Journal of Statistical Software

August 2013, Volume 54, Issue 8. *http://www.jstatsoft.org/*

A Fortran 90 Program for the Generalized Order-Restricted Information Criterion

Rebecca M. Kuiper
Utrecht University

Herbert Hoijtink
Utrecht University

Abstract

The generalized order-restricted information criterion (GORIC) is a generalization of the Akaike information criterion such that it can evaluate hypotheses that take on specific, but widely applicable, forms (namely, closed convex cones) for multivariate normal linear models. It can examine the traditional hypotheses H_0: $\beta_{1,1} = \cdots = \beta_{t,k}$ and H_u: $\beta_{1,1}, \ldots, \beta_{t,k}$ and hypotheses containing simple order restrictions H_m: $\beta_{1,1} \geq \ldots \geq \beta_{t,k}$, where any "$\geq$" may be replaced by "$=$" and m is the model/hypothesis index; with $\beta_{h,j}$ the parameter for the h-th dependent variable and the j-th predictor in a t-variate regression model with k predictors (which might include the intercept). But, the GORIC can also be applied to restrictions of the form H_m: $R_1\beta = r_1,\ R_2\beta \geq r_2$, with β a vector of length tk, R_1 a $c_{m1} \times tk$ matrix, r_1 a vector of length c_{m1}, R_2 a $c_{m2} \times tk$ matrix, and r_2 a vector of length c_{m2}. It should be noted that $[R_1^\top, R_2^\top]^\top$ should be of full rank when $[r_1^\top, r_2^\top]^\top \neq 0$. In practice, this implies that one cannot examine range restrictions (e.g., $0 < \beta_{1,1} < 2$ or $\beta_{1,2} < \beta_{1,1} < 2\beta_{1,2}$) with the GORIC. A Fortran 90 program is presented, which enables researchers to compute the GORIC for hypotheses in the context of multivariate regression models. Additionally, an R package called **goric** is made by Daniel Gerhard and the first author.

Keywords: Fortran 90, inequality constraint, model selection, order restriction, R, regression model.

1. Introduction

Researchers often have hypotheses with respect to the relation among model parameters. Consider, for example, the simple regression model $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \epsilon$, where ϵ is normally distributed with mean 0 and variance σ^2. Hypotheses of interest could be H_1: $\beta_1 = \beta_2 = \beta_3$, H_2: $\beta_1 = \beta_2,\ \beta_3$, and H_3: $\beta_1,\ \beta_2 = \beta_3$. One can employ information criteria to select the best of a set of hypotheses. The Akaike information criterion (AIC; Akaike

1973) is one such criterion. However, a more flexible class of hypotheses can be evaluated if, in addition to equality constraints, order restrictions can be used in the formulation of hypotheses (e.g., H_1: $\beta_1 \geq \ldots \geq \beta_k$ and H_2: $\beta_1 = \ldots = \beta_{k'} \geq \ldots \geq \beta_k$ for $1 < k' < k$). The AIC is not suited for the evaluation of order-constrained hypotheses. In the context of analysis of variance (i.e., $y_{ij} = \beta_j + \epsilon_{ij}$, with $i = 1, \ldots, N_j$, $j = 1, \ldots, k$, β_j the mean for group j, and $\epsilon_{ij} \sim N(0, \sigma^2)$), the order-restricted information criterion (ORIC), proposed by Anraku (1999), can be used to select the best of a set of hypotheses that can be written as simple order restrictions (i.e., H_m: $\beta_1 \geq \ldots \geq \beta_k$, where any "$\geq$" may be replaced by "="). Kuiper, Hoijtink, and Silvapulle (2011) generalized the ORIC, called the GORIC, such that it can be applied to a more general form of order restrictions, namely H_m: $R\beta \geq 0$ for $m \in \mathcal{M}$, where $\mathcal{M}$ is the set of hypothesis indices, β a vector of length k, and R a $c_m \times k$ matrix. Special cases of these matrix order restrictions are the simple order (i.e., H_m: $\beta_1 \geq \ldots \geq \beta_k$) and the tree order (i.e., H_m: $\beta_1 \geq \beta_2, \ldots, \beta_1 \geq \beta_k$). Notably, simple order restrictions can be written as

$$H_m : \begin{bmatrix} 1 & -1 & 0 & 0 & \ldots & 0 & 0 \\ 0 & 1 & -1 & 0 & \ldots & 0 & 0 \\ \vdots & & & & \ldots & & \vdots \\ 0 & 0 & 0 & 0 & \ldots & 1 & -1 \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \\ \beta_4 \\ \vdots \\ \beta_{k-1} \\ \beta_k \end{bmatrix} \geq \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix},$$

which equals $\beta_1 - \beta_2 \geq 0$, $\beta_2 - \beta_3 \geq 0$, $\ldots$, $\beta_{k-1} - \beta_k \geq 0$ and thus H_m: $\beta_1 \geq \ldots \geq \beta_k$; and tree order restrictions can be written as

$$H_m : \begin{bmatrix} 1 & -1 & 0 & 0 & \ldots & 0 & 0 \\ 1 & 0 & -1 & 0 & \ldots & 0 & 0 \\ \vdots & \vdots & & & \ldots & & \\ 1 & 0 & 0 & 0 & \ldots & 0 & -1 \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \\ \beta_4 \\ \vdots \\ \beta_{k-1} \\ \beta_k \end{bmatrix} \geq \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix},$$

which equals H_m: $\beta_1 \geq \beta_2, \ldots, \beta_1 \geq \beta_k$. Kuiper, Hoijtink, and Silvapulle (2012) extended the use of the GORIC to univariate and multivariate normal linear models with not only hypotheses of the type H_m: $R\beta \geq 0$ (closed convex cone), but also H_m: $R\beta \geq r$ (relocated closed convex cone), where β is a vector of length tk containing the parameters in a t-variate normal linear model, with k the number of predictors (which can include an intercept), as elaborated below. The more general expression for these two types of hypotheses is H_m: $\beta \in \mathcal{C}_m$, where $\mathcal{C}_m$ is a closed convex cone or a relocated one. The hypotheses of interest and therewith the closed convex cones are further discussed in Section 2.2.

