelligence factor (i.e., $\gamma_4 < 0$). For this set of hypotheses, the structural parameters are $\theta = (\gamma_1, \gamma_2, ..., \gamma_4)$ and the nuisance parameters are $\xi = (\lambda_1, \lambda_2, ..., \lambda_{10}, \omega_{11}, \omega_{21}, \omega_{22}, \zeta, \Phi)$.

The last set of hypotheses is about the covariance between the two latent variables:

$$\begin{aligned} H_6 &: \omega_{21} = 0, \\ H_7 &: \omega_{21} > 0, \\ H_8 &: \omega_{21} < 0. \end{aligned} \tag{5.12}$$

Hypothesis H_6 states that the covariance (and thus also the correlation) between the crystallized and fluid intelligence factors is zero, while hypotheses H_7 and H_8 specify that this covariance is positive and negative, respectively. Note that hypotheses H_6, H_7, and H_8 together cover all possible values for the covariance between the latent variables, that is, equal to zero, bigger than zero, and smaller than zero, respectively. Therefore, the unconstrained hypothesis $H_{u3} : \omega_{21}$ is not needed in the set of hypotheses. The three hypotheses above are formulated by the structural parameter $\theta = \omega_{21}$, where the nuisance parameters are $\xi = (\lambda_1, \lambda_2, ..., \lambda_{10}, \gamma_1, \gamma_2, ..., \gamma_4, \omega_{11}, \omega_{22}, \zeta, \Phi)$.

Estimates and their standard errors for the factor loadings, regression coefficients, and the covariance between the two latent variables in terms of standardized observed and latent variables are shown in Table 5.7. Note that the choice of scaling might be of importance when obtaining these estimates and their covariance matrix (see Supplementary material, pp.132-136 for more details).

Table 5.7: Estimates and standard errors of the parameters for the SEM example.

	Maximum likelihood estimation		Nonparametric bootstrapping		Gibbs sampling	
θ	$\hat{\theta}$	SE($\hat{\theta}$)	$\hat{\theta}$	SE($\hat{\theta}$)	$\hat{\theta}$	SE($\hat{\theta}$)
λ_1	0.886	0.006	0.886	0.007	0.886	0.006
λ_2	0.725	0.021	0.726	0.023	0.725	0.021
λ_3	0.510	0.022	0.510	0.024	0.510	0.022
λ_4	0.938	0.005	0.938	0.006	0.938	0.005
λ_5	0.132	0.024	0.131	0.026	0.133	0.024
λ_6	0.393	0.023	0.393	0.025	0.394	0.023
λ_7	0.826	0.009	0.825	0.011	0.826	0.009
λ_8	0.817	0.010	0.816	0.011	0.817	0.010
λ_9	0.781	0.011	0.781	0.011	0.781	0.011
λ_{10}	0.753	0.012	0.753	0.014	0.753	0.012
γ_1	0.580	0.017	0.579	0.018	0.581	0.016
γ_2	0.147	0.022	0.146	0.022	0.149	0.021
γ_3	0.387	0.020	0.387	0.022	0.389	0.019
γ_4	-0.368	0.020	-0.368	0.022	-0.366	0.021
ω_{21}	0.763	0.015	0.763	0.020	0.762	0.016

The misfit and complexity parts, the values of the GORICA and GORICA weights for hypotheses H_1, H_2, and H_3 with the unconstrained hypothesis $H_{u1} : \lambda_1, ..., \lambda_{10}$, hypotheses H_4 and H_5 with the unconstrained hypothesis $H_{u2} : \gamma_1, ..., \gamma_4$, and hypotheses H_6, H_7, and H_8 are given in Table 5.8. Based on the GORICA weights, very similar results are obtained using each estimation method. Hypothesis H_3 is much better supported by the data than hypotheses H_1 and H_2 and the unconstrained hypothesis H_{u1}. That is, for example, when using maximum likelihood estimation, hypothesis H_3 is $0.768/0.113 \approx 6.80$ times better than hypothesis H_1, $0.768/0.104 \approx 7.38$ times better than hypothesis H_2, and $0.768/0.015 \approx 51.20$ times better than the unconstrained hypothesis H_{u1}. For the second set of hypotheses, hypotheses H_4 and H_5 are both strong hypotheses. For example, when using nonparametric bootstrapping, hypothesis H_4 is $0.223/0.066 \approx 3.38$ times better and hypothesis H_5 is $0.711/0.066 \approx 10.77$ times better than the unconstrained hypothesis H_{u2}. However, hypothesis H_5 is better supported by the data when compared to hypothesis H_4, that is, hypothesis H_5 is $0.711/0.223 \approx 3.19$ times better than hypothesis H_4. When looking at the third set of hypotheses, hypotheses H_6 and H_8 are not supported by the data at all, which have zero weight using each estimation method. Hypothesis H_7 is selected as the best hypothesis for this set of hypotheses, since it has a weight of one using each estimation method. Therefore, it is concluded that hypothesis H_3 for the factor loadings, hypothesis H_5 for the regression coefficients, and hypothesis H_7 for the covariance between the latent variables are selected as the best hypotheses from the three sets of informative hypotheses.

Table 5.8: The misfit and complexity parts, the values of the GORICA and GORICA weights of hypotheses H_1, H_2, H_3, and H_{u1}; H_4, H_5, and H_{u2}; and H_6, H_7, and H_8 for the SEM example.

		Maximum likelihood estimation		
H_m	misfit	complexity	$\mathrm{GORICA_m}$	w_m
H_1	-73.673	16.028	-57.645	0.113
H_2	-73.673	16.195	-57.479	0.104
H_3	-73.673	12.195	-61.479	0.768
H_{u1}	-73.673	20.000	-53.673	0.015
H_4	-25.288	5.575	-19.713	0.225
H_5	-25.288	3.284	-22.004	0.708
H_{u2}	-25.288	8.000	-17.288	0.067
H_6	2576.823	0.000	2576.823	0.000
H_7	-6.560	0.998	-5.562	1.000
H_8	2576.823	1.000	2577.824	0.000
		Nonparametric bootstrapping		
H_m	misfit	complexity	$\mathrm{GORICA_m}$	w_m
H_1	-71.626	16.161	-55.465	0.107
H_2	-71.626	16.050	-55.576	0.113
H_3	-71.626	12.217	-59.409	0.765
H_{u1}	-71.626	20.000	-51.626	0.015
H_4	-24.703	5.584	-19.119	0.223
H_5	-24.703	3.264	-21.439	0.711
H_{u2}	-24.703	8.000	-16.703	0.066
H_6	1422.494	0.000	1422.494	0.000
H_7	-5.968	0.998	-4.970	1.000
H_8	1422.494	1.001	1423.495	0.000
		Gibbs sampling		
H_m	misfit	complexity	$\mathrm{GORICA_m}$	w_m
H_1	-73.528	16.046	-57.482	0.107
H_2	-73.528	16.132	-57.396	0.103
H_3	-73.528	12.094	-61.434	0.775
H_{u1}	-73.528	20.000	-53.528	0.015
H_4	-25.034	5.590	-19.444	0.226
H_5	-25.034	3.304	-21.730	0.707
H_{u2}	-25.034	8.000	-17.034	0.067
H_6	2461.992	0.000	2461.992	0.000
H_7	-6.517	0.998	-5.519	1.000
H_8	2461.992	1.001	2462.994	0.000

5.5.4 Example 4: Contingency tables

Evaluation of informative hypotheses for contingency tables is illustrated by the study of Agresti (2007, p.50). This study investigates whether racial characteristics of criminals convicted of murder influences receiving the death penalty in Florida. The outcome, the death

penalty, is defined as DP $\in$ {1 = Death penalty, 2 = No death penalty}. Similarly, both the race of victims RV and the race of defendants RD are described in two levels as {1 = White, 2 = Black}. For this study, three fictional hypotheses including the unconstrained hypothesis H_u are:

H_1 : Irrespective of the victims being white or black, black defendants are sentenced to death more often than white defendants,

H_2 : In the case the victims were white, black defendants are sentenced to death more often than white defendants, but when the victims were black, the difference on being sentenced to death with respect to the race of defendants was not apparent,

H_u : Any relationship between the race of victims and defendants and being sentenced to death or not can occur.

Hypotheses H_1 and H_2, and the unconstrained hypothesis H_u can be formulated by means of using conditional cell probabilities. These conditional cell probabilities are specified for four different groups of people. That is, the conditional cell probability for white victims and defendants (i.e., $\theta_{\mathrm{w|w}}$), for white victims and black defendants (i.e., $\theta_{\mathrm{b|w}}$), for black victims and white defendants (i.e., $\theta_{\mathrm{w|b}}$), and for black victims and defendants (i.e., $\theta_{\mathrm{b|b}}$). Here, for example, $\theta_{\mathrm{b|w}}$ represents the probability of a black defendant being sentenced to death given that the victim was white.

The hypotheses specify linear restrictions on functions of cell probabilities, namely on conditional cell probabilities:

$$\begin{aligned} H_1 &: \theta_{\mathrm{b|w}} > \theta_{\mathrm{w|w}}, \theta_{\mathrm{b|b}} > \theta_{\mathrm{w|b}}, \\ H_2 &: \theta_{\mathrm{b|w}} > \theta_{\mathrm{w|w}}, \theta_{\mathrm{b|b}} = \theta_{\mathrm{w|b}}, \\ H_u &: \theta_{\mathrm{w|w}}, \theta_{\mathrm{b|w}}, \theta_{\mathrm{w|b}}, \theta_{\mathrm{b|b}}, \end{aligned} \tag{5.13}$$

with $\theta = (\theta_{\mathrm{w|w}}, \theta_{\mathrm{b|w}}, \theta_{\mathrm{w|b}}, \theta_{\mathrm{b|b}})^T = (\frac{\pi_{111}}{\pi_{111}+\pi_{121}}, \frac{\pi_{112}}{\pi_{112}+\pi_{122}}, \frac{\pi_{211}}{\pi_{211}+\pi_{221}}, \frac{\pi_{212}}{\pi_{212}+\pi_{222}})^T$. The hypotheses in (14) are specified in terms of a subclass of restrictions on conditional cell probabilities. Note that hypotheses containing restrictions from different subclasses cannot be evaluated in the same set of hypotheses (see Altınışık, Hessels, and Kuiper, unpublished, pp.5-7).

After formulating informative hypotheses using (functions of) cell probabilities, the GORICA can be used to evaluate them. Script file GoricaCont.R (which is discussed in the Supplementary material, pp.31-34) utilizes nonparametric bootstrapping to evaluate informative hypotheses for contingency tables. This script can be used to evaluate informative hypotheses for contingency tables with and without empty cell(s). One needs to adjust two text files (called data.txt and input.txt) before executing the R code in GoricaCont.R. In data.txt, one should insert the data set at hand. In input.txt, one needs to specify functions of cell probabilities and the hypotheses under evaluation. Once these two files are created, the results can easily be obtained using the source function in R as source("GoricaCont.R") (see Supplementary material, pp.148-151).

For this study, population probabilities for the combinations of the race of victims and defendants and the death penalty (together with the observed numbers of individuals) are

Table 5.9: Population probabilities π_{ijv} for the combinations of the race of victims and defendants and the death penalty (for $i = 1, 2$; $j = 1, 2$; and $v = 1, 2$) and observed numbers of individuals between brackets.

		RD	
RV	DP	White	Black
White	Yes	π_{111} (53)	π_{112} (11)
	No	π_{121} (414)	π_{122} (37)
Black	Yes	π_{211} (0)	π_{212} (4)
	No	π_{221} (16)	π_{222} (139)

Table 5.10: Estimates and standard errors of the conditional cell probabilities using nonparametric bootstrapping.

θ	$\hat{\theta}$	$\mathrm{SE}(\hat{\theta})$
$\theta_{\mathrm{w\|w}}$	0.114	0.015
$\theta_{\mathrm{b\|w}}$	0.228	0.061
$\theta_{\mathrm{w\|b}}$	0.000	0.000
$\theta_{\mathrm{b\|b}}$	0.028	0.014

displayed in Table 5.9. As mentioned before, the θ parameters and their covariance matrix are estimated using nonparametric bootstrapping. In Table 5.10, these estimates and their standard errors resulting from source("GoricaCont.R") are displayed. Note that the estimate for the conditional cell probability $\theta_{\mathrm{w|b}}$ and its standard error are zero, because of the empty cell in the contingency table, namely, $\hat{\theta}_3 = \hat{\theta}_{\mathrm{w|b}} = 0$, and therefore, $\mathrm{SE}(\hat{\theta}_{\mathrm{w|b}}) = 0$. The interested reader is referred to Altınışık, Hessels, and Kuiper, unpublished, pp.17-23).

In Table 5.11, we display the misfit and complexity parts, the values of the GORICA and GORICA weights for the hypotheses H_1, H_2, and the unconstrained hypothesis H_u. Based on the GORICA weights, hypothesis H_1 is $0.623/0.144 \approx 4.33$ times more supported by the data than hypothesis H_2, and it is a strong hypothesis since it is $0.623/0.233 \approx 2.67$ times better supported by the data when compared to the unconstrained hypothesis H_u. Therefore, it is concluded that hypothesis H_1 is selected as the best hypothesis in the set of the three hypotheses under evaluation.

Table 5.11: The misfit and complexity parts, the values of the GORICA and GORICA weights of hypotheses H_1, H_2, and H_u using nonparametric bootstrapping for the death penalty data.

H_m	misfit	complexity	$\mathrm{GORICA_m}$	w_m
H_1	-15.223	4.033	-11.191	0.623
H_2	-11.254	2.999	-8.256	0.144
H_u	-15.223	6.000	-9.223	0.233

5.6 Discussion

We showed how the GORICA can be applied to evaluate informative hypotheses by using the three estimation methods. One of these estimation techniques should be chosen on beforehand, rather than looking for one of them that gives the "wanted" results. It is recommended to use the usual technique maximum likelihood estimation when model assumptions are satisfied (e.g., the response variable is normally distributed) for moderate and large samples. When sample size is adequate but model assumptions are not satisfied we suggested to utilize nonparametric bootstrapping (together with the GORICA) to evaluate informative hypotheses. When researchers have small samples for which model assumptions are often not satisfied, we recommended to use gibbs sampling for the evaluation of informative hypotheses. In the Supplementary material, we elaborate on the functions *ormle* and *gorica* and provide the R code to duplicate the results presented in our examples.

5.A Supplementary material: Evaluation of informative hypotheses using gorica

Throughout this document, we elaborate on how the results in the paper can be obtained using R functions/code. The outline of the document is as follows. First, we elaborate on how to estimate (standardized) structural model parameters and their covariance matrix in R using maximum likelihood estimation, nonparametric bootstrapping, and gibbs sampling for the poisson regression model, two-level multilevel regression model, and structural equation model as being representatives of GLMs, GLMMs, and SEMs, respectively. Second, we introduce the functions **ormle** and **gorica** in R and elaborate on how to use these functions to duplicate our results for the three types of models above. Third, we elaborate how to evaluate the hypotheses for the $2 \times 2 \times 2$ contingency table using script GoricaCont.R, which is provided by Altınışık, Hessels, and Kuiper (unpublished, pp.46-57).

5.B Estimating (standardized) structural parameters and their covariance matrix

In this section, we elaborate on how to estimate (standardized) structural parameters and their covariance matrix for the poisson regression model, two-level multilevel regression model, and structural equation model. Three estimation methods are used to estimate these parameters and their covariance matrix, which are maximum likelihood estimation, nonparametric bootstrapping, and gibbs sampling. First, we show how to fit these models to the data sets at hand. Once these models are fitted, one needs to choose the estimates (and their covariance matrix) corresponding to the structural parameters, which are used in the specification of the hypotheses under evaluation. The rest of the parameters are nuisance parameters, which are used in model fitting but not primary of interest by researchers.

