AMERICAN PSYCHOLOGICAL ASSOCIATION

Psychological Methods

ISSN: 1082-989X

2021, Vol. 26, No. 5, 599–621
https://doi.org/10.1037/met0000406

Evaluation of Inequality Constrained Hypotheses Using a Generalization of the AIC

Yasin Altinisik[1], Caspar J. Van Lissa[2], Herbert Hoijtink[2,3], Albertine J. Oldehinkel[4], and Rebecca M. Kuiper[2]

[1] Department of Statistics, Sinop University
[2] Department of Methodology and Statistics, Utrecht University
[3] CITO Institute for Educational Measurement, Arnhem, the Netherlands
[4] Department of Psychiatry, University of Groningen, University Medical Center Groningen

Abstract

In the social and behavioral sciences, it is often not interesting to evaluate the null hypothesis by means of a *p*-value. Researchers are often more interested in quantifying the evidence in the data (as opposed to using *p*-values) with respect to their own expectations represented by equality and/or inequality constrained hypotheses (as opposed to the null hypothesis). This article proposes an Akaike-type information criterion (AIC; Akaike, 1973, 1974) called the generalized order-restricted information criterion approximation (GORICA) that evaluates (in)equality constrained hypotheses under a very broad range of statistical models. The results of five simulation studies provide empirical evidence showing that the performance of the GORICA on selecting the best hypothesis out of a set of (in)equality constrained hypotheses is convincing. To illustrate the use of the GORICA, the expectations of researchers are investigated in a logistic regression, multilevel regression, and structural equation model.

Translational Abstract

Evaluation of Inequality Constrained Hypotheses Using a Generalization of the AIC: Researchers are interested in evaluating equality and/or inequality constrained hypotheses in the context not only of normal linear models, but also of the families outside of normal linear models using a suitable information criterion. However, the available information criteria in the literature are not capable of evaluating (in) equality constrained hypotheses under such a broad range of statistical models. The main aim of this paper is to close this research gap by proposing a new information criterion named the GORICA which can be utilized to evaluate these hypotheses for generalized linear (mixed) models and structural equation models. The GORICA enables researchers to quantify the evidence in the data for two or more (in) equality constrained hypotheses. Like all the other information criteria, the GORICA has the log likelihood and penalty parts. The superiority of the GORICA over the other information criteria lies behind the use of a simple formula when calculating its log likelihood. We investigated the performance of the GORICA on choosing the true hypothesis out of a set of competing hypotheses using simulation studies

Yasin Altinisik https://orcid.org/0000-0002-9248-2339
Albertine J. Oldehinkel https://orcid.org/0000-0003-3925-3913
Rebecca M. Kuiper https://orcid.org/0000-0002-0267-5197

Yasin Altinisik, Herbert Hoijtink, and Rebecca M. Kuiper developed the study concept. All authors contributed to the study design. Data collection was performed by Albertine J. Oldehinkel and the writing, programming, analysis, and interpretation was performed by Yasin Altinisik under the supervision of Rebecca M. Kuiper and Herbert Hoijtink. The accompanying R package GORICA is established by Caspar J. Van Lissa under the supervision of Yasin Altinisik and Rebecca M. Kuiper. All authors read, commented, and approved the final article. We thank Esther Nederhof for fruitful discussions and her comments on the article.

This research was supported by The Republic of Turkey, Ministry of National Education, the Netherlands Organization for Scientific Research (NWO; VENI Grant 451-16-019), 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 024.001.003). Some ideas in the article were presented at the CID symposium in 2015, the 25th and 26th Interuniversity Graduate School of Psychometrics and Sociometrics (IOPS) conferences in 2015 and 2016, respectively, and International Society for Bayesian Analysis (ISBA) in 2016.

The data used in the logistic regression example were collected as part of the Tracking Adolescent's Individual Lives Survey (TRAILS). Participating centers of TRAILS include various departments of the University Medical Center and University of Groningen, the University of Utrecht, the Radboud Medical Center Nijmegen, and the Parnassia Bavo group, all in the Netherlands. TRAILS has been financially supported by various grants from the Netherlands Organization for Scientific Research (NWO), ZonMW, GB-MaGW, the Dutch Ministry of Justice, the European Science Foundation, BBMRI-NL, and the participating universities. We are grateful to everyone who participated in this research or worked on this project to make it possible.

Correspondence concerning this article should be addressed to Yasin Altinisik, Department of Statistics, Sinop University, P.O. Box 57000, Nasuhbasoglu Street, Sinop, Turkey. Email: yaltinisik@sinop.edu.tr

for logistic regression, multilevel regression, and structural equation model. The findings in these simulation studies suggest that the GORICA has a convincing performance on choosing the true hypothesis. The use of the GORICA is illustrated for (real) data sets in line with these simulation studies.

Keywords: AIC, Akaike weights, GORICA, (in)equality constrained hypotheses, model selection

Supplemental materials: https://doi.org/10.1037/met0000406.supp

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 a p-value is controversial. First, the null hypothesis is often not 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 μ_1 and μ_2 are exactly equal, that is, $H_0 : \mu_1 = \mu_2$. In addition, Royall (1997, pp. 79–81) elaborates that the null hypothesis can never be accepted. For instance, failing to reject the null hypothesis $H_0 : \mu_1 = \mu_2$ against hypothesis $H_1 : \mu_1 > \mu_2$ essentially means that hypothesis H_1 is incorrect and not that the null hypothesis is correct. Second, p-values cannot measure the evidence in the data for H_0 or any other hypothesis (Wagenmakers, 2007). After rejecting the null hypothesis (e.g., p = .04), it is still not quantifiable to what degree the alternative hypothesis is better than the null hypothesis.

In our opinion, such hypotheses should only be used if researchers think they provide a plausible description of the relations in the population of interest. Clear cut examples for this can be found in Wainer (1999) and Bem (2011). Beside these examples, when there is a theory stating that the means of two groups are equal, then this should be incorporated (but only then). Notably, in many instances, we expect either order constraints or no constraints. Evaluation of (in)equality constrained hypotheses, also called informative hypotheses (Hoijtink, 2012) representing the researchers' expectations may be more interesting for researchers than evaluating null-hypotheses. To illustrate how to formulate (in) equality constrained hypotheses, consider the study conducted on a replication (N = 310) of the study in Nederhof et al. (2014), in which 11-year-old participants are divided into three 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: 0 = no depressive episode, 1 = experienced an episode) is predicted by the categorical variable early life stress (ES: 0 = low, 1 = high), the standardized continuous variable recent stress, RS, and the interaction between both predictors. The continuous variable RS is standardized because of one main reason: to improve the interpretation of main effects when interactions exist (Gelman, 2008). The resulting logistic regression model for *group j* and person *i* is:

$$f(\hat{D}_{ji}) = \theta_{j0} + \theta_{j1}\mathrm{RS}_{ji} + \theta_{j2}\mathrm{ES}_{ji} + \theta_{j3}\mathrm{RS}_{ji}\mathrm{ES}_{ji}, \tag{1}$$

where $f(.)$ denotes the logit link function, θ_{j0} denotes the intercept for *group j*, and θ_{j1}, θ_{j2}, and θ_{j3} are the group dependent coefficients of the values of the three predictors for j = 1, 2, 3 and $i = 1, 2, \ldots, N_j$, where N_j denotes the number of persons in *group j* with $\sum_{j=1}^{3} N_j = N$. Based on the values of variable ES, the model in Equation 1 can be converted to:

$$f(\hat{D}_{ji}) = \begin{cases} \theta_{j0} + \theta_{j1}\mathrm{RS}_{\mathrm{ji}} & \text{if ES} = 0\,(\text{low}) \\ (\theta_{j0} + \theta_{j2}) + (\theta_{j1} + \theta_{j3})\mathrm{RS}_{\mathrm{ji}} & \text{if ES} = 1\,(\text{high}). \end{cases} \tag{2}$$

Our hypotheses of interest are based on the studies in Nederhof and Schmidt (2012) and Nederhof et al. (2014). Both studies investigate the relationship between stress and depression based on the mismatch and cumulative stress expectations. The mismatch expectation states that the risk of depression for adolescents with low levels of early life stress increases with high recent stress levels (i.e., $\theta_{j1} > 0$) while adolescents with high levels of early life stress are not affected by high recent stress levels (i.e., $\theta_{j1} + \theta_{j3} = 0$). The cumulative stress expectation states that there is no interaction between early and recent life stress (i.e., $\theta_{j3} = 0$) and that only the main effect of recent stress predicts depression and, furthermore, that this relation is positive (i.e., $\theta_{j1} > 0$).

Based on the theory in Nederhof and Schmidt (2012), we formulated together with the authors an (in)equality constrained hypothesis, referred to as H_1. It states that the mismatch expectation applies to the sustainers and the shifters: $\theta_{j1} > 0$ if ES is low and $\theta_{j1} + \theta_{j3} = 0$ if ES is high for j = 1, 2. In contrast, the cumulative stress expectation applies to the comparison groups: $\theta_{j1} > 0$ if ES is low and $\theta_{j3} = 0, \theta_{j1} > 0$ if ES is high for j = 3. Hypothesis H_2 is based on the results presented in Nederhof et al. (2014, p. 689). It expresses that the mismatch expectation applies to the sustainers: $\theta_{j1} > 0$ if ES is low and $\theta_{j1} + \theta_{j3} = 0$ if ES is high for j = 1. Because the mismatch and the cumulative expectations do not apply, there is no interaction effect and no main effect for the shifters: $\theta_{j1} = 0$ if ES is low and $\theta_{j1} + \theta_{j3} = 0$ if ES is high for j = 2. The cumulative stress expectation applies to the comparison group: $\theta_{j1} > 0$ if ES is low and $\theta_{j3} = 0, \theta_{j1} > 0$ if ES is high for j = 3. Note that, as a safeguard against choosing a weak hypothesis as the best hypothesis out of the set of weak hypotheses, we include the unconstrained hypothesis H_3 in the set, where there are no restrictions on model parameters. In summary,

$$\begin{array}{lll} & (\textbf{Sustainers}) & (\textbf{Shifters}) & (\textbf{Comparison}) \\ H_1: & \theta_{11} + \theta_{13} = 0, \theta_{11} > 0, & \theta_{21} + \theta_{23} = 0, \theta_{21} > 0, & \theta_{33} = 0, \theta_{31} > 0, \\ H_2: & \theta_{11} + \theta_{13} = 0, \theta_{11} > 0, & \theta_{21} = \theta_{23} = 0, & \theta_{33} = 0, \theta_{31} > 0, \\ H_3: & \theta_{11}, \theta_{13}, & \theta_{21}, \theta_{23}, & \theta_{31}, \theta_{33}. \end{array} \tag{3}$$

The use of information criteria (ICs) is a well-known alternative technique against the traditional null hypothesis testing when evaluating the hypotheses of interest. An information criterion selects the best hypothesis in a set of candidate hypotheses which are evaluated for the same data set. The AIC (Akaike, 1973, 1974) is one of the most commonly used ICs to evaluate the hypotheses under consideration. The AIC selects the best of a set of hypotheses, that is, the hypothesis that has the shortest distance to the true (but unknown) 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 $\mu_1 < \ldots < \mu_J$, where "< " may be replaced by "=", or "," and J is the number of groups in the context of ANOVA. 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 following hypotheses: $H_4 : \theta_1 > 0, \theta_2 > 0, \theta_1 > \theta_2$ (two regression coefficients are larger than zero and the first is larger than the second); $H_5 : 0.5(\theta_1 + \theta_3) > \theta_2$ (the average of the means for the first and third groups $0.5(\theta_1 + \theta_3)$ is higher than the mean for the second group θ_2 in an analysis of variance); $H_6 : (\theta_1 - \theta_2) > (\theta_3 - \theta_4), \theta_2 = \theta_3$ (which specifies an interaction effect in a 2×2 analysis of variance and additionally states that the means in Group 2 and Group 3 are the same); $H_7 : \theta_1 - \theta_2 > 2$ (which specifies that the mean in Group 1 is at least 2 points larger than the mean in Group 2); and $H_8 : -2 < \theta_1 < 2, \theta_2$ (which specifies a range restriction on θ_1, which is restricted in the interval [−2, 2], and no restriction on parameter θ_2). However, the GORIC can only be applied to univariate and multivariate normal linear models.

In this article, we propose an AIC-type information criterion named the generalized order-restricted information criterion approximation (GORICA) 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; Bollen, 1989). Like other confirmatory methods (e.g., F-Bar tests, Silvapulle & Sen, 2005, pp. 25–42; the Bayes factor, Kass, 1993; Kass & Raftery, 1995; the GORIC, etc.), the GORICA requires researchers to formulate theories and/or expectations. Although researchers need to make an extra effort to formulate these theories and expectations into hypotheses, we strongly encourage them to pursue this worthwhile effort toward good research practice. Evaluating a reasonable set of (in)equality constrained hypotheses is powerful, because it increases the probability of selecting the correct hypotheses from the set (comparable to an increase in power in hypothesis testing; Kuiper & Hoijtink, 2010). The GORICA is based on large-sample theory and uses the property that the likelihood function of a statistical model can be approximated by a normal distribution (Fisher, 1922). The interpretation of the GORICA is the same as that of the AIC. The GORICA uses the GORICA weights when evaluating hypotheses, which have the same functional characteristics as the AIC weights (Burnham & Anderson, 2002, p. 75). In contrast to p-values, the GORICA weights improve the interpretability of analysis results. The GORICA weights are between 0 and 1 and they sum to one for a set of hypotheses of interest. Researchers can quantify the support in the data for a hypothesis against another hypothesis in the set. The larger the GORICA weight for hypothesis H_m, the smaller the Kullback-Leibler divergence (Kullback & Leibler, 1951) between this hypothesis and the true hypothesis when compared with that between other hypotheses in the set and the true hypothesis, and thus, the better the support by the data for hypothesis H_m. We perform simulations to compare the performance of the GORIC and GORICA for normal linear models on choosing the best hypothesis out of a set of hypotheses under evaluation. Further, we investigate the performance of the GORICA for a broader range of models, namely, for a logistic regression, multilevel regression, and structural equation model as being representatives of generalized linear models, generalized linear mixed models, and structural equation models, respectively.

Like with hypothesis testing, the study design plays a pivotal role when evaluating (in)equality constrained hypotheses using all information criteria. Campbell and Machin (1999) state that the study design is more important than analyzing the data, because the mistake made in data analysis can always be corrected while the problems that occur by implementing the wrong study design can never be rectified. The study design is strongly associated with the data quality which is influenced by data definition, collection, processing, and representation mechanisms (Richesson & Andrews, 2012, p. 177). Thus, each stage of this process also plays an essential role on correctly evaluating the hypotheses of interest. Moreover, Ioannidis (2005) elaborates on the factors causing many studies to produce false results which bias the scientific literature such as the power of study, the number of studies answering the same research question, and more than enough flexibility in designs and outcomes and so forth. Additionally, replication crisis (Pashler & Wagenmakers, 2012; Maxwell et al., 2015), publication bias (Stern & Simes, 1997; Schimmack, 2012), and some questionable research practices (Agnoli et al., 2017) such as ad-hoc specification of hypotheses of interest (Kerr, 1998) and utilizing methods which improves the probability of publication (Wagenmakers et al., 2012) also require a worthwhile attention when evaluating hypotheses using information criteria. Although all of these aspects are important, the focus of this article is only one particular aspect among these, that is, the GORICA as an alternative to hypothesis testing using p-values. Our contribution is based on the same arguments as Wagenmakers (2007) uses for the Bayesian information criterion (BIC; Raftery, 1995), but note that the GORICA can evaluate a broader class of hypotheses (namely, also order and range constraints).

The outline of the article is as follows. First, we introduce the GORIC and GORICA, respectively. Second, we elaborate on the conceptual similarities and dissimilarities between the GORIC and GORICA. Third, we illustrate how to evaluate (in)equality constrained hypotheses using the GORICA in the context of GLMs, GLMMs, and SEMs, respectively, in three consecutive sections. The illustrations in these three sections are accompanied by simulation studies to further examine the performance of the GORICA on choosing the best hypothesis. Fourth, we elaborate on how to evaluate hypotheses containing range restrictions using the GORIC and GORICA. The article concludes with a discussion.

GORIC and GORICA

GORIC

We will first introduce the GORIC (Kuiper et al., 2011) and then generalize it to the GORICA. The GORIC will be introduced using the following univariate normal linear model:

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}, \tag{4}$$

where $\boldsymbol{y} = (y_1, y_2, \ldots, y_N)^T \in \mathbb{R}^{N \times 1}$ denotes the outcome, $\boldsymbol{X} = (\boldsymbol{x}_0, \boldsymbol{x}_1, \ldots, \boldsymbol{x}_p, \ldots, \boldsymbol{x}_{P-1}) \in \mathbb{R}^{N \times P}$ with $\boldsymbol{x}_p = (x_{p1}, x_{p2}, \ldots, x_{pN})^T$

$\in \mathbb{IR}^{N\times 1}$ for $p = 0, 1, \ldots, P-1$ contains the predictors, $\boldsymbol{\beta} = (\beta_0, \beta_1, \ldots, \beta_{(P-1)})^T \in \mathbb{IR}^{P\times 1}$ are the regression coefficients, and $\boldsymbol{\epsilon} \sim N(\mathbf{0}, \sigma^2 \boldsymbol{I}_N)$ represents the vector of residuals with mean vector $\mathbf{0}$ and covariance matrix $\sigma^2 \boldsymbol{I}_N$, where σ^2 is the variance term and $\boldsymbol{I}_N$ denotes the $N \times N$ identity matrix.