In the next section, the GORIC will be presented in the context of multivariate regression models. The GORIC comprises a likelihood part and a penalty part. The likelihood is computed using order-restricted maximum likelihood estimators (MLEs), that is, MLEs in agreement with the (order-restricted) hypothesis at hand. The iteration process employed to obtain the order-restricted MLEs is described in Section 3. In Section 4, we will elaborate on the penalty part. Section 5 illustrates the application of the GORIC in the context of univariate and multivariate analysis of variance. Subsequently, Section 6 discusses GORIC weights, which are easier to interpret than the GORIC values themselves. We end, in Appendix A, with a user manual for the software in Fortran 90. Running the Fortran 90 files result in a stand-alone program, namely an `.exe` file, which can also be found on `http://www.uu.nl/staff/RMKuiper`. In addition, an R package, called **goric**, is made available (Gerhard and Kuiper 2011). This package will not be discussed here, more details can be found in its reference manual.

2. The GORIC

In this section, we provide the GORIC applicable to a wide range of hypotheses (namely, those of the form H_m: $\beta \in \mathcal{C}_m$) formulated for a t-variate regression model. The derivation is shown in Kuiper *et al.* (2012). First, we briefly discuss the t-variate regression model. Then, we give the expression of the GORIC. Finally, we elaborate on the hypotheses that can be evaluated by it.

2.1. The t-variate regression model

A multivariate regression model with t dependent variables can be written as

$$\begin{aligned} y_{1i} &= \beta_{1,1}x_{1i} + \ldots + \beta_{1,k}x_{ki} + \epsilon_{1i} \\ &\vdots \\ y_{ti} &= \beta_{t,1}x_{1i} + \ldots + \beta_{t,k}x_{ki} + \epsilon_{ti} \end{aligned} \tag{1}$$

where y_{hi} denotes the score of the i-th person on the h-th dependent variable for $i = 1, \ldots, N$ and $h = 1, \ldots, t$. The x variables are predictors. They can be dummy variables representing group membership and/or continuous predictors, where x_{ji} then reflects the score of the i-th person on the j-th predictor for $j = 1, \ldots, k$. The relationship between x_{ji} and y_{hi} (controlled for the other predictors) is denoted by $\beta_{h,j}$. Furthermore, it is assumed that

$$\begin{bmatrix} \epsilon_{1i} \\ \vdots \\ \epsilon_{ti} \end{bmatrix} \sim \mathcal{N}_t \left(\begin{bmatrix} 0 \\ \vdots \\ 0 \end{bmatrix}, \Sigma = \begin{bmatrix} \sigma_1^2 & \cdots & \sigma_{1t} \\ \vdots & \ddots & \vdots \\ \sigma_{1t} & \cdots & \sigma_t^2 \end{bmatrix} \right).$$

It is noteworthy that the βs associated with x variables regarding the same dependent variable are only comparable when the corresponding x variables are standardized. Moreover, βs associated with x variables belonging to different dependent variables can solely be examined if both the dependent variables and the x variables are standardized.

2.2. The hypotheses of interest and (relocated) closed convex cones

Let $\beta = (\beta_{1,1}, \ldots, \beta_{1,k}, \ldots, \beta_{t,1}, \ldots, \beta_{t,k})$ and β_l the l-th element of β for $l = 1, \ldots, tk$. The GORIC can be applied to hypotheses that are closed convex cones or relocated ones; both denoted by $\mathcal{C}_m$. In this article, we will focus on

$$H_m: \quad R_1\beta = r_1, \; R_2\beta \geq r_2, \tag{2}$$

where R_1 is a $c_{m1} \times tk$ matrix, R_2 a $c_{m2} \times tk$ matrix, r_1 a vector of length c_{m1}, and r_2 a vector of length c_{m2}. Equation 2 is a closed convex cone, when $r_1 = r_2 = 0$. Special cases of closed convex cones are the simple order, the tree order, and the matrix order (Silvapulle and Sen 2005, p. 82). For Equation 2 to be a relocated closed convex cone, it should not only hold true that $[r_1^\top, r_2^\top]^\top \neq 0$ but also that $R = [R_1^\top, R_2^\top]^\top$ is of full rank (see Kuiper *et al.* (2012) and Section 4). Note that full rank of R may be obtained by discarding redundant restrictions. For example, a hypothesis containing $\beta_l \geq r_{21},\ \beta_l \leq r_{22}$ is not a relocated closed convex cone for $r_{21} \neq r_{22}$, since R is not of full rank and there are no redundant restrictions. When the hypothesis is $\beta_l \geq r_{21},\ \beta_{l'} \geq r_{22},\ \beta_l + \beta_{l'} \geq r_{23}$ with $l \neq l'$, R is not of full rank either. However, when $r_{21} + r_{22} \geq r_{23}$, the constraint $\beta_l + \beta_{l'} \geq r_{23}$ is redundant. In case this redundant restriction is discarded, R is of full rank, that is, H_m: $\beta_l \geq r_{21},\ \beta_{l'} \geq r_{22}$ is a relocated closed convex cone.

2.3. The GORIC

The GORIC is, like the AIC and the ORIC, based on the Kullback–Leibler (KL) discrepancy (Kullback and Leibler 1951). The KL discrepancy is ideally estimated by the maximum log-likelihood subject to the hypothesis at hand, that is, the log-likelihood evaluated at the MLEs which are in agreement with the (order) restrictions in Hypothesis H_m (referred to as order-restricted MLEs and denoted by $\tilde{\beta}^m$ and $\tilde{\Sigma}^m$). Since this is not a good estimator, a bias results which is adjusted for by a penalty part (denoted by PT_m). More precisely, the penalty is based on the expectation of the difference between the maximum log-likelihood subject to the hypothesis at hand and the expected log-likelihood at $(\tilde{\beta}^m, \tilde{\Sigma}^m)$; more details can be found in Kuiper *et al.* (2012). In case of the AIC, where solely equality restrictions (of the form $\beta_{hj} = \beta_{h'j'}$ for $h' = 1, \ldots, t$ and $j' = 1, \ldots, k$) are evaluated, the penalty equals the number of distinct parameters. When order restrictions are examined, the bias does not reduce to a constant, but to a term with a certain null distribution, namely the chi-bar-square distribution (Kuiper *et al.* 2011, 2012). This is a weighed chi-square distribution, where the weights are called chi-bar-square weights or level probabilities. A level probability (denoted by $w_l(tk, W, H_m)$ for level l) is the probability that there are l levels among the tk order-restricted MLEs of β, given that the parameters β are generated from its null distribution: a normal distribution with a mean vector of zeros and covariance matrix W (see also Anraku (1999); Silvapulle and Sen (2005, pp. 77–83); Robertson, Wright, and Dykstra (1988, p. 69)). Stated otherwise, it is the probability that the β parameter space in accordance with the active constraints in H_m is of dimension l. Notably, equality restrictions are always active constraints and each (non-redundant) one reduces the dimension by one. Hence, in case there are tk regression parameters, as in Equation 1, $\beta_{hj} = \beta_{h'j'}$ lowers the dimension of the order-restricted MLEs of β to $tk - 1$. Note that the same holds for equalities like $\beta_{hj} = 2\beta_{h'j'}$ even though the order-restricted MLEs of β do not have the same value. Thus, the penalty is not based on the number of distinct values, but on active/binding constraints. The level is the number of β parameters minus the number of active constraints in the hypothesis of interest. In case of order restrictions, there are certain probabilities that the restriction is binding/active, which is the case when an order restriction does not hold. For instance, in a univariate regression model with three regression parameters and H: $\beta_{11} \geq \beta_{12},\ \beta_{13}$, the parameter β_{11} will (under the null distribution) half of the time be greater than β_{12} (i.e., half of the time there are 3 levels); in the other half, β_{11} will be lower than β_{12} in which case the order-restricted MLEs of β will be set equal (in this example) such that they are in agreement with the hypothesis, that is, the constraint will be active (i.e., in the other half, there are 2 levels). This yields an expected β dimension (of the order-restricted MLEs of β)