First of all, the relavant libraries have to be installed. Using the package easypackages, these libraries can be automatically installed as displayed below.

```
library(easypackages)
list.of.packages <- c("base", "blavaan", "boot", "coda", "foreign",
 "FRACTION", "lavaan", "lme4", "MASS", "matrixcalc", "mvtnorm", "nlme",
 "quadprog", "R2OpenBUGS")
new.packages <- list.of.packages[!(list.of.packages %in%
 installed.packages()[, "Package"])]
if (length(new.packages)) install.packages(new.packages)
libraries(list.of.packages)
```

In the next three subsections, we elaborate on how to estimate (standardized) structural parameters and their covariance matrix in our examples using the three estimation methods.

5.B.1 Example 1: Poisson regression modeling

This example illustrates evaluation of informative hypotheses using a poisson regression model, as being a representative of GLMs. The academic_awards data ($N = 200$) used for this example is available on the Web `https://stats.idre.ucla.edu/stat/data/poisson_sim.csv`. The outcome the number of awards earned by the students (defined in the article as A; and in the data set and thus here as num_awards) is a count variable and is predicted by the standardized math scores of the students (MS; zmath), the academic programs (P; prog), and their interaction (P * MS; prog $*$ zmath). The predictor prog is a categorical variable with the three levels general, academic, and vocational, which represent three academic programs in the school. This categorical variable is defined by the academic program general as being the reference category.

Maximum likelihood estimation

The simplest approach in estimating structural parameters and their covariance matrix is maximum likelihood estimation. For GLMs, this estimation technique is implemented using the glm function in the package stats:

```
academic_awards <- read.csv("academic_awards.csv")
#Below, the categorical variable prog is defined with three levels
#such that the academic program general is the reference category
academic_awards <- within(academic_awards, {
 prog <- factor(prog, levels = 1:3, labels = c("General", "Academic",
 "Vocational") ) } )
#Standardizing math scores of the students
zmath <- scale(academic_awards$math)
#Fitting the poisson regression model which indicates the relationship
#between the variables in the data
model <- glm(num_awards ~ prog + zmath + prog * zmath, family = "poisson",
 data = academic_awards)
#Obtaining the MLEs of the standardized structural parameters.
#The regression coefficient for zmath denotes the case for which the
#reference category is general, namely, P = General
strest <- model$coefficients[c(4, 5, 6)]
strest
```

```
     zmath   progAcademic:zmath progVocational:zmath
0.4121808            0.2661545            0.2145273
```

```
#Obtaining the covariance matrix of these MLEs
strcovmtrx <- vcov(model)[c(4, 5, 6), c(4, 5, 6)]
strcovmtrx
```

```
                          zmath progAcademic:zmath progVocational:zmath
zmath                  0.195636         -0.1956360           -0.1956360
progAcademic:zmath    -0.195636          0.2081928            0.1956360
progVocational:zmath  -0.195636          0.1956360            0.2579104
```

Note that the regression coefficient for zmath denotes the case for which the reference category is general, namely, P = General. The structural parameters used to specify the set of four hypotheses for the poisson regressin example are $\theta = \{\beta_3, \beta_4, \beta_5\}$ and the nuisance parameters that are not used to specify the four hypotheses are $\xi = \{\beta_0, \beta_1, \beta_2\}$. Note that the estimates and their covariance matrix corresponding to these standardized structural parameters are chosen by using the index numbers 4, 5, and 6, which correspond to parameters β_3, β_4, and β_5, respectively.

Nonparametric bootstrapping

A more computer intensive estimation method that can be used to estimate (standardized) structural parameters and their covariance matrix is nonparametric bootstrapping. To attain the estimates of the standardized structural parameters and their covariance matrix using nonparametric bootstrapping for the poisson regression example we utilize the **boot** function in the package boot. In **boot**, R = 1000 bootstrap samples are created using the ordinary nonparametric bootstrapping technique, by stating sim = "ordinary". These bootstrap samples can be displayed by simply typing boot_sim$t on R console. Below, set.seed(111) is used as the random number generator to duplicate our results for this example.

```
set.seed(111)
#Below in boot, the function boot.fn is used to obtain the MLEs of the
#standardized structural parameters in B = 1000 bootstrap samples
#using the ordinary nonparametric bootstrapping technique
boot.fn <- function(data, index) {
 return(coef(glm(num_awards ~ prog + zmath + prog * zmath,
 family = c("poisson"), data = academic_awards, subset = index) ) ) }
boot_sim <- boot(academic_awards, boot.fn, R = 1000, sim = "ordinary")
#Below, boot_sim$t denotes a matrix containing the MLEs of
#(standardized) model parameters in all bootstrap samples.
#The names of the columns in this matrix are determined to make the
# output more understandable
colnames(boot_sim$t) <- names(boot_sim$t0)
#Obtaining the MLEs of the standardized structural parameters
strest <- apply(boot_sim$t, 2, mean)[c(4, 5, 6)]
strest
```

```
     zmath   progAcademic:zmath progVocational:zmath
0.46747081           0.21980556           0.06697369
```

```
#Obtaining the covariance matrix of these MLEs
strcovmtrx <- cov(boot_sim$t)[c(4, 5, 6), c(4, 5, 6)]
strcovmtrx
```

```
                          zmath progAcademic:zmath progVocational:zmath
zmath                 0.2280587         -0.2280500           -0.2176842
progAcademic:zmath   -0.2280500          0.2415237            0.2157397
progVocational:zmath -0.2176842          0.2157397            0.3901629
```

Similar to maximum likelihood estimation, the index numbers 4, 5, and 6 are used to select the estimates of the standardized structural parameters and their covariance matrix for nonparametric bootstrapping.

Gibbs sampling

The third estimation method used to estimate (standardized) structural parameters and their covariance matrix is gibbs sampling. The software OpenBUGS needs to be downloaded to the computer to estimate these parameters and their covariance matrix for the poisson regression example (and for the two-level multilevel regression model elaborated in the next subsection), which is available on the Web at `http://www.openbugs.net/w/Downloads`. OpenBUGS is implemented within R using the function **bugs** in the package R2OpenBUGS. To be able to use **bugs**, one needs to create a model file named bugsmodelglm.txt containing the code below. This file is used to specify the model and its parameters with their noninformative priors to generate gibbs samples, which are often called posterior samples.

```
#Do not run the code below and change the directory such that it
#reaches the model file bugsmodelglm.txt which is used in the function
#bugs and the data set academic_awards.csv
model{
#Specifying the model used for the poisson regression example
for(i in 1:N) {
 num_awards[i] ~ dpois(lambda[i])
 log(lambda[i]) <- beta0 + beta[1] * academic[i] + beta[2] * vocational[i] +
 beta[3] * zmath[i] + beta[4] * academic[i] * zmath[i] +
 beta[5] * vocational[i] * zmath[i] }
#Specifying noninformative priors for model parameters
beta0 ~ dnorm(0,0.0001)
beta[1]  ~ dnorm(0,0.0001)
beta[2] ~ dnorm(0,0.0001)
beta[3]  ~ dnorm(0,0.0001)
beta[4] ~ dnorm(0,0.0001)
beta[5] ~ dnorm(0,0.0001)
}
```

Once the model file is created as shown above, sample size N and the variables in the model have to be specified. The model parameters must be initialized to trigger the procedure in obtaining the gibbs samples as displayed below.

```
#Introducing the sample size N = 200 and the variables in the model file
N <- nrow(academic_awards)
prog <- academic_awards$prog
academic <- model.matrix(~ prog)[, 2]
vocational <- model.matrix(~ prog)[, 3]
zmath <- as.numeric(zmath)
num_awards <- as.numeric(academic_awards$num_awards)
data <- list("N", "academic", "vocational", "zmath", "num_awards")
#Initializing the model parameters
beta0 <- c(rnorm(1, 0, 1e-05))
beta <- c(rnorm(5, 0, 1e-05))
inits <- function(){
 list(beta0 = beta0, beta = beta) }
```

After introducing the data set and initializing the model parameters, the results of gibbs sampling for the logistic regression example are obtained using the command:

```
gibbs.sim <- bugs(data, inits, model.file = "bugsmodelglm.txt",
 parameters = c("beta0", "beta"), n.chains = 3, n.iter = 30000,
 n.burnin = 3000, debug = TRUE, codaPkg = TRUE)
```

In gibbs.sim, one needs to specify the numbers of the chains and iterations, and a burn-in period to generate the gibbs samples. We generate n.iter $=$ 30000 samples, with n.chains $=$ 3 chains and a burn-in period of n.burnin $=$ 3000. The debug $=$ TRUE is used to follow up the procedure in OpenBUGS. An OpenBUGS screen pops up after executing the code above. Once gibbs sampling procedure is completed, the results are displayed on the same screen with the iteration process. This screen must be closed to proceed in estimating model parameters and their covariance matrix. The codaPkg $=$ TRUE in gibbs.sim makes it easy to estimate model parameters and their covariance matrix. To do this, the function **read.bugs** is utilized as follows.

```
#The results of executing the function bugs are read using the function
#read.bugs. Below, codaobject denotes a matrix containing the parameter
#estimates for n.iter = 30000 samples
codaobject <- as.matrix(read.bugs(gibbs.sim) )
#Below, we attach the names to the columns of this matrix based on the
#model at hand and the intercept and the deviance
colnames(codaobject) <- c("progAcademic", "progVocational", "zmath",
 "progAcademic:zmath", "progVocational:zmath", "Intercept", "deviance")
#Obtaining the estimates of the standardized structural parameters
strest <- apply(codaobject, 2, mean)[c(3, 4, 5)]
strest
```

```
     zmath   progAcademic:zmath progVocational:zmath
 0.4476426             0.2337392             0.1527306
```

```
#Obtaining the covariance of these estimates
strcovmtrx <- cov(codaobject)[c(3, 4, 5), c(3, 4, 5)]
strcovmtrx
```

```
                          zmath progAcademic:zmath progVocational:zmath
zmath                 0.2005855         -0.2002213           -0.2013066
progAcademic:zmath   -0.2002213          0.2124619            0.2010467
progVocational:zmath -0.2013066          0.2010467            0.2680433
```

Note that in contrast to maximum likelihood estimation and nonparametric bootstrapping, one needs to choose the index numbers 3, 4, and 5 to be able to select the estimates of the standardized structural parameters ($\hat{\theta}$; strest) and their covariance matrix ($\hat{\Sigma}_{\hat{\theta}}$; strcovmtrx) for the poisson regression example using gibbs sampling.

5.B.2 Example 2: Multilevel regression modeling

Evaluation of informative hypotheses for GLMMs is illustrated by using the two-level multilevel regression model involving the reading_achievement data. The outcome (RA; zgeread) represents the standardized test scores of $N = 10320$ students for a reading achievement test in $J = 160$ schools. This outcome is predicted by the standardized test scores of the students in a vocabulary achievement test (defined in the article as V; and here as zgevocab), the standardized measures of the age of students (A; zage), the nominal variable gender (G; gender), and their two-way interactions (i.e., A * V; zage $*$ zgevocab, A * G; zage $*$ gender, and V * G; zgevocab $*$ gender). The vocabulary test scores of the students may be influenced by the school they belong. Therefore, these scores are considered to be nested within the schools using $(1 + \text{zgevocab} \mid \text{school})$, where the number "1" denotes the random intercept across the schools and the operator "|" specifies that the standardized vocabulary test achievement scores are nested within the schools.

Maximum likelihood estimation

Maximum likelihood estimation for this model is performed using the **lmer** function in the package lm4. Note that, in **lmer**, one needs to specify REML = FALSE to obtain maximum likelihood estimates of (standardized) structural parameters and their covariance matrix.

```
set.seed(111)
reading_ach <- read.csv("reading_achievement.csv")
#Standardizing the continuous variables
zgeread <- scale(reading_ach$geread)
zage <- scale(reading_ach$age)
zgevocab <- scale(reading_ach$gevocab)
#The variable gender is defined as a factor variable
```

```
reading_ach$gender <- as.factor(reading_ach$gender)
#The multilevel regression model is fitted to the data using maximum
#likelihood estimation
model <- lmer(zgeread ~ zage + zgevocab + gender + zage * zgevocab +
 zage * gender + zgevocab * gender + (1 + zgevocab | school),
 data = reading_ach, REML = FALSE)
#Obtaining the MLEs of the standardized structural parameters
strest <- summary(model)$coefficients[c(2, 3, 5, 6, 7), 1]
strest
```

```
        zage          zgevocab    zage:zgevocab     zage:gender2 zgevocab:gender2
-0.007818124      0.519438764      0.030511093     -0.011227740      0.016414439
```

```
#Obtaining the covariance matrix of these MLEs
strcovmtrx <- vcov(model)[c(2, 3, 5, 6, 7), c(2, 3, 5, 6, 7)]
strcovmtrx
```

```
                      zage      zgevocab zage:zgevocab  zage:gender2 zgevocab:gender2
zage             1.434839e-04  8.854509e-06  1.631158e-05 -1.387815e-04    -1.067631e-05
zgevocab         8.854509e-06  2.803465e-04  5.536735e-06 -8.243253e-06    -1.364774e-04
zage:zgevocab    1.631158e-05  5.536735e-06  8.427495e-05  2.386934e-06    -1.147775e-05
zage:gender2    -1.387815e-04 -8.243253e-06  2.386934e-06  2.686163e-04     1.742067e-05
zgevocab:gender2 -1.067631e-05 -1.364774e-04 -1.147775e-05  1.742067e-05     2.800482e-04
```

The standardized structural parameters for this example are $\theta = (\beta_1, \beta_2, \beta_4, \beta_5, \beta_6)$ and the nuisance parameters are $\xi = (\beta_0, \beta_3, \sigma^2, \psi_0^2, \psi_1^2, \psi_{01})$. The estimates of the standardized structural parameters ($\hat{\theta}$; strest) and their covariance matrix ($\hat{\Sigma}_{\hat{\theta}}$; strcovmtrx) are chosen by using the index numbers 2, 3, 5, 6, and 7.

Nonparametric bootstrapping

Estimating (standardized) structural parameters and their covariance matrix for GLMMs using nonparametric bootstrapping is complicated. Many bootstrapping methods involve different sampling strategies, which are applicable in certain situations. One common idea is that each bootstrap sample should have the same sample size of N and the same number of groups J. Based on Roberts and Fan (2004), we sample the observations for each group in the original sample with replacement. Afterwards, these samples are combined to create B bootstrap samples of size N each containing J groups (Altınışık, Nederhof, et al., unpublished).

There is not a function that implements this procedure in R. Therefore, we provide the **covglmm** function that can be used to estimate model parameters and their covariance matrix, using nonparametric bootstrapping for two-level multilevel regression models. **covglmm** can be utilized to obtain both standardized and unstandardized parameter estimates. To be able to obtain standardized parameter estimates, continuous dependent and independent variables should be standardized in each bootstrap sample. Moreover, as is shown above in estimating model parameters using maximum likelihood estimation, categorical variables must be introduced to R, for example, using the function **fact**. Therefore, the categorical variables and/or continuous variables that are considered to be standardized in the data must be introduced

to **covglmm**. The arguments of **covglmm** with their classes and descriptions are displayed below.

Inputs in *covglmm*	Classes	Description
formula	"call"	The formula which calls the **lmer** function with REML = FALSE and data = Dataset.
B	"numeric"	The number of bootstrap samples to be generated.
cluster	"numeric"	The column index of the cluster variable in the data.
fact	"numeric"	The column index(es) of the categorical variable(s) in the data.
std	"numeric"	The column index(es) of the continuous variable(s) that one wants to standardize in the data.