In the remainder of this article, we will make a distinction between structural parameters ($\boldsymbol{\theta}$ of length $K \leq P$ for a univariate normal linear model), that is, the parameters that appear in the hypotheses of interest, and nuisance parameters (ξ of length $P - K + 1$, where the number "1" denotes the variance term σ^2 for a univariate normal linear model), that is, the parameters that do *not* appear in the hypotheses of interest. Let, for the univariate normal linear model, $\boldsymbol{\theta} = \boldsymbol{\beta}$ (and thus, $K = P$) and $\boldsymbol{\xi} = \sigma^2$. The hypotheses of interest in this paper are of the form:

$$H_m : \boldsymbol{S}_m\boldsymbol{\theta} = \boldsymbol{s}_m, \boldsymbol{R}_m\boldsymbol{\theta} > \boldsymbol{r}_m, \tag{5}$$

where m denotes the index number for the hypotheses under evaluation. Here, S_m is a $q_s \times K$ matrix and R_m is a $q_r \times K$ matrix representing the equality and inequality constraints of hypothesis H_m, respectively, and $\boldsymbol{s}_m$ is a q_s-vector and $\boldsymbol{r}_m$ is a q_r-vector containing the constants in H_m, respectively. For example, for hypothesis $H_6 : (\theta_1 - \theta_2) > (\theta_3 - \theta_4), \theta_2 = \theta_3$ in the introduction, it holds that $\boldsymbol{\theta} = (\theta_1, \theta_2, \theta_3, \theta_4)^T$, $P = K = 4$ (i.e., the number of parameters used to formulate hypothesis H_6), $q_s = 1, q_r = 1$, and

$$\begin{aligned} \boldsymbol{S}_6 &= \begin{pmatrix} 0 & 1 & -1 & 0 \end{pmatrix}, \boldsymbol{s}_6 = 0, \\ \boldsymbol{R}_6 &= \begin{pmatrix} 1 & -1 & -1 & 1 \end{pmatrix}, \boldsymbol{r}_6 = 0. \end{aligned}$$

Two important aspects need attention on formulating hypotheses using Equation 5. First, the parameters have to be measured on the same scale for a meaningful comparison among them. For example, while group means will be comparable with each other (e.g., in an ANOVA model), regression coefficients should often be standardized for comparison. This can be achieved by standardizing the variables to which they correspond (e.g., the predictors in a regression). In each of the simulations and examples presented in this article, it will be highlighted which variables were standardized. Second, $\begin{bmatrix} \boldsymbol{S}_m \\ \boldsymbol{R}_m \end{bmatrix}$ should be of full (row) 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. In case $\begin{bmatrix} \boldsymbol{S}_m \\ \boldsymbol{R}_m \end{bmatrix}$ does not contain a redundant restriction but it is still not of full (row) rank, hypotheses should be evaluated with caution using the GORIC and GORICA. A special case is when the hypotheses of interest contain range restrictions. Hypothesis $H_8 : -2 < \theta_1 < 2, \theta_2$ containing a simple range constraint is formulated by specifying

$$\boldsymbol{R}_8 = \begin{pmatrix} 1 & 0 \\ -1 & 0 \end{pmatrix}$$

with

$$\boldsymbol{r}_8 = (-2, -2).$$

Because the two rows of $\boldsymbol{R}_8$ are linearly dependent on each other, H_8 is not of full rank. In the section Range Restrictions, we describe how to evaluate hypotheses containing range restrictions using the GORIC and GORICA.

For the univariate normal linear model, the GORIC for hypothesis H_m is defined as:

$$\text{GORIC}_m = -2\,L(\tilde{\boldsymbol{\theta}}^m, \tilde{\boldsymbol{\xi}}^m \mid \boldsymbol{y}, \boldsymbol{X}) + 2\,[PT_m(\boldsymbol{\theta}) + PT_m(\boldsymbol{\xi})], \tag{6}$$

where, $\tilde{\boldsymbol{\theta}}^m$ and $\tilde{\boldsymbol{\xi}}^m$ maximize the log likelihood, $L(\boldsymbol{\theta}, \boldsymbol{\xi} \mid \boldsymbol{y}, \boldsymbol{X})$, subject to the restrictions in hypothesis H_m. The maximum log likelihood stands for the fit of hypothesis H_m to the data at hand and is defined for the univariate normal linear model as:

$$L(\tilde{\boldsymbol{\theta}}^m, \tilde{\boldsymbol{\xi}}^m \mid \boldsymbol{y}, \boldsymbol{X}) = -\frac{N}{2}\log(2\pi) - \frac{1}{2}\log|\tilde{\boldsymbol{\xi}}^m \boldsymbol{I}_N| - \frac{1}{2} [(\boldsymbol{y} - \boldsymbol{X}\tilde{\boldsymbol{\theta}}^m)^T(\tilde{\boldsymbol{\xi}}^m \boldsymbol{I}_N)^{-1}(\boldsymbol{y} - \boldsymbol{X}\tilde{\boldsymbol{\theta}}^m)]. \tag{7}$$

Like all ICs, the GORIC evaluates a hypothesis using its fit and specificity. A hypothesis has a larger fit if the maximum of the log likelihood for this hypothesis (obtained by considering the restrictions in the hypothesis) is larger than that of the other hypotheses. The specificity for hypothesis H_m is accounted for using the penalty term PT_m. It consists of two parts: $PT_m(\boldsymbol{\theta})$ denotes the expected number of distinct regression parameters after taking into account the order constraints in H_m; and $PT_m(\boldsymbol{\xi}) = 1$ because the vector $\boldsymbol{\xi}$ only contains the unconstrained residual variance σ^2 (see Appendix A).

In, for example, an ANOVA model with three means, for hypotheses $H_0 : \mu_1 = \mu_2 = \mu_3$ and $H_1 : \mu_1 > \mu_2 > \mu_3$, and the unconstrained hypothesis $H_2 : \mu_1, \mu_2, \mu_3$, it holds that $\boldsymbol{\theta} = (\theta_1, \theta_2, \theta_3)^T$, $P = K = 3$, and $PT_m(\boldsymbol{\theta})$ is, 1, 1.834, and 3, respectively (see the end of Appendix B for more details on the calculation of the penalty value of 1.834 by taking into account the restrictions in H_1). A hypothesis that has large fit (and thus a small misfit) and a small penalty will render a small GORIC value which means that this hypothesis has a relatively small Kullback-Leibler divergence to the unknown true hypothesis. Therefore, the hypothesis with the smallest GORIC value is the best in a set of hypotheses.

The GORIC has specifically been developed for hypotheses evaluation in (multivariate) normal linear models (Kuiper et al., 2011, 2012). It cannot be applied for hypotheses evaluation in the context of other statistical models. In fact, it must be rederived for each statistical model of interest. To avoid these derivations, we propose the GORICA, which can be computed using a normal approximation of the log likelihood function, which only requires the maximum likelihood estimates (MLEs) of the structural parameters and their covariance matrix. The GORICA is superior than the GORIC in terms of the simplicity of its formula and the ease of its applicability when evaluating (in)equality constrained hypotheses outside of normal linear models. In the remainder of this section, the GORICA will be presented (to keep the presentation accessible, the reader will be referred to the Appendices of this article for derivations).

GORICA

In this subsection, we show the expression of the GORICA. Its derivation can be found in Appendix A. Here, we use a normal approximation of the likelihood in Equation 6 and we can leave out all the parts related to the nuisance parameters:

$$\text{GORICA}_m = -2\,L(\tilde{\boldsymbol{\theta}}^m \mid \hat{\boldsymbol{\theta}}, \hat{\boldsymbol{\Sigma}}_{\hat{\theta}}) + 2\,PT_m(\boldsymbol{\theta}), \quad (8)$$

where $\hat{\boldsymbol{\theta}}$ and $\hat{\boldsymbol{\Sigma}}_{\hat{\theta}}$ denote the maximum likelihood estimates of the structural parameters and their covariance matrix, respectively, that are used to construct a normal approximation of the likelihood:

$$L(\tilde{\boldsymbol{\theta}}^m \mid \hat{\boldsymbol{\theta}}, \hat{\boldsymbol{\Sigma}}_{\hat{\theta}}) = -\frac{K}{2}\log(2\pi) - \frac{1}{2}\log|\hat{\boldsymbol{\Sigma}}_{\hat{\theta}}| - \frac{1}{2} (\hat{\boldsymbol{\theta}} - \tilde{\boldsymbol{\theta}}^m)^T (\hat{\boldsymbol{\Sigma}}_{\hat{\theta}})^{-1} (\hat{\boldsymbol{\theta}} - \tilde{\boldsymbol{\theta}}^m). \quad (9)$$

Note that, for many statistical models, the (standardized) estimates and their covariance matrix are very easy to obtain, for example, using the glm (R Core Team, 2012) and glmer (Bates et al., 2015) procedures for GLMs and GLMMs, and the lavaan package (Rosseel, 2012) for SEMs in R. For each H_m, the order-restricted maximum likelihood estimates $\tilde{\boldsymbol{\theta}}^m$ (that is, the best estimates of $\boldsymbol{\theta}$ when maximizing the log likelihood in Equation 9 accounting for the restrictions in H_m) have to be obtained. Note that the calculation of these estimates of the GORICA strongly resembles that of the GORIC, because both methods minimize similar objective functions taking into account the restrictions of each hypothesis under evaluation. In this sense, the GORICA is an adequate approximation of the GORIC for normal linear models when researchers have enough data to estimate model parameters, because the sampling distribution of these parameters will be approximately normal. Like the AIC and the likelihood ratio test, the GORICA is based on the central limit theory and it is known that such asymptotic methods perform well in practice for medium and large data sets. The calculation of the order-restricted MLEs for the GORICA is described in Appendix B.

The GORICA is superior than the GORIC because it can easily be applied to any statistical model due to the normal approximation used in the log likelihood part. In contrast, there is no (relative) difference between the GORIC and GORICA in computing the penalty part. In the calculation, the GORICA leaves out the number of nuisance parameters (that is, $PT_m(\boldsymbol{\xi})$ in Equation 6), which is the same for each hypothesis under consideration (e.g., $PT_m(\boldsymbol{\xi}) = 1$ for the univariate normal linear model). As shown in Appendix A, this does not influence the differences in penalties nor the comparison (i.e., relative support) of the hypotheses of interest.

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

$$w_m = \frac{\exp\{-\frac{1}{2}\,\text{GORICA}_m\}}{\sum_{m'=1}^{M} \exp\{-\frac{1}{2}\,\text{GORICA}_{m'}\}}. \quad (10)$$

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 set of hypotheses: Although H_1 is a better hypothesis than H_2, it is a weak hypothesis, because it is outperformed by the unconstrained hypothesis 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 both H_1 and H_2 are not supported by the data.

We next provide guidance on how to create appropriate sets of (in)equality constrained hypotheses along with the procedure to evaluate the hypotheses of interest. The steps below should be taken to evaluate a set of (in)equality constrained hypotheses using the GORICA.

1. Check the number of theories under consideration.

   a. If there is only one hypothesis under consideration, it should be evaluated against the complement of the hypothesis of interest representing all possible theories except the hypothesis of interest (Vanbrabant et al., 2020). Alternatively, one can use the unconstrained hypothesis representing all possible theories including the hypothesis of interest. Note that one should not include any fabricated hypothesis in the set when there is only one theory under evaluation.

   b. If there are multiple hypotheses in the set, check whether the restrictions of hypotheses cover all possible relations between parameters.

      i. When the restrictions of the hypotheses of interest cover all possible relations between parameters, do not include the unconstrained hypothesis (or the complement of a hypothesis) in the set. For example, there is no need to include the unconstrained hypothesis (or the complements of the hypotheses) in the set when evaluating hypothesis $H_1: \theta_1 > \theta_2$ against hypothesis $H_2: \theta_2 > \theta_1$, because the restrictions of these hypotheses together cover the whole parameter space.

      ii. When the restrictions of hypotheses do not cover the whole parameter space, then include a safeguard hypothesis to avoid choosing a weak hypothesis as the best hypothesis among a set of weak hypotheses. For instance, the unconstrained hypothesis (or the complements of the hypotheses) should be included in the set when evaluating hypothesis $H_1: \theta_1 > \theta_2 > \theta_3$ against hypothesis $H_2: \theta_3 > \theta_2 > \theta_1$, since H_1 and H_2 do not specify all possible relations between parameters θ_1, θ_2, and θ_3 and both hypotheses may be weak hypotheses. The gorica package (Van Lissa et al., 2020) in R provides a user-friendly way to specify the hypotheses of interest. We elaborate on how to specify the restrictions of the hypotheses under consideration for each example presented in this article using the gorica package in the online supplementary material. In the section "Range restrictions", we elaborate on how to evaluate hypotheses containing range restrictions using the goric function in the restriktor package (Vanbrabant, 2020) in R. Additionally, we compare the GORIC and GORICA in an ANOVA example for which the set also contains the complement of a hypothesis.

2. If necessary, standardize model parameters to compare them on the same scale. For example, if hypotheses contain parameters compared to only constants and/or the data contain only categorical variables, then there might not be a reason for standardizing model parameters. In contrast, if

hypotheses consist of directional relationships between model parameters and/or the model contains interactions on continuous variables, then the model parameters should be standardized to compare them on the same scale.

3. Estimate structural parameters ($\hat{\boldsymbol{\theta}}$) and their covariance matrix ($\hat{\boldsymbol{\Sigma}}_{\hat{\theta}}$). Sometimes the model fit object is sufficient to obtain these estimates and their covariance matrix, when all parameters in this object are used to formulate the hypotheses of interest.

The procedure in Steps 4–8 below is fully automatic in the gorica package and does not require intervention by users.

4. Obtain the order-restricted MLEs (see Appendix B). Notably, the gorica package checks whether the restriction matrices $\boldsymbol{S}_m$ and/or $\boldsymbol{R}_m$ for each hypothesis under evaluation are of full (row) rank. If all the restriction matrices are not of full (row) rank, our function in the package stops and gives the error message: "The restriction matrices for the hypotheses of interest have to be of full (row) rank." In such a case, we cannot obtain the order-restricted MLEs, and consequently, the results.

5. Calculate the value of the log likelihood function in Equation 9 for each hypothesis under evaluation by using the estimates of structural parameters and their covariance matrix ($\hat{\boldsymbol{\theta}}$ and $\hat{\boldsymbol{\Sigma}}_{\hat{\theta}}$) in Step 3 and the order-restricted MLEs ($\tilde{\boldsymbol{\theta}}^m$) in Step 4.

6. Compute the penalty part, $PT_m(\boldsymbol{\theta})$, for each hypothesis under consideration (see Appendix B).

7. Calculate the GORICA values for the hypotheses of interest using the equation in Equation 8.

8. Calculate the GORICA weights to determine the extent of support by the data for each hypothesis under consideration using the equation in Equation 10.

The GORIC and GORICA show some similarities and dissimilarities in evaluating hypotheses containing (in)equality constraints, which will be elaborated in the next section.

GORIC and GORICA: How They Work and How They Differ

The GORIC and GORICA share a common background depending on the AIC with a common goal: to evaluate (in)equality constrained hypotheses of the form in Equation 5. All information criteria need the data to evaluate hypotheses of interest, even though their calculations are different. The AIC uses the data and equality restrictions in the hypotheses to calculate the log likelihood part and uses the number of distinct parameters in these hypotheses to compute the penalty part. The GORICA needs the (unrestricted) MLEs and their covariance matrix from summary statistics using the data, besides the equality and/or inequality restrictions on parameters, to evaluate the hypotheses under consideration. The GORICA is an approximation of the GORIC assuming large sample sizes. The formal proof indicating that the GORICA is an adequate approximation of the GORIC is provided in Appendix A. Using the (unrestricted) MLEs and their covariance matrix, the GORICA approximates the GORIC retaining only the first and second order terms of the Taylor series expansion of the log likelihood function given in the first equation of Appendix A. Notably, the second order Taylor approximation of the log likelihood renders a normal approximation of the likelihood. Maximizing this log likelihood function for a study requires the estimates of model parameters and their covariance matrix. For example, the lm() object in R can be used to obtain the estimates of model parameters and their covariance matrix for normal linear models such as the ANOVA and multiple regression. Similarly, the glm(), glmer(), and sem() objects can be used for GLMs, GLMMs, and SEMs, respectively. In addition, we have performed two simulation studies to show that the performance of the GORIC and GORICA in the context of normal linear models are comparable (see Section 1 of the online supplementary material). The results of the simulation studies indicate that the performance of the GORIC and GORICA with respect to the probability of choosing the true hypothesis from a set of competing hypotheses are identical for large samples in the context of normal linear models. Note that the GORIC and GORICA results may be slightly different for small samples. Because the GORICA is an approximation procedure assuming large sample sizes, we recommend using the GORIC instead of the GORICA for small samples in the context of normal linear models.

The distinction between the GORIC and GORICA lies in the definition of the log likelihood part for hypothesis H_m. Suppose that a researcher aims to apply both the GORIC and GORICA to the logistic regression model in Equation 1 to evaluate the hypotheses in Equation 3. Note that the structural parameters addressed in hypotheses H_1 and H_2 are $\boldsymbol{\theta} = (\theta_{11}, \theta_{13}, \theta_{21}, \theta_{23}, \theta_{31}, \theta_{33})$ and the nuisance parameters are $\boldsymbol{\xi} = (\theta_{10}, \theta_{12}, \theta_{20}, \theta_{22}, \theta_{30}, \theta_{32})$. The GORIC needs to maximize the order-restricted log likelihood $L(\boldsymbol{\theta}, \boldsymbol{\xi} \mid \boldsymbol{y}, \boldsymbol{X})$, which is conditional on the data (i.e., $\boldsymbol{y}$), and to determine this log likelihood, it needs the conditional covariance matrix of the data from the logistic regression model, $\text{cov}(\boldsymbol{y} \mid \boldsymbol{X\beta})$, where $\boldsymbol{\beta} = \{\boldsymbol{\theta}, \boldsymbol{\xi}\}$. This is not an easy task to achieve, because the expected value of the binary data ($\boldsymbol{y}$) is related to the linear predictor ($\boldsymbol{X\beta}$) via a logit link function. Similarly, derivation of the GORIC for the other models in the general class of models in this article is complicated and cumbersome, because it involves different sets of formulas for different types of models. In contrast, the GORICA has only one simple formula for maximizing its log likelihood function, which can be used not only for the logistic regression model in Equation 1, but also for the other types of models in the general class of models.

In the sequel, we exemplify how to evaluate (in)equality constrained hypotheses using the GORICA for the logistic regression, multilevel regression, and SEM examples. For each of these examples, we used the following strategy.

1. And formulate the hypotheses under consideration.

2. Simulation: Before analyzing the data with the GORICA, we investigate the performance of the GORICA on evaluating these hypotheses for the specific statistical model in Step 1 using a simulation study.

3. Example (continued): When we know that the performance of the GORICA is good, we evaluate the example data.

Logistic Regression Modeling

In this section, we exemplify how to evaluate (in)equality constrained hypotheses with the GORICA using the study in the introduction. Before analyzing the data for this study, we want to make sure that the GORICA performs well on choosing the best hypothesis out of the three hypotheses in Equation 3. Therefore, in the next subsection, we will conduct a simulation study to evaluate the performance of the GORICA in the context of logistic regression.

Performance of the GORICA for the Logistic Regression Example

The logistic regression model in the simulation is the same model as in Equation 1. We create three different cases, where in Cases 1 and 2, the population values of the parameters are in accordance with H_1 and H_2, respectively. In Case 3, the θ's are only in accordance with the unconstrained hypothesis H_3. Table 1 displays our choices for the θ's for which it holds that if we generate $D_{ji} \sim \text{Bernoulli}(\hat{D}_{ji})$ for $j = 1, 2, 3$ and $i = 1, 2, \ldots, N_j$, the expected correct classification rate (CCR) is equal to 60%, 65%, 70%, or 75%. The correct classification rate is based on a cut-off value of 0.5 for the expected value of the outcome for the *i*th person in the *j*th group. We choose the sample size in each group as $N_j = 50, 100, 150$, or 200 for $j = 1, 2, 3$. For more details on data generation in the logistic regression simulation see the section "Simulation Steps" in the online supplementary material.