of $0.5 \times 3 + 0.5 \times 2 = 2.5$, that is, there is a reduction of a half a parameter in the β parameter space. Next, we will give the expression for the GORIC.

Let

$$\begin{aligned}
Y &= \begin{bmatrix} y_{11} & \dots & y_{t1} \\ \vdots & & \vdots \\ y_{1N} & \dots & y_{tN} \end{bmatrix}, \\
y_i &= [y_{1i}, \ \dots, \ y_{ti}]^\top, \\
X &= \begin{bmatrix} x_{11} & \dots & x_{k1} \\ \vdots & & \vdots \\ x_{1n} & \dots & x_{kn} \end{bmatrix}, \\
x_i &= [x_{1i}, \ \dots, \ x_{ki}]^\top, \text{ and} \\
B &= \begin{bmatrix} \beta_{1,1} & \dots & \beta_{t,1} \\ \vdots & & \vdots \\ \beta_{1,k} & \dots & \beta_{t,k} \end{bmatrix}.
\end{aligned} \tag{3}$$

According to Kuiper *et al.* (2012), it holds true for t-variate regression models with H_m: $\beta \in \mathcal{C}_m$ that

$$\begin{aligned}
GORIC_m &= -2 \ \log f(Y|X, \tilde{B}^m, \tilde{\Sigma}^m) + 2 \ PT_m, \qquad (4) \\
\text{with} & \\
\log f(Y|X, \tilde{B}^m, \tilde{\Sigma}^m) &= -\frac{tN}{2} \ \log(2\pi) - \frac{N}{2} \log|\tilde{\Sigma}^m| - \frac{1}{2} \sum_{i=1}^{N} \epsilon_i^\top \left(\tilde{\Sigma}^m\right)^{-1} \epsilon_i, \\
\text{and} & \\
PT_m &= 1 + \sum_{l=1}^{tk} w_l(tk, W, H_m) \ l,
\end{aligned}$$

where $\log f(Y|X, \tilde{B}^m, \tilde{\Sigma}^m)$ is the log-likelihood, $\tilde{B}^m$ and $\tilde{\Sigma}^m$ are the order-restricted MLEs of B and Σ, respectively, PT_m is the penalty part, $w_l(tk, W, H_m)$ denotes the level probability for level l, and

$$\begin{aligned}
\epsilon_i &= y_i - \tilde{B}^{m\top} x_i, \\
W &= \hat{\Sigma} \otimes [X^\top X]^{-1}, \qquad (5) \\
\text{with} & \\
\hat{\Sigma} &= N^{-1}(Y - X\hat{B})^\top (Y - X\hat{B}) \qquad (6) \\
\text{and} & \\
\hat{B} &= (X^\top X)^{-1} X^\top Y,
\end{aligned}$$

where $\otimes$ denotes the Kronecker product. Hence, $\hat{\Sigma}$ and $\hat{B}$ are the (unrestricted) MLEs of Σ and B, respectively. The derivation of the penalty can be found in Kuiper *et al.* (2012). In that, Σ is assumed to be known up to a positive constant, that is, $\Sigma = \sigma^2 S$ with S a known $t \times t$ matrix and σ^2 an unknown constant which represents the variance when $t = 1$. Since often there exists no known S, Σ is estimated by $\hat{\Sigma}$, see Equation 6.

The GORIC is easily applied, namely the hypothesis/model (see H_m in Equation 2) with the lowest GORIC value (see Equation 4) is the preferred one.

In the next two sections, we will subsequently elaborate upon the order-restricted MLEs $\tilde{B}^m$ and $\tilde{\Sigma}^m$ and the penalty term PT_m.

3. The order-restricted maximum likelihood estimators

The order-restricted MLEs, $\tilde{B}^m$ and $\tilde{\Sigma}^m$, are obtained by

$$\arg \min_{\beta \in H_m,\ \Sigma} \sum_{i=1}^{N} (y_i - \tilde{B}^{m\top} x_i)^\top \Sigma^{-1} (y_i - \tilde{B}^{m\top} x_i).$$

From this, it follows that

$$\tilde{B}^m = \arg \min_{\beta \in H_m} \sum_{i=1}^{N} (y_i - B^\top x_i)^\top \left(\tilde{\Sigma}^m\right)^{-1} (y_i - B^\top x_i), \tag{7}$$

$$\tilde{\Sigma}^m = N^{-1} (Y - X\tilde{B}^m)^\top (Y - X\tilde{B}^m). \tag{8}$$

It should be stressed that, in univariate regression (i.e., for $t = 1$), $\tilde{B}^m$ does not depend on $\tilde{\Sigma}^m = \tilde{\sigma}^2_m$. In multivariate regression (i.e., $t > 1$), on the other hand, $\tilde{B}^m$ does depend on the unknown $\tilde{\Sigma}^m$ and, in addition, $\tilde{\Sigma}^m$ depends on the unknown $\tilde{B}^m$. Therefore, iterations are required to calculate them. The iteration process comprises the following steps, where $\tilde{B}^m_p$ and $\tilde{\Sigma}^m_p$ are the values of $\tilde{B}^m$ and $\tilde{\Sigma}^m$, respectively, in Iteration p:

1. Set $\tilde{B}^m_0$ equal to $\hat{B} = (X^\top X)^{-1} X^\top Y$, the (unrestricted) MLE of B.
   Note that any value for $\tilde{B}^m_0$ can be chosen. We employ $\hat{B}$ to increase the speed of convergence and, therefore, to reduce computing time.