Using the arguments above, **covglmm** is used to estimate the standardized structural parameters and their covariance matrix for the reading achievement data as follows.

```
set.seed(111)
reading_ach <- read.csv("reading_achievement.csv")
#Below, the source code in covglmm is reached, which is saved in
#file glmmcov.txt
source("glmmcov.txt")
#The function covglmm is utilized to estimate (standardized) model
#parameters and their covariance matrix
covar <- covglmm(lmer(geread ~ age + gevocab + gender + age * gevocab +
 age * gender + gevocab * gender + (1 + gevocab | school),
 data = reading_ach, REML = FALSE), B = 1000, cluster = c(1), fact = c(2),
 std = c(3, 4, 5) )
#Obtaining the MLEs of the standardized structural parameters
colnames(covar$boot) <- names(summary(model)$coefficients[, 1] )
strest <- apply(covar$boot, 2, mean)[c(2, 3, 5, 6, 7)]
strest
```

```
        zage          zgevocab     zage:zgevocab     zage:gender2 zgevocab:gender2
-0.006173408      0.524566421      0.030987949     -0.010877684      0.018241465
```

```
#Obtaining the covariance matrix of these MLEs
strcovmtrx <- cov(covar$boot)[c(2, 3, 5, 6, 7), c(2, 3, 5, 6, 7)]
strcovmtrx
```

```
                       zage      zgevocab zage:zgevocab  zage:gender2 zgevocab:gender2
zage              1.227524e-04  9.268654e-06  6.059932e-05 -1.009743e-04    -2.217920e-05
zgevocab          9.268654e-06  2.046284e-04  4.849501e-06  3.673119e-07    -2.164760e-04
zage:zgevocab     6.059932e-05  4.849501e-06  1.217255e-04  4.912284e-06    -5.973925e-06
zage:gender2     -1.009743e-04  3.673119e-07  4.912284e-06  2.071148e-04     2.719023e-05
zgevocab:gender2 -2.217920e-05 -2.164760e-04 -5.973925e-06  2.719023e-05     4.331091e-04
```

In **covglmm**, the variables that will and will not be standardized within the method are introduced to the model. Because the variables are standardized in each bootstrap sample and not in the original sample, using the column indexes of the variables in std. For example, $\text{std} = \text{c}(3, 4, 5)$ specifies that the continuous variables (namely, age, gread, and gevocab) are standardized in $\text{B} = 1000$ bootstrap samples, which have the column index numbers 3, 4, and 5, respectively. Similarly, $\text{cluster} = \text{c}(1)$ states that the clustering variable (namely, school) is in the first column of the data and $\text{fact} = \text{c}(2)$ specifies that the categorical variable (namely, gender) is in the second column of the data. Note that the $\text{B} = 1000$ bootstrap samples can be displayed on the R console by simply typing covar$boot. Once the model parameters are estimated, similar to maximum likelihood estimation, the estimates of the standardized structural parameters ($\hat{\theta}$; strest) and their covariance matrix ($\hat{\Sigma}_{\hat{\theta}}$; strcovmtrx) are chosen using the index numbers 2, 3, 5, 6, and 7.

Gibbs sampling

Estimating the standardized structural parameters and their covariance matrix for the two-level multilevel regression model using gibbs sampling is performed by the function **bugs** in the package R2OpenBUGS. The use of this function is explained for the poisson regression model representing GLMs in the previous subsection. The same estimation procedure can be utilized for the reading_ach data. To avoid repetition, the model file for this data set is not displayed, but instead only provided in the text file bugsmodelglmm.txt. Here, we provide the rest of the code in estimating the standardized structural parameters and their covariance matrix for the two-level multilevel regression model.

```
set.seed(111)
reading_ach <- read.csv("reading_achievement.csv")
#Introducing the sample size N = 10320 and the variables in the
#model file
N <- nrow(reading_ach)
school <- reading_ach[, 1]
gender <- model.matrix(~ factor(reading_ach$gender))[, 2]
zage <- as.numeric(scale(reading_ach$age))
zgevocab <- as.numeric(scale(reading_ach$gevocab))
zgeread <- as.numeric(scale(reading_ach$geread))
data <- list("N", "school", "gender", "zage", "zgevocab", "zgeread")
#Initializing the model parameters
beta0 <- c(rnorm(1, 0, 1e-05))
beta <- c(rnorm(6, 0, 1e-05))
inits <- function(){
 list(beta0 = beta0, beta = beta, prec.sigma2 = 1, prec.tau2 = 1) }
#Performing gibbs sampling based on the data and model parameters
gibbs.sim <- bugs(data, inits, model.file = "bugsmodelglmm.txt",
 parameters = c("beta0", "beta", "prec.sigma2", "prec.tau2"),
 n.iter = 10000, n.burnin = 1000, n.chains = 3, debug = TRUE,
```

```
 codaPkg = TRUE)
#Obtaining the estimates of the standardized structural parameters
codaobject <- as.matrix(read.bugs(gibbs.sim))
colnames(codaobject) <- c("zage", "zgevocab", "gender2",
 "zage:zgevocab", "zage:gender2", zgevocab:gender2", "Intercept",
 "deviance", "prec.sigma2", "prec.tau2")
strest <- apply(codaobject, 2, mean)[c(1, 2, 4, 5, 6)]
strest
```

```
       zage         zgevocab    zage:zgevocab     zage:gender2 zgevocab:gender2
-0.009939674      0.511902893      0.024692233     -0.010516558      0.018092626
```

```
#Obtaining the covariance matrix of these estimates
strcovmtrx <- cov(codaobject)[c(1, 2, 4, 5, 6), c(1, 2, 4, 5, 6)]
strcovmtrx
```

```
                        zage       zgevocab zage:zgevocab  zage:gender2 zgevocab:gender2
zage              1.448195e-04  9.997490e-06  1.459698e-05 -1.394528e-04    -1.116991e-05
zgevocab          9.997490e-06  1.379665e-04  7.464077e-06 -9.108301e-06    -1.327628e-04
zage:zgevocab     1.459698e-05  7.464077e-06  8.251127e-05  2.819134e-06    -1.272293e-05
zage:gender2     -1.394528e-04 -9.108301e-06  2.819134e-06  2.705450e-04     1.650929e-05
zgevocab:gender2 -1.116991e-05 -1.327628e-04 -1.272293e-05  1.650929e-05     2.743202e-04
```

Gibbs sampling is a more computer extensive estimation method than maximum likelihood estimation and nonparametric bootstrapping. Moreover, the sample size $\mathsf{N} = 10320$ for the reading achievement data is quite large. Performing gibbs sampling for the two-level multilevel regression model to estimate the standardized structural parameters and their covariance matrix, using 30000 samples with a burn-in period of 3000 requires a considerable amount of time. For the sake of time, instead of using 30000 samples with a burn-in period of 3000, we used 10000 samples with a burn-in period of 1000 above in **bugs**. Note that, different from maximum likelihood estimation and nonparametric bootstrapping, index numbers 1, 2, 4, 5, and 6 have to be used to obtain the estimates of the standardized structural parameters and their covariance matrix using gibbs sampling, because the intercept is separately defined from the other model parameters.

5.B.3 Example 3: Structural equation modeling

This example focuses on the wechsler data to illustrate the evaluation of informative hypotheses for SEMs. The structural equation model used to relate the observed and unobserved variables and the dependent and independent variables for this data set is shown below:

```
SEM.model <- '
Cry = ~ y1 + y2 + y3 + y4
Fld =  ~ y2 + y3 + y5 + y6 + y7 + y8
Cry ~ edc + age
Fld ~ edc + age
'
```

Note that, based on the lavaan notation in SEM.model, unobserved (latent) variables (i.e., Cry and Fld) and indicators (i.e., the y's) are related to each other by the operator "= ∼" and these latent variables and independent variables (i.e., the categorical variable edc with two levels non high school graduate and high school graduate and the continuous variable age) are related to each other by the operator "∼".

Maximum likelihood estimation

Maximum likelihood estimation is performed using the function **cfa** in the package lavaan. This function uses SEM.model to obtain maximum likelihood estimates and their covariance matrix for the wechsler data. Both observed and unobserved (latent) variables are standardized to ensure the comparability of the parameters on the same scale. These standardized estimates are obtained using the function **standardizedSolution**. Note that, this function standardizes both observed and unobserved (latent) variables automatically, and raw values of the indicators and independent variables should be introduced in SEM.model as shown above.

The SEM is performed for three sets of informative hypotheses. These hypotheses contain factor loadings, regression coefficients, and the covariance (and therefore also the correlation) between the two latent variables. Below, it is displayed how to estimate the standardized structural parameters and their covariance matrix for the three types of parameters, respectively.

```
wechsler <- read.csv("wechsler.csv")
fit <- cfa(SEM.model, data = wechsler, std.lv = TRUE)
#Estimating factor loadings and their covariance matrix
strest1 <- standardizedSolution(fit)[1:10]
strcovmtrx1 <- lavInspect(fit, "vcov.std.all")[1:10, 1:10]
names(strest1) <- colnames(strcovmtrx1)
strest1
```

```
  Cry=~y1   Cry=~y2   Cry=~y3   Cry=~y4   Fld=~y2   Fld=~y3   Fld=~y5   Fld=~y6   Fld=~y7   Fld=~y8
0.8855355 0.7250211 0.5100878 0.9382192 0.1322106 0.3933931 0.8256388 0.8168043 0.7813435 0.7531984
```

```
strcovmtrx1
```

```
             Cry=~y1      Cry=~y2      Cry=~y3      Cry=~y4      Fld=~y2      Fld=~y3      Fld=~y5      Fld=~y6      Fld=~y7      Fld=~y8
Cry=~y1  4.107501e-05  1.055471e-05  8.660294e-06 -5.332108e-07  4.836609e-07 -1.256708e-06 6.999081e-06 7.237563e-06 8.095467e-06 8.665745e-06
Cry=~y2  1.055471e-05  4.420730e-04  5.945684e-05 -6.450562e-06 -4.674774e-04 -5.874568e-05 6.626202e-06 6.682066e-06 6.967029e-06 7.198359e-06
Cry=~y3  8.660294e-06  5.945684e-05  4.700270e-04 -2.889223e-06 -5.709730e-05 -4.542653e-04 4.179139e-06 3.776406e-06 2.597125e-06 1.947664e-06
Cry=~y4 -5.332108e-07 -6.450562e-06 -2.889223e-06  2.272956e-05  1.263566e-05  7.102624e-06 4.318660e-06 4.447553e-06 4.920254e-06 5.238977e-06
Fld=~y2  4.836609e-07 -4.674774e-04 -5.709730e-05  1.263566e-05  5.854693e-04  7.082613e-05 3.275643e-06 3.552616e-06 4.467238e-06 5.034435e-06
Fld=~y3 -1.256708e-06 -5.874568e-05 -4.542653e-04  7.102624e-06  7.082613e-05  5.073110e-04 5.026750e-06 5.750155e-06 8.079547e-06 9.491802e-06
Fld=~y5  6.999081e-06  6.626202e-06  4.179139e-06  4.318660e-06  3.275643e-06  5.026750e-06 8.600375e-05 1.449676e-05 1.740701e-05 1.924305e-05
Fld=~y6  7.237563e-06  6.682066e-06  3.776406e-06  4.447553e-06  3.552616e-06  5.750155e-06 1.449676e-05 9.187618e-05 1.831782e-05 2.018514e-05
Fld=~y7  8.095467e-06  6.967029e-06  2.597125e-06  4.920254e-06  4.467238e-06  8.079547e-06 1.740701e-05 1.831782e-05 1.174981e-04 2.343257e-05
Fld=~y8  8.665745e-06  7.198359e-06  1.947664e-06  5.238977e-06  5.034435e-06  9.491802e-06 1.924305e-05 2.018514e-05 2.343257e-05 1.398284e-04
```

```
#Estimating regression coefficients and their covariance matrix
strest2 <- standardizedSolution(fit)[11:14]
strcovmtrx2 <- lavInspect(fit, "vcov.std.all")[11:14, 11:14]
names(strest2) <- colnames(strcovmtrx2)
strest2
```

```
  Cry~edc    Cry~age    Fld~edc    Fld~age
0.5797923  0.1469647  0.3871731 -0.3680861
```

```
strcovmtrx2
```

```
             Cry~edc      Cry~age      Fld~edc      Fld~age
Cry~edc 2.777597e-04 5.777265e-05 1.986075e-04 5.896618e-05
Cry~age 5.777265e-05 4.672375e-04 9.720086e-05 2.607484e-04
Fld~edc 1.986075e-04 9.720086e-05 4.049795e-04 1.965670e-04
Fld~age 5.896618e-05 2.607484e-04 1.965670e-04 4.143894e-04
```

```
#Estimating the covariance between the latent variables and its variance
strest3 <- standardizedSolution(fit)[25]
strcovmtrx3 <- lavInspect(fit, "vcov.std.all")[23, 23]
names(strest3) <- colnames(lavInspect(fit, "vcov.std.all"))[23]
rownames(strcovmtrx3) <- colnames(lavInspect(fit, "vcov.std.all"))[23]
colnames(strcovmtrx3) <- colnames(lavInspect(fit, "vcov.std.all"))[23]
strest3
```

```
 Cry~~Fld
0.7628486
```

```
strcovmtrx3
```

```
              Cry~~Fld
Cry~~Fld 0.0002252619
```

One needs to use std.lv = TRUE in **cfa** and vcov.std.all in **lavInspect** to obtain the covariance matrix of all standardized estimates. The estimates and their covariance matrix for the factor loadings and regression coefficients are chosen by the index numbers from 1 to 10 and from 11 to 14, respectively. The estimate of the covariance between the latent variables is chosen by the index number 25 and its variance estimate is chosen by the index number 23.