Figure 1 shows the proportion of times each of the hypotheses was selected by the GORICA in 1000 independent samples generated from each of the populations displayed in Table 1. The probability of correctly preferring hypothesis H_1 and the unconstrained hypothesis H_3 increases with the correct classification rate. The probability of correctly preferring hypothesis H_2 does not depend much on the values of the correct classification rate for this set of hypotheses. This is because hypotheses H_1 and H_2 differ from each other only with respect to parameter θ_{21}, that is, $\theta_{21} > 0$ for hypothesis H_1 and $\theta_{21} = 0$ for hypothesis H_2. Note that the performance of the GORICA is still satisfactory, that is, the true hypothesis H_2 is chosen at least 81% of the times and at most 85% of the times.

Like with increasing the correct classification rate, the probability of correctly preferring hypothesis H_1 and the unconstrained hypothesis H_3 increases with the sample size, but this increase does not depend much on the sample size when hypothesis H_2 is the true hypothesis. When sampling from a population based on the restrictions in H_2, even in large samples, the estimate of θ_{21} is not exactly equal to zero. In that case, the data may or may not be in accordance with H_1. If the data are in accordance with H_1, then the (log) likelihood for H_1 is larger than that for H_2, and if the data are not in agreement with H_1, both hypotheses have the same (log) likelihood. Thus, hypothesis H_2 is never exactly true and never better than hypothesis H_1 in terms of fit. Although hypothesis H_2 always has a lower penalty than hypothesis H_1, sometimes when the data are in

Table 1
Logistic Regression Simulation: Population Values of the Regression Coefficients for CCR =60%, 65%, 70%, and 75% for the Three Cases

Case 1: $H_1: \theta_{11}+\theta_{13}=\theta_{21}+\theta_{23}=\theta_{33}=0, \{\theta_{11}, \theta_{21}, \theta_{31}\} > 0$ is true

	Group 1				Group 2				Group 3			
CCR	θ_{10}	θ_{11}	θ_{12}	θ_{13}	θ_{20}	θ_{21}	θ_{22}	θ_{23}	θ_{30}	θ_{31}	θ_{32}	θ_{33}
60%	−0.15	0.62	−0.10	−0.62	−0.15	0.62	−0.10	−0.62	−0.15	0.61	−0.10	0.00
65%	−0.15	1.09	−0.10	−1.09	−0.15	1.09	−0.10	−1.09	−0.15	1.20	−0.10	0.00
70%	−0.15	2.22	−0.10	−2.22	−0.15	2.22	−0.10	−2.22	−0.15	1.61	−0.10	0.00
75%	−0.15	3.22	−0.10	−3.22	−0.15	3.22	−0.10	−3.22	−0.15	3.19	−0.10	0.00

Case 2: $H_2: \theta_{11}+\theta_{13}=\theta_{21}=\theta_{23}=\theta_{33}=0, \{\theta_{11}, \theta_{31}\} > 0$ is true

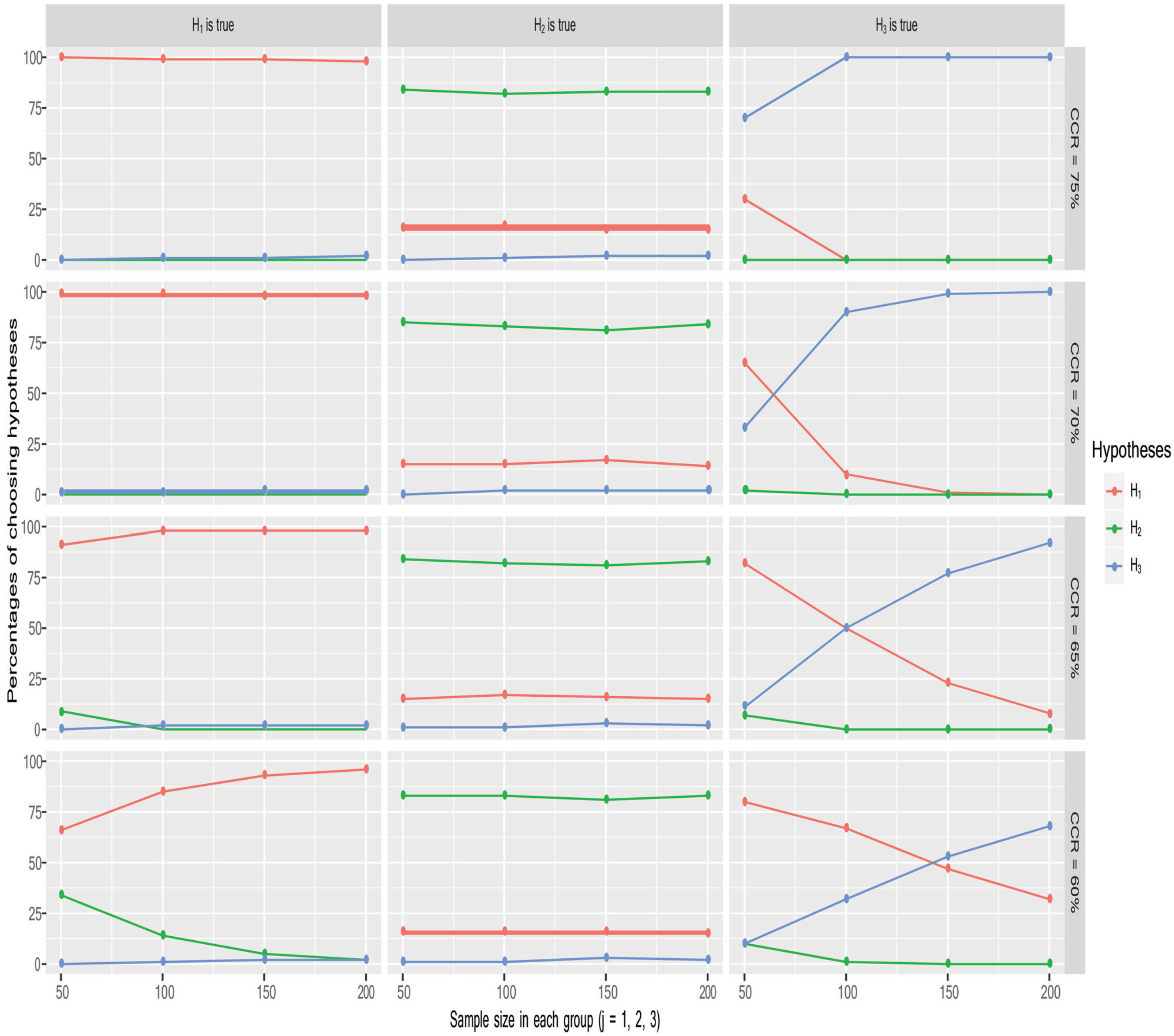
	Group 1				Group 2				Group 3			
CCR	θ_{10}	θ_{11}	θ_{12}	θ_{13}	θ_{20}	θ_{21}	θ_{22}	θ_{23}	θ_{30}	θ_{31}	θ_{32}	θ_{33}
60%	0.66	0.62	−0.55	−0.62	0.66	0.00	−0.55	0.00	0.66	0.31	−0.55	0.00
65%	0.66	0.62	−0.55	−0.62	0.66	0.00	−0.55	0.00	0.66	1.71	−0.55	0.00
70%	0.97	1.90	−0.65	−1.90	0.97	0.00	−0.65	0.00	0.97	1.70	−0.65	0.00
75%	0.97	2.90	−0.65	−2.90	0.97	0.00	−0.65	0.00	0.97	4.78	−0.65	0.00

Case 3: $H_3: \theta_{11}, \theta_{13}, \theta_{21}, \theta_{23}, \theta_{31}, \theta_{33}$ is true

	Group 1				Group 2				Group 3			
CCR	θ_{10}	θ_{11}	θ_{12}	θ_{13}	θ_{20}	θ_{21}	θ_{22}	θ_{23}	θ_{30}	θ_{31}	θ_{32}	θ_{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
65%	0.14	1.18	−0.02	−0.55	0.14	1.10	−0.02	−0.51	0.14	0.92	−0.02	−0.30
70%	0.14	1.58	−0.02	−0.55	0.14	1.40	−0.02	−0.51	0.14	1.32	−0.02	−0.30
75%	0.14	2.09	−0.02	−0.55	0.14	1.90	−0.02	−0.51	0.14	1.74	−0.02	−0.30

Note. CCR = correct classification rate.

Figure 1
Percentages of Choosing Hypotheses for the Logistic Regression Simulation



Note. Percentage of times that the hypotheses H_1, H_2, and H_3 are selected by the GORICA for the four classification rates (CCR = 60%, 65%, 70%, and 75%) for the three cases in which H_1, H_2, and H_3 are the true hypothesis, respectively. See the online article for the color version of this figure.

agreement with H_1, the (log) likelihood differences between H_1 and H_2 outweigh their penalty differences. This causes H_2 not be selected 100% of the times for even large samples while it is the true hypothesis.

The performance of the GORICA when correctly preferring the unconstrained hypothesis H_3 is less convincing than that when the other hypotheses are the true hypotheses. This is because the θ's used to simulate the data for the true unconstrained hypothesis H_3 are more or less in accordance with the restrictions of hypothesis H_1. Therefore, sometimes the effect size and/or sample size is not large enough for correctly preferring the unconstrained hypothesis H_3 over hypothesis H_1. We could have chosen a population which does not agree with the restrictions of hypotheses H_1 and H_2 at all. Assuming that a strong hypothesis (e.g., H_1) is not completely true but at the same time is not completely different from the full reality, it is more reasonable to believe that the restrictions of a hypothesis representing the theory are partly in agreement with the restrictions of the true hypothesis. This makes it harder to distinguish the strong hypothesis from the unconstrained hypothesis. However, even in this case, the probability of correctly preferring the unconstrained hypothesis H_3 is higher with increasing values of the effect size and/or sample size.

Example 1 (Continued): Logistic Regression Modeling

In the previous subsection, we have shown by means of a simulation study that the performance of the GORICA is satisfactory in our logistic regression model. Now, we analyze the replication ($N = 310$) of the study in Nederhof et al. (2014, p. 689) to evaluate hypotheses H_1 and H_2, and the unconstrained hypothesis H_3. Parameter estimates and their covariance matrix are displayed in Table 2. The order-restricted MLEs for the three hypotheses H_1, H_2, and the unconstrained hypothesis H_3 are given in Table 3. Note that the order-restricted MLEs are always in agreement with the constraints of the corresponding hypothesis. The values of the likelihood and penalty parts, GORICA values and GORICA weights are displayed in Table 4. The hypothesis with the smallest GORICA value and, therefore, the highest GORICA weight is the best one among the three hypotheses. Both hypotheses H_1 and H_2 have higher GORICA weights when compared to that of the unconstrained hypothesis H_3. Therefore, it is concluded that both hypotheses H_1 and H_2 are supported by the data. In the replication data, the theory in Nederhof and Schmidt (2012); which is represented by hypothesis H_1, has more support than hypothesis H_2, specified based on the results in Nederhof et al., (2014; p. 689). Based on the GORICA weights, hypothesis H_1 has $0.761/0.207 \approx 3.68$ times more support than hypothesis H_2. Note that hypothesis H_2 is more specific than hypothesis H_1, and thus, it has a lower penalty part when compared to hypothesis H_1. However, also note that hypothesis H_1 fits the data better than hypothesis H_2 such that the difference between their fits outweighs the difference between their penalties in favor of hypothesis H_1. Thus, based on the data at hand, we concluded that the mismatch expectation applies to both the sustainers and the shifters, and the cumulative stress expectation applies to the comparison groups.

Multilevel Regression Modeling

In this section, we illustrate how (in)equality constrained hypotheses can be evaluated with the GORICA and investigate its performance in the context of a multilevel regression model.

Example 2: Multilevel Regression Modeling

Multilevel regression models are an extension of ordinary regression models that involve populations of interest with hierarchical data structures. We utilize the study in Hox (2010, p. 16) to illustrate how to evaluate (in)equality constrained hypotheses for GLMMs using a two-level multilevel regression model. 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 pupil level predictors gender (G: 0 = boy, 1 = girl) and pupil extraversion scores (PE) that range from 1 (*introversion*) to 10 (*extraversion*), a class level predictor teacher experience (TE), and the cross-level interaction between PE and TE. Because standardization is recommended when the model contains interactions (Gelman, 2008), we standardize PS, PE, and TE by means of utilizing grand mean centering (Algina & Swaminathan, 2011, pp. 285–312). That is, we first subtract the overall means of the continuous variables PS, PE, and TE from each of their values, before dividing these values by their standard deviations (i.e., $\mathrm{PS}^S_{ji} = \frac{\mathrm{PS}_{ji} - \overline{\mathrm{PS}}}{\mathrm{sd}(\mathrm{PS}_{ji})}$, $\mathrm{PE}^S_{ji} = \frac{\mathrm{PE}_{ji} - \overline{\mathrm{PE}}}{\mathrm{sd}(\mathrm{PE}_{ji})}$, and $\mathrm{TE}^S_{j} = \frac{\mathrm{TE}_{j} - \overline{\mathrm{TE}}}{\mathrm{sd}(\mathrm{TE}_{j})}$). The resulting multilevel regression model for the ith student in the jth class is:

$$\mathrm{PS}^S_{ji} = \theta_{00} + \theta_{10}\mathrm{G}_{ji} + \theta_{20}\mathrm{PE}^S_{ji} + \theta_{01}\mathrm{TE}^S_{j} + \theta_{11}\mathrm{G}_{ji}\mathrm{TE}^S_{j} + \theta_{21}\mathrm{PE}^S_{ji}\mathrm{TE}^S_{j} + \mu_{0j} + \mu_{2j}\mathrm{PE}^S_{ji} + \epsilon_{ji}, \quad (11)$$

where θ_{00} is the intercept, θ_{10}, θ_{20}, θ_{01}, and θ_{21} are the regression slopes for the pupil level variables gender and pupil extraversion, class level variable teacher experience, and the cross-level interaction between pupil extraversion and teacher experience, respectively, μ_{0j} and μ_{2j} are the random effects at the class level, and $\epsilon_{ji} \sim N(0, \sigma^2)$ is the pupil level error for $j = 1, 2, \ldots, J$ and $i = 1, 2, \ldots, N_j$. The random effects μ_{0j} and μ_{2j} are assumed to have a bivariate normal distribution with mean vector $\mathbf{0}$ and covariance matrix:

$$\Psi = \begin{bmatrix} \psi_0^2 & \psi_{02} \\ \psi_{02} & \psi_2^2 \end{bmatrix}.$$

Table 2
Estimates $(\hat{\boldsymbol{\theta}})$ and Covariance Matrix $(\hat{\boldsymbol{\Sigma}}_{\hat{\boldsymbol{\theta}}})$ of the Structural Parameters for the Logistic Regression Example

	$\hat{\boldsymbol{\Sigma}}_{\hat{\boldsymbol{\theta}}}$					
$\hat{\boldsymbol{\theta}}$	$\hat{\theta}_{11}$	$\hat{\theta}_{13}$	$\hat{\theta}_{21}$	$\hat{\theta}_{23}$	$\hat{\theta}_{31}$	$\hat{\theta}_{33}$
$\hat{\theta}_{11} = 0.48$	0.17					
$\hat{\theta}_{13} = 0.28$	−0.17	0.63				
$\hat{\theta}_{21} = 0.87$	0.00	0.00	0.21			
$\hat{\theta}_{23} = -0.44$	0.00	0.00	−0.21	0.37		
$\hat{\theta}_{31} = 0.81$	0.00	0.00	0.00	0.00	0.14	
$\hat{\theta}_{33} = -0.25$	0.00	0.00	0.00	0.00	−0.14	0.18

Note. The respective terms are made bold because they are either vector or matrix.

Table 3
The Order-Restricted MLEs, $\tilde{\boldsymbol{\theta}}^m = (\tilde{\theta}^m_{11}, \tilde{\boldsymbol{\theta}}^m_{13}, \ldots, \tilde{\boldsymbol{\theta}}^m_{33})^T$, for Hypothesis H_m (m = 1, 2, 3) for the Logistic Regression Example

$\mathbf{H_m}$	$\tilde{\theta}^m_{11}$	$\tilde{\theta}^m_{13}$	$\tilde{\theta}^m_{21}$	$\tilde{\theta}^m_{23}$	$\tilde{\theta}^m_{31}$	$\tilde{\theta}^m_{33}$
H_1	0.48	−0.48	0.87	−0.87	0.62	0.00
H_2	0.48	−0.48	0.00	0.00	0.62	0.00
H_3	0.48	0.28	0.87	−0.44	0.81	−0.25

Note. MLE = maximum likelihood estimates. The respective terms are made bold because they are either vector or matrix.

Table 4
The Likelihood $(L(\tilde{\boldsymbol{\theta}}^m \mid \hat{\boldsymbol{\theta}}, \hat{\boldsymbol{\Sigma}}_{\hat{\boldsymbol{\theta}}}))$ and Penalty (PT_m) Parts, GORICA Values $(GORICA_m)$, and GORICA Weights (w_m) of Hypothesis H_m (m = 1, 2, 3) for the Logistic Regression Example

H_m	$L(\tilde{\boldsymbol{\theta}}^m \mid \hat{\boldsymbol{\theta}}, \hat{\boldsymbol{\Sigma}}_{\hat{\boldsymbol{\theta}}})$	PT_m	$GORICA_m$	w_m
H_1	−1.373	1.499	5.743	0.761
H_2	−3.168	1.004	8.343	0.207
H_3	−0.045	6.000	12.089	0.032

Note. GORICA = generalized order-restricted information criterion approximation. The respective terms are made bold because they are either vector or matrix.

The model in Equation 11 can be converted into separate models for boys and girls:

$$PS^S_{ji} = \begin{cases} \theta_{00} + \theta_{20} PE^S_{ji} + \theta_{01} TE^S_j + \theta_{21} PE^S_{ji} TE^S_j + \mu_{0j} + \mu_{2j} PE^S_{ji} + \epsilon_{ji} & \text{if } G = 0 \text{ (boy)} \\ (\theta_{00} + \theta_{10}) + \theta_{20} PE^S_{ji} + (\theta_{01} + \theta_{11}) TE^S_j + \theta_{21} PE^S_{ji} TE^S_j + \mu_{0j} + \mu_{2j} PE^S_{ji} + \epsilon_{ji} & \text{if } G = 1 \text{ (girl).} \end{cases}$$

For this example, we do not have a solid theoretical background, like in the sustainers-versus-shifters hypotheses in the logistic regression example. This example serves to illustrate the applicability of the GORICA in the context of multilevel regression. Nevertheless, we did base the hypotheses under evaluation on theory as well.