2. Optimize $\tilde{\Sigma}^m_p$ by substituting $\tilde{B}^m_{p-1}$ for $\tilde{B}^m$ in Equation 8 for Iteration $p = 1, \ldots, P$.

3. Optimize $\tilde{B}^m_p$ by replacing $\tilde{\Sigma}^m$ with $\tilde{\Sigma}^m_p$ in Equation 7 for $p = 1, \ldots, P$.
   For the calculation of $\tilde{B}^m_p$, one can use a quadratic programming algorithm like the **IMSL** subroutine `QPROG` (Visual Numerics 2003, pp. 1307–1310) in Fortran 90.

4. Continue Steps 2 and 3 until convergence is reached (at Iteration P) and set $\tilde{B}^m$ and $\tilde{\Sigma}^m$ equal to $\tilde{B}^m_P$ and $\tilde{\Sigma}^m_P$, respectively.
   In the software, we base the convergence criterion on the values of the parameter estimates. Namely, we stop iterating when the absolute values of the elements of $\tilde{B}^m_p - \tilde{B}^m_{p-1}$ and $\tilde{\Sigma}^m_p - \tilde{\Sigma}^m_{p-1}$ are less than 10^{-10}.

4. The penalty part

In this section, we elaborate on the calculation of the penalty term. We first assume that Σ is known up to the positive constant σ^2: $\Sigma = \sigma^2 S$ with S a known matrix. In that case, $\hat{\Sigma}$

in Equation 5 is replaced by Σ. After that, we discuss the consequences of estimating Σ from the data by $\hat{\Sigma}$.

The calculation of the level probabilities can be done via simulation (Silvapulle and Sen 2005, pp. 78–81). Herein, we use the property that all closed convex cones (i.e., Equation 2 for $r_1 = r_2 = 0$) and relocated ones (i.e., Equation 2 for $r = [r_1^\top, r_2^\top]^\top \neq 0$ and $R = [R_1^\top, R_2^\top]^\top$ of full rank) can be written in the form H_m: $R_1\beta^* = 0,\ R_2\beta^* \geq 0$ (Kuiper *et al.* 2012), with $\beta^* = \beta$ when $r_1 = r_2 = 0$ and $\beta^* = \beta - q$ with $Rq = r$ when $r \neq 0$, respectively. Note that q only exist when R is of full rank (after discarding redundant restrictions). The simulation consists of 5 steps:

1. Generate z (of length tk) from $\mathcal{N}_{tk}(\beta^0, W)$, with $\beta^0 = 0$ and $W = \sigma^2 S \otimes [X^\top X]^{-1}$, where S is a known matrix.
   Silvapulle and Sen (2005, p. 86) and Robertson *et al.* (1988, p. 69) prove that the calculation of the level probabilities does not depend on the mean value β^0 for closed convex cones. Furthermore, Robertson *et al.* (1988, p. 69) demonstrate for closed convex cones that the calculation of the level probabilities are invariant for positive constants like σ^2 and N. However, there is one exception, which is discussed below.

2. Compute $\tilde{z}_m$ via $\tilde{z}_m = \arg\ \min_{\beta^* \in \{\beta^* \in \mathbb{R}^{tk}:\ R_1\beta^*=0,\ R_2\beta^* \geq 0\}} (z - \beta^*)^\top W^{-1}(z - \beta^*)$, such that $\tilde{z}_m$ is in accordance with H_m: $R_1\beta^* = 0,\ R_2\beta^* \geq 0$, the hypothesis of interest.
   To implement this in software, one requires a quadratic programming algorithm, where one minimizes $1/2\ \tilde{z}_m^\top H \tilde{z}_m +\ c^\top \tilde{z}_m$ with respect to $\tilde{z}_m$, with $H = 2\ W^{-1}$ and $c^\top = -2\ z^\top W^{-1}$. For example, one can use the **IMSL** subroutine `QPROG` (Visual Numerics 2003, pp. 1307–1310) in Fortran 90. Since $H = 2W^{-1}$ is positive definite, the objective is a convex function and the problem has a feasible solution which is a unique global minimizer.

3. Determine the number of levels in $\tilde{z}_m$ and denote this by L_m.
   Let restriction a be denoted by $R_{2a}\beta^* \geq 0$ for $a = 1, \ldots, c_{m1}$, $A = \{a\text{: } R_{2a}\tilde{z}_m = 0\}$, that is, the set of restriction indices for which the restriction is binding, and $\phi = \{\beta\text{: } R_1\beta^* = 0,\ R_{2a}\beta^* = 0\ \forall\ a \in A\}$. Then, L_m is the dimension of ϕ.

4. Repeat the previous steps T (e.g., $T = 100,000$) times.
   To examine the stability of the penalty term, one could calculate it a second time with another seed value. If the two penalties are dissimilar, one should increase the value of T.

5. Estimate the level probability $w_l(tk, W, H_m)$ by the proportion of times L_m is equal to l $(l = 1, \ldots, tk)$ in the T simulations.

As discussed in the first simulation step, the level probabilities are invariant for the mean value β^0 and the variance term σ^2. This holds almost always true for closed convex cones (i.e., H_m: $R_1\beta = 0,\ R_2\beta \geq 0$) and relocated ones (i.e., H_m: $R_1\beta = r_1,\ R_2\beta \geq r_2$ where $r = [r_1^\top, r_2^\top]^\top \neq 0$ and $R = [R_1^\top, R_2^\top]^\top$ is of full rank after discarding redundant restrictions). There is one exception, namely for restrictions of the type $\beta_l \geq r_{21}$ (including $r_{21} = 0$) for $l = 1, \ldots, tk$. When the hypothesis of interest contains this type of restriction, one must use $\beta^0 = 0$. This results in level probabilities that are invariant for the value of σ^2.