Note that both observed and unobserved (latent) variables are standardized in the function standardizedSolution automatically, which gives the standardized parameter estimates for factor loadings, regression coefficients, the covariances between latent variables, and the variances of these latent variables. However, sometimes the function lavInspect does not provide the covariance matrix of these estimates based on the parameters used in the specification of the hypotheses under evaluation. The covariance matrices of factor loadings, regression coefficients, and the covariances between latent variables can be obtained as shown above. However, when the hypotheses contain the variances of latent variables, a different standardization must be used to obtain their covariance matrix. For example, the covariance matrix containing the variances and covariance of the two latent variables for the example above using maximum likelihood estimation can be obtained as follows.

```
fit <- cfa(SEM.model, data = wechsler)
covmtrx <- as.matrix(lavInspect(fit, "vcov.std.all"))[21:23, 21:23]
covmtrx
```

```
             Cry~~Cry     Fld~~Fld     Cry~~Fld
Cry~~Cry 3.361585e-04 1.410198e-04 3.649966e-05
Fld~~Fld 1.410198e-04 3.658837e-04 3.488875e-05
Cry~~Fld 3.649966e-05 3.488875e-05 2.252619e-04
```

Nonparametric bootstrapping

We utilize the **bootstrapLavaan** function in the package lavaan to estimate model parameters and their covariance matrix using nonparametric bootstrapping for SEMs. In **bootstrapLavaan**, R = 1000 bootstrap samples are created using the ordinary nonparametric bootstrapping method with type = "nonparametric".

```
boot <- bootstrapLavaan(fit, R = 1000, type = "nonparametric",
 FUN = function(x) { standardizedSolution(x)$est }, verbose = TRUE,
 warn = TRUE)
#Estimating factor loadings and their covariance matrix
strest1 <- apply(boot, 2, mean)[1:10]
strcovmtrx1 <- cov(boot)[1:10, 1:10]
#Giving names to estimates for readibility
names(strest1) <- colnames(as.matrix(lavInspect(fit,
 "vcov.std.all")[1:10,1:10] ))
rownames(strcovmtrx1) <- colnames(as.matrix(lavInspect(fit,
 "vcov.std.all")[1:10,1:10] ))
colnames(strcovmtrx1) <- colnames(as.matrix(lavInspect(fit,
 "vcov.std.all")[1:10,1:10] ))
strest1
```

```
  Cry=~y1   Cry=~y2   Cry=~y3   Cry=~y4   Fld=~y2   Fld=~y3   Fld=~y5   Fld=~y6   Fld=~y7   Fld=~y8
0.8855181 0.7258680 0.5103059 0.9381830 0.1306195 0.3929143 0.8254179 0.8162671 0.7808728 0.7526907
```

```
strcovmtrx1
```

```
             Cry=~y1      Cry=~y2      Cry=~y3      Cry=~y4      Fld=~y2      Fld=~y3      Fld=~y5      Fld=~y6      Fld=~y7      Fld=~y8
Cry=~y1  4.744296e-05 3.666030e-06 1.247571e-05 1.738177e-06 8.833648e-06 -4.959260e-06 1.508574e-05 4.983509e-06 1.491882e-05 7.922728e-06
Cry=~y2  3.666030e-06 5.174419e-04 1.764009e-04 -1.497351e-05 -5.421392e-04 -1.733360e-04 9.938031e-06 2.180733e-05 2.576630e-05 7.236012e-06
Cry=~y3  1.247571e-05 1.764009e-04 5.941198e-04 3.236106e-06 -1.686294e-04 -5.613832e-04 1.244862e-05 3.140470e-05 6.080451e-06 2.857373e-05
Cry=~y4  1.738177e-06 -1.497351e-05 3.236106e-06 3.325242e-05 2.420582e-05 7.730115e-06 6.477122e-06 6.709585e-06 6.815175e-06 9.274457e-06
Fld=~y2  8.833648e-06 -5.421392e-04 -1.686294e-04 2.420582e-05 6.931870e-04 1.936501e-04 1.082089e-06 -2.543151e-06 -5.194236e-06 1.368607e-05
Fld=~y3 -4.959260e-06 -1.733360e-04 -5.613832e-04 7.730115e-06 1.936501e-04 6.070378e-04 4.035230e-07 -1.608530e-05 7.807759e-06 -6.740697e-06
Fld=~y5  1.508574e-05 9.938031e-06 1.244862e-05 6.477122e-06 1.082089e-06 4.035230e-07 1.124518e-04 3.479980e-06 1.508682e-05 3.754491e-05
Fld=~y6  4.983509e-06 2.180733e-05 3.140470e-05 6.709585e-06 -2.543151e-06 -1.608530e-05 3.479980e-06 1.288953e-04 2.211789e-05 5.905314e-05
Fld=~y7  1.491882e-05 2.576630e-05 6.080451e-06 6.815175e-06 -5.194236e-06 7.807759e-06 1.508682e-05 2.211789e-05 1.304608e-04 1.116156e-05
Fld=~y8  7.922728e-06 7.236012e-06 2.857373e-05 9.274457e-06 1.368607e-05 -6.740697e-06 3.754491e-05 5.905314e-05 1.116156e-05 1.975515e-04
```

```
#Estimating regression coefficients and their covariance matrix
strest2 <- apply(boot, 2, mean)[11:14]
strcovmtrx2 <- cov(boot)[11:14, 11:14]
#Giving names to estimates for readibility
names(strest2)<-colnames(as.matrix(lavInspect(fit,
```

```
 "vcov.std.all")[11:14,11:14] ))
rownames(strcovmtrx2)<-colnames(as.matrix(lavInspect(fit,
 "vcov.std.all")[11:14,11:14] ))
colnames(strcovmtrx2)<-colnames(as.matrix(lavInspect(fit,
 "vcov.std.all")[11:14,11:14] ))
strest2
```

```
   Cry~edc    Cry~age    Fld~edc    Fld~age
 0.5789066  0.1462441  0.3870952 -0.3679593
```

```
strcovmtrx2
```

```
             Cry~edc      Cry~age      Fld~edc      Fld~age
Cry~edc 3.409729e-04 0.0000388883 0.0002636541 6.847774e-05
Cry~age 3.888830e-05 0.0004968670 0.0001021448 2.730563e-04
Fld~edc 2.636541e-04 0.0001021448 0.0004982744 2.286612e-04
Fld~age 6.847774e-05 0.0002730563 0.0002286612 4.742253e-04
```

```
#Estimating the covariance between the latent variables and its variance
strest3 <- apply(boot, 2, mean)[25]
strcovmtrx3 <- cov(boot)[23, 23]
#Giving names to estimates for readibility
names(strest3) <- colnames(lavInspect(fit, "vcov.std.all"))[23]
rownames(strcovmtrx3) <- colnames(lavInspect(fit, "vcov.std.all"))[23]
colnames(strcovmtrx3) <- colnames(lavInspect(fit, "vcov.std.all"))[23]
strest3
```

```
 Cry~~Fld
0.7629687
```

```
strcovmtrx3
```

```
             Cry~~Fld
Cry~~Fld 0.0004075161
```

Note that the same index numbers are used to choose the estimates for the factor loadings, regression coefficients, and the covariance between the latent variables.

Gibbs sampling

Gibbs sampling for SEMs is implemented in R using the function **bsem** in the package blavaan. In **bsem**, we generate 27000 samples along with 3 chains after discarding a burn-in period of 3000.

```
fit <- bsem(SEM.model, data = wechsler, n.chains = 3, burnin = 3000,
 sample = 27000, std.lv = TRUE)
```

```
#Estimating factor loadings and their covariance matrix
strest1 <- standardizedSolution(fit)[1:10]
strcovmtrx1 <- lavInspect(fit, "vcov.std.all")[1:10, 1:10]
#Giving names to estimates for readibility
names(strest1) <- colnames(strcovmtrx1)
strest1
```

```
  Cry=~y1   Cry=~y2   Cry=~y3   Cry=~y4   Fld=~y2   Fld=~y3   Fld=~y5   Fld=~y6   Fld=~y7   Fld=~y8
0.8859250 0.7247043 0.5108712 0.9387127 0.1331300 0.3929819 0.8256774 0.8168072 0.7811679 0.7533454
```

```
strcovmtrx1
```

```
             Cry=~y1      Cry=~y2      Cry=~y3      Cry=~y4      Fld=~y2      Fld=~y3      Fld=~y5      Fld=~y6      Fld=~y7      Fld=~y8
Cry=~y1  4.042776e-05 9.624032e-06 7.969514e-06 8.089500e-07 -4.322644e-07 -2.110906e-06 7.768578e-06 7.593084e-06 9.232299e-06 9.066359e-06
Cry=~y2  9.624032e-06 4.531463e-04 7.682969e-05 -8.570526e-06 -4.783259e-04 -7.820886e-05 2.085381e-06 1.080010e-05 8.713979e-06 9.360411e-06
Cry=~y3  7.969514e-06 7.682969e-05 4.848823e-04 -2.156366e-06 -7.284366e-05 -4.687075e-04 2.600699e-06 1.203725e-05 -7.788089e-07 1.520519e-05
Cry=~y4  8.089500e-07 -8.570526e-06 -2.156366e-06 2.286881e-05  1.558558e-05  7.635001e-06 4.114173e-06 4.807945e-06 4.072127e-06 5.404009e-06
Fld=~y2 -4.322644e-07 -4.783259e-04 -7.284366e-05 1.558558e-05  5.963787e-04  9.046443e-05 8.516243e-06 -1.222252e-06 1.766972e-06 4.477832e-06
Fld=~y3 -2.110906e-06 -7.820886e-05 -4.687075e-04 7.635001e-06  9.046443e-05  5.217378e-04 5.096734e-06 -2.608237e-06 1.019347e-05 -1.703211e-06
Fld=~y5  7.768578e-06 2.085381e-06 2.600699e-06 4.114173e-06  8.516243e-06  5.096734e-06 8.861255e-05  1.281546e-05 1.773032e-05 2.470985e-05
Fld=~y6  7.593084e-06 1.080010e-05 1.203725e-05 4.807945e-06 -1.222252e-06 -2.608237e-06 1.281546e-05 9.588256e-05 1.384951e-05 3.387411e-05
Fld=~y7  9.232299e-06 8.713979e-06 -7.788089e-07 4.072127e-06  1.766972e-06  1.019347e-05 1.773032e-05 1.384951e-05 1.202175e-04 1.501590e-05
Fld=~y8  9.066359e-06 9.360411e-06 1.520519e-05 5.404009e-06  4.477832e-06 -1.703211e-06 2.470985e-05 3.387411e-05 1.501590e-05 1.463407e-04
```

```
#Estimating regression coefficients and their covariance matrix
strest2 <- standardizedSolution(fit)[11:14]
strcovmtrx2 <- lavInspect(fit, "vcov.std.all")[11:14, 11:14]
#Giving names to estimates for readibility
names(strest2) <- colnames(strcovmtrx2)
strest2
```

```
  Cry~edc    Cry~age    Fld~edc    Fld~age
0.5813934  0.1517420  0.3904284 -0.3628171
```

```
strcovmtrx2
```

```
             Cry~edc      Cry~age      Fld~edc      Fld~age
Cry~edc 2.707919e-04 3.953203e-05 1.895937e-04 0.0000473055
Cry~age 3.953203e-05 4.501127e-04 8.908486e-05 0.0002505971
Fld~edc 1.895937e-04 8.908486e-05 4.037316e-04 0.0001923255
Fld~age 4.730550e-05 2.505971e-04 1.923255e-04 0.0004088266
```

```
#Estimating the correlation between the latent variables and its variance
strest3 <- standardizedSolution(fit)[25]
strcovmtrx3 <- lavInspect(fit, "vcov.std.all")[23, 23]
#Giving names to estimates for readibility
names(strest3) <- colnames(lavInspect(fit, "vcov.std.all"))[23]
rownames(strcovmtrx3) <- colnames(lavInspect(fit, "vcov.std.all"))[23]
colnames(strcovmtrx3) <- colnames(lavInspect(fit, "vcov.std.all"))[23]
strest3
```

```
 Cry~~Fld
0.7615458
```

```
strcovmtrx3
```

```
               Cry~~Fld
Cry~~Fld 0.0002364444
```

5.C Evaluation of informative hypotheses using GORICA

The GORICA evaluation of informative hypotheses is performed using two functions: **ormle** and **gorica**. In this section, we elaborate on these two functions by means of evaluating the sets of hypotheses for the three examples.

5.C.1 Example 1: Poisson regression modeling (Continued)

After selecting the estimates and their covariance matrix for structural parameters, the function **ormle** is used to obtain the order-restricted estimates, which are the estimates that are in accordance with the restrictions of the hypotheses under evaluation. For the poisson regression example, the four informative hypotheses including the unconstrained hypothesis H_u are:

$$
\begin{aligned}
H_1 &: \beta_3 = (\beta_3 + \beta_4) = (\beta_3 + \beta_5) = 0,\\
H_2 &: (\beta_3 + \beta_4) > (\beta_3 + \beta_5) > \beta_3,\\
H_3 &: (\beta_3 + \beta_4) - (\beta_3 + \beta_5) > (\beta_3 + \beta_4) - \beta_3,\\
H_u &: \beta_3, \beta_4, \beta_5.
\end{aligned}
$$

Using the arguments above, **ormle** should be applied analogously regardless of the estimation technique used to evaluate the hypotheses. Note that the arguments constr, rhs, and nec in **ormle** are identical for each estimation method. The only difference is in est and covmtrx, which are obtained using different estimation techniques. For simplicity, we only show the results with respect to maximum likelihood estimation below.

```
#Below, the source code in ormle is reached, which is saved in
#file restrictedest.txt
source("restrictedest.txt")

#Hypothesis 1
constr <- matrix(c(1, 0, 0,
                   1, 1, 0,
                   1, 0, 1), nrow = 3, ncol = 3, byrow = TRUE)
rhs <- rep(0, 3)
nec <- 3
H1 <- ormle(est = strest, covmtrx = strcovmtrx, const = constr,
 nec = nec, rhs = rhs)
H1
```

```
$est
                zmath   progAcademic:zmath progVocational:zmath
            0.4121808             0.2661545             0.2145273


$restrictedest
                zmath   progAcademic:zmath progVocational:zmath
         0.000000e+00        -2.775558e-16         0.000000e+00


#Hypothesis 2
constr <- matrix(c(0, 1, -1,
                   0, 0, 1), nrow = 2, ncol = 3, byrow = TRUE)
rhs <- rep(0, 2)
nec <- 0
H2 <- ormle(est = strest, covmtrx = strcovmtrx, const = constr,
 nec = nec, rhs = rhs)
H2

$est
                zmath   progAcademic:zmath progVocational:zmath
            0.4121808             0.2661545             0.2145273


$restrictedest
                zmath   progAcademic:zmath progVocational:zmath
            0.4121808             0.2661545             0.2145273


#Hypothesis 3
constr <- matrix(c(0, 0, -1), nrow = 1, ncol = 3, byrow = TRUE)
rhs <- rep(0, 1)
nec <- 0
H3 <- ormle(est = strest, covmtrx = strcovmtrx, const = constr,
 nec = nec, rhs = rhs)
H3

$est
                zmath   progAcademic:zmath progVocational:zmath
            0.4121808             0.2661545             0.2145273


$restrictedest
                zmath   progAcademic:zmath progVocational:zmath
            0.5749089             0.1034265             0.0000000
```

```
#Below, constr contains only zeros for the unconstrained hypothesis
#$H_u$, because it does not consist of any restriction on
#the structural parameters
constr <- matrix(c(rep(0, 3)), nrow = 1, ncol = 3, byrow = TRUE)
rhs <- rep(0, 1)
nec <- 0
Hu <- ormle(est = strest, covmtrx = strcovmtrx, const = constr,
 nec = nec, rhs = rhs)
Hu
```

```
$est
              zmath   progAcademic:zmath progVocational:zmath
          0.4121808            0.2661545            0.2145273


$restrictedest
              zmath   progAcademic:zmath progVocational:zmath
          0.4121808            0.2661545            0.2145273
```

The two lines for each hypothesis under evaluation display the results for the maximum likelihood estimates and the order-restricted maximum likelihood estimates, respectively. Note that in case of the hypothesis under evaluation contains both equality and inequality restrictions, the equality restrictions have to be specified before the inequality restrictions. Thus, based on the number of equality restrictions specified by nec, **ormle** knows which lines in constr and which elements in rhs correspond to equality and inequality restrictions, respectively.

Once the restricted estimates are obtained as shown above, the function **gorica** is used to obtain the misfits and complexities, the values of the GORICA, and GORICA weights for hypotheses H_1, H_2, H_3, and the unconstrained hypothesis H_u. There are two arguments in **gorica** which are shown below with their descriptions.

Inputs in *gorica*	Classes	Description
H1, H2, ..., Hm	"ormle"	Provides the order-restricted estimates which are in agreement with the restrictions of the hypotheses under evaluation.
iter	"numeric"	The number of iterations performed to obtain the complexity part of the GORICA.