The study in Hox (2010) regards mainly whether teacher's experience moderates the impact of gender and pupil extraversion on popularity scores. Here, we assume that the researcher has two competing hypotheses/theories regarding the sign and size of the population parameters:

$$\begin{aligned} &H_1: \theta_{10} > 0, \theta_{20} > 0, \theta_{11} < 0, \theta_{21} < 0, \\ &H_2: \theta_{10} < 0, \theta_{20} > 0, \theta_{11} = \theta_{21} = 0, \\ &H_3: \theta_{10}, \theta_{20}, \theta_{11}, \theta_{21}. \end{aligned} \quad (12)$$

Hypothesis H_1 states that girls and extraverted pupils have higher popularity scores than boys and introverted pupils, respectively, when the teacher in the jth class has no teaching experience (i.e., $TE_j = 0$ for $j = 1, 2, \ldots, 100$; and thus $\theta_{10} > 0, \theta_{20} > 0$) and that the difference in popularity scores between boys and girls and between extraverted and introverted pupils become smaller with more experienced teachers (i.e., $\theta_{11} < 0, \theta_{21} < 0$). Namely, this hypothesis states that teacher experience moderates the association between popularity scores and both gender and pupil extraversion. Hypothesis H_2 specifies that boys and extraverted pupils have higher popularity scores than girls and introverted pupils (i.e., $\theta_{10} < 0, \theta_{20} > 0$), but teacher's experience does not moderate the effects of gender and pupil extraversion on popularity scores (i.e., $\theta_{11} = \theta_{21} = 0$). In other words, teacher experience moderates the association between popularity scores and gender, but has no impact on the association between popularity scores and pupil extraversion. The unconstrained hypothesis H_3 states that there is no restriction on the model parameters, which is included in the set as a safeguard against choosing a weak hypothesis as the best hypothesis. Thus, the structural parameters of the model are $\boldsymbol{\theta} = (\theta_{10}, \theta_{20}, \theta_{11}, \theta_{21})$ and the nuisance parameters are $\xi = (\theta_{00}, \theta_{01}, \psi_0^2, \psi_2^2, \psi_{02}, \sigma^2)$. We will first conduct a simulation study to evaluate the performance of the GORICA on choosing the best hypothesis out of the set of the three hypotheses above in the context of multilevel regression, and then, we will analyze the data for this study.

Performance of the GORICA for the Multilevel Regression Example

We conduct a simulation study to evaluate the performance of the GORICA in the context of multilevel regression. The model for the multilevel regression simulation, as being a representative of GLMMs, is the same model as in Equation 11. The simulation contains three different cases, where in Cases 1 and 2, the θ's are in accordance with H_1 and H_2, respectively. In Case 3, the population values of the parameters are only concordant with the unconstrained hypothesis H_3. Choose the θ's using Cohen's f^2 (Cohen, 1992, p. 157) which is an adequate effect size measure for the fixed part of the model in the context of multilevel regression (Lorah, 2018). Table 5 displays our choices for the θ's where we use only the medium effect size $f^2 = 0.15$ when determining population values for the θ's in line with Maas and Hox (2005). Following the suggestions provided in Hox (2010, pp. 244–249), we choose the population values for the intercept variance and the slope variance: (a) $\psi_0^2 = 0.05$ and $\psi_2^2 = 0.05$ for small variances, (b) $\psi_0^2 = 0.11$ and $\psi_2^2 = 0.10$ for medium variances, and (c) $\psi_0^2 = 0.18$ and $\psi_2^2 = 0.15$ for large variances, with $\psi_{02} = 0$, so that the random effects are independent from each other. We choose the number of groups as $J = 30, 50$, or 100 and the number of observations in each group as $N_j = 5, 30$, or 50, in line with the suggestions made in the simulation design presented in Maas and Hox (2005, p. 88). For more details on data generation in the two-level multilevel regression simulation see the section "Simulation Steps" in the online supplementary material.

Figure 2 displays the proportion of times each of the hypotheses was chosen by the GORICA in 1,000 independent samples. The GORICA performs well with respect to choosing the true hypothesis out of the three hypotheses for the model in Equation 11. Increasing the random variation in the data (i.e., more noise) by increasing both intercept and slope variances does not exert much influence on the probability of choosing the true hypothesis. The GORICA performs better for the data sets containing more groups (J) and/or more observations in each group (N_j) no matter the values of the intercept and slope variances. In many cases where N_j = 30 or 50, the GORICA chooses the correct hypothesis 100% of the times no matter the values of the number of groups, the values of the intercept and slope variances, and which hypothesis is the true hypothesis. Thus, the performance of the GORICA on correctly preferring the true hypothesis is convincing.

Example 2 (Continued)

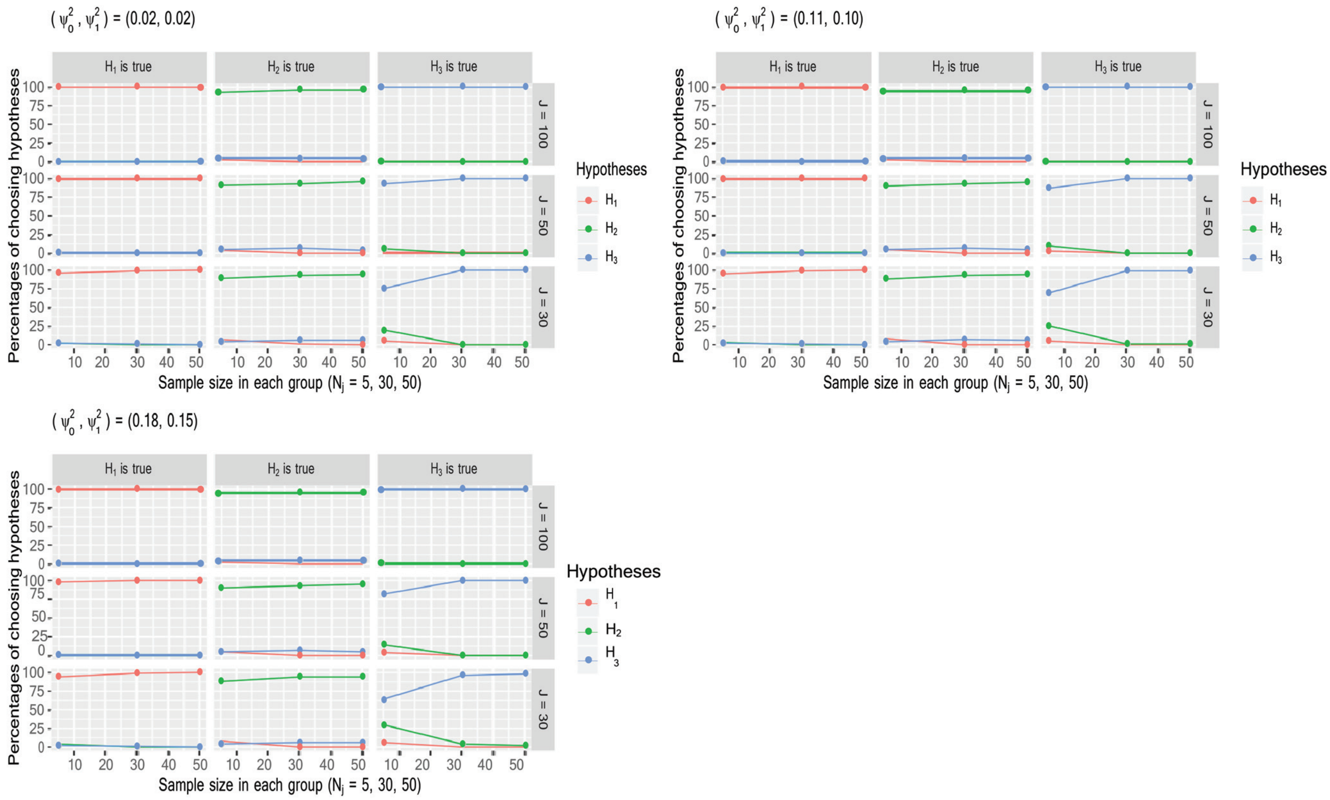
For the study in Hox (2010, p. 16), the parameter estimates and their covariance matrix are displayed in Table 6. Because these estimates are in agreement with the constraints in hypotheses H_1 and the unconstrained hypothesis $H_3: \theta_{10}, \theta_{20}, \theta_{11}, \theta_{21}$, the order-restricted MLEs of these hypotheses are equal to the MLEs (see Table 7). Thus, these

Table 5

Multilevel Regression Simulation: Population Values of the Regression Coefficients When $f^2 = 0.15$ for the Three Cases

Case 1: $H_1: \theta_{10} > 0, \theta_{20} > 0, \theta_{11} < 0, \theta_{21} < 0$ is true			
θ_{10}	θ_{20}	θ_{11}	θ_{21}
0.070	0.070	−0.265	−0.265
Case 2: $H_2: \theta_{10} < 0, \theta_{20} > 0, \theta_{11} = \theta_{21} = 0$ is true			
θ_{10}	θ_{20}	θ_{11}	θ_{21}
−0.245	0.247	0.000	0.000
Case 3: $H_3: \theta_{10}, \theta_{20}, \theta_{11}, \theta_{21}$ is true			
θ_{10}	θ_{20}	θ_{11}	θ_{21}
−0.219	−0.227	0.128	−0.128

Figure 2
Percentages of Choosing Hypotheses for the Multilevel Regression Simulation



Note. Percentage of times that hypotheses H_1, H_2, and H_3 are chosen by the GORICA using varying values of the number of groups J, number of observations in each group N_j $(j = 1, 2, \ldots, J)$, and the intercept and slope variance $\psi_0^2$and $\psi_1^2$for the three cases in which H_1, H_2, and H_3 are the true hypothesis, respectively. See the online article for the color version of this figure.

two hypotheses have the same log likelihood and, therefore, the GORICA distinguishes among them only through the penalty term (see Table 8). Based on the GORICA weights in Table 8, hypothesis H_1 is a strong hypothesis (i.e., it has $0.889/0.111 \approx 8$ times more support than the unconstrained hypothesis H_3), while hypothesis H_2 is a weak hypothesis (i.e., it has $0/0.111 = 0$ times more support than H_3). Because at least one of the hypotheses of interest is not a weak hypothesis, one can compare their strengths: Hypothesis H_1 is infinite times better than hypothesis H_2. Thus, we concluded that girls and extraverted pupils have higher popularity scores than boys and introverted pupils and teacher's experience moderates the influences of gender and pupil extraversion on the popularity scores.

Structural Equation Modeling (SEM)

In this section, we illustrate how to evaluate (in)equality constrained hypotheses for a structural equation model and investigate the performance of the GORICA for this model.

Example 3: Structural Equation Modeling

In this example, the specification of (in)equality constrained hypotheses for structural equation models is exemplified based on the study in Stevens (1999, p. 596). This study evaluates the effects of the first year of the Sesame Street TV series in a sample of 3- to 5-year-old children in the U.S. (N = 240). Two latent variables can be constructed using these data. The first latent variable prewatch (x_1) is the knowledge before watching the Sesame Street TV series for a year, using 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 postwatch (x_2) is the knowledge after watching the Sesame Street TV series for a year, using 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 are regressed on two observed predictors: age in months (A) and the Peabody score (P), that is, a mental age score which is obtained from the Peabody Picture Vocabulary test. Table 9 displays the descriptives for each observed variable. Moreover, Figure 3 shows the path diagram indicating the relationship between these variables. To be able to compare the structural parameters to each other, we used the R package lavaan (Rosseel, 2012) to obtain the standardized structural parameter estimates and their covariance matrix.

Structural equation models typically contain two submodels: the measurement model and the structural model. The measurement

Table 6

Estimates ($\hat{\boldsymbol{\theta}}$) and Covariance Matrix ($\hat{\boldsymbol{\Sigma}}_{\hat{\boldsymbol{\theta}}}$) of the Structural Parameters in the Multilevel Regression Example

$\hat{\boldsymbol{\theta}}$	$\hat{\boldsymbol{\Sigma}}_{\hat{\boldsymbol{\theta}}}$			
	$\hat{\theta}_{10}$	$\hat{\theta}_{20}$	$\hat{\theta}_{11}$	$\hat{\theta}_{21}$
$\hat{\theta}_{10} = 0.90$	6.9e−4			
$\hat{\theta}_{20} = 0.41$	−3.4e−5	2.5e−4		
$\hat{\theta}_{11} = -0.01$	−1.6e−5	1.4e−5	7.6e−4	
$\hat{\theta}_{21} = -0.15$	1.4e−5	3.4e−6	−4.9e−5	2.3e−4

Note. The respective terms are made bold because they are either vector or matrix.

model represents the connections between unobserved latent variables and observed indicators, while the structural model describes the relations between latent variables and between latent variables and observed predictors. Let $\boldsymbol{y}_i = (y_1, y_2, \ldots, y_{12})^T$ and $\boldsymbol{\eta}_i = (\eta_1, \eta_2)^T$ denote the vectors of indicators and latent variables, respectively, for person $i = 1, \ldots, 240$. The measurement model for these indicators and latent variables is defined as:

$$\boldsymbol{y}_i = \boldsymbol{\Lambda}\boldsymbol{\eta}_i + \boldsymbol{\upsilon}_i, \tag{13}$$

where

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

is a 12×2 matrix of factor loadings relating these indicators and latent variables, $\boldsymbol{\upsilon}_i$ is a vector of measurement errors in the indicators for person i, which are normally distributed with the 12×1 mean vector $\mathbf{0}$ and 12×12 covariance matrix $\boldsymbol{\Theta}_{\upsilon}$, that is, $\boldsymbol{\upsilon}_i \sim N(0, \boldsymbol{\Theta}_{\upsilon})$.

The structural model for the Sesame Street data that relates these two latent variables to the two predictors, namely, age and Peabody score, is defined as:

$$\boldsymbol{\eta}_i = \boldsymbol{\theta}\boldsymbol{x}_i + \boldsymbol{\zeta}_i, \tag{14}$$

where

$$\boldsymbol{\theta} = \begin{bmatrix} \theta_1 & \theta_2 \\ \theta_3 & \theta_4 \end{bmatrix}$$

is a 2×2 matrix that relates the two latent variables to the two predictors, $\boldsymbol{x}_i = (A_i, P_i)^T$ is a 2×1 vector of the two predictors,

Table 7

The Order-Restricted MLEs, $\tilde{\boldsymbol{\theta}}^m = (\tilde{\theta}^m_{10}, \tilde{\boldsymbol{\theta}}^m_{20}, \tilde{\boldsymbol{\theta}}^m_{11}, \tilde{\boldsymbol{\theta}}^m_{21})^T$, for Hypothesis H_m (m = 1, 2, 3) for the Multilevel Regression Example

H_m	$\tilde{\theta}^m_{10}$	$\tilde{\theta}^m_{20}$	$\tilde{\theta}^m_{11}$	$\tilde{\theta}^m_{21}$
H_1	0.90	0.41	−0.01	−0.15
H_2	0.00	0.46	0.00	0.00
H_3	0.90	0.41	−0.01	−0.15

Note. MLE = maximum likelihood estimates. The respective terms are made bold because they are either vector or matrix.

Table 8

The Likelihood ($L(\tilde{\boldsymbol{\theta}}^m \mid \hat{\boldsymbol{\theta}}, \hat{\boldsymbol{\Sigma}}_{\hat{\boldsymbol{\theta}}})$) and Penalty ($PT_m$)Parts, GORICA Values ($GORICA_m$), and GORICA Weights ($w_m$) of Hypothesis H_m (m = 1, 2, 3) for the Multilevel Regression Example

H_m	$L(\tilde{\boldsymbol{\theta}}^m \mid \hat{\boldsymbol{\theta}}, \hat{\boldsymbol{\Sigma}}_{\hat{\boldsymbol{\theta}}})$	PT_m	$GORICA_m$	w_m
H_1	11.911	1.918	−19.986	0.889
H_2	−635.462	1.030	1,272.984	0.000
H_3	11.911	4.000	−15.822	0.111

Note. GORICA = generalized order-restricted information criterion approximation. The respective terms are made bold because they are either vector or matrix.

and $\boldsymbol{\zeta}_i \sim N(0, \boldsymbol{\Theta}_{\zeta})$ is the vector of residuals with $\boldsymbol{\Theta}_{\zeta}$ a 2×2 covariance matrix. Note that we only need the estimates of structural parameters and their covariance matrix when evaluating hypotheses using the GORICA. The estimates of all parameters used in the model are displayed in Figure 3. Notably, the off-diagonals of the covariance matrix for measurement errors in the indicators, $\boldsymbol{\Theta}_{\upsilon}$, are not displayed in this figure, since they are set equal to zero in the model.

For this example, we do not have a theoretical background. The hypotheses are formulated to illustrate the applicability of the GORICA for SEM. We focus on the comparison of the effects of age and Peabody on each latent variable. The following hypotheses can be of interest regarding the regression relations between latent variables and predictors:

$$\begin{aligned} &H_1\!: \ \theta_2 > \theta_1, \theta_4 > \theta_3, \\ &H_2\!: \ \theta_1 > \theta_2, \theta_3 > \theta_4, \\ &H_3\!: \ \theta_1, \theta_2, \theta_3, \theta_4. \end{aligned} \tag{15}$$

Hypothesis H_1 states that Peabody has more impact on each latent variable than age; hypothesis H_2 specifies that the ordering of parameters for these variables should be reversed, namely, the effect of age for each latent variable is larger than that of Peabody; and the unconstrained hypothesis H_3 states the absence of information about the relationship between model parameters, which is included in the set as a fail-safe hypothesis against choosing H_1 or H_2 in the case both of them are weak hypotheses. Here, the θ's that are used to formulate the hypotheses above are the structural parameters, while all the other parameters (i.e., factor loadings and covariance parameters) are the nuisance parameters.

One may consider to test measurement invariance in the context of SEM, where hypotheses of interest contain only equality restrictions between factor loadings (i.e., the λ's). Because these hypotheses contain only equality restrictions between the λ's, the GORICA reduces to the AIC. Therefore, we will not investigate the evaluation of measurement invariance using the GORICA. For more details on how to obtain the estimates of factor loadings and their covariance matrix in R, which can be used to evaluate measurement invariance, see the section "Structural Equation Modeling Example" in the online supplementary material. We will first perform a simulation study investigating the performance of the GORICA on choosing the best hypothesis out of the set of hypotheses containing regression coefficients in the context of structural equation model, and afterward, we will analyze the data for the example.