Notably, the level probabilities for H_m: $\beta_l \geq r_{21}$ are the same as for H_m: $\beta_l \geq 0$, that is, here is no difference in complexity for these two hypothesis, since $\beta_l \geq r_{21}$ equals $\beta_l^* = \beta_l - r_{21} \geq 0$.

When sampling z from $\mathcal{N}_1(0, W)$ with W a scalar, half the time H_m: $z \geq 0$ is valid and $\tilde{z}_m$ has one level, and half the time H_m: $z \geq 0$ will be invalid and $\tilde{z}_m$ has zero levels. As a consequent, the expected dimension of the order-restricted MLE of β_l for H_m: $\beta_l \geq r_{21}$ is a half.

The penalty term

$$PT_m \quad = \quad 1 + \sum_{l=1}^{tk} w_l(tk, W, H_m)\ l$$

can be seen as the expected dimension of the parameters in accordance with H_m. That is, it reflects the expected dimension of the order-restricted MLE of β values plus 1 because of the unknown variance term σ^2 in $\Sigma = \sigma^2 S$ with S a known matrix.

Until now, we have assumed in the calculation of the level probabilities that Σ is known up to the constant σ^2. Often Σ is unknown, in that case one should estimate it to determine the level probabilities. However, when $t = 1$, no estimation of $\Sigma = \sigma^2$ is required, since the level probabilities are invariant of positive constants like σ^2 (see Step 1). In contrast, Σ needs to be estimated for $t > 1$. One can estimate Σ by $\hat{\Sigma}$, see Equation 6; as is done in the software.

If Σ is estimated from the data, the dimension of Σ, which is the number of unknown distinct elements of Σ, is $(t+1)t/2$ instead of 1. Since the restrictions are always on the β parameters and never on the elements of Σ, the number of unknown distinct elements (of the order-restricted MLE of Σ) is equal for all hypotheses of interest (H_m). So, although the penalty should then (perhaps) be corrected, the correction is equal for all H_m for $m \in \mathcal{M}$. Therefore, it has no impact on the model selection process.

In the next section, we will demonstrate evaluating hypotheses with the GORIC for different types of models.

5. The GORIC illustrated

5.1. Analysis of variance (ANOVA)

In this section, we will illustrate the GORIC supported by real data for which the descriptive statistics are available in Lievens and Sanchez (2007). They investigated the effect of training on the quality of ratings made by consultants. One variable of interest is the signal detection accuracy index, which "refers to the extent to which individuals were accurate in discerning essential from nonessential competencies for a given job" and is measured by "standardized proportion of hits – standardized proportion of false alarms" (Lievens and Sanchez 2007, p. 817). Three groups of consultants are distinguished: (1) expert, (2) training, and (3) control. There are 21 raters in the expert group, 25 in the training group, and 26 in the control group. Hence, the ANOVA model can be written as Equation 1 with $t = 1$, $k = 3$, and $N = \sum_{j=1}^{k} n_j = 21 + 25 + 26 = 72$, where x_1, x_2, and x_3 denote group membership variables. Since $t = 1$, we will drop the first subscript for ease of notation and use, for example, β_j instead of $\beta_{1,j}$. Note that for $t = 1$ no iteration is required between $\tilde{B}^m$ and $\tilde{\Sigma}^m$ (see Section 3), and that Σ does not need to be estimated to calculate the level probabilities (see Section 4).

The authors expected that accuracy of competency ratings would be higher among experts and trained raters than among raters in the control group (i.e., $\beta_1 \geq \beta_3$ and $\beta_2 \geq \beta_3$) and

m	$\tilde{\beta}_1^m$	$\tilde{\beta}_2^m$	$\tilde{\beta}_3^m$	$\log f(Y\|X, \tilde{B}^m, \tilde{\Sigma}^m)$	PT_m	$GORIC_m$
1	0.79	0.64	0.29	−24.85	2.84	*55.38*
2	0.79	0.64	0.29	−24.85	2.90	55.50
u	0.79	0.64	0.29	−24.85	4.00	57.70

Table 1: The GORIC values for the three specified hypotheses in the ANOVA example (with lowest value emphasized).

furthermore, that it would be highest among raters who already had competency modeling experience (i.e., $\beta_1 \geq \beta_2$). These expectations can be represented by the hypothesis H_1: $\beta_1 \geq \beta_2 \geq \beta_3$. Another theory could be that the accuracy of the training group is at least twice as high as the one in the control group and that of the export group is higher than that of the training group. This leads to H_2: $\beta_1 \geq \beta_2 \geq 2\ \beta_3$. Since both can be bad/weak hypotheses, it is informative to evaluate the unconstrained hypothesis (H_u) as well, in which there are no restrictions on the parameters. Namely, its inclusion ensures that no weak hypothesis is selected, since H_u will be preferred if the other two hypotheses are weak / do not fit the data. The set of hypotheses, therefore, consists of

$$
\begin{aligned}
H_1: &\quad \beta_1 \geq \beta_2 \geq \beta_3,\\
H_2: &\quad \beta_1 \geq \beta_2 \geq 2\ \beta_3,\\
H_u: &\quad \beta_1,\ \beta_2,\ \beta_3.
\end{aligned}
$$

Table 1 displays the order-restricted means $\tilde{\beta}_j^m$ (Equation 7), the log-likelihood values log $f(Y|X, \tilde{B}^m, \tilde{\Sigma}^m)$, the penalty terms PT_m, and the GORIC values (Equation 4) for the three hypotheses of interest. Since the sample means are in accordance with the restrictions in all the three hypotheses, the order-restricted means equal the sample means for each of these hypotheses. Therefore, the three hypotheses render the same log-likelihood and the distinction between the three is based on the penalty, that is, the complexity of the hypotheses (i.e., the expected dimension of the order-restricted MLEs). Since H_1 is less complex than H_2 and H_u (i.e, $PT_1 < PT_2$ and $PT_1 < PT_u$), H_1 is the preferred hypothesis. As a result, the first theory is preferred over the second and it is not a weak theory.

5.2. Multivariate analysis of variance (MANOVA)

In this section, we will illustrate the GORIC supported by real data which are available on page 10 of Silvapulle and Sen (2005) and in a report prepared by Litton Bionetics Inc in 1984. These data were obtained from an experiment to find out whether vinylidene fluoride gives rise to liver damage. Since increased levels of serum enzyme are inherent in liver damage, the focus was on whether enzyme levels are affected by vinylidene fluoride.