The command lines below can be utilized to duplicate our results using each estimation technique.

```
set.seed(111)
#Below, the source code in gorica is reached, which is saved in
#file Gorica.txt
```

```
source("Gorica.txt")
#Below, gorica is utilized to obtain the misfits, complexities,
#the values of the GORICA, and GORICA weights for the four hypotheses
#under evaluation
gorica(H1, H2, H3, Hu, iter = 100000)
      misfit complexity    gorica gorica_weights
H1 40.548291    0.00000 40.548291          0.0000
H2 -3.271562    3.70344  0.431878          0.5567
H3 -3.093120    4.99712  1.904000          0.2667
Hu -3.271562    6.00000  2.728438          0.1766
```

5.C.2 Example 2: Multilevel regression modeling (Continued)

For the two-level multilevel regression example, the four informative hypotheses including the unconstrained hypothesis H_u are:

$$\begin{aligned}
H_1 &: \beta_1 = 0, && \beta_2 > 0, \\
&\quad \beta_1 + \beta_5 = 0, && \beta_2 + \beta_6 > 0,\ \beta_4 > 0, \\
H_2 &: \beta_1 < 0, && \beta_2 > 0, \\
&\quad \beta_1 + \beta_5 < 0, && \beta_2 + \beta_6 > 0,\ \beta_4 > 0, \\
H_3 &: \beta_1 > 0, && \beta_2 > 0, \\
&\quad \beta_1 + \beta_5 < 0, && \beta_2 + \beta_6 < 0,\ \beta_4 = 0, \\
H_u &: \beta_1, \beta_2, \beta_4, \beta_5, \beta_6.
\end{aligned} \tag{5.14}$$

The order-restricted estimates of the standardized β's in these hypotheses with respect to each estimation method are obtained using the R code below. Similarly, the results are displayed for only maximum likelihood estimation. But the same code in which est and covmtrx are obtained using nonparametric bootstrapping and gibbs sampling can be used to attain the corresponding order-restricted estimates.

```
source("restrictedest.txt")
#Hypothesis 1
constr <- matrix(c(1, 0, 0, 0, 0,
                   1, 0, 0, 1, 0,
                   0, 1, 0, 0, 0,
                   0, 0, 1, 0, 0,
                   0, 1, 0, 0, 1), nrow = 5, ncol = 5, byrow = TRUE)
rhs <- rep(0, 5)
nec <- 2
H1 <- ormle(est = strest, covmtrx = strcovmtrx, const = constr,
 nec = nec, rhs = rhs)
H1
```

```
$est
           zage         zgevocab    zage:zgevocab     zage:gender2 zgevocab:gender2
   -0.007818124      0.519438764      0.030511093     -0.011227740      0.016414439


$restrictedest
           zage         zgevocab    zage:zgevocab     zage:gender2 zgevocab:gender2
   0.000000e+00     5.199661e-01     3.393959e-02    -1.734723e-18     1.682464e-02
```

```
#Hypothesis 2
constr <- matrix(c(-1, 0, 0, 0, 0,
                    0, 1, 0, 0, 0,
                    0, 0, 1, 0, 0,
                   -1, 0, 0, -1, 0,
                    0, 1, 0, 0, 1), nrow = 5, ncol = 5, byrow = TRUE)
rhs <- rep(0, 5)
nec <- 0
H2 <- ormle(est = strest, covmtrx = strcovmtrx, const = constr,
 nec = nec, rhs = rhs)
H2
```

```
$est
           zage         zgevocab    zage:zgevocab     zage:gender2 zgevocab:gender2
   -0.007818124      0.519438764      0.030511093     -0.011227740      0.016414439


$restrictedest
           zage         zgevocab    zage:zgevocab     zage:gender2 zgevocab:gender2
   -0.007818124      0.519438764      0.030511093     -0.011227740      0.016414439
```

```
#Hypothesis 3
constr <- matrix(c(0, 0, 1, 0, 0,
                    1, 0, 0, 0, 0,
                    0, 1, 0, 0, 0,
                   -1, 0, 0, -1, 0,
                    0, -1, 0, 0, -1), nrow = 5, ncol = 5, byrow = TRUE)
rhs <- rep(0, 5)
nec <- 1
H3 <- ormle(est = strest, covmtrx = strcovmtrx, const = constr,
 nec = nec, rhs = rhs)
H3
```

```
$est
          zage         zgevocab    zage:zgevocab     zage:gender2 zgevocab:gender2
-0.007818124       0.519438764      0.030511093     -0.011227740       0.016414439


$restrictedest
          zage         zgevocab    zage:zgevocab     zage:gender2 zgevocab:gender2
0.00000000       0.24774971       0.00000000      -0.04197598       -0.24774971

#The unconstrained hypothesis
constr <- matrix(c(rep(0, 5)), nrow = 1, ncol = 5, byrow = TRUE)
rhs <- rep(0, 1)
nec <- 0
Hu <- ormle(est = strest, covmtrx = strcovmtrx, const = constr,
 nec = nec, rhs = rhs)
Hu
$est
          zage         zgevocab    zage:zgevocab     zage:gender2 zgevocab:gender2
-0.007818124       0.519438764      0.030511093     -0.011227740       0.016414439


$restrictedest
          zage         zgevocab    zage:zgevocab     zage:gender2 zgevocab:gender2
-0.007818124       0.519438764      0.030511093     -0.011227740       0.016414439
```

Similarly, the misfits, complexities, the values of the GORICA, and GORICA weights are obtained as follows:

```
set.seed(111)
source("Gorica.txt")
gorica(H1, H2, H3, Hu, iter = 100000)
      misfit complexity    gorica gorica_weights
H1 -31.59713     3.3401 -28.25703         0.3273
H2 -34.65032     5.1200 -29.53032         0.6187
H3 985.95985     3.7381 989.69795         0.0000
Hu -34.65032    10.0000 -24.65032         0.0539
```

5.C.3 Example 3: Structural equation modeling (Continued)

The SEM is performed for the three sets of informative hypotheses which contain factor loadings, regression coefficients, and the covariance between the two latent variables. The first set of informative hypotheses containing factor loadings are:

$$\begin{aligned} H_1 &: \lambda_2 > \lambda_5,\ \lambda_3 > \lambda_6, \lambda_1 > 0, \lambda_4 > 0, \\ H_2 &: \lambda_7 > \lambda_8 > \lambda_9 > \lambda_{10}, \\ H_3 &: H_1 \,\&\, H_2, \\ H_{u1} &: \lambda_1, \lambda_2, \lambda_3, \lambda_4, \lambda_5, \lambda_6, \lambda_7, \lambda_8, \lambda_9, \lambda_{10}. \end{aligned} \tag{5.15}$$

Below, the R code that can be implemented to reproduce our results for the SEM example is given. Similarly, the results are displayed in terms of only maximum likelihood estimation, but the same commands with different values of est and covmtrx can be used to evaluate the hypotheses using nonparametric bootstrapping and gibbs sampling.

```
#Obtaining order-restricted estimates
source("restrictedest.txt")
#Hypothesis 1
constr <- matrix(c(0, 1, 0, 0, -1, 0, 0, 0, 0, 0,
                   0, 0, 1, 0, 0, -1, 0, 0, 0, 0,
                   1, 0, 0, 0, 0, 0, 0, 0, 0, 0,
                   0, 0, 0, 1, 0, 0, 0, 0, 0, 0), nrow = 4,
 ncol = 10, byrow = TRUE)
rhs <- rep(0, 4)
nec <- 0
H1 <- ormle(est = strest1, covmtrx = strcovmtrx1, const = constr,
 nec = nec, rhs = rhs)
H1
```

```
$est
  Cry=~y1   Cry=~y2   Cry=~y3   Cry=~y4   Fld=~y2   Fld=~y3   Fld=~y5   Fld=~y6   Fld=~y7   Fld=~y8
0.8855355 0.7250211 0.5100878 0.9382192 0.1322106 0.3933931 0.8256388 0.8168043 0.7813435 0.7531984


$restrictedest
  Cry=~y1   Cry=~y2   Cry=~y3   Cry=~y4   Fld=~y2   Fld=~y3   Fld=~y5   Fld=~y6   Fld=~y7   Fld=~y8
0.8855355 0.7250211 0.5100878 0.9382192 0.1322106 0.3933931 0.8256388 0.8168043 0.7813435 0.7531984
```

```
#Hypothesis 2
constr <- matrix(c(0, 0, 0, 0, 0, 0, 1, -1, 0, 0,
                   0, 0, 0, 0, 0, 0, 0, 1, -1, 0,
                   0, 0, 0, 0, 0, 0, 0, 0, 1, -1), nrow = 3,
 ncol = 10, byrow = TRUE)
rhs <- rep(0, 3)
nec <- 0
H2 <- ormle(est = strest1, covmtrx = strcovmtrx1, const = constr,
 nec = nec, rhs = rhs)
H2
```

```
$est
  Cry=~y1   Cry=~y2   Cry=~y3   Cry=~y4   Fld=~y2   Fld=~y3   Fld=~y5   Fld=~y6   Fld=~y7   Fld=~y8
0.8855355 0.7250211 0.5100878 0.9382192 0.1322106 0.3933931 0.8256388 0.8168043 0.7813435 0.7531984


$restrictedest
  Cry=~y1   Cry=~y2   Cry=~y3   Cry=~y4   Fld=~y2   Fld=~y3   Fld=~y5   Fld=~y6   Fld=~y7   Fld=~y8
0.8855355 0.7250211 0.5100878 0.9382192 0.1322106 0.3933931 0.8256388 0.8168043 0.7813435 0.7531984
```

```
#Hypothesis 3
constr <- matrix(c(0, 1, 0, 0, -1, 0, 0, 0, 0, 0,
                   0, 0, 1, 0, 0, -1, 0, 0, 0, 0,
```

```
                    1, 0, 0, 0, 0, 0, 0, 0, 0, 0,
                    0, 0, 0, 1, 0, 0, 0, 0, 0, 0,
                    0, 0, 0, 0, 0, 0, 1, -1, 0, 0,
                    0, 0, 0, 0, 0, 0, 0, 1, -1, 0,
                    0, 0, 0, 0, 0, 0, 0, 0, 1, -1), nrow = 7,
 ncol = 10, byrow = TRUE)
rhs <- rep(0, 7)
nec <- 0
H3 <- ormle(est = strest1, covmtrx = strcovmtrx1, const = constr,
 nec = nec, rhs = rhs)
H3
```

```
$est
  Cry=~y1   Cry=~y2   Cry=~y3   Cry=~y4   Fld=~y2   Fld=~y3   Fld=~y5   Fld=~y6   Fld=~y7   Fld=~y8
0.8855355 0.7250211 0.5100878 0.9382192 0.1322106 0.3933931 0.8256388 0.8168043 0.7813435 0.7531984

$restrictedest
  Cry=~y1   Cry=~y2   Cry=~y3   Cry=~y4   Fld=~y2   Fld=~y3   Fld=~y5   Fld=~y6   Fld=~y7   Fld=~y8
0.8855355 0.7250211 0.5100878 0.9382192 0.1322106 0.3933931 0.8256388 0.8168043 0.7813435 0.7531984
```

```
#The unconstrained hypothesis
constr <- matrix(c(rep(0, 10)), nrow = 1, ncol = 10, byrow = TRUE)
rhs <- rep(0, 1)
nec <- 0
Hu1 <- ormle(est = strest1, covmtrx = strcovmtrx1, const = constr,
 nec = nec, rhs = rhs)
Hu1
```

```
$est
  Cry=~y1   Cry=~y2   Cry=~y3   Cry=~y4   Fld=~y2   Fld=~y3   Fld=~y5   Fld=~y6   Fld=~y7   Fld=~y8
0.8855355 0.7250211 0.5100878 0.9382192 0.1322106 0.3933931 0.8256388 0.8168043 0.7813435 0.7531984

$restrictedest
  Cry=~y1   Cry=~y2   Cry=~y3   Cry=~y4   Fld=~y2   Fld=~y3   Fld=~y5   Fld=~y6   Fld=~y7   Fld=~y8
0.8855355 0.7250211 0.5100878 0.9382192 0.1322106 0.3933931 0.8256388 0.8168043 0.7813435 0.7531984
```

```
#Evaluation of the hypotheses containing factor loadings
set.seed(111)
source("Gorica.txt")
gorica(H1, H2, H3, Hu1, iter = 100000)
```

```
       misfit complexity    gorica gorica_weights
H1   -73.67345   16.02832 -57.64513         0.1129
H2   -73.67345   16.19468 -57.47877         0.1039
H3   -73.67345   12.19462 -61.47883         0.7677
Hu1  -73.67345   20.00000 -53.67345         0.0155
```

The second set of informative hypotheses containing regression coefficients are:

$$\begin{aligned} H_4 &: \gamma_1 > \gamma_3 > 0, \\ H_5 &: \gamma_1 > \gamma_3 > 0,\ \gamma_2 > 0,\ \gamma_4 < 0, \\ H_{u2} &: \gamma_1, \gamma_2, \gamma_3, \gamma_4. \end{aligned} \tag{5.16}$$

The functions **ormle** and **gorica** for these hypotheses are executed respectively as follows:

```
#Obtaining order-restricted estimates
source("restrictedest.txt")
#Hypothesis 4
constr <- matrix(c(1, 0, 0, 0,
                   0, 0, 1, 0,
                   1, 0, -1, 0), nrow = 3, ncol = 4, byrow = TRUE)
rhs <- rep(0, 3)
nec <- 0
H4 <- ormle(est = strest2, covmtrx = strcovmtrx2, const = constr,
 nec = nec, rhs = rhs)
H4
```

```
$est
   Cry~edc    Cry~age    Fld~edc    Fld~age
 0.5797923  0.1469647  0.3871731 -0.3680861


$restrictedest
   Cry~edc    Cry~age    Fld~edc    Fld~age
 0.5797923  0.1469647  0.3871731 -0.3680861
```

```
#Hypothesis 5
constr <- matrix(c(1, 0, 0, 0,
                   0, 0, 1, 0,
                   1, 0, -1, 0,
                   0, 1, 0, 0,
                   0, 0, 0, -1), nrow = 5, ncol = 4, byrow = TRUE)
rhs <- rep(0, 5)
nec <- 0
H5 <- ormle(est = strest2, covmtrx = strcovmtrx2, const = constr,
 nec = nec, rhs = rhs)
H5
```

```
$est
   Cry~edc    Cry~age    Fld~edc    Fld~age
 0.5797923  0.1469647  0.3871731 -0.3680861


$restrictedest
   Cry~edc    Cry~age    Fld~edc    Fld~age
 0.5797923  0.1469647  0.3871731 -0.3680861

#The unconstrained hypothesis
constr <- matrix(c(rep(0, 4)), nrow = 1, ncol = 4, byrow = TRUE)
rhs <- rep(0, 1)
nec <- 0
Hu2 <- ormle(est = strest, covmtrx = strcovmtrx, const = constr,
 nec = nec, rhs = rhs)
Hu2

$est
   Cry~edc    Cry~age    Fld~edc    Fld~age
 0.5797923  0.1469647  0.3871731 -0.3680861


$restrictedest
   Cry~edc    Cry~age    Fld~edc    Fld~age
 0.5797923  0.1469647  0.3871731 -0.3680861

#Evaluation of the hypotheses containing regression coefficients
source("Gorica.txt")
gorica(H4, H5, Hu2, iter = 100000)

       misfit complexity    gorica gorica_weights
H4   -25.28808    5.57478 -19.71330         0.2252
H5   -25.28808    3.28422 -22.00386         0.7078
Hu2  -25.28808    8.00000 -17.28808         0.0670
```

The third set of hypotheses containing the correlation coefficient between the two latent variables are:

$$\begin{aligned} H_6 &: \omega_{21} = 0, \\ H_7 &: \omega_{21} > 0, \\ H_8 &: \omega_{21} < 0. \end{aligned} \tag{5.17}$$

In the same manner, the functions **ormle** and **gorica** are executed respectively as follows:

```
#Obtaining order-restricted estimates
source("restrictedest.txt")
#Hypothesis 6
constr <- matrix(c(1), nrow = 1, ncol = 1, byrow = TRUE)
rhs <- rep(0,1)
nec <- 1
H6 <- ormle(est = strest3, covmtrx = strcovmtrx3, const = constr,
 nec = nec, rhs = rhs)
H6
```

```
$est
[1] 0.7628486


$restrictedest
[1] 0
```

```
#Hypothesis 7
constr <- matrix(c(1), nrow = 1, ncol = 1, byrow = TRUE)
rhs <- rep(0,1)
nec <- 0
H7 <- ormle(est = strest3, covmtrx = strcovmtrx3, const = constr,
 nec = nec, rhs = rhs)
H7
```

```
$est
[1] 0.7628486


$restrictedest
[1] 0.7628486
```

```
#Hypothesis 8
constr <- matrix(c(-1), nrow = 1, ncol = 1, byrow = TRUE)
rhs <- rep(0, 1)
nec <- 0
H8 <- ormle(est = strest3, covmtrx = strcovmtrx3, const = constr,
 nec = nec, rhs = rhs)
H8
```

```
$est
[1] 0.7628486


$restrictedest
[1] 0

#Evaluation of the hypotheses containing the correlation coefficient
#between the latent variables
set.seed(111)
source("Gorica.txt")
gorica(H5, H6, Hu, iter = 100000)
```

```
       misfit complexity     gorica gorica_weights
H6 2576.82308    0.00000 2576.82308              0
H7   -6.56037    0.99794   -5.56243              1
H8 2576.82308    1.00112 2577.82420              0
```

5.D Contingency tables

Script file GoricaCont.R utilizes nonparametric bootstrapping to evaluate the hypotheses presented in the contingency table example for the death penalty data. One needs a data file data.txt and an input file input.txt to execute the code in GoricaCont.R, which are available together with GoricaCont.R in the research archive sent with this dissertation. This section elaborates on how to create the input.txt file to duplicate the results for the contingency table example.