Table 9
Descriptives for the Variables in the SEM Example

Type of descriptive	Prebody	Prelet	Preform	Prenumb	Prerelat	Preclas	Postbody
Min:	6.000	1.000	2.000	1.000	2.000	0.000	11.000
M:	21.400	15.940	9.921	20.900	9.938	12.240	25.260
Max:	32.000	55.000	19.000	52.000	17.000	24.000	32.000
SD:	6.390	8.536	3.737	10.685	3.074	4.658	5.412

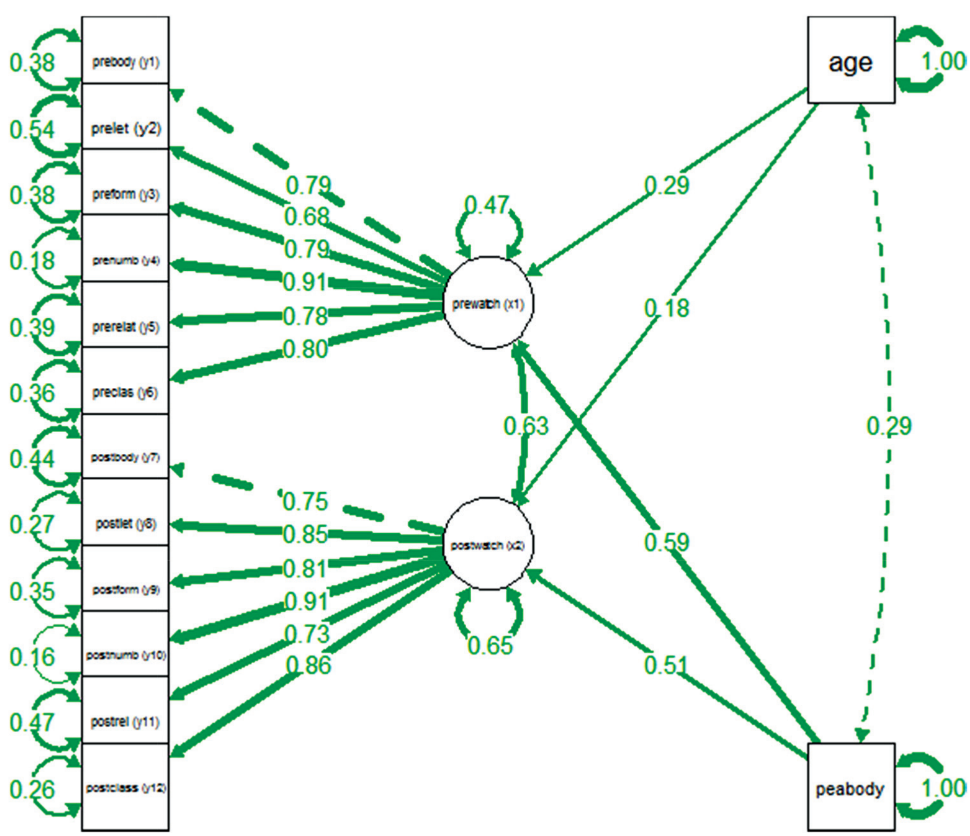
	Postlet	Postform	Postnumb	Postrel	Postclas	Age	Peabody
Min:	0.000	0.000	0.000	0.000	0.000	34.000	27.000
M:	26.700	13.740	30.050	11.650	15.740	51.520	46.770
Max:	54.000	20.000	54.000	17.000	24.000	69.000	99.000
SD:	13.272	4.001	12.846	2.832	5.151	6.281	15.987

Note. SEM = structural equation models.

Performance of the GORICA for the SEM Example

We conduct a simulation study to investigate the performance of the GORICA when evaluating hypotheses containing regression coefficients in the context of SEMs. The measurement model and the structural model for the SEM simulation are the same models as in Equations 13 and 14, respectively. We create three different cases, where in Cases 1 and 2, the θ's are in agreement with H_1 and H_2, respectively. In Case 3, the population values of the parameters are only in agreement with the unconstrained hypothesis H_3. Table 10 gives our choices for the θ's such that it holds that $f^2 = 0.02, 0.15$, or .35 (Cohen, 1992, p. 157). Because the standardized factor loadings have little effect on choosing the true hypothesis out of the set of

Figure 3
Structural Equation Model for the Sesame Street Data



Note. See the online article for the color version of this figure.

Table 10
SEM Simulation: Population Values of the Regression Coefficients With $f^2 = 0.02$, 0.15, and 0.35 for the Three Cases

	Latent variable η_1		Latent variable η_2	
f^2	θ_1	θ_2	θ_3	θ_4
Case 1: $H_1: \theta_2 > \theta_1, \theta_4 > \theta_3$ is true				
0.02	0.000	0.140	0.000	0.140
0.15	0.000	0.361	0.000	0.361
0.35	0.000	0.509	0.000	0.509
Case 2: $H_2: \theta_1 > \theta_2, \theta_3 > \theta_4$ is true				
0.02	0.140	0.000	0.140	0.000
0.15	0.361	0.000	0.361	0.000
0.35	0.509	0.000	0.509	0.000
Case 3: $H_3: \theta_1, \theta_2, \theta_3, \theta_4$ is true				
0.02	0.140	0.000	0.000	0.140
0.15	0.361	0.000	0.000	0.361
0.35	0.509	0.000	0.000	0.509

Note. SEM = structural equation models. The effect size f^2 is Cohen's f^2 (Cohen, 1992, p. 157). The population value of each factor loading is chosen as .65. The correlation between the two predictors for latent variables η_1 and η_2 is chosen as $\rho = 0.3$.

hypotheses when they concern regression coefficients, we choose $\lambda = 0.65$ as population values of the factor loadings to relate the indicators with the latent variables in the simulation in line with Wolf et al. (2013, p. 917). We choose a sample of size $N = 120, 150$, or 200 and the correlation between latent variables $\rho = 0.3$ or 0.5 in line with Wolf et al. (2013). We use the correlation between the two predictors age and Peabody score for each latent variable as $\rho = 0.3$. For more details on data generation in the SEM simulation see the section "Simulation Steps" in the online supplementary material.

Figure 4 shows the proportion of the times each of the hypotheses containing regression coefficients was chosen by the GORICA for 1,000 independent samples created from the population displayed in Table 10. Increasing the effect size and/or sample size (almost in all simulation conditions) improves the performance of the GORICA with respect to selecting the true hypothesis out of the three hypotheses under evaluation. Therefore, we conclude that the GORICA performs well on choosing the true hypothesis for the hypotheses under evaluation.

Example 3 (Continued)

For the study in Stevens (1999, p. 596), the standardized parameter estimates and their covariance matrix for (structural) regression coefficients are shown in Table 11.[1] The order-restricted MLEs for hypotheses H_1, H_2, and the unconstrained hypothesis H_3 are given in Table 12. Based on the GORICA weights in Table 13, hypothesis H_1 is a strong hypothesis, while hypothesis H_2 is a weak hypothesis, because it has a GORICA weight of zero. Note that, because the restrictions of hypothesis H_1 are in agreement with the data, the log likelihood for H_1 is equal to the log likelihood for the unconstrained hypothesis H_3. Additionally, the former has a smaller penalty than the latter. Therefore, the relative weight between these hypotheses is at its maximum value (i.e., H1 is $0.693/0.307 \approx 2.26$ times better than the unconstrained hypothesis H_3). These results imply that when assessing social and mental intelligence of children before and after watching the Sesame Street TV series, the Peabody score (P) is a more important predictor for intelligence than age (A).

In the next section, we describe how the GORIC and GORICA evaluate hypotheses containing range restrictions (e.g., $H_8: -2 < \theta_1 < 2, \theta_2$ in the section "GORIC and GORICA").

Range Restrictions

The fit part of the GORIC and GORICA can be uniquely calculated under range restrictions, the challenge is in determining the penalty part. The penalty part of a hypothesis is uniquely defined for restrictions of the type $\boldsymbol{S}_m\boldsymbol{\theta} = \boldsymbol{s}_m \boldsymbol{R}_m\boldsymbol{\theta} > \boldsymbol{r}_m$ in Equation 5 where $\begin{bmatrix} \boldsymbol{S}_m \\ \boldsymbol{R}_m \end{bmatrix}$ is of full rank (after discarding redundant restrictions). As shown in the section "GORIC and GORICA", the restriction matrix for hypothesis $H_8: -2 < \theta_1 < 2, \theta_2$ is not of full rank because there is a linear dependency between its two rows, and thus, the penalty part for this hypothesis is not uniquely defined. Stated otherwise, many solutions exist to determine the space of range restriction $-2 < \theta_1 < 2$ in H_8. The scaling of the covariance matrix of estimates is the main factor when determining the space of range restrictions in these solutions. Figure 5 shows how this scaling influences determining the space of range restriction $-2 < \theta_1 < 2$ in the whole parameter space. Each of the six plotted choices would lead to a different penalty value. In Plot (2, 3), one can see that the range restriction is like a line. Hence, when inspecting the complete parameter space, the range restriction is like an equality restriction. This is the way the penalty is calculated in current software. In that case, when calculating the penalty, θ_1 is set equal to a constant and thus leads to zero expected distinct parameters. When inspecting the full hypothesis, $H_8: -2 < \theta_1 < 2, \theta_2$, its penalty is 1, because θ_2 is one free parameter and θ_1 is a constant and thus no parameter.

Next, we exemplify how the GORIC and GORICA deal with range restrictions in the context of an ANOVA model using the PlantGrowth data in R. The outcome the weight of plant (N = 30) is measured in a control and two treatment groups (1: control, 2: Treatment 1, 3: Treatment 2). The ANOVA model for comparing the means of these groups is defined as:

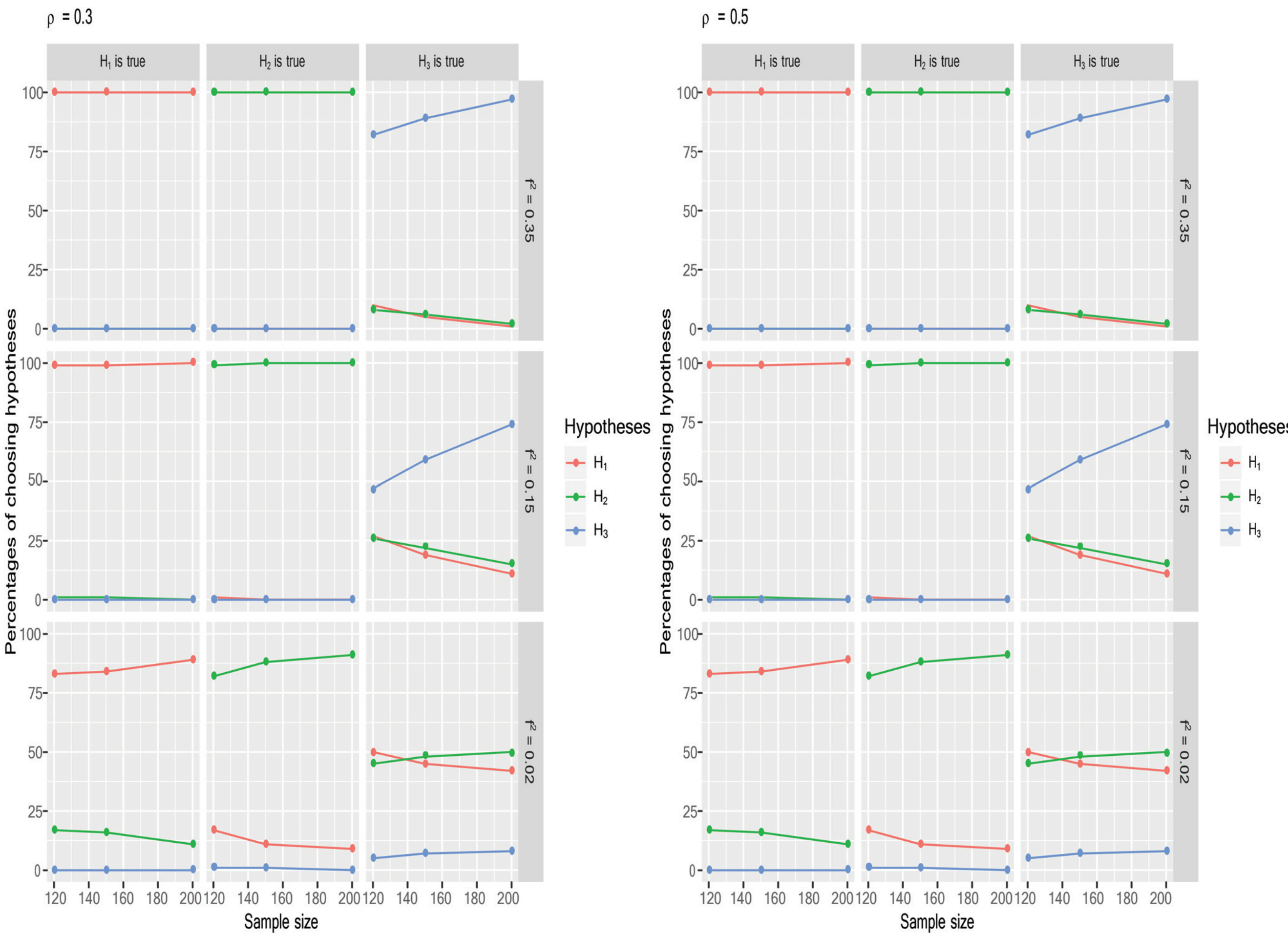
$$y_{ji} = \theta_j + \epsilon_{ji}, \qquad (16)$$

where y_{ji} is the weight of the ith plant in the jth group, θ_j is the mean of the weights of the plants in the jth group and $\epsilon_{ji} \sim N(0, \sigma^2)$ is corresponding residual with variance σ^2 for j = 1, 2, 3 and $i = 1, 2, \ldots, 10$. To illustrate the applicability of the GORIC and GORICA for range restrictions, we utilize Cohen's d (Cohen, 1992) effect size measure where 0.2, 0.5, and 0.8 indicate small, medium, and large effect sizes, respectively.

[1] For more details on how to obtain these estimates and their covariance matrix (together with the estimates of the factor loadings and their covariance matrix) using R, see the section "Structural Equation Modeling Example" in the online supplementary material.

Figure 4

Percentages of Choosing Hypotheses for the SEM Simulation



Note. Percentage of times that the hypotheses H_1, H_2, and H_3 are selected by the GORICA with the effect size values $f^2 = 0.02, 0.15$, and 0.35; sample sizes N = 120, 150, and 200; and the correlation between latent variables $\rho = 0.3$ and 0.5 for the three cases in which H_1, H_2, and H_3 are the true hypothesis, respectively. SEM = structural equation model. See the online article for the color version of this figure.

We compare the means of plant weight in control and treatment groups and we evaluate whether the expected effect size is between medium and large when comparing the means in treatment groups. Since we have only one hypothesis of interest, we evaluate this hypothesis against its complement (see Step 1(a) on the guidance presented in the section "GORIC and GORICA"):

$$H_1:\ \theta_1 > \theta_2, \theta_3 > \theta_1, 0.5\sigma_p < \{\theta_3 - \theta_2\} < 0.8\sigma_p,$$
$$H_2:\ \text{The complement of hypothesis}\ H_1, \quad (17)$$

where H_1 specifies that the first treatment distorts the plant growth (i.e., $\theta_1 > \theta_2$) while the second treatment improves the plant growth (i.e., $\theta_3 > \theta_1$) relative to the control group and the corresponding Cohen's d is between medium and large when comparing the means of these treatment groups (i.e., $0.5\sigma_p < \{\theta_3 - \theta_2\} < 0.8\sigma_p$). Here, because the pooled standard error (σ_p) is unknown, we use an estimate of it based on the sample, that is, the pooled standard deviation ($\hat{\sigma}_p = 0.623$). The estimates of population means and their covariance matrix are displayed in Table 14. The order-restricted MLEs for hypothesis H_1 and its complement hypothesis H_2 are displayed in Table 15. This table shows that the GORIC and GORICA produce the same order-restricted MLEs for the hypotheses of interest of the data. The values of the likelihood and penalty parts, GORIC and GORICA values, and GORIC and GORICA weights are displayed in Table 16. Notably, the restrictions in H_1 are a two-dimensional plane because the difference in means θ_2 and θ_3 are set equal to a constant. Therefore, the restrictions in the complement of H_1 practically cover the three-dimensional whole parameter space, and thus, the penalty for the complement is 3. Notably, the range restriction $0.5\sigma_p < \{\theta_3 - \theta_2\} < 0.8\sigma_p$ in H_1 is almost a line in a two-dimensional whole parameter space (see Plot (2, 3) in Figure 6). The results show that hypothesis H_1 is $0.745/0.255 \approx 2.92$ and $0.757/0.243 \approx 3.12$ times more supported by the data using the GORIC and GORICA, respectively, when compared with its complement H_2. This again shows that the GORICA is an adequate approximation of the

Table 11

Estimates ($\hat{\boldsymbol{\theta}}$) and Covariance Matrix ($\hat{\boldsymbol{\Sigma}}_{\hat{\boldsymbol{\theta}}}$) of the Standardized Structural Regression Coefficients for the SEM Example

	$\hat{\boldsymbol{\Sigma}}_{\hat{\boldsymbol{\theta}}}$			
$\hat{\boldsymbol{\theta}}$	$\hat{\theta}_1$	$\hat{\theta}_2$	$\hat{\theta}_3$	$\hat{\theta}_4$
$\hat{\theta}_1 = 0.29$	2.3e − 3			
$\hat{\theta}_2 = 0.59$	− 1.2e − 3	1.6e − 3		
$\hat{\theta}_3 = 0.18$	1.5e − 3	− 7.6e − 4	3.2e − 3	
$\hat{\theta}_4 = 0.51$	− 7.2e − 4	9.8e − 4	− 1.3e − 3	2.3e − 3

Note. SEM = structural equation models. The vector of the estimates for factor loadings is $(\hat{\lambda}_1, \hat{\lambda}_2, \hat{\lambda}_3, \hat{\lambda}_4, \hat{\lambda}_5, \hat{\lambda}_6, \hat{\lambda}_7, \hat{\lambda}_8, \hat{\lambda}_9, \hat{\lambda}_{10}, \hat{\lambda}_{11}, \hat{\lambda}_{12})$ = (0.79, 0.68, 0.79, 0.91, 0.78, 0.80, 0.75, 0.85, 0.81, 0.91, 0.73, 0.86). For more details on how to obtain these estimates and their covariance matrix using R, see the section "Structural Equation Modeling Example" in the online supplementary material. The respective terms are made bold because they are either vector or matrix.

GORIC, because these ratios are reasonably close to each other. Although the ratios are close to each other, we recommend the reader to rely on the results for the GORIC in this example, because each group contains only 10 observations when comparing the means of plant growth across these groups. The working scheme for calculation of the penalty term to obtain GORICA weights are displayed in Figure 7. For more details on how to evaluate the set of hypotheses containing the range restrictions in Equation 17 using the restriktor package, see the section "ANOVA Example with Range Restrictions" in the online supplementary material.

Conclusion and Discussion

The GORIC evaluation of (in)equality constrained hypotheses requires different formulations for each family of models outside the family of normal linear models. Thus, the computation and application of the GORIC for these model families is complicated. Therefore, we have proposed an approximation, the GORICA, to extend the applicability of the GORIC to a large class of statistical models, that is, GLMs, GLMMs, and SEMs. The GORICA is a simple function that requires parameter estimates and their covariance matrix to evaluate (in)equality constrained hypotheses for GLMs, GLMMs, and SEMs. Researchers can specify theories and their expectations for the GORICA and evaluate them using the accompanying gorica package in R.