Hence, the variable of interest is the serum enzyme level. Three types of enzymes are inspected, namely SDH, SGOT, and SGPT. To study whether vinylidene fluoride has an influence on the three serum enzymes, four dosages of this substance were examined. In each of these four treatment groups, ten male Fischer-344 rats received the substance. The MANOVA model can be written as Equation 1 with $t = 3$, $k = 4$, and $N = 10$. Hence, $(y_{1i}, y_{2i}, y_{3i})^\top$ denotes the observations on the three enzymes for rat i, x_1 to x_4 are the group membership variables, and $\beta_{h,j}$ denote the mean response for dose j and dependent variable h.

	SDH				SGOT				SGPT			
m	$\tilde{\beta}^m_{1,1}$	$\tilde{\beta}^m_{1,2}$	$\tilde{\beta}^m_{1,3}$	$\tilde{\beta}^m_{1,4}$	$\tilde{\beta}^m_{2,1}$	$\tilde{\beta}^m_{2,2}$	$\tilde{\beta}^m_{2,3}$	$\tilde{\beta}^m_{2,4}$	$\tilde{\beta}^m_{3,1}$	$\tilde{\beta}^m_{3,2}$	$\tilde{\beta}^m_{3,3}$	$\tilde{\beta}^m_{3,4}$
0	24.13	24.13	24.13	24.13	105.38	105.38	105.38	105.38	59.70	59.70	59.70	59.70
1	24.13	24.13	24.13	24.13	105.37	105.37	105.37	105.37	63.00	63.00	60.64	52.16
u	22.70	22.80	23.70	27.30	99.30	108.40	100.90	112.90	61.90	63.80	60.20	52.90

Table 2: The order-restricted means ($\tilde{\beta}^m_{h,j}$) for dependent variable h, predictor j, and Hypothesis H_m in the MANOVA example.

If vinylidene fluoride induces liver damage, we expect that each serum level increases with the dosage of the substance, see H_1 below. Another theory could be that there is no effect of dosage, see H_0 below. Since both can be bad/weak hypotheses, it is informative to evaluate the unconstrained hypothesis (H_u) in which there are no restrictions on the parameters. The set of hypotheses, therefore, comprises

$$\begin{aligned} H_0: &\quad \beta_{h,1} = \beta_{h,2} = \beta_{h,3} = \beta_{h,4} \text{ for all } h = 1,2,3, \\ H_1: &\quad \beta_{h,1} \geq \beta_{h,2} \geq \beta_{h,3} \geq \beta_{h,4} \text{ for all } h = 1,2,3, \\ H_u: &\quad \beta_{h,1},\ \beta_{h,2},\ \beta_{h,3},\ \beta_{h,4} \text{ for all } h = 1,2,3. \end{aligned}$$

Note that, in total, there are $tk = 12$ β parameters.

Since the covariance matrix Σ is unknown, it is estimated from the data by the MLE of Σ (Equation 6):

$$\hat{\Sigma} = \begin{bmatrix} 10.79750 & -0.85750 & -0.07000 \\ -0.85750 & 226.75750 & 21.00500 \\ -0.07000 & 21.00500 & 24.67500 \end{bmatrix}.$$

This estimate is used in determining the level probabilities (see Section 4).

Table 2 displays the order-restricted means $\tilde{\beta}^m_{h,j}$ (Equation 7). Furthermore, Table 3 presents the log-likelihood values ($\log f(Y|X, \tilde{B}^m, \tilde{\Sigma}^m)$), the penalty terms ($PT_m$), and the GORIC values (Equation 4) for the three hypotheses of interest. The penalty values for both H_0 and H_1 are low(er), whereas the fit of H_u is high(er). The support in the data for H_u is that much higher that it renders the lowest GORIC value. Therefore, it is concluded that H_u is the preferred hypothesis. Notably, although H_1 is preferred over H_0, H_1 is a weak theory, since it is not preferred over the unconstrained hypothesis H_u.

m	$\log f(Y\|X, \tilde{B}^m, \tilde{\Sigma}^m)$	PT_m	$GORIC_m$
0	−406.54	4.00	821.09
1	−396.85	7.48	808.66
u	−388.80	13.00	*803.61*

Table 3: The GORIC values for the three specified hypotheses in the MANOVA example (with lowest value emphasized).

6. The GORIC weights

As can be seen from the two examples, the relevant information is not contained in the GORIC values themselves but in their differences. To improve the interpretation, we introduce GORIC weights (w_m), comparable to the Akaike weights (Burnham and Anderson 2002, p. 75–79, 302–305, 438–439), with

$$w_m = \frac{\exp\{-1/2(GORIC_m - GORIC_{min})\}}{\sum_{m' \in \mathcal{M}} \exp\{-1/2(GORIC_{m'} - GORIC_{min})\}}, \tag{9}$$

where $\mathcal{M}$ denotes the set of hypothesis indices and $GORIC_{min}$ the lowest GORIC value, that is, the GORIC value of the preferred model. The GORIC weights are numbers on a scale from 0 to 1 that sum to 1 over the set of hypotheses under investigation. These numbers can be interpreted as the relative weight of evidence in the data of each hypothesis.

For the two examples, the GORIC weights are given in Table 4. From these weights, one can also determine the relative evidence for Hypothesis m compared to m'. For instance, in the example of Lievens and Sanchez (2007), H_1 is $0.44/0.14 \approx 3.18$ more likely than H_u. Therefore, it is not a weak hypothesis. On the other hand, H_1 and H_2 receive (about) the same amount of support (and are not weak), namely $0.44/0.42 \approx 1.06$. Thus, although H_1 is the preferred hypothesis in the set (and not weakly supported by the data), there is no compelling evidence, since H_2 receives more or less the same support. Hence, both H_1 and H_2 can be preferred in this set (we will elaborate on this below). In contrast, there is eminent support for one hypothesis in the example of Silvapulle and Sen (2005). Namely, H_u is preferred and it has $0.93/0.07 \approx 12.52$ times more support than H_1.