The input file for the contingency table example is shown below and elaborated on next:

#D	K	Seed	B	Iterations
8	4	111	1000	100000

#Parameters used in evaluation

x[1] / (x[1] + x[3])

x[2] / (x[2] + x[4])

x[5] / (x[5] + x[7])

x[6] / (x[6] + x[8])

#Number of models to be compared

3

#Number of equality and inequality constraints per model

```
 0          2
 1          1
 0          0
#Model 1
#Restriction matrix
-1  1  0  0
 0  0 -1  1
#Constants
 0  0

#Model 2
#Restriction matrix
 0  0 -1  1
-1  1  0  0
#Constants
 0  0

#Model 3 (unconstrained hypothesis)
#Restriction matrix
 0  0  0  0

#Constants
 0
```

The arguments of input.txt file with their descriptions are elaborated below.

Arguments in input.txt file	Description
$D = 8$	The number of cells in the contingency table.
$K = 4$	The number of conditional cell probabilities used in evaluation.
Seed = 111	The index number used in the random number generator.
B = 1000	The number of bootstrap samples used when obtaining the MLEs and their covariance matrix.
Iterations = 100000	The number of iterations performed to calculate the complexities of the hypotheses under evaluation.
#Parameters used in evaluation x[1] / (x[1] + x[3]) x[2] / (x[2] + x[4]) x[5] / (x[5] + x[7]) x[6] / (x[6] + x[8])	Four conditional cell probabilities are expressed as functions of the $D = 8$ cell probabilities.
#Number of models to be compared 3	The number of $M = 3$ hypotheses under evaluation, namely, hypotheses H_1, H_2, and the unconstrained hypothesis H_u.
#Number of equality and inequality constraints per model 0 2 1 1 0 0	These numbers represent the number of equality and inequality restrictions in the hypotheses under evaluation, respectively. For example, the numbers 0 and 2 in the first line mean that hypothesis H_1 does not contain any equality restrictions and it contains 2 inequality restrictions.
#Model 1 #Restriction matrix -1 1 0 0 0 0 -1 1 #Constants 0 0 ...	Labels Model 1, 2, and 3 (unconstrained hypothesis) denote hypotheses H_1, H_2, and the unconstrained hypothesis H_u, respectively. Label Restriction matrix reflects S_m and/or R_m and label Constants denotes s_m and/or r_m in equation (4) in the article for the corresponding hypothesis.

The evalution of informative hypotheses using script GoricaCont.R contains three basic steps: (1) change the working directory to the folder that contains the files data.txt and input.txt and script GoricaCont.R, (2) enter the data set into data.txt and specify the entries of the file input.txt as shown above, and (3) use **source** function to execute the code in script GoricaCont.R by source("GoricaCont.R"). Utilizing source("GoricaCont.R") based on the files data.txt and input.txt produces the following results in R:

```
NOTE: Some of the eta parameters are estimated as zero.
The hypotheses under evaluation are rewritten due to empty cell(s).

$MLEs
            p1       p2 p3         p4
[1,] 0.1137123 0.228132  0 0.02823122

$Covariance_matrix
               p1             p2 p3            p4
p1  2.131010e-04 -4.491622e-05  0 -1.450185e-06
p2 -4.491622e-05  3.683139e-03  0  3.295651e-05
p3  0.000000e+00  0.000000e+00  0  0.000000e+00
p4 -1.450185e-06  3.295651e-05  0  2.008079e-04

$Restricted_MLEs
          p1        p2 p3         p4
H1 0.1137123 0.2281320  0 0.02823122
H2 0.1139162 0.2234987  0 0.00000000
H3 0.1137123 0.2281320  0 0.02823122

   LogLikelihood Penalty GORICA_value GORICA_weight
H1      7.611729 2.01628   -11.190897     0.6232822
H2      5.627240 1.49933    -8.255820     0.1436617
H3      7.611729 3.00000    -9.223457     0.2330561
```

In $MLEs, the third conditional cell probability is $\hat{\theta}_{\mathrm{w|b}} = 0$ because of an empty cell in the contingency table. Empty cells often complicate parameter estimation when evaluating informative hypotheses in the context of contingency tables. Nevertheless, this problem is solved in script GoricaCont.R such that the hypotheses under evaluation are inspected by rewriting these hypotheses and taking into account empty cells. Readers are referred to Altınışık, Hessels, and Kuiper (unpublished, pp.17-23) who elaborate the GORICA evaluation informative hypotheses for contingency tables in the presence of empty cells.

Note that the resulting values of misfits and complexities for the hypotheses H_1, H_2, and the unconstrained hypothesis H_u (represented by hypothesis H_3 above) are implicitly there, that is, misfit = -2 * LogLikelihood and complexity = 2 * Penalty.

Chapter 6

Discussion of the GORICA evaluation of informative hypotheses[1]

This chapter provides a discussion of the GORICA evaluation of informative hypotheses. First, we discuss the position of the GORICA among other methods for the evaluation of informative hypotheses. Second, we elaborate on the strengths and limitations of the GORICA evaluation of informative hypotheses. Third, in light of these strengths and weaknesses, we compare the use of the GORICA with other approaches that can be found in the literature. This chapter will be concluded with a brief discussion.

6.1 Assessment of the position of the GORICA in evaluating hypotheses

In recent years, null hypothesis significance testing (NHST) has attracted criticism from many researchers, because it comes with conceptual problems. These problems are caused by the use of the null hypothesis and the use of p-values. The null hypothesis H_0 : "Nothing is going on" rarely represents the (directional) relationships between variables in the population of interest (Cohen, 1994). For example, the null hypothesis $H_0 : \theta_1 = \theta_2$ could state that the average heights of two groups of people in a population are exactly the same, despite this being empirically impossible. This raises the question: Why then can we not reject the null hypothesis against the alternative hypothesis H_a in many situations if it is an empirical impossibility? The answer to this question lies in the definition of p-values, which represent the probability of observing the data at hand (or data that deviate more from the null hypothesis) when the null hypothesis is assumed to be correct. Failing to reject the null hypothesis using p-values does not mean that the null hypothesis is correct, actually it means that the alternative hypothesis H_a is wrong (Royall, 1997, p.79-81). Furthermore, Cohen (1994, p.998)

[1] This chapter is written by Yasin Altınışık.

states that researchers are not interested in the probability of obtaining the observed or more extreme results when the null hypothesis is true, instead what they want to know is the probability that the null hypothesis H_0 (or the alternative hypothesis H_a) is correct based on the data at hand.

In addition to these conceptual problems, limitations of NHST often prevent researchers from directly evaluating the hypotheses that are of interest to them. Three limitations are especially important. First of all, evaluation of hypotheses using NHST is limited to only two hypotheses, namely, H_0 is evaluated against H_a. However, researchers are often interested in the evaluation of more than two hypotheses and the null hypothesis may not even be among them, because it may not be of primary interest to researchers since it does not represent one of their expectations. Secondly, the information obtained from a p-value close to 0.05 is vague. For example, researchers cannot reliably use any of the following p-values to decide whether the data set at hand is in favor of H_0 or H_a: p = .040, .051, .065, .080, etc. Thirdly, p-values cannot be used to quantify the relative importance of H_0 over H_a given the data, or vice versa. For example, $p < .01$ does not provide any information about the extent to which H_a is better than H_0.

The conceptual problems and statistical limitations of NHST have motivated researchers to develop alternative methodologies. These methodologies will enable researchers to evaluate their expectations represented in the form of informative hypotheses. There are two such methodologies: Bayesian evaluation of informative hypotheses and evaluation of informative hypotheses using a (suitable) information criterion. Bayesian evaluation of informative hypotheses uses the model selection criterion called the Bayes factor (Kass, 1993; Kass & Raftery, 1995). The Bayes factor quantifies the evidence in the data for each hypothesis (that is, *not* a dichotomous decision as is obtained from p-values) in a set of two or more competing hypotheses (which does not need to include the null hypothesis). Evaluation of informative hypothesis using the Bayes factor can be applied to a broad range of statistical models. Readers are referred to Klugkist, Laudy, and Hoijtink (2005) for analysis of variance and covariance models; Klugkist et al. (2010) for contingency tables; Mulder et al. (2009) for repeated measures analysis; Mulder, Hoijtink, and Klugkist (2010) for multivariate normal linear models; and Gu, Mulder, Deković, and Hoijtink (2014) and Gu, Mulder, and Hoijtink (in press) for a rather general class of models among which logistic regression models, contingency tables, and structural equation models. The second approach is developed in this dissertation and uses the GORICA, an information criterion, to evaluate informative hypotheses containing equality and/or inequality restrictions. Similar to the Bayes factor, the GORICA is applicable to a broad range of models: generalized linear models (GLMs) (McCullagh & Nelder, 1989), generalized linear mixed models (GLMMs) (McCullogh & Searle, 2001), structural equation models (SEMs) (Bollen, 1989), and contingency tables, and also quantifies the support in the data (like for the Bayes factor not a dichtomous decision) for each hypothesis under consideration (the null hypothesis may but does not have to be included).

6.2 Strengths and weaknesses of the GORICA

The GORICA is the end point of a sequence of developments that started with the AIC. The AIC can only be used to evaluate hypotheses specified using equality constraints among the parameters of interest, while the ORIC (Anraku, 1999) which is a modified version of the AIC can be used to evaluate informative hypotheses containing simple order restrictions (e.g., $H_1 : \theta_1 = \theta_2 = \theta_3$, $H_2 : \theta_1 = \theta_2 < \theta_3$, $H_3 : \theta_1 < \theta_2 < \theta_3$). The GORIC (Kuiper et al., 2011, 2012) is a generalization of the ORIC that can be used to evaluate informative hypotheses containing linear restrictions on model parameters. These hypotheses are of the form $H_m : R\theta \geq r$, where R is the restriction matrix imposing equality and/or inequality restrictions on parameter vector θ and r is the vector of constants in hypothesis H_m. The ORIC and GORIC produce exactly the same results when evaluating informative hypotheses containing simple order restrictions in the context of ANOVA models. Currently, our approach is the most general information criterion compared to the AIC, ORIC, and GORIC when evaluating informative hypotheses. The GORICA mimics the performance of the GORIC for normal linear models and has the advantage that it can be applied to a broader range of model, namely, GLMs, GLMMs, SEMs, and contingency tables.

The GORICA does not suffer from the limitations of NHST using p-values. The GORICA enables researchers to simultaneously evaluate two or more (informative) hypotheses. The evidence of support in the data for these hypotheses can be quantified using the GORICA weights which are comparable to the AIC weights (Burnham & Anderson, 2002, p.75). The relative importance of the hypotheses under evaluation can be obtained via the calculation of pairwise ratios across these weights. For more details on how the GORICA is used to evaluate informative hypotheses for GLMs, GLMMs, and SEMs, see Chapter 2, pp.13-43.

The GORICA can be employed for the evaluation of complicated hypotheses containing linear functions of model parameters or (nonlinear) transformations of them. This feature is especially useful in the context of contingency tables for which hypotheses are often built using nonlinear functions of cell probabilities like odds ratios. Furthermore, the GORICA can be applied to (sparse) high-dimensional contingency tables which often contain small and/or zero cell frequencies. These small and/or zero cell frequencies often cause estimation problems when using other methods. To illustrate, consider the evaluation of hypothesis $H_1 : \eta_1 > \eta_2$, which is formulated by using two local odds ratios with $\eta = (\eta_1, \eta_2)^T = (\frac{\pi_1\pi_6}{\pi_2\pi_5}, \frac{\pi_3\pi_8}{\pi_4\pi_7})^T$. The classical log-linear analysis (Agresti, 2007; Azen & Walker, 2010) using p-values cannot be utilized to reliably evaluate this hypothesis against the null hypothesis, for example, when $\hat{\pi}_3 = 0$ and/or $\hat{\pi}_8 = 0$, and consequently, $\hat{\eta}_2 = 0$ because of empty cell(s). One may consider to add a small number to each cell of a contingency table in order to avoid estimation problems. However, these fictional observations can exert a substantial influence on the results. The GORICA provides an ideal solution to this matter without altering the observed data. This solution relies on rewriting hypotheses by taking into account empty cell(s) in contingency tables. For example, hypothesis $H_1 : \eta_1 > \eta_2$ where $\hat{\eta}_2 = 0$ because of empty cell(s) can be rewritten as $H_1 : \eta_1 > 0$, which can be easily evaluated using the GORICA. For more details about this solution and how the GORICA is used to evaluate

informative hypotheses in the context of contingency tables, see Chapter 3, pp.45-81.

The GORICA in the context of contingency tables can be used to address the problem of scale dependency across replication studies. Dependent variables in the original study and replication studies are often measured on different scales for different groups of participants. These scale dependencies across studies often prevent researchers from replicating their initial results. The GORICA is used to evaluate the same set of informative hypotheses in the context of contingency tables to solve the problem of scale dependency between the original study and replication studies. When studies involve continuous variables, these variables are discretized into categories to make them compatible across studies. Two replication strategies exist: researchers are interested in evaluating hypotheses originating from a theory with observed data; or, to verify the results of original study findings translated into informative hypotheses using the data from a replication study. To attain these objectives, this dissertation provided two coherent and useful replication strategies in both of which the GORICA is utilized to evaluate hypotheses. However, note that, the averaged GORICA weights across replication studies is used as an ad-hoc solution to determine the overall support in the original and replicated data sets for each hypothesis under evaluation. Future studies should aim to find a better solution for combining the GORICA weights across studies both in the context of contingency tables and the other models elaborated in this dissertation. For more details on how the GORICA is used to solve the problem of scale dependency across studies for non-replication, see Chapter 4, pp.83-95.

Although the GORICA has many desirable features, there are still a number of features that require further investigation. First of all, although a simulation study (see Chapter 2, pp.23-26) showed that the GORICA mimics the GORIC when the goal is to select the best of a set of competing informative hypotheses in the context of the normal linear model, there is no formal proof of this empirical evidence in or outside the context of the normal linear model. Secondly, the performance of the GORICA on selecting the best hypothesis is dependent on the choice of technique used in estimating model parameters and their covariance matrix. We utilized three methods in this dissertation: maximum likelihood estimation (MLE) (Fisher, 1922), nonparametric bootstrapping (Efron & Tibshirani, 1993), and Gibbs sampling (D. Spiegelhalter, 1994). For more details on how to implement these estimation techniques together with the GORICA to evaluate informative hypotheses in R, see Chapter 5, pp.97-151. Because the GORICA is an approximation of the GORIC assuming large sample sizes, one should be cautious when using these techniques in conjunction with the GORICA for small samples since the performance of the GORICA for small samples has not been fully investigated yet. Thirdly, another situation in which the performance of the GORICA has not been thoroughly evaluated is when the GORICA is used to evaluate (co)variance components of a model. For example, researchers should be careful when using the GORICA to evaluate (co)variance components of a multilevel regression model, because their standard errors for small samples may be biased (Mass & Hox, 2005, p.86). In such a case nonparametric bootstrapping with bias correction (Efron & Tibshirani, 1993; Hesterberg, 2015) may be considered as a reasonable alternative to the three estimation methods elaborated above. However, using bias corrected nonparametric bootstrapping estimates together with

the GORICA for the inspection of informative hypotheses has not been evaluated yet.