Table 12

The Order-Restricted MLEs, $\tilde{\boldsymbol{\theta}}^m = (\tilde{\theta}_1^m, \tilde{\boldsymbol{\theta}}_2^m, \tilde{\boldsymbol{\theta}}_3^m, \tilde{\boldsymbol{\theta}}_4^m)^T$, of the Standardized Structural Regression Coefficients for Hypothesis H_m (m = 1, 2, 3) for the SEM Example

H_m	$\tilde{\theta}_1^m$	$\tilde{\theta}_2^m$	$\tilde{\theta}_3^m$	$\tilde{\theta}_4^m$
H_1	0.29	0.59	0.18	0.51
H_2	0.45	0.45	0.37	0.37
H_3	0.29	0.59	0.18	0.51

Note. MLE = maximum likelihood estimates; SEM = structural equation models. The respective terms are made bold because they are either vector or matrix.

Table 13

The Likelihood ($L(\tilde{\boldsymbol{\theta}}^m \mid \hat{\boldsymbol{\theta}}, \hat{\boldsymbol{\Sigma}}_{\hat{\boldsymbol{\theta}}})$) and Penalty ($PT_m$) Parts, GORICA Values ($GORICA_m$), and GORICA Weights ($w_m$) of Hypothesis H_m (m = 1, 2, 3) for the SEM Example

H_m	$L(\tilde{\boldsymbol{\theta}}^m \mid \hat{\boldsymbol{\theta}}, \hat{\boldsymbol{\Sigma}}_{\hat{\boldsymbol{\theta}}})$	PT_m	$GORICA_m$	w_m
H_1	9.209	3.186	−12.047	0.693
H_2	0.470	3.185	5.429	0.000
H_3	9.209	4.000	−10.419	0.307

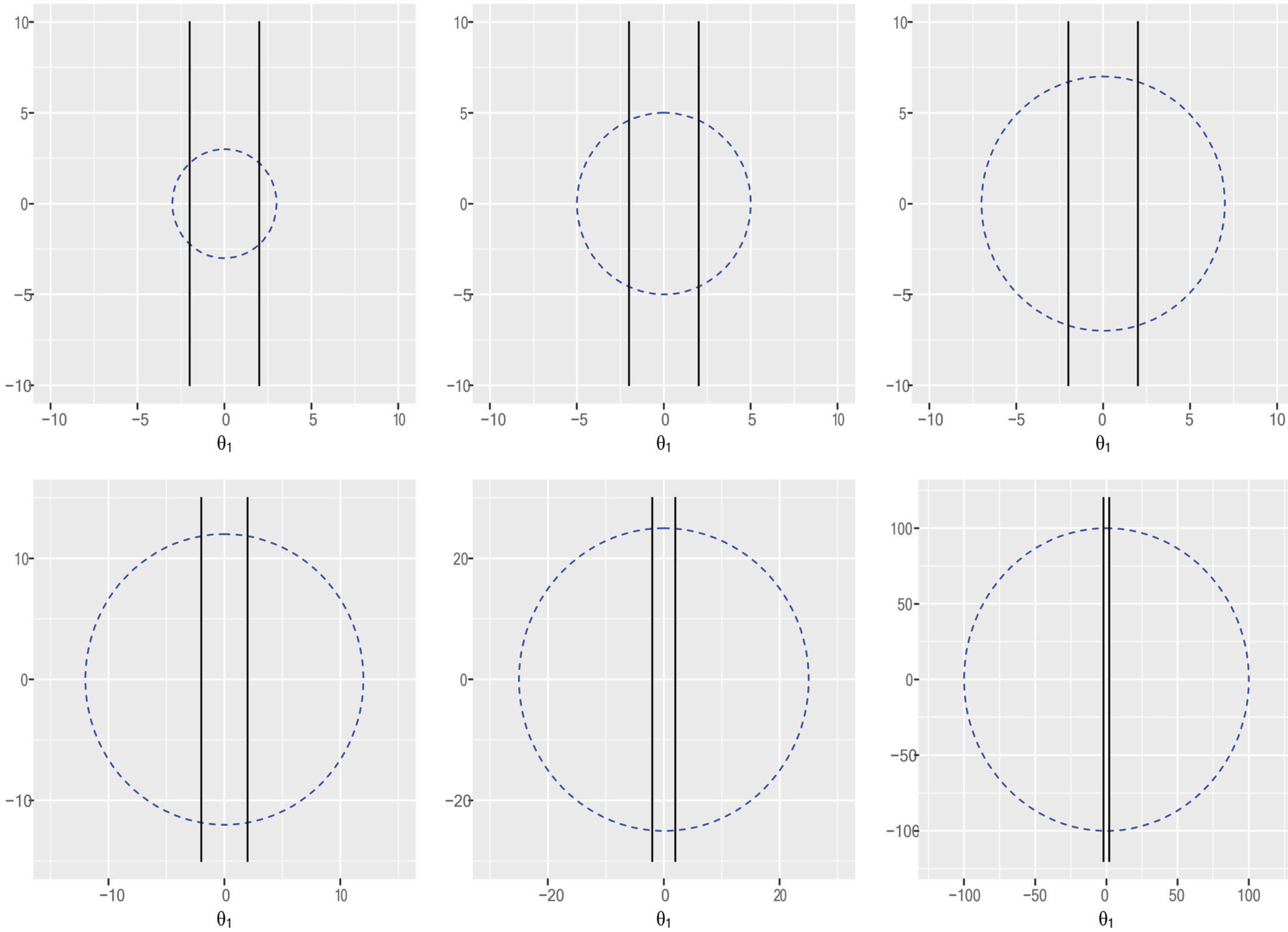
Note. SEM = structural equation models; GORICA = generalized order-restricted information criterion approximation. The respective terms are made bold because they are either vector or matrix.

We performed two simulation studies to compare the performance of the GORIC and GORICA in the context of normal linear models (see Section 1 of the online supplementary material). We have empirically verified that the performance of the GORICA is similar to that of the GORIC on choosing the true hypothesis out of a set of hypotheses for the ANOVA and multiple regression models. Moreover, we showed that the GORIC and GORICA weights are asymptotically the same (see Appendix A). Then, we performed three more simulation studies in the context of a logistic regression, multilevel regression, and structural equation models, respectively, to further investigate the performance of the GORICA (see the main text and some additional information in Section 2 of the online supplementary material). Empirical evidence obtained from each simulation study shows that the GORICA can be used to evaluate (in)equality constrained hypotheses in the context of GLMs, GLMMs, and SEMs. Evidently, the GORICA performed better when increasing the sample size and/or effect size unless one or more hypotheses in the set behave like a null hypothesis. Subsequently, we illustrated how to evaluate (in) equality constrained hypotheses using the GORICA for logistic regression, multilevel regression, and structural equation models (see the main text and some additional information in Section 3 of the online supplementary material). The examples rendered GORICA weights that quantified the evidence in the data with respect to hypotheses under evaluation. In all our examples and simulations, the hypothesis of interest contained limited numbers of restrictions on structural parameters. It is known that the Bayes factor cannot easily be determined in high dimensional settings (Berger et al., 2003; Shang & Clayton, 2011). This is because the area in agreement with the restrictions of the hypothesis of interest is too small, which could also influence the calculation of the penalty in information criteria like the GORICA. Therefore, we investigated the calculation of the penalty for the GORICA under a high dimensional setting in the context of multiple regression (see Section 4 of the online supplementary material). We found that the penalty for the GORICA can still be accurately determined when the hypothesis of interest contains restrictions on 100 structural parameters.

Although the GORICA is a more easy-to-apply and flexible information criterion than the GORIC, it has some limitations when evaluating (in)equality constrained hypotheses under the general class of models. Here, we elaborate on three main limitations of the GORICA. First, the GORICA is an approximation of the GORIC which assumes that there is an adequate sample size

Figure 5

The Effect of Scaling the Covariance Matrix of Estimates on Determining the Space of Range Restriction $-2 < \theta_1 < 2$



Note. See the online article for the color version of this figure.

when evaluating hypotheses. Therefore, caution should be taken when using the GORICA in case of small samples. In this case, the GORICA sometimes was not able to correctly prefer the unconstrained hypothesis in the set when the other hypotheses under evaluation were not correct in the population. Note that this issue is not related only to the GORICA, but also to the GORIC. Moreover, this problem vanishes when sample size and/or effect size are large enough, that is, it asymptotically chooses the correct hypothesis (and in case of multiple correct hypotheses the most parsimonious one). Second, like all ICs,

Table 14

Estimates ($\hat{\boldsymbol{\theta}}$) and Covariance Matrix ($\hat{\boldsymbol{\Sigma}}_{\hat{\boldsymbol{\theta}}}$) of the Population Means for the ANOVA Example

	$\hat{\boldsymbol{\Sigma}}_{\hat{\boldsymbol{\theta}}}$		
$\hat{\boldsymbol{\theta}}$	$\hat{\theta}_1$	$\hat{\theta}_2$	$\hat{\theta}_3$
$\hat{\theta}_1 = 5.03$	0.04		
$\hat{\theta}_2 = 4.66$	0.00	0.04	
$\hat{\theta}_3 = 5.53$	0.00	0.00	0.04

Note. The respective terms are made bold because they are either vector or matrix.

Table 15

The Order-Restricted MLEs, $\tilde{\boldsymbol{\theta}}^m = (\tilde{\theta}_1^m, \tilde{\boldsymbol{\theta}}_2^m, \tilde{\boldsymbol{\theta}}_3^m)^T$, for Hypothesis H_m (m = 1, 2) for the ANOVA Example

Method	H_m	$\tilde{\theta}_1^m$	$\tilde{\theta}_2^m$	$\tilde{\theta}_3^m$
GORIC	H_1	5.03	4.84	5.34
	H_2	5.03	4.66	5.53
GORICA	H_1	5.03	4.84	5.34
	H_2	5.03	4.66	5.53

Note. MLE = maximum likelihood estimates; GORICA = generalized order-restricted information criterion approximation. The respective terms are made bold because they are either vector or matrix.

Table 16

The Likelihood ($L(\tilde{\boldsymbol{\theta}}^m|\hat{\boldsymbol{\theta}}, \hat{\boldsymbol{\Sigma}}_{\hat{\boldsymbol{\theta}}})$) and Penalty ($PT_m$) Parts, GORIC and GORICA Values ($GORIC_m, GORICA_m$), and GORIC and GORICA Weights (w_m) of Hypothesis H_m (m = 1, 2) for the ANOVA Example

H_m	$L(\tilde{\boldsymbol{\theta}}^m\|\hat{\boldsymbol{\theta}}, \hat{\boldsymbol{\Sigma}}_{\hat{\boldsymbol{\theta}}})$	PT_m	$GORIC_m$	w_m
H_1	−27.739	2.000	59.478	0.745
H_2	−26.810	4.000	61.619	0.255
H_m	$L(\tilde{\boldsymbol{\theta}}^m\|\hat{\boldsymbol{\theta}}, \hat{\boldsymbol{\Sigma}}_{\hat{\boldsymbol{\theta}}})$	PT_m	$GORICA_m$	w_m
H_1	1.252	1.000	−0.503	0.757
H_2	2.115	3.000	1.770	0.243

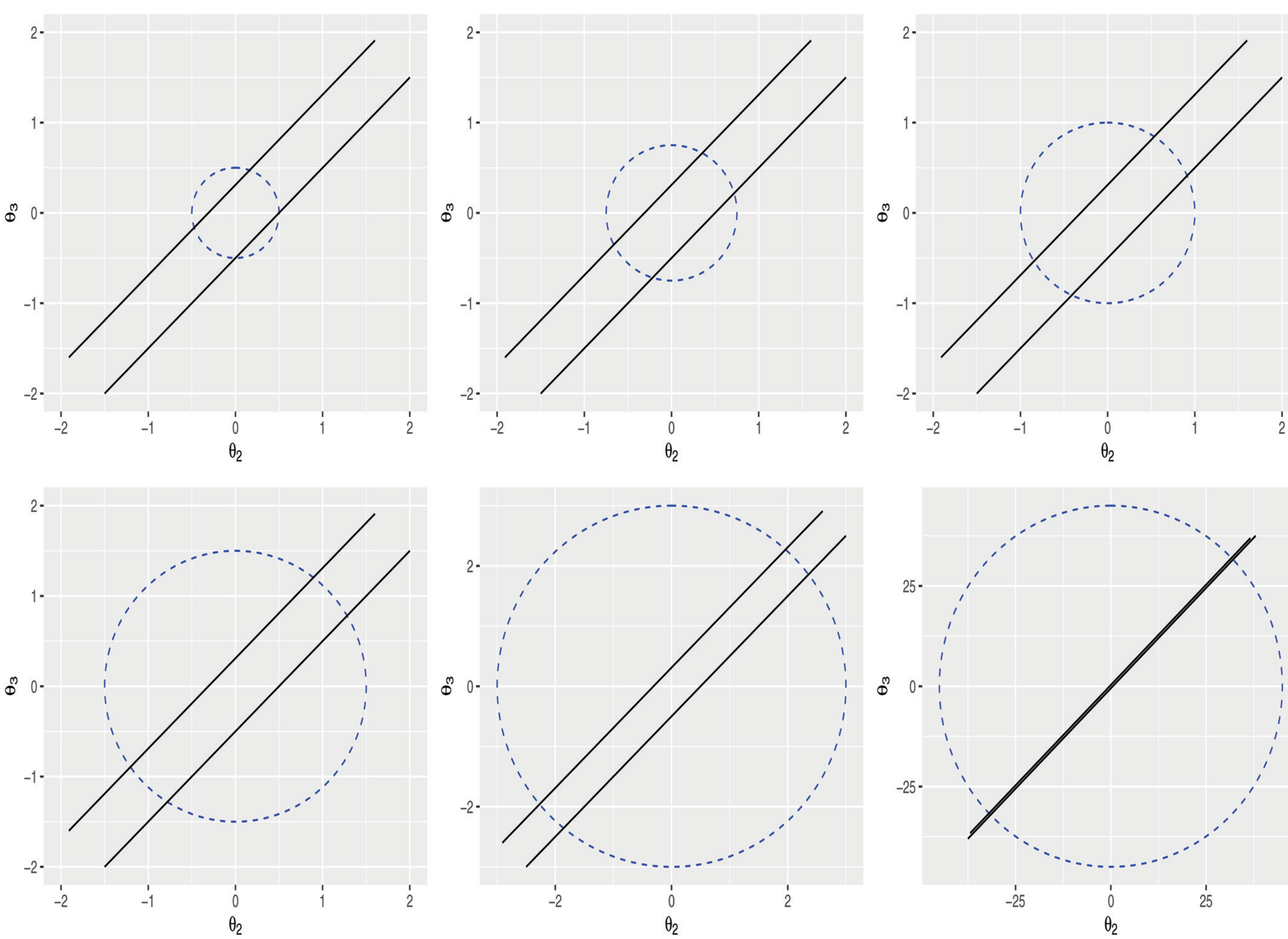
Note. GORICA = generalized order-restricted information criterion approximation. The respective terms are made bold because they are either vector or matrix.

the results of the GORICA is dependent of the method used for estimation. Occasionally, researchers may prefer to use another estimation technique than the MLE in conjunction with the GORICA. For example, it is well known that the MLE may produce bias parameter estimates and their asymptotic standard errors when the data at hand contain outliers, which influences the performance of the GORICA. In such a case, one may consider to apply an outlier robust method to estimate structural model parameters and their covariance matrix. The third limitation is a general limitation that applies not only to the GORICA but also to other statistical methods. Like with hypothesis testing or using other information criteria, the GORICA results depend on many other aspects of the study such as study design, the power of study, data quality and so forth. The focus of this study is to investigate the GORICA itself as an alternative to hypothesis testing and other information criteria. When applying the method in practice, these aspects should of course be taken into account.

The GORICA is implemented in the gorica package in R for each example presented in this article which is elaborated in the online supplementary material. This enables researchers to evaluate (in)equality constrained hypotheses in a wide range of statistical models.

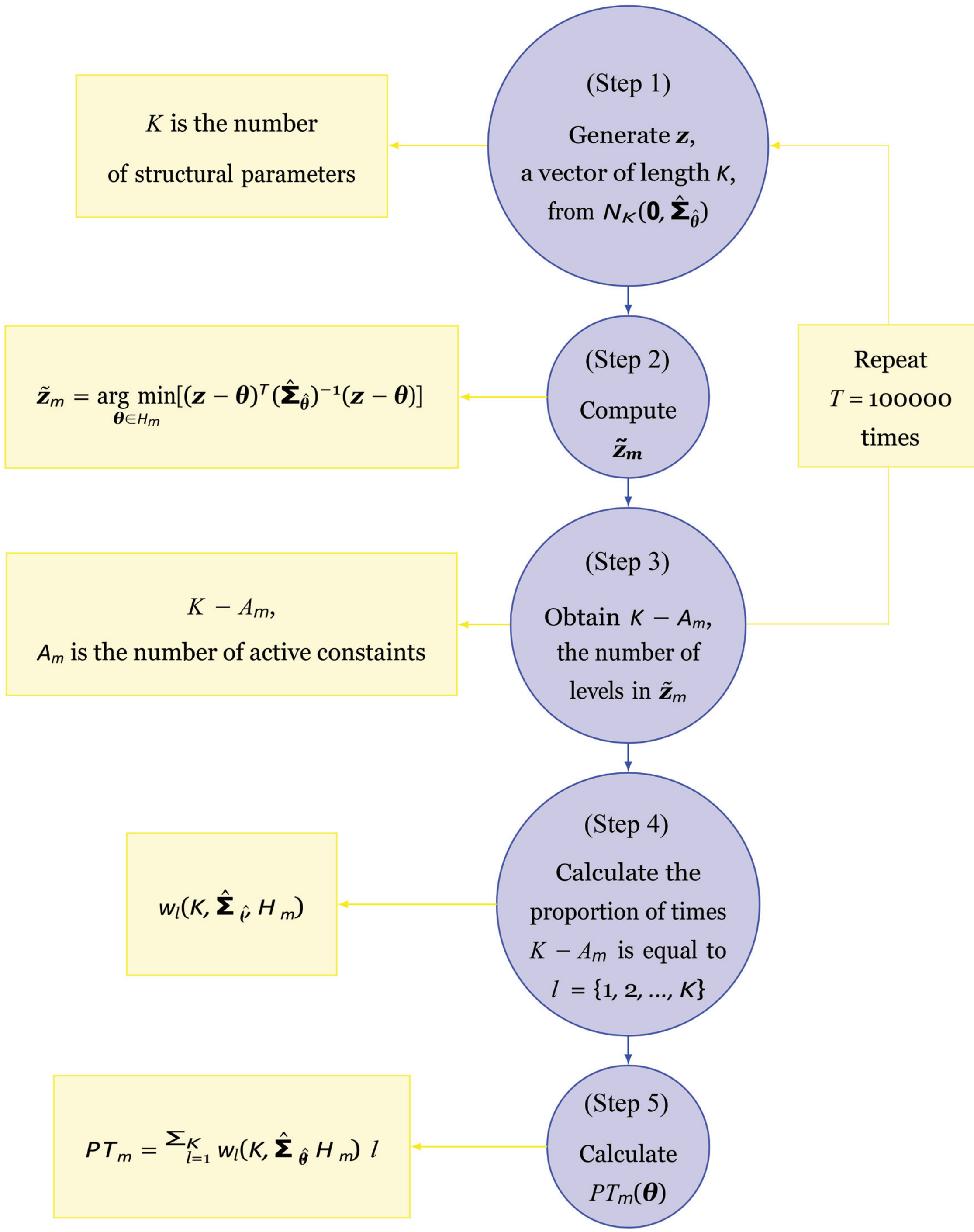
Figure 6

The Effect of Scaling the Covariance Matrix of Estimates on Determining the Space of Range Restriction $0.5\sigma_p < \{\theta_3 - \theta_2\} < 0.8\sigma_p$



Note. See the online article for the color version of this figure.