It should be stressed that, in the first example, the differences in GORIC values for H_1, H_2, and H_u equal the differences in penalty term values, since the data are in accordance with all three hypotheses (rendering the same likelihood). Therefore, increasing the number of observations (in the same ratio) does not affect the relative evidences (assuming that the data are then still in agreement with all the hypotheses). Bear in mind that H_2: $\beta_1 \geq \beta_2 \geq 2\ \beta_3$ and H_1: $\beta_1 \geq \beta_2 \geq \beta_3$ strongly resemble each other. Thus, both hypotheses can be concluded to be the correct/best one (simultaneously), but one can also say that H_1 is because of the lower penalty for the model. One should perhaps take into account the maximum value of the relative evidence for two hypotheses, when the data are in accordance with these two or when their likelihood values are the same, which is likely to occur when the two hypotheses resemble each other. Therefore, more research might be required regarding the performance of the GORIC weights. Notably, in another data set, it is possible that H_1 is (more) supported by the data whereas H_2 is not (or less), in that case the evidence for H_1 will be more compelling

Example	m	$GORIC_m$	w_m
Lievens and Sanchez (2007, see Section 5.1)	1	55.38	0.44
$n_1 = 21, n_2 = 25, n_3 = 26$	2	55.50	0.42
	u	57.70	0.14
Silvapulle and Sen (2005, see Section 5.2)	0	821.09	0.00
$n_1 = n_2 = n_3 = n_4 = 10$	1	808.66	0.07
	u	803.61	0.93

Table 4: The GORIC weights (w_m) for all the hypotheses (H_m) in the two examples.

(at least asymptotically). In general, as Kuiper *et al.* (2011) show, if the true β parameter lies in one model/hypothesis and not in the other, the correct model will be chosen by the GORIC with a probability going to one when the number of observations (N) goes to infinity. Based on Burnham and Anderson (2002, p. 75–79, 302–305, 438–439), we conclude that the GORIC weights in Equation 9 represent the weight of evidence for the corresponding hypothesis (H_m) to be the best of the set for the data at hand.

Acknowledgments

This research is funded by the Netherlands Organization for Scientific Research NWO-VICI-453-05-002.

References

Akaike H (1973). "Information Theory and an Extension of the Maximum Likelihood Principle." In B Petrov, F Cźaki (eds.), *Second International Symposium on Information Theory*, pp. 267–281. Akademiai Kiadó, Budapest, Hungary.

Anraku K (1999). "An Information Criterion for Parameters under a Simple Order Restriction." *Biometrika*, **86**, 141–152.

Burnham KP, Anderson DR (2002). *Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach*. 2nd edition. Springer-Verlag.

Gerhard D, Kuiper RM (2011). ***goric**: Generalized Order-Restricted Information Criterion*, R package version 0.0-7 edition. URL http://CRAN.R-project.org/package=goric.

Kuiper RM, Hoijtink H, Silvapulle MJ (2011). "An Akaike-Type Information Criterion for Model Selection Under Inequality Constraints." *Biometrika*, **98 (2)**, 495–501.

Kuiper RM, Hoijtink H, Silvapulle MJ (2012). "Generalization of the Order-Restricted Information Criterion for Multivariate Normal Linear Models." *Journal of Statistical Planning and Inference*, **142**, 2454–2463.

Kullback S, Leibler RA (1951). "On Information and Sufficiency." *The Annals of Mathematical Statistics*, **22**, 79–86.

Lievens F, Sanchez JI (2007). "Can Training Improve the Quality of Inferences Made by Raters in Competence Modeling? A Quasi-Experiment." *Journal of Applied Psychology*, **92**, 812–819.

Robertson T, Wright FT, Dykstra RL (1988). *Order Restricted Statistical Inference*. John Wiley & Sons.

Silvapulle MJ, Sen PK (2005). *Constrained Statistical Inference*. John Wiley & Sons.

Visual Numerics (2003). ***IMSL** Fortran Library User's Guide MATH/LIBRARY Volume 2 of 2: Mathematical Functions in Fortran*.

A. GORIC.exe user manual

This user manual will describe and illustrate the options available in `GORIC.exe` (published along with this article and also available at http://www.uu.nl/staff/RMKuiper/). It also includes a directory with the input and output files of the ANOVA and MANOVA example given in this article. This program is made in Fortran 90 using the Intel Visual Fortran Compiler 10.0 for Windows. This compiler uses **IMSL** 5.0.

`GORIC.exe` is free, however, when results obtained with this program are published, please refer to this article, Kuiper *et al.* (2011), and Kuiper *et al.* (2012).

A.1. GORIC.exe

In the software, we use a $N \times k$ matrix X, like in Equation 3, where the x_{ji} variables can be continuous predictors and grouping/dummy variables. The order of the (types of) predictors is not of importance. Note that a variable of group membership is obtained by filling in ones and zeros at the appropriate places in a predictor/vector. In addition, when there are no grouping variables, one should include an intercept by adding a vector of ones in X. Like explained in Section 2.2, the parameters are taken together, leading to a vector, β, of tk parameters with indices 1 to tk (or, when $k = 0$, to a vector, θ, of t variable / group means). The order of the parameters corresponds to the order of the k predictors and the order of the t dependent variables. Namely, the first k parameters belong to the first dependent variable, ..., and the last k parameters belong to the last one. Stated differently, $(\beta_1, \ldots, \beta_k, \ldots, \beta_{(t-1)k+1}, \ldots, \beta_{tk})$ corresponds to $(\beta_{1,1}, \ldots, \beta_{1,k}, \ldots, \beta_{t,1}, \ldots, \beta_{t,k})$. Bear in mind that $\beta_1, \beta_{k+1}, \ldots$, and $\beta_{(t-1)k+1}$ reflect the intercepts when the first column of X consists of ones.

As discussed in Step 4 in Section 3, we stop iterating when the absolute values of the elements of $\tilde{B}_p^m - \tilde{B}_{p-1}^m$ and $\tilde{\Sigma}_p^m - \tilde{\Sigma}_{p-1}^m$ are less than $C = 10^{-10}$. But, to increase computing time, C is lowered to $C = 10^{-9}$ after $50,000$ iterations and to $C = 10^{-8}$ after $100,000$ iterations. When still no convergence is achieved after $200,000$ iterations, the program uses the current estimates $\tilde{B}_P^m$ and $\tilde{\Sigma}_P^m$ and displays these estimates together with $\tilde{B}_{P-1}^m$ and $\tilde{\Sigma}_{P-1}^m$ in the DOS box and the output file. The consequence of lowering C is that the procedure might not result in good approximations of $\tilde{B}^m$ and $\tilde{\Sigma}^m$. However, slow convergence only occurs when the hypothesis of interest does not fit the data.