6.3 Comparison of the GORICA with other approaches

This section elaborates on the comparison of the GORICA with some other approaches in the literature that can be used to evaluate (informative) hypotheses, which are chi-bar-squared type approaches (Silvapulle & Sen, 2004, pp.75-78), the BIC (Raftery, 1995), DIC (D. J. Spiegelhalter, Best, Carlin, & van der Linde, 2002), and the Bayes factor (Kass, 1993; Kass & Raftery, 1995). The GORICA evaluation of informative hypotheses is superior over evaluating them using chi-bar-squared based approaches. Chi-bar-squared based approaches can be used to evaluate the null hypothesis $H_0 : R\theta = 0$ against the informative hypothesis $H_a : R\theta > r$, where R is the restriction matrix corresponding to both hypotheses, θ is the vector of parameters, and r is the vector of constants in H_a. This test statistic relies on p-values and thus suffers from the limitations of the classical NHST, while the GORICA does not have these limitations. For example, researchers need to conduct multiple testing procedures when they consider to use chi-bar-squared based approaches to evaluate two informative hypotheses. Furthermore, direct comparison of both informative hypotheses is not possible. However, the GORICA evaluates the set of these hypotheses directly and can be used to quantify the support in the data for each one of them without the need for multiple testing procedures and with an unproblematic comparison of both informative hypotheses.

The GORICA evaluation of informative hypotheses is also superior to the use of other information criteria such as the BIC and the DIC. Similar to the limitation of the AIC, both the BIC and the DIC can be utilized to evaluate equality constrained hypotheses but not inequality constrained hypotheses. The prior-adapted BIC (PBIC) (Romeijn, Van de Schoot, & Hoijtink, 2010) is a generalization of the BIC that can be used to compare encompassing hypotheses with constrained hypotheses that are nested within these encompassing hypotheses (e.g., evaluation of constrained hypotheses $H_1 : \theta_1 < \theta_2 < \theta_3$ and $H_2 : \theta_1 < \theta_2, \theta_3$ and encompassing hypothesis $H_3 : \theta_1, \theta_2, \theta_3$). These hypotheses differ in terms of inequality constraints between the parameters (but not on functions of them) and do not differ in terms of dimensionality, that is, the numbers of parameters in hypotheses are the same. More recently, van de Schoot, Hoijtink, Romeijn, and Brugman (2012) elaborate that both the DIC and a modification called the prior predictive DIC fail in evaluating inequality constrained hypotheses. They propose the prior information criterion (PIC) that can be used to evaluate inequality constrained hypotheses. However, the class of these inequality restrictions is limited in the sense that only regression coefficients themselves (but not functions of them) can be compared to each other and constants. The PBIC and the PIC have not attracted much attention from researchers. The GORICA which is a generalization of the GORIC (Kuiper et al., 2011, 2012) should be preferred over these information criteria, because it has been well-researched, applied, and made available to researchers.

The GORICA is more robust than any of these information criteria in the sense that it can be used to evaluate equality and/or inequality constrained hypotheses. The class of re-

strictions in hypotheses evaluated using the GORICA can be in the form of not only model parameters themselves but also linear or nonlinear functions of them. Furthermore, the prior-adapted BIC, the DIC, and the PIC are Bayesian information criteria that require the specifications of appropriate prior distributions before using them to evaluate hypoheses. The GORICA has a straightforward and easy-to-follow procedure to evaluate hypotheses in which such prior specifications are not necessary.

In this paragraph, we compare the GORICA to the Bayesian approach using the Bayes factor presented in Gu et al. (2014, in press). First of all, both of these methods are approximate procedures and they can be used to evaluate informative hypotheses under the same class of statistical models. Both methods require the estimates of model parameters used in hypotheses and their covariance matrix in order to be able to evaluate the hypotheses. However, these estimates, their covariance matrix, and the hypotheses under evaluation are all that is needed for the GORICA, while for the Bayes factor, additionally, a prior distribution for the parameters needs to be specified. Although often default prior distributions are used, these are not undisputed (see, for example, Hoijtink, van Kooten, and Hulsker (2016), and the discussions that follow). The GORICA does not need the specification of prior distributions. Furthermore, caution should be taken when the parameters are bounded (e.g., probabilities in the context of contingency tables are bounded between the numbers 0 and 1). Although Gu et al. (2014, p.515) state that their Bayes factor takes into account the bounded nature of probabilities if the hypotheses belong to an equivalent set (Hoijtink, 2012, p.202), this has not been fully investigated for inequality constrained hypotheses that do not belong to equivalent sets. In this respect, the GORICA is further developed because it can be used to evaluate very complicated hypotheses containing (functions of) cell probabilities in the context of contingency tables (for more details see Chapter 3, pp.47-50). All in all, both the GORICA and the Bayes factor can be valuable for the evaluation of informative hypotheses. The choice of which method used to evaluate informative hypotheses is determined by whether one is more a frequentist or a Bayesian statistician, and this choice is one major dispute between researchers that we cannot and did not solve in this dissertation.

6.4 Discussion

The GORICA only needs the estimates of model parameters used in hypotheses and their covariance matrix as input for the evaluation of informative hypotheses. When these parameters are regression coefficients, they need to be standardized to make them comparable. This dissertation focuses more on nonparametric bootstrapping for estimation of the parameters of interest and their covariance matrix, rather than focusing on maximum likelihood estimation and Gibbs sampling. Because the standard softwares (e.g., MPlus, R, SAS, Amos, EQS, Lisrel) using MLE as the default option do not always produce the covariance matrix of standardized parameter estimates, while Gibbs sampling does not produce MLEs when the posterior distribution is not normally distributed. The evaluation of informative hypotheses using nonparametric bootstrapping is a reliable choice in many cases, because it provides the MLEs

of (standardized) model parameters and their covariance matrix even when model assumptions are not met. Nonparametric bootstrapping is especially attractive when the asymptotic results are invalid and/or the data are not normally distributed (Hox, 2010, p.44). However, in some circumstances matters need to be approached cautiously. As mentioned, small samples and comparison of (co)variance parameters are two issues needing further research when using the GORICA to evaluate informative hypotheses. There is one more circumstance in which researchers should be careful when using the GORICA in combination with nonparametric bootstrapping to evaluate informative hypotheses. This problem is particularly related to logistic and poisson regression models for which the (quasi) complete separation problem (Albert & Anderson, 1984) may occur in some bootstrap samples causing large parameter estimates and standard errors. The (quasi) complete separation problem occurs when the response variable separates explanatory variable(s) completely, complete separation, or to a certain level, quasi complete separation (for more details see Chapter 2, pp.39-40). Because we encountered the separation problem in only 30 out of 1000 bootstrap samples, we discarded these 30 samples when estimating the standardized model parameters and their covariance matrix for the logistic regression model. However, this ad-hoc solution should not be used automatically especially when the percentage of bootstrap samples in which the separation problem occur is high. Such cases require a better solution.

All in all, the GORICA is a great method for the evaluation of informative hypotheses in a large class of statistical models. Researchers are well advised to use it (see Chapter 2, pp.40-43; Chapter 3, pp.72-81; Chapter 5, pp.122-151) if they keep the limitations highlighted in this discussion in the back of their minds.

References

Agresti, A. (2007). *An introduction to categorical data analysis*. Hoboken, NJ: John Wiley and Sons.

Akaike, H. (1973). Information theory and an extension of the maximum likelihood principle. In B. N. Petrov & F. Csáki (Eds.), *Proc. 2nd int. symp. information theory* (p. 267-281). Budapest: Akademiai kiado.

Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transaction on Automatic Control*, *19*, 716-723.

Akkerman, S., Blokland, A., Reneerkens, O., van Goethem, N. P., Bollen, E., Gijselaers, H. J., ... Prickaerts, J. (2012). Object recognition testing: methodological considerations on exploration and discrimination measures. *Behav. Brain Res.*, *232*, 335-347.

Albert, A., & Anderson, J. A. (1984). On the existence of maximum likelihood estimates in logistic regression models. *Biometrika*, *71*, 1-10.

Algina, J., & Swaminathan, H. (2011). *The handbook of advanced multilevel data analysis*. New York, NY: Routledge.

Altınışık, Y., Hessels, R. S., & Kuiper, R. M. (unpublished). An AIC-based information criterion evaluating (in)equality constrained hypotheses for contingency tables.

Altınışık, Y., Nederhof, E., Hoijtink, H., Oldehinkel, A. J., & Kuiper, R. M. (unpublished). Evaluation of inequality constrained hypotheses using a generalization of the AIC.

Anderson, S. F., & Maxwell, S. E. (2016). There's more than one way to conduct a raplication study. Beyond statistical significance. *Psychological Methods*, *21*, 1-12.

Anraku, K. (1999). An information criterion for parameters under a simple order restriction. *Biometrika*, *86*, 141-152.

Azen, R., & Walker, C. M. (2010). *Categorical data analysis for the behavioral and social sciences*. New York, NY: Routledge: Taylor and Francis Group.

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1), 1–48. doi: 10.18637/jss.v067.i01

Bollen, K. (1989). *Structural equations with latent variables*. New York, NY: John Wiley and Sons.

Boulle, M. (2003). Khiops: A discretization method of continuous attributes with guaranteed resistance to noise. *Third International Conference, MLDM 2003 Leipzig, Germany, July 5-7, 2003 Proceedings*, *2734*, 50-64.

Bring, J. (1994). How to standardize regression coefficients. *The American Statistician*, *48*, 209-213.

Burnham, K., & Anderson, D. (2002). *Model selection and multimodel inference: A practical information-theoretic approach* (2nd ed.). New York, NY: Springer-Verlag.

Button, K., loannidis, J., Mokrysz, C., Nosek, B., Flint, J., Robinson, E., & Munafò, M. (2013). Power failure: why small sample size undermines the reliability of neuroscience. *Natura Reviews Neuroscience*, *14*, 365-376.

Catlett, J. (2005). On changing continuous attributes into ordered discrete attributes. *Machine Learning-EWSL-91*, *482*, 164-178.

Cohen, J. (1992). A power primer. *Psychological Bulletin*, *112*, 155-159.

Cohen, J. (1994). The earth is round, $p < .05$. *American Psychologist*, *49*, 997-1003.

Efron, B., & Tibshirani, R. J. (1993). *An introduction to the bootstrap*. New York: Chapman & Hall.

Eliason, S. R. (1993). *Maximum likelihood estimation*. Newbury Park, CA: Sage.

Finch, W. H., Bolin, J. E., & Kelley, K. (2014). *Multilevel modeling using r*. CRC Press 2014.

Fisher, R. A. (1922). On the mathematical foundations of theoretical statistics. *Philosophical Transactions of the Royal Society*, *222*, 309-368.

Gelman, A. (2008). Scaling regression inputs by dividing by two standard deviations. *Statis-*

tics in Medicine, *27*, 2865-2873.

Gu, X., Mulder, J., Deković, M., & Hoijtink, H. (2014). Bayesian evaluation of inequality constrained hypotheses. *Psychological Methods*, *19*, 511-527.

Gu, X., Mulder, J., & Hoijtink, H. (in press). Approximated adjusted fractional bayes factors: A general method for testing informative hypotheses. *British Journal of Mathematical and Statistical Psychology*.

Hendrickx, J. (2004). Using standardized tables for interpreting loglinear models. *Quality & Quantity*, *38*, 603-620.

Hessels, R. S., Cornelissen, T. H. W., Hooge, I. T. C., & Kemner, C. (2017). Gaze behavior to faces in dyadic interaction. *Canadian Journal of Experimental Psychology/Revue canadienne de psychologie expèrimentale*, 1-18.

Hesterberg, T. C. (2015). What teachers should know about the bootstrap: Resampling in the undergraduate statistics curriculum. *The American Statistician*, *69*, 371-386.

Hoijtink, H. (2012). *Informative hypotheses: Theory and practice for behavioural and social scientists*. Boca Raton, FL: Chapman and Hall/CRC.

Hoijtink, H., van Kooten, P., & Hulsker, K. (2016). Why Bayesian psychologists should change the way they use the Bayes factor. *Multivariate Behavioral Research*, *51*, 2-10.

Hox, J. J. (2010). *Multilevel analysis: Techniques and applications* (2nd ed.). New York, NY: Routledge.

Hubbard, R. (2015). *Corrupt research: The case for reconceptualizing empirical management and social science*. Sage Publications.

Kamakura, T. (2011). Inference for the coefficient parameters of the logistic regression from a small sample. In *Int. statistical inst: Proc. 58th world statistical congress* (p. 5340-5343). Dublin.

Kanatsou, S., Karst, H., Kortesidou, D., van den Akker, R. A., den Blaauwen, J., Harris, A. P., ... Joëls, M. (2017). Overexpression of mineralocorticoid receptors in the mouse forebrain partly alleviates the effects of chronic early life stress on spatial memory, neurogenesis and synaptic function in the dentate gyrus. *Frontiers in Cellular Neuroscience*, *11*, 132.

Kanatsou, S., Kuil, L. E., Arp, M., Oitzl, M. S., Harris, A. P., Seckl, J. R., ... Joëls, M. (2015). Overexpression of mineralocorticoid receptors does not affect memory and anxiety-like behavior in female mice. *Front. Behav. Neurosci.*, *9*, 182.

Kanatsou, S., Ter Horst, J. P., Harris, A. P., Seckl, J. R., Krugers, H. J., & Joëls, M. (2016). Effects of mineralocorticoid receptor overexpression on anxiety and memory after early life stress in female mice. *Front. Behav. Neurosci.*, *9*, 374.

Kass, R. (1993). Bayes factors in practice. *Journal of the Royal Statistical Society, Series D, The Statistician*, *42*, 551-560.

Kass, R., & Raftery, A. (1995). Bayes factors. *Journal of the American Statistical Association*, *90*, 773-795.

Kline, R. B. (2010). *Principles and practice of structural equation modeling* (3rd ed.). New York, NY: The Guilford Press.

Klugkist, I., Laudy, O., & Hoijtink, H. (2005). Inequality constrained analysis of variance: A Bayesian approach. *Psychological Methods*, *10*, 447-493.

Klugkist, I., Laudy, O., & Hoijtink, H. (2010). Bayesian evaluation of inequality and equality constrained hypotheses for contingency tables. *Psychological Methods*, *15*, 281-299.

König, A. (2009). *Cognitive framing and incumbent inertia in response to technological discontinuities: A replication and extension of the gilbert study in the german book retailing industry*. Berlin: Pro Business.

Kuiper, R. M., & Hoijtink, H. (2013). A fortran90 program for the generalized order-restricted information criterion. *Journal of Statistical Software*, *54(8)*.

Kuiper, R. M., Hoijtink, H., & Silvapulle, M. J. (2011). An akaike-type information criterion for model selection under inequality constraints. *Biometrika*, *98*, 495-501.

Kuiper, R. M., Hoijtink, H., & Silvapulle, M. J. (2012). Generalization of the order-restricted information criterion for multivariate normal linear models. *Journal of Statistical Planning and Inference*, *142*, 2454-2463.

Lindsay, D. S. (2015). Replication in psychological science. *Psychol Sci.*, *12*, 1827-1832.

loannidis, J. P. A. (2005). Why most published research findings are false. *PLoS Med*, *2*(8), e124.