Figure 7
Working Scheme for Calculation of the Penalty Term



Note. The respective terms are made bold because they are either vector or matrix. See the online article for the color version of this figure.

References

Agnoli, F., Wicherts, J. M., Veldkamp, C. L., Albiero, P., & Cubelli, R. (2017). Questionable research practices among Italian research psychologists. *PloS ONE*, *12*(3), e0172792. https://doi.org/10.1371/journal.pone.0172792

Akaike, H. (1973). Information theory and an extension of the maximum likelihood principle. In B. N. Petrov & F. Csaki (Eds.), *Proceedings of the Second International Symposium on Information Theory* (pp. 267–281). Akailseoniai-Kuido. https://doi.org/10.1007/978-1-4612-1694-0_15

Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transaction on Automatic Control*, *19*(6), 716–723. https://doi.org/10.1109/TAC.1974.1100705

Algina, J., & Swaminathan, H. (2011). *The handbook of advanced multilevel data analysis*. Routledge.

Anraku, K. (1999). An information criterion for parameters under a simple order restriction. *Biometrika*, *86*(1), 141–152. https://doi.org/10.1093/biomet/86.1.141

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1), 1–48. https://doi.org/10.18637/jss.v067.i01

Bem, D. J. (2011). Feeling the future: Experimental evidence for anomalous retroactive influences on cognition and affect. *Journal of Personality and Social Psychology*, *100*(3), 407–425. https://doi.org/10.1037/a0021524

Berger, J. O., Ghosh, J. K., & Mukhopadhyay, N. (2003). Approximations and consistency of Bayes factors as model dimension groups. *Journal of Statistical Planning and Inference*, *112*(1), 241–258. https://doi.org/10.1016/S0378-3758(02)00336-1

Bollen, K. (1989). *Structural equations with latent variables*. Wiley and Sons.

Burnham, K. P., & Anderson, D. R. (2002). *Model selection and multimodel inference: A practical information-theoretic approach* (2nd ed.). Springer-Verlag.

Campbell, M. J., & Machin, D. (1999). *Medical statistics: A commonsense approach* (3rd ed.). Wiley.

Cohen, J. (1992). A power primer. *Psychological Bulletin*, *112*(1), 155–159. https://doi.org/10.1037/0033-2909.112.1.155

Cohen, J. (1994). The earth is round (p <.05). *American Psychologist*, *49*(12), 997–1003. https://doi.org/10.1037/0003-066X.49.12.997

Fisher, R. A. (1922). On the mathematical foundations of theoretical statistics. *Philosophical Transactions of the Royal Society*, *222*(594), 309–368. https://doi.org/10.1098/rsta.1922.0009

Gelman, A. (2008). Scaling regression inputs by dividing by two standard deviations. *Statistics in Medicine*, *27*(15), 2865–2873. https://doi.org/10.1002/sim.3107

Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A., & Rubin, D. B. (2013). *Bayesian data analysis* (3rd ed.). Chapman & Hall/CRC.

Gu, X., Mulder, J., & Hoijtink, H. (2018). Approximated adjusted fractional Bayes factors: A general method for testing informative hypotheses. *British Journal of Mathematical and Statistical Psychology*, *71*(2), 229–261. https://doi.org/10.1111/bmsp.12110

Hoijtink, H. (2012). *Informative hypotheses: Theory and practice for behavioral and social scientists* (2nd ed.). Chapman & Hall/CRC.

Hox, J. J. (2010). *Multilevel analysis: Techniques and applications* (2nd ed.). Routledge.

Kass, R. (1993). Bayes factors in practice. *Journal of the Royal Statistical Society, Series D, the Statistician*, *42*(5), 551–560. https://doi.org/10.2307/2348679

Kass, R., & Raftery, A. (1995). Bayes factors. *Journal of the American Statistical Association*, *90*(430), 773–795. https://doi.org/10.1080/01621459.1995.10476572

Kerr, N. L. (1998). HARKing: Hypothesizing after the results are known. *Personality and Social Psychology Review*, *2*(3), 196–217. https://doi.org/10.1207/s15327957pspr0203_4

Kuiper, R. M., Hoijtink, H., & Silvapulle, M. J. (2011). An Akaike-type information criterion for model selection under inequality constraints. *Biometrika*, *98*(2), 495–501. https://doi.org/10.1093/biomet/asr002

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*(8), 2454–2463. https://doi.org/10.1016/j.jspi.2012.03.007

Kuiper, R. M., & Hoijtink, H. (2010). Comparisons of means using explanatory and confirmatory approaches. *Psychological Methods*, *15*(1), 69–86. https://doi.org/10.1037/a0018720

Kuiper, R. M., & Hoijtink, H. (2013). A Fortran 90 program for the generalized order-restricted information criterion. *Journal of Statistical Software*, *54*(8), 1–19. https://doi.org/10.18637/jss.v054.i08

Kullback, S., & Leibler, R. A. (1951). On information and sufficiency. *Annals of Mathematical Statistics*, *22*(1), 79–86. http://www.jstor.org/stable/2236703

Ioannidis, J. P. (2005). Why most published research findings are false. *PLoS Med*, *2*(8), e124. https://doi.org/10.1371/journal.pmed.0020124

Lorah, J. (2018). Effect size measures for multilevel models: definition, interpretation, and TIMMSS example. *Large-Scale Assessments in Education*, *6*(8), 2018. https://doi.org/10.1186/s40536-018-0061-2

Maas, C. J. M., & Hox, J. J. (2005). Sufficient sample sizes for multilevel modeling. *Methodology: European Journal of Research Methods for the Behavioral and Social Sciences*, *1*(3), 86–92. https://doi.org/10.1027/1614-2241.1.3.86

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*(6), 487–498. https://doi.org/10.1037/a0039400

McCullagh, P., & Nelder, J. (1989). *Generalized linear models* (2nd ed.). Chapman & Hall/CRC.

McCullogh, C. E., & Searle, S. R. (2001). *Generalized linear and mixed models*. Wiley.

Nederhof, E., Ormel, J., & Oldehinkel, A. J. (2014). Mismatch or cumulative stress: The pathway to depression is conditional on attention style. *Psychological Science*, *25*(3), 684–692. https://doi.org/10.1177/0956797613513473

Nederhof, E., & Schmidt, M. V. (2012). Mismatch or cumulative stress: Toward an integrated hypothesis of programming effects. *Physiology & Behavior*, *106*(5), 691–700. https://doi.org/10.1016/j.physbeh.2011.12.008

Pashler, H., & Wagenmakers, E.-J. (2012). Eds.' introduction to the special section on replicability in psychological science: A crisis of confidence? *Perspectives on Psychological Science*, *7*(6), 528–530. https://doi.org/10.1371/journal.pone.0033423

R Core Team. (2012). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. http://www.R-project.org

Raftery, A. E. (1995). Bayesian model selection in social research. *Sociological Methodology*, *25*, 111–163. https://doi.org/10.2307/271063

Richesson, R. L., & Andrews, J. E. (2012). *Clinical research informatics*. Springer-Verlag London Limited.

Rosseel, Y. (2012). Lavaan: An R package for structural equation modeling. *Journal of Statistical Software*, *48*(2), 1–36. https://doi.org/10.18637/jss.v048.i02

Royall, R. M. (1997). *Statistical evidence: A likelihood paradigm*. Chapman & Hall.

Schimmack, U. (2012). The ironic effect of significant results on the credibility of multiple-study articles. *Psychological Methods*, *17*(4), 551–566. https://doi.org/10.1037/a0029487

Shang, Z., & Clayton, M. K. (2011). Consistency of Bayesian linear model selection with a growing number of parameters. *Journal of Statistical Planning and Inference*, *141*(11), 3463–3474. https://doi.org/10.1016/j.jspi.2011.05.002

Silvapulle, M. J., & Sen, P. K. (2005). *Constrained statistical inference: Inequality order and shape restrictions*. Wiley.

Stern, J. M., & Simes, R. J. (1997). Publication bias: Evidence of delayed publication in a cohort study of clinical research projects. *British Medical Journal*, *315*(7109), 640–645. https://doi.org/10.1136/bmj.315.7109.640

Stevens, J. (1999). *Applied multivariate statistics for the social sciences*. (3rd ed.). Lawrence Erlbaum Associates, Inc.

Turlach, B. A. (2014). quadprog: Functions to solve quadratic programming problems (R Package Version 1.5-5). http://CRAN.R-project.org/package=quadprog

Van Lissa, C., Altinisik, Y., & Kuiper, R. M. (2020). *gorica: Evaluation of inequality constrained hypotheses using generalized AIC (R package Version 0.1.0)*. https://informative-hypotheses.sites.uu.nl/software/goric/

Vanbrabant, L. (2020). restriktor: Restricted statistical estimation and inference for linear models *(R Package Version 0.2-800)*. http://CRAN.R-project.org/package=restriktor

Vanbrabant, L., Van Loey, N., & Kuiper, R. M. (2020). Evaluating a theory-based hypothesis against its complement using an AIC-type information criterion with an application to facial burn injury. *Psychological Methods*, *25*(2), 129–142. https://doi.org/10.1037/met0000238

Venables, W. N., & Ripley, B. D. (2002). *Modern applied statistics with S. Fourth edition*. Springer. http://www.stats.ox.ac.uk/pub/MASS4

Wagenmakers, E. J. (2007). A practical solution to the pervasive problem of p values. *Psychonomic Bulletin & Review*, *14*(5), 779–804. https://doi.org/10.3758/bf03194105

Wagenmakers, E.-J., Wetzels, R., Borsboom, D., van der Maas, H. L. J., & Kievit, R. A. (2012). An agenda for purely confirmatory research. *Perspectives on Psychological Science*, *7*(6), 632–638. https://doi.org/10.1177/1745691612463078

Wainer, H. (1999). One cheer for null hypothesis significance testing. *Psychological Methods*, *4*(2), 212–213. https://doi.org/10.1037/1082-989X.4.2.212

Willis, C. E., & Perlack, R. D. (1978). Multicollinearity: Effects, symptoms, and remedies. *Journal of the Northeastern Agricultural Economics Council*, *7*(1), 55–61. https://doi.org/10.1017/S0163548400001989

Wolf, E. J., Harrington, K. M., Clark, S. H., & Miller, M. W. (2013). Sample size requirements for structural equation models: An evaluation of power, bias, and solution propriety. *Educational and Psychological Measurement*, *76*(6), 913–934. https://doi.org/10.1177/0013164413495237

Appendix A

Derivation of the GORICA

In this appendix, we will derive the GORICA, which is the asymptotic expression of the GORIC (when leaving out the nuisance parameters). Hence, the derivation of the GORICA is based on that of the GORIC (Kuiper et al., 2012, pp. 2455–2458), in which it is assumed that the density of the data follows a normal distribution. For the more general class of models that we inspect in this article (e.g., logistic regression models and mixed models), this normality assumption holds asymptotically (Gelman et al., 2013, pp. 83–88). When the likelihood is unimodal, roughly symmetric and twice differentiable, we can usually accurately approximate it by a normal distribution centered at the MLE. The second order Taylor expansion of the log likelihood centered at the MLE $\hat{\boldsymbol{\eta}}$ is given by:

$$\log L(\boldsymbol{\eta} \mid Y) \approx \log L(\hat{\boldsymbol{\eta}} \mid Y) + \frac{1}{2}(\boldsymbol{\eta} - \hat{\boldsymbol{\eta}})^T \left[\frac{d^2}{d\boldsymbol{\eta}^2} \log L(\boldsymbol{\eta} \mid Y)\right]_{\boldsymbol{\eta}=\hat{\boldsymbol{\eta}}} (\boldsymbol{\eta} - \hat{\boldsymbol{\eta}}),$$

with $\boldsymbol{\eta}$ is the vector of the model parameters and Y denotes the data at hand, where we have used that the first derivative of the log likelihood is zero at its MLE.

This equation yields:

$$L(\boldsymbol{\eta} \mid Y) \approx N(\boldsymbol{\eta} \mid \hat{\boldsymbol{\eta}}, \hat{\boldsymbol{\Sigma}}_{\hat{\boldsymbol{\eta}}}). \tag{18}$$

with $\hat{\boldsymbol{\Sigma}}_{\hat{\boldsymbol{\eta}}}$ the covariance matrix of $\hat{\boldsymbol{\eta}}$.[2]

In the asymptotic expression for the GORIC, we will maximize this likelihood (i.e., the asymptotic likelihood) under the restrictions in hypothesis H_m for $m = 1, \ldots, M$.

The asymptotic expression for the GORIC is thus given by:

$$\begin{aligned} &-2\log L(\tilde{\boldsymbol{\eta}}^m \mid \hat{\boldsymbol{\eta}}, \hat{\boldsymbol{\Sigma}}_{\hat{\boldsymbol{\eta}}}) + 2\,PT_m(\boldsymbol{\eta}), \text{with} \\ &\tilde{\boldsymbol{\eta}}^m = \arg\max_{\boldsymbol{\eta} \in H_m} \log L(\boldsymbol{\eta} \mid \hat{\boldsymbol{\eta}}, \hat{\boldsymbol{\Sigma}}_{\hat{\boldsymbol{\eta}}}), \end{aligned} \tag{19}$$

where $PT_m(\boldsymbol{\eta})$ denotes the penalty term for Hypothesis H_m and $\tilde{\boldsymbol{\eta}}^m$ is the order-restricted MLE of $\boldsymbol{\eta}$, that is, the value of $\boldsymbol{\eta}$ that maximizes the likelihood under the restrictions in Hypothesis H_m.

Notably, $\boldsymbol{\eta}$ in Equation 19 contains all the model parameters: both structural ($\boldsymbol{\theta}$) and nuisance parameters ($\boldsymbol{\xi}$). In the next section, we will show for the asymptotic expression of the GORIC in Equation 19 that we can leave out the nuisance parameters. We will do this for both the (log) likelihood part and the penalty part. Afterwards, we will define the expression of the GORICA.

When using ICs like the AIC, GORIC or GORICA, the values of these ICs themselves are not of importance but differences between two IC values are. To clarify, let the IC values for hypotheses H_m and $H_{m'}$ be

$$\begin{aligned} IC_m &= -2\log L_m + 2\,PT_m, \\ IC_{m'} &= -2\log L_{m'} + 2\,PT_{m'}, \end{aligned}$$

respectively, where $\log L$ stands for the log likelihood and PT for the penalty. Using this notation, if IC_m is smaller than $IC_{m'}$, then the support for H_m is stronger than that for $H_{m'}$. Stated otherwise, the support for H_m is stronger if $IC_m - IC_{m'} < 0$, that is,

[2] The covariance matrix in the normal approximation equals the inverse of the expected Fisher information $(I(\hat{\boldsymbol{\eta}})^{-1})$, with $I(\eta) = -E\left[\frac{d^2}{d\boldsymbol{\eta}^2} \log L(\boldsymbol{\eta} \mid Y)\right]$, where the random variable Y has been averaged out. Based on the Cramér-Rao bound, $I(\hat{\boldsymbol{\eta}})^{-1}$ is a lower bound on the covariance matrix of $\hat{\boldsymbol{\eta}}$, that is, $\hat{\boldsymbol{\Sigma}}_{\hat{\boldsymbol{\eta}}} \geq I(\hat{\boldsymbol{\eta}})^{-1}$; where the bound is attained in case of a consistent estimator. Consequently, since the MLE is consistent, we can use $I(\hat{\boldsymbol{\eta}})^{-1} = \hat{\boldsymbol{\Sigma}}_{\hat{\boldsymbol{\eta}}}$.

(*Appendices continue*)

$\log L_m + PT_m - (\log L_{m'} + PT_{m'}) = \log L_m - \log L_{m'} + PT_m - PT_{m'} < 0$. Hence, parts that are constant over hypotheses will cancel out when comparing hypotheses.