Mass, C. J. M., & Hox, J. J. (2005). Sufficient sample sizes for multilevel modeling. *Methodology*, *1*, 86-92.

Maxwell, S. E., Lau, M. Y., & Howard, G. S. (2015). Is psychology suffering from a replication crisis? What does"failure to replicate" really mean? *American Psychologist*, *70*, 487-498.

McArdle, J. J., & Prescott, C. A. (1992). Age-based construct validation using stuctural equation modeling. *Experimental Aging Research*, *18*, 87-115.

McCullagh, P., & Nelder, J. (1989). *Generalized linear models* (2nd ed.). Boca Raton, FL: Chapman and Hall/CRC.

McCullogh, C. E., & Searle, S. R. (2001). *Generalized linear and mixed models*. New York, NY: Wiley.

Moore, D. S., McCabe, G. P., & Craig, B. A. (2009). *Introduction to the practice of statistics* (6th ed.). New York: W. H. Freeman and Company.

Muhlenbach, F., & Rakotomalala, R. (2005). Discretization for continuous attributes. *In J Wang Editor. Encyclopedia of Data Warehousing and Mining, Idea Group Reference*, 397-402.

Mulder, J., Hoijtink, H., & Klugkist, I. (2010). Equality and inequality constrained multivariate linear models: Objective model selection using constrained posterior priors. *Journal of Statistical Planning and Inference*, *140*, 887-906.

Mulder, J., Klugkist, I., van de Schoot, R., Meeus, W., Selfhout, M., & Hoijtink, H. (2009). Bayesian model selection of informative hypotheses for repeated measurements. *Journal of Mathematical Psychology*, *53*, 530-546.

Mumby, D. G., Gaskin, S., Glenn, M. J., Schramek, T. E., & Lehmann, H. (2002). Hippocampal damage and exploratory preferences in rats: memory for objects, places and contexts. *Learn. Mem.*, *9*, 49-57.

Naninck, E. F., Hoeijmakers, L., Kakava-Georgiadou, N., Meesters, A., Lazic, S. E., Lucassen, P. J., & Korosi, A. (2015). Chronic early life stress alters developmental and adult neurogenesis and impairs cognitive function in mice. *Hippocampus*, *25*, 309-328.

Nederhof, E., Ormel, J., & Oldehinkel, A. J. (2014). Mismatch or cumulative stress: The pathway to depression is conditional on attention style. *Psychological Science*, *25*,

684-692.

Nederhof, E., & Schmidt, M. V. (2012). Mismatch or cumulative stress: Toward an integrated hypothesis of programming effects. *Physiology & Behavior*, *106*, 691-700.

Nosek, B. A., & Lakens, D. (2014). A method to increase the credibility of published results. *Social Psychology*, *45*, 137-141.

Pashler, H., & Wagenmakers, E.-J. (2012). Ediators' introduction to the special section on replicability in psychological science. A crisis of confidence? *Perspectives on Psychological Science*, *7*, 528-530.

Porte, G. (2012). *Replication research in applied linguistics*. Cambridge University Press,.

Qureshi, T., & Zighed, D. A. (2009). A soft discretization technique for fuzzy decision trees using resampling. *In: Corchado E., Yin H. (eds) Intelligent Data Engineering and Automated Learning-IDEAL 2009. IDEAL 2009. Lecture Notes in Computer Science*, *5788*.

R Development Core Team. (2012). R: A language and environment for statistical computing [Computer software manual]. Vienna, Austria. Retrieved from `http://www.R-project.org` (ISBN 3-900051-07-0)

Raftery, A. E. (1995). Bayesian model selection in social research. *Sociological Methodology*, *25*, 111-163.

Roberts, J. K., & Fan, X. (2004). Bootstrapping within the multilevel/hierarchical linear modeling framework: A primer for use with SAS and S-PLUS. *Multiple Linear Regression Viewpoints*, *30*, 23-34.

Romeijn, J. W., Van de Schoot, R., & Hoijtink, H. (2010). One size does not fit all: Derivation of an adapted bic. *Probabilities, Laws, and Structures*, *3*, 87-106.

Rosseel, Y. (2012). lavaan: an R package for structural equation modeling. *Journal of Statistical Software*, *48*(2), 1-36.

Royall, R. (1997). *Statistical evidence: A likelihood paradigm*. New York: NY: Chapman and Hall/CRC.

Rutherford, A. (2001). *Introducing anova and ancova*. London: SAGE.

Silvapulle, M., & Sen, P. (2004). *Constrained statistical inference; order, inequality, and*

shape constraints. New York: NY: Wiley.

Soetart, K., Van den Meersche, K., & Van Oevelen, D. (2014). limsolve: Solving linear inverse models [Computer software manual]. Retrieved from `http://CRAN.R-project.org/package=limSolve`

Spiegelhalter, D. (1994). *Bugs: Bayesian inference using gibbs sampling*. Cambridge, UK: MRC Biostatistics Unit.

Spiegelhalter, D. J., Best, N. G., Carlin, B. P., & van der Linde, A. (2002). Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *64*, 583-639.

Stevens, J. (1999). *Applied multivariate statistics for the social sciences* (3rd ed.). New Jersey: Lawrance Erlbaum Associates, Inc.

Stroebe, W., & Strack, F. (2014). The alleged crisis and the illusion of exact replication. *Perspectives on Psychological Science*, *9*, 59-71.

Tsai, C. J., Lee, C. I., & Yang, W. P. (2008). A discretization algorithm based on class-attribute contingency coefficient. *Information Sciences*, *178*, 714-731.

Turlach, B. A. (2014). quadprog: Functions to solve quadratic programming problems [Computer software manual]. Retrieved from `http://CRAN.R-project.org/package=quadprog`

van Bavel, J. J., Mende-Siedlecki, P., Brady, W. J., & Reinero, D. A. (2016). Contextual sensitivity in scientific reproducibility. *Proceedings of the National Academy of Sciences of the United States of America*, *113*, 6454-6459.

van de Schoot, R., Hoijtink, H., Romeijn, J.-W., & Brugman, D. (2012). A prior predictive loss function for the evaluation of inequality constrained hypotheses. *Journal of Mathematical Psychology*, *56*, 13-23.

van de Schoot, R., & Strohmeier, D. (2011). Testing informative hypotheses in SEM increases power: An illustration contrasting classical hypothesis testing with a parametric bootstrap approach. *International Journal of Behavioral Development*, *35*, 180-190.

Varadhan, R. (2015). alabama: Constrained nonlinear optimization [Computer software manual]. Retrieved from `http://CRAN.R-project.org/package=alabama`

Wagenmakers, E.-J. (2007). A practical solution to the pervasive problems of p values. *Psychonomic Bulletin and Review*, *14*, 779-804.

Webb, M. C., Wilson, J. R., & Chong, J. (2004). An analysis of quasi-complete binary data with logistic models: Applications to alcohol abuse data. *Journal of Data Science*, *2*, 273-285.

Wechsler, D. (1981). *Wais-r: Wechsler adult intelligence scale revised*. New York: Harcourt, Brace, & Jovanovich.

Weiss, D. J. (2006). *Analysis of variance and functional measurement: A practical guide*. Oxford: Oxford University Press.

Yuan, K. H., & Hayashi, K. (2006). Standard errors in covariance structure models: Asymptotic versus bootstrap. *British Journal of Mathematical and Statistical Psychology*, *59*, 397-417.

Zighed, D. A., Rabaseda, S., & Rakotomalala, R. (1998). Fusinter: A method for discretization of continuous attributes for supervised learning. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, *6*, 307-326.

Zorn, C. (2005). A solution to separation in binary response models. *Political Analysis*, *13*, 157-170.

Summary

In the social and behavioral sciences, the support in the data regarding theories and researchers' own expectations can be quantified by evaluating (in)equality constrained hypotheses. This dissertation proposes an AIC-type information criterion named the GORICA to evaluate these hypotheses under very broad range of statistical models, that is, GLMs, GLMMs, SEMs, and contingency tables. The evaluation of (in)equality constrained hypotheses requires the estimates of structural parameters which are the parameters that are used in the specification of the hypotheses under evaluation and their covariance matrix.

Chapter 2 elaborates the evaluation of (in)equality constrained hypotheses using the GORICA for GLMs, GLMMs, and SEMs for which structural parameters and their covariance matrix are estimated using nonparametric bootstrapping. In this chapter the (in)equality constrained hypotheses contain linear restrictions on the structural parameters. Chapter 3 extends the use of the GORICA such that it can be applied to evaluate (in)equality constrained hypotheses in the context of contingency tables for hypotheses containing linear and non-linear restrictions on cell probabilities. Chapter 4 elaborates the GORICA evaluation of (in)equality constrained hypotheses on a different perspective. We provided a coherent strategy to quantify replication success across studies by evaluating (in)equality constrained hypotheses using the GORICA. The same set of hypotheses are evaluated across the original study and replication studies to quantify the support in the data sets for each hypothesis under evaluation. In Chapter 5 two more alternative estimation techniques (i.e., maximum likelihood estimation and gibbs sampling) together with nonparametric bootstrapping are applied in the GORICA to evaluate (in)equality constrained hypotheses. Similar results in terms of selecting the best in a set of hypotheses are obtained using each estimation technique for a poisson regression model, multilevel regression model, and structural equation model as being representatives of GLMs, GLMMs, and SEMs, respectively. In this Chapter (in)equality constrained hypotheses are also evaluated in the context of contingency tables but only using nonparametric bootstrapping and not using maximum likelihood estimation and gibbs sampling as an illustration of the GORICA presented in Chapter 3.

We provided two easy-to-use R script files Gorica.R for GLMs, GLMMs, and SEMs and GoricaCont.R for contingency tables which can be used to duplicate the results presented in this dissertation. The user manuals explaning how to use script files are included in the dissertation. However, note that these script files apply only nonparametric bootstrapping to

estimate structural parameters and their covariance matrix. Nevertheless, the Supplementary Material elaborated in Chapter 5 can be used to apply the GORICA with each one of the three estimation techniques, namely, maximum likelihood estimation, nonparametric bootstrapping, and gibbs sampling.

Samenvatting

In de sociale en gedragswetenschappen kan de support in data voor de theorie en de eigen verwachtingen van onderzoekers gekwantificeerd worden door het evalueren van (on)gelijkheidsbeperkte hypotheses. In dit proefschrift wordt een AIC-type informatie criterium voorgesteld, genaamd de GORICA, om deze hypotheses te evalueren onder een breed scala aan statistische modellen, zijnde: GLM's, GLMM's, SEM's en kruistabellen. Voor de evaluatie van (on)gelijkheidsbeperkte hypotheses zijn schattingen nodig van structurele parameters welke gebruikt worden in de specificatie van de te evalueren hypotheses en hun covariantiematrices.

In hoofdstuk 2 wordt uitgeweid over de evaluatie van (on)gelijkheidsbeperkte hypotheses met behulp van de GORICA voor GLM's, GLMM's en SEM's waarbij hun structurele parameters en covariantiematrices geschat worden door middel van de non-parametrische bootstrap. In dit hoofdstuk bevatten de (on)gelijkheidsbeperkte hypotheses lineaire restricties op de structurele parameters. Hoofdstuk 3 breidt het gebruik van de GORICA uit zodat deze toegepast kan worden om (on)gelijkheidsbeperkte hypotheses te evalueren in de context van kruistabellen voor hypotheses die lineaire en non-lineaire restricties op celwaarschijnlijkheden bevatten. In hoofdstuk 4 wordt de GORICA evaluatie van (on)gelijkheidsbeperkte hypotheses vanuit een ander perspectief beschouwd. We noemen een coherente strategie om het replicatie-succes van studies te kwantificeren door het evalueren van (on)gelijkheidsbeperkte hypotheses met behulp van de GORICA. Dezelfde verzameling hypotheses worden geëvalueerd over de originele en replicatiestudies om de ondersteuning van elke te evalueren hypothese door de data te kwantificeren. In hoofdstuk 5 wordt aanvullend twee alternatieve schattingstechnieken (zijnde de maximum likelihood methode en gibbs sampling) samen met non-parametrische bootstrapping toegepast in de GORICA om (on)gelijkheidsbeperkte hypotheses te evalueren. Vergelijkbare resultaten met betrekking tot het selecteren van de beste in een verzameling hypotheses worden behaald door elke schattingstechniek te gebruiken voor een poisson regressiemodel, een multilevel regressiemodel en een structureel vergelijkingsmodel als zijnde representatief voor respectievelijk GLM's, GLMM's en SEM's. In dit hoofdstuk worden daarnaast (on)gelijkheidsbeperkte hypotheses geëvalueerd in de context van kruistabellen, daarbij enkel gebruik makende van de non-parametrische bootstrap, ter illustratie van de GORICA uit hoofdstuk 3.

Er worden twee toegankelijke R scripts gegeven, Gorica.R voor GLM's, GLMM's en

SEM's en GoricaCont.R voor kruistabellen, welke gebruikt kunnen worden om de resultaten uit dit proefschrift te dupliceren. De gebruikshandleiding voor deze scripts worden bij dit proefschrift gegeven. Echter, merk op dat deze scripts enkel de non-parametrische bootstrap gebruiken om structurele parameters en hun covariantiematrices te schattten. Desondanks kan het aanvullende materiaal uit hoofdstuk 5 gebruikt worden om de GORICA toe te passen met elk van de drie schattingstechnieken, zijnde: de maximum likelihood methode, de non-parametrische bootstrap en gibbs sampling.

Acknowledgement

I would like to convey my gratefulness to many people for helping me to complete my PhD study. I would especially like to thank my supervisor Herbert Hoijtink and co-supervisors Rebecca M. Kuiper and Rinke H. Klein Entink. Herbert, I will never forget your patience and understanding when things get complicated due to comprehensive analysis procedures. I am very happy that I had the opportunity to work with you. You encouraged me a lot through my study and your suggestions in the Action plan were very effective. Rebecca, you always had clear ideas on how to solve the problems that I have encountered during my PhD project. Sometimes I spent hours on a single sentence to make it easily understandable for readers. It was not always successful. But I was always sure that your suggestions would solve the issue (and I was right). You were always available to me whenever I needed your help. Thank you for creating an environment of warmth and geniality for me in your office. Rinke, I would like to thank you for your bright ideas into my study during Friday meetings. Thanks also for your clear guidance into one of my papers.

Then thanks to all members of the department of Methodology and Statistics. It was a pleasure to share the same working environment with my Chinese colleague Xin for one year. I express deep and sincere gratitude to one of the most kind-hearted persons I have ever met, Marielle, with whom I worked together at various stages of my study period. Also a great thanks to my other colleagues Fayette, Kees, Frank, Sanne, Jolien, Noemi, and Susanna for your friendship. Kevin, Flip, and Chantal, I appreciate the technical and administrative support that you provided me during my PhD study.

Finally, I would like to thank my mother and father for their love, compassion, and support which I feel during my whole life.

Yasin Altınışık
Utrecht, September, 2017

About the author

Yasin Altınışık was born on April 11, 1986 in Trabzon, Turkey. He obtained his BSc with specialization in Statistics and Computer Science at Black Sea Technical University, Trabzon in June 2009. During the same summer he won a full scholarship from the Ministry of National Education of Turkey to pursue an M.S. and Ph.D abroad. After studying English at University of California, Davis in 2010, he became a research master student at Penn State University in 2011 and obtained his MSc in 2013. During the last year of his MSc, he focused on intrablock, interblock, and combined estimates in incomplete block designs. In February 2014, he started his PhD research in the Department of Methodology and Statistics at Utrecht University in the Netherlands. His PhD work focused on developing a generalized information criterion based on the AIC that can be used to evaluate equality and/or inequality constrained hypotheses for a broad range of statistical models.