In this section, we will demonstrate that the nuisance parameters can be left out in the log likelihood part and in the penalty part of the asymptotic expression of the GORIC in Equation 19. Let $\hat{\boldsymbol{\Sigma}}_{\hat{\boldsymbol{\eta}}}$ consist of four block matrices:

$$\hat{\boldsymbol{\Sigma}}_{\hat{\boldsymbol{\eta}}} = \left[\begin{array}{c|c} \hat{\boldsymbol{\Sigma}}_{\hat{\theta}\hat{\theta}} & \hat{\boldsymbol{\Sigma}}_{\hat{\theta}\hat{\xi}} \\ \hline \hat{\boldsymbol{\Sigma}}_{\hat{\xi}\hat{\theta}} & \hat{\boldsymbol{\Sigma}}_{\hat{\xi}\hat{\xi}} \end{array}\right].$$

The likelihood $L(\boldsymbol{\eta} \mid \hat{\boldsymbol{\eta}}, \hat{\boldsymbol{\Sigma}}_{\hat{\boldsymbol{\eta}}})$ that needs to be maximized under the restrictions in H_m can now be written as:

$$\begin{aligned} L(\eta|\hat{\eta}, \hat{\boldsymbol{\Sigma}}_{\hat{\eta}}) &= L(\theta, \xi|\hat{\theta}, \hat{\xi}, \hat{\boldsymbol{\Sigma}}_{\hat{\theta}\hat{\theta}}, \hat{\boldsymbol{\Sigma}}_{\hat{\theta}\hat{\xi}}, \hat{\boldsymbol{\Sigma}}_{\hat{\xi}\hat{\theta}}, \hat{\boldsymbol{\Sigma}}_{\hat{\xi}\hat{\xi}}) \\ &= L(\theta|\hat{\theta}, \hat{\xi}, \hat{\boldsymbol{\Sigma}}_{\hat{\theta}\hat{\theta}}, \hat{\boldsymbol{\Sigma}}_{\hat{\theta}\hat{\xi}}, \hat{\boldsymbol{\Sigma}}_{\hat{\xi}\hat{\theta}}, \hat{\boldsymbol{\Sigma}}_{\hat{\xi}\hat{\xi}}) \cdot L(\xi|\theta, \hat{\theta}, \hat{\xi}, \hat{\boldsymbol{\Sigma}}_{\hat{\theta}\hat{\theta}}, \hat{\boldsymbol{\Sigma}}_{\hat{\theta}\hat{\xi}}, \hat{\boldsymbol{\Sigma}}_{\hat{\xi}\hat{\theta}}, \hat{\boldsymbol{\Sigma}}_{\hat{\xi}\hat{\xi}}) \\ &= L(\theta|\hat{\theta}, \hat{\boldsymbol{\Sigma}}_{\hat{\theta}\hat{\theta}}) \cdot L(\xi|\theta, \hat{\theta}, \hat{\xi}, \hat{\boldsymbol{\Sigma}}_{\hat{\theta}\hat{\theta}}, \hat{\boldsymbol{\Sigma}}_{\hat{\theta}\hat{\xi}}, \hat{\boldsymbol{\Sigma}}_{\hat{\xi}\hat{\theta}}, \hat{\boldsymbol{\Sigma}}_{\hat{\xi}\hat{\xi}}) \end{aligned} \tag{20}$$

By definition, this likelihood is maximized for the order-restricted MLEs $\tilde{\boldsymbol{\theta}}^m$ and $\tilde{\boldsymbol{\xi}}^m$. Hence, the maximized likelihood can be written as

$$\begin{aligned} L(\tilde{\boldsymbol{\eta}}^m|\hat{\boldsymbol{\eta}}, \hat{\boldsymbol{\Sigma}}_{\hat{\boldsymbol{\eta}}}) &= L(\tilde{\boldsymbol{\theta}}^m|\hat{\boldsymbol{\theta}}, \hat{\boldsymbol{\Sigma}}_{\hat{\theta}\hat{\theta}}) \cdot L(\tilde{\boldsymbol{\xi}}^m|\boldsymbol{\theta}, \hat{\boldsymbol{\theta}}, \hat{\boldsymbol{\xi}}, \hat{\boldsymbol{\Sigma}}_{\hat{\theta}\hat{\theta}}, \hat{\boldsymbol{\Sigma}}_{\hat{\theta}\hat{\xi}}, \hat{\boldsymbol{\Sigma}}_{\hat{\xi}\hat{\theta}}, \hat{\boldsymbol{\Sigma}}_{\hat{\xi}\hat{\xi}}) \\ &= L(\tilde{\boldsymbol{\theta}}^m|\hat{\boldsymbol{\theta}}, \hat{\boldsymbol{\Sigma}}_{\hat{\theta}\hat{\theta}}) \cdot L(\tilde{\boldsymbol{\xi}}^m|\hat{\boldsymbol{\mu}}_{\xi|\theta=\tilde{\theta}^m}, \hat{\boldsymbol{\Sigma}}_{\xi|\theta=\tilde{\theta}^m}), \end{aligned} \tag{21}$$

with $\hat{\boldsymbol{\mu}}_{\xi\,|\,\theta=\tilde{\theta}^m}$ and $\hat{\boldsymbol{\Sigma}}_{\xi\,|\,\theta=\tilde{\theta}^m}$, the conditional mean and conditional covariance matrix of $\boldsymbol{\xi}$, respectively:

$$\begin{aligned} \hat{\boldsymbol{\mu}}_{\xi\,|\,\theta=\tilde{\theta}^m} &= \hat{\boldsymbol{\xi}} + \hat{\boldsymbol{\Sigma}}_{\hat{\xi}\hat{\theta}} \hat{\boldsymbol{\Sigma}}_{\hat{\theta}\hat{\theta}}^{-1} (\tilde{\boldsymbol{\theta}}^m - \hat{\boldsymbol{\theta}}), \\ \hat{\boldsymbol{\Sigma}}_{\xi\,|\,\theta=\tilde{\theta}^m} &= \hat{\boldsymbol{\Sigma}}_{\hat{\xi}\hat{\xi}} - \hat{\boldsymbol{\Sigma}}_{\hat{\xi}\hat{\theta}} \hat{\boldsymbol{\Sigma}}_{\hat{\theta}\hat{\theta}}^{-1} \hat{\boldsymbol{\Sigma}}_{\hat{\theta}\hat{\xi}} \\ &= \hat{\boldsymbol{\Sigma}}_{\xi}. \end{aligned}$$

Note that, when maximizing the likelihood under the restrictions in H_m, the order-restricted MLE of the nuisance parameters ($\tilde{\boldsymbol{\xi}}^m$) will equal $\hat{\boldsymbol{\mu}}_{\xi\,|\,\theta}$, which value is dependent on $\tilde{\boldsymbol{\theta}}^m$, while the conditional covariance matrix is not (hence the notation $\hat{\boldsymbol{\Sigma}}_{\xi}$). Note further that

Writing out the likelihoods in Equation 21 as log likelihoods gives:

$$\begin{aligned} \log L(\tilde{\boldsymbol{\theta}}^m|\hat{\boldsymbol{\theta}}, \hat{\boldsymbol{\Sigma}}_{\hat{\theta}}) &= constant + \log|\hat{\boldsymbol{\Sigma}}_{\hat{\theta}\hat{\theta}}| - \frac{1}{2}(\hat{\boldsymbol{\theta}} - \tilde{\boldsymbol{\theta}}^m) T \hat{\boldsymbol{\Sigma}}_{\hat{\theta}\hat{\theta}}^{-1} (\hat{\boldsymbol{\theta}} - \tilde{\boldsymbol{\theta}}^m), \\ \log L(\tilde{\boldsymbol{\xi}}^m|\hat{\boldsymbol{\mu}}_{\xi|\theta=\tilde{\theta}^m}, \hat{\boldsymbol{\Sigma}}_{\xi}) &= constant + \log|\hat{\boldsymbol{\Sigma}}_{\xi}| - \frac{1}{2}(\hat{\boldsymbol{\mu}}_{\xi|\theta=\tilde{\theta}\mathrm{m}-} \tilde{\boldsymbol{\xi}} m) T \hat{\boldsymbol{\Sigma}}_{\xi}^{-1} (\hat{\boldsymbol{\mu}}_{\xi|\theta=\tilde{\theta}^m} - \tilde{\boldsymbol{\xi}}^m) \\ &= constant + \log|\hat{\boldsymbol{\Sigma}}_{\xi}|, \end{aligned}$$

using $\tilde{\boldsymbol{\xi}}^m = \hat{\boldsymbol{\mu}}_{\xi\,|\,\theta=\tilde{\theta}^m}$ in the last line. This implies that the maximized joint log order-restricted likelihood, used in the asymptotic GORIC expression in Equation 19, can be written as:

$$\log L(\tilde{\boldsymbol{\eta}}^m \mid \hat{\boldsymbol{\eta}}, \hat{\boldsymbol{\Sigma}}_{\hat{\boldsymbol{\eta}}}) = constant + \log|\hat{\boldsymbol{\Sigma}}_{\hat{\theta}\hat{\theta}}| - \frac{1}{2}(\hat{\boldsymbol{\theta}} - \tilde{\boldsymbol{\theta}}^m) T \hat{\boldsymbol{\Sigma}}_{\hat{\theta}\hat{\theta}}^{-1} (\hat{\boldsymbol{\theta}} - \tilde{\boldsymbol{\theta}}^m) + constant + \log|\hat{\boldsymbol{\Sigma}}_{\xi}|.$$

Note that both *constant*s, $\log|\hat{\boldsymbol{\Sigma}}_{\hat{\theta}\hat{\theta}}|$, and $\log|\hat{\boldsymbol{\Sigma}}_{\xi}|$ are independent of the hypothesis of interest (H_m).

We will first focus on the log likelihood part, that is, $\log L_m - \log L_{m'}$

$$\begin{aligned} \log L_m - \log L_{m'} &= \log L(\tilde{\eta}^m|\hat{\eta}, \hat{\boldsymbol{\Sigma}}_{\hat{\eta}}) - \log L(\tilde{\eta}^{m'}|\hat{\eta}, \hat{\boldsymbol{\Sigma}}_{\hat{\eta}}) \\ &= constant + \log|\hat{\boldsymbol{\Sigma}}_{\hat{\theta}\hat{\theta}}| - \frac{1}{2}(\hat{\theta} - \tilde{\theta}^m)^T \hat{\boldsymbol{\Sigma}}_{\hat{\theta}\hat{\theta}}^{-1} (\hat{\theta} - \tilde{\theta}^m) + constant + \log|\hat{\boldsymbol{\Sigma}}_{\xi}| - \\ &\left(constant + \log|\hat{\boldsymbol{\Sigma}}_{\hat{\theta}\hat{\theta}}| - \frac{1}{2}(\hat{\theta} - \tilde{\theta}^{m'})^T \hat{\boldsymbol{\Sigma}}_{\hat{\theta}\hat{\theta}}^{-1} (\hat{\theta} - \tilde{\theta}^{m'}) + constant + \log|\hat{\boldsymbol{\Sigma}}_{\xi}| \right) \\ &= -\frac{1}{2}(\hat{\theta} - \tilde{\theta}^m)^T \hat{\boldsymbol{\Sigma}}_{\hat{\theta}\hat{\theta}}^{-1} (\hat{\theta} - \tilde{\theta}^m) - (-\frac{1}{2}(\hat{\theta} - \tilde{\theta}^{m'})^T \hat{\boldsymbol{\Sigma}}_{\hat{\theta}\hat{\theta}}^{-1} (\hat{\theta} - \tilde{\theta}^{m'})), \end{aligned}$$

where the last line does not contain nuisance parameters anymore, but only the structural ones and their covariance matrix.

Thus, when comparing hypotheses using the asymptotic GORIC expression, only the differences for H_m and $H_{m'}$ with respect to the first term on the last line of Equation 21 is needed, that is, only the normal approximation of the likelihood of the structural parameters is needed. Stated otherwise, when comparing hypotheses and inspecting the log likelihood, one can leave out the nuisance parameters.

The same line of reasoning holds for the penalty part. The penalty part with respect to the nuisance parameters, $PT_m(\boldsymbol{\xi})$, equals the number of distinct nuisance parameters, even though the values of the nuisance parameters are dependent on the hypothesis of interest, they are all unrestricted and thus distinct. Hence, the penalty term for hypothesis H_m is:

$$\begin{aligned} PT_m &= PT_m(\boldsymbol{\eta}) \\ &= PT_m(\boldsymbol{\theta}, \boldsymbol{\xi}) \\ &= PT_m(\boldsymbol{\theta}) + PT_m(\boldsymbol{\xi}). \end{aligned}$$

Thus, $PT_m(\boldsymbol{\xi}) = PT(\boldsymbol{\xi})$ is constant over all hypotheses and, when comparing hypotheses, it will cancel out:

$$PT_m - PT_{m'} = PT_m(\boldsymbol{\theta}) - PT_{m'}(\boldsymbol{\theta}).$$

Hence, we can leave out the nuisance parameters also in the penalty term.

In sum, we took the following steps:

- We use the asymptotic expression for the likelihood, which is a normal distribution: $L(\boldsymbol{\eta} \mid \hat{\boldsymbol{\eta}}, \boldsymbol{\Sigma}_{\hat{\eta}})$, with $\boldsymbol{\eta} = (\boldsymbol{\theta}, \boldsymbol{\xi})$;
- Because we compare models, we can replace this likelihood by a likelihood based on just the structural parameters: $L(\boldsymbol{\theta} \mid \hat{\boldsymbol{\theta}}, \boldsymbol{\Sigma}_{\hat{\theta}\hat{\theta}})$;
- Note that the log likelihood is maximized under the restrictions under H_m, which then results in: $\log L(\tilde{\boldsymbol{\theta}}^m \mid \hat{\boldsymbol{\theta}}, \boldsymbol{\Sigma}_{\hat{\theta}\hat{\theta}})$;
- Because we compare models, we can also replace the penalty part by the one based on just the structural parameters: $PT_m(\boldsymbol{\theta})$.

This leads to the following GORICA expression:

$$\begin{aligned} GORICA_m &= \log L_m + PT_m(\boldsymbol{\theta}), \text{ with} \\ \log L_m &= \log L(\tilde{\eta}^m \mid \hat{\boldsymbol{\theta}}, \hat{\boldsymbol{\Sigma}}_{\hat{\theta}\hat{\theta}}). \end{aligned}$$

(*Appendices continue*)

For readability, we use $\hat{\boldsymbol{\Sigma}}_{\hat{\theta}}$ instead of $\hat{\boldsymbol{\Sigma}}_{\hat{\theta}\hat{\theta}}$ in the main text and in online supplemental material.

When comparing this to the GORIC expression in Equation 19, we see that this expression is no longer asymptotically the same in case there are nuisance parameters. However, when comparing two hypotheses, the difference in GORICA values is asymptotically the same as the difference in GORIC values. This implies that the GORIC weights and the GORICA weights will asymptotically equate.

Appendix B

The GORICA (Continued)

In this appendix, we elaborate on the GORICA in terms of how to obtain the order-restricted MLEs, $\tilde{\boldsymbol{\theta}}^m$, and how to compute the penalty part, $PT_m(\boldsymbol{\theta})$, respectively. These calculations are similar to that of the GORIC.

The Order-Restricted MLEs

The order-restricted MLEs of the structural parameters $\tilde{\boldsymbol{\theta}}^m \in \mathrm{IR}^{K\times 1}$ are obtained by

$$\tilde{\boldsymbol{\theta}}^m = \underset{\boldsymbol{\theta}\in H_m}{\arg\min}\ [(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})(\hat{\boldsymbol{\Sigma}}_{\hat{\theta}})^{-1}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})^T], \tag{22}$$

which maximizes the log likelihood in Equation 9 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 $\tilde{\theta}_1 = 0$ and $\tilde{\theta}_2 = \hat{\theta}_2$. When the hypotheses under evaluation are more complex, the calculations of the order-restricted MLEs become more complex as well. In our software, a quadratic programming algorithm the solve.QP subroutine of the quadprog package (Turlach, 2014, pp. 2–4) in R is used to calculate the values of $\tilde{\boldsymbol{\theta}}^m$.

Computing the Penalty

The penalty part $PT_m(\theta)$ is used to penalize hypothesis H_m while taking into account the equality and/or inequality constraints imposed on the parameters in H_m. Next, we describe the steps to be taken when computing the penalty part of the GORICA.

First, a vector $\boldsymbol{z} = (z_1, z_2, \ldots, z_K)^T \in \mathrm{IR}^{K\times 1}$, with K is the number of the structural model parameters that are used to formulate the hypotheses under evaluation, is sampled from a normal distribution with mean vector 0 and covariance matrix $\hat{\Sigma}\hat{\theta} \in \mathrm{IR}^{K\times K}$ (see Step 1 in Figure 7). Second, the vector of order-restricted estimates $\tilde{\boldsymbol{z}}_m = (\tilde{z}_1^m, \tilde{z}_2^m, \ldots, \tilde{z}_K^m)^T \in \mathrm{IR}^{K\times 1}$ is computed using the values of $\boldsymbol{z}$ (see Step 2 in Figure 7). Third, the number of the levels in $\tilde{\boldsymbol{z}}_m$ is calculated, that is, $K - A_m$, where K is the number of structural parameters and A_m is the number of active constraints in H_m (see Step 3 in Figure 7). Note that a constraint becomes active when it is imposed on $\boldsymbol{z}$. For example, if $z_1 < 0$ and $z_2 > 0$ for hypothesis $H_1: \theta_1 > 0, \theta_2 > 0$ where $K = 2$, then the first constraint in H_1 is an active (or a violated) constraint but the second constraint is inactive (or in agreement). The first three steps are repeated to calculate the level probabilities ($w_l(.)$; Silvapulle & Sen, 2005, pp. 78–81), that is, the probabilities that the vector of order-restricted estimates $\tilde{\boldsymbol{z}}_m$ has l levels (see Step 4 in Figure 7), with $l = \{1, 2, \ldots, K\}$ and K is the number of the structural parameters in hypothesis H_m, for example, for T = 100, 000 times (which is an adequate number of iterations). The penalty part $PT_m(\boldsymbol{\theta})$ (see Step 5 in Figure 7) is defined as:

$$PT_m(\boldsymbol{\theta}) = \sum_{l=1}^{K} w_l(K, \hat{\boldsymbol{\Sigma}}_{\hat{\theta}}, H_m)\, l. \tag{23}$$

For example, suppose $K = 3$ and $\hat{\boldsymbol{\Sigma}}_{\hat{\theta}} = \boldsymbol{I}_3$, where $\boldsymbol{I}_3$ denotes a 3×3 identity matrix. Then for hypothesis $H_1: \theta_1, \theta_2, \theta_3$, $PT_1 = w_1\times 1 + w_2\times 2 + w_3\times 3 = 3$, with $w_1 = w_2 = 0$ and $w_3 = 1$. Note that, in general, the unconstrained hypothesis H_M implies no restriction on the parameters, which is represented by $w_K = 1$ and $K - A_M = K$; consequently, $PT_M(\boldsymbol{\theta}) = K$. When calculating the level probabilities for hypothesis $H_2: \theta_1 > \theta_2 > \theta_3$, there are six possible orders between the structural model parameters:

$$\theta_1 > \theta_2 > \theta_3,$$
$$\theta_1 > \theta_3 > \theta_2,$$
$$\theta_2 > \theta_1 > \theta_3,$$
$$\theta_2 > \theta_3 > \theta_1,$$
$$\theta_3 > \theta_1 > \theta_2,$$
$$\theta_3 > \theta_2 > \theta_1.$$

Analogously, suppose we sample from a population for which $\theta_1 = \theta_2 = \theta_3$ and $\hat{\boldsymbol{\Sigma}}_{\hat{\theta}} = \boldsymbol{I}_3$, where $\boldsymbol{I}_3$ indicates a 3×3 identity matrix. Note that, when looking at the first level, θ_1 is bigger than both θ_2 and θ_3 as in hypothesis H_2 in two out of the six cases, and thus, $w_1 = 0.333$. When looking at the second level, θ_1 is bigger than θ_2 in three out of the six cases, and thus, $w_2 = 0.500$. When looking at the third level, θ_1 is bigger than θ_2 which in turn is bigger than θ_3 in only one out of the six cases, and thus, $w_3 = 0.167$. Thus, for hypothesis $H_2: \theta_1 > \theta_2 > \theta_3$, $PT_2 = w_1\times 1 + w_2\times 2 + w_3\times 3 = 1.834$.

Received May 29, 2019
Revision received March 9, 2021
Accepted March 19, 2021 ■