A.2. Modification of the input files

Irrespectively of what analysis should be performed, two text files have to be modified (such that they apply to your data), namely `Input.txt` and `Data.txt`.

It should be noted that:

- The names of the text files are fixed and cannot be changed.
  These files have to be ANSI or ASCII files. When you open or write your input and/or data in **Notepad** (++), you should save it as a ANSI file (not a unicode or utf-8 file). In **Word**, you should save it as a `.txt` (ASCII) file.

- The format of these files should not be changed, that is, do not add empty lines and do not delete lines containing labels.

- The data in `Data.txt` should be complete, that is, missing data are not allowed. Furthermore, a dot (".") is used as decimal separator, not a comma (",").

`Data.txt`

The file `Data.txt` looks as follows (in the MANOVA example):

```
18 101 65 1 0 0 0
...
27  88 56 1 0 0 0
25 113 65 0 1 0 0
...
27  98 65 0 1 0 0
22  88 54 0 0 1 0
...
21 107 61 0 0 1 0
31 104 57 0 0 0 1
...
29  99 48 0 0 0 1
```

In the data file, a $N \times (t + k)$ matrix must be given. The t dependent variables must be given first, followed by the k predictors. In this example, the predictors only consist of group membership variables. In case there are no group membership variables, a vector of ones should be included, which represents the intercept. This can be done by specifying it in the input (see below) or by adding a column of ones to your data file.

It should be stressed that a dot (".") should be used as decimal separator. When a comma (",") is used, only the number proceeding it is read (e.g., "1,9" is read as "1"). Furthermore, no text nor additional hard returns should be included in `Data.txt`.

`Input.txt`

The file `Input.txt` looks as follows (in the MANOVA example):

```
t k intercept N  Stand x Stand y
3 4 0         40 0       0

Seed T
123  100000

M
3

Number of Equality (c_e) and Order (c_o) Restrictions for Each Model
(resulting in M lines with 2 numbers)
9 0
0 9
0 0

R for Model 1
```

```
1 -1  0  0  0  0  0  0  0  0  0  0
...
0  0  0  0  0  0  0  0  0  0  1 -1
R for Model 2
1 -1  0  0  0  0  0  0  0  0  0  0
...
0  0  0  0  0  0  0  0  0  0  1 -1
R for Model 3
r for Model 1
0
...
0
r for Model 2
0
...
0
r for Model 3
```

t, k, and N: t is the number of dependent variable, k the number of predictors, and N the number of observations, see Section 2.1; for k see also the item below.

intercept: This should be a 1 if you want the software to incorporate the intercept and a 0 when you do not.
Suppose your input of k, the number of predictors, is k'. When you want the software to include a vector of ones to the set of predictors, the software will change k into $k'+1$. Consequently, the restrictions should be given for $tk = t(k'+1)$ β parameters as opposed to tk'. Note that the first β parameter (for every dependent variable) will represent the intercept.
When your data (represented by the $N \times k$ matrix X) includes a vector of ones, your input for the number of predictors (k) should already include the intercept (see Section 2.1). In that case, "intercept" must be set to 0, otherwise the program will fail to continue.

Stand x and Stand y: If you set "Stand x" to 1, the predictors (X) will be standardized. The analogue holds true for "Stand y".
Notably, the β parameters regarding the same dependent variable are only comparable when the x variables are standardized. Additionally, the β parameters belonging to different dependent variables can solely be examined if both the dependent variables and the corresponding x variables (if any) are standardized.

Seed and T: The seed value is represented by "Seed" and the number of iterations required for computing the penalty part of the GORIC by T. These are discussed in Simulation step 4 in Section 4.

M, c_e, and c_o: M denotes the number of models/hypotheses and c_e and c_o the number of equality (c_1) and order restrictions (c_2), respectively; see Section 2.2.

R and r: R is the restriction matrix and equals $[R_1^\top, R_2^\top]^\top$ and r the right hand side and equals $[r_1^\top, r_2^\top]^\top$. Notably, the models are of the form H_m: $R_1\beta = r_1,\ R_2\beta \geq r_2$, see

Section 2.2. Furthermore, R should be of full rank when $r \neq 0$ and this is (partly) tested in the software, see Section A.3.
It should be stressed that the order of the restrictions are of importance: the $c\_e = c_1$ equality restrictions must be given first and the $c\_o = c_2$ order restrictions second.
One must give a restriction matrix ($R = [R_1^\top, R_2^\top]^\top$) and a right hand side ($r = [r_1^\top, r_2^\top]^\top$) for each model. Hence, you need to fill in M restriction matrices with each a heading and then M right hand side vectors with each a heading. Note that there is only a heading when there are no restrictions, that is, in case of the unconstrained model. Bear in mind that the ordering of the columns in the restriction matrix depend on the ordering of the β parameters. In the software, the first k parameters belong to the first dependent variable ($h = 1$), ..., and the last k to the last dependent variable ($h = t$). Hence, in the example, β_1 corresponds to $\beta_{1,1}$, β_2 to $\beta_{1,2}, \ldots, \beta_5$ to $\beta_{2,1}$, ..., and β_{12} to $\beta_{3,4}$.

As in `Data.txt`, no text nor additional hard returns should be added to `Input.txt`, except for (after) headings for supplementary models.

A.3. Error messages

In the program `GORIC.exe`, error messages are incorporated to detect wrongly stated input. However, it is possible to make a mistake that we have not foreseen. In that case, check the input and compare it to the data. If you cannot solve the problem, send the input and data file to `r.m.kuiper@uu.nl`.

The requirement that $R = [R_1^\top, R_2^\top]^\top$ should be of full rank when $r = [r_1^\top, r_2^\top]^\top \neq 0$ (see Kuiper *et al.* (2012) and Section 4) is investigated in the software. However, note that R is not examined on redundant restrictions. A warning appears when R is not of full rank when $r \neq 0$ and the user is asked to investigate whether the additional restrictions are redundant. By pressing the enter button, the program proceeds. It should be stressed that the program stops without a warning in case of conflicting restrictions (e.g., H_m: $\beta_l \leq -r_{21}$, $\beta_l \geq r_{21}$ for $r_{21} > 0$). Moreover, the GORIC is calculated in presence of non-redundant restrictions, like range restrictions (e.g., H_m: $\beta_l \geq -r_{21}$, $\beta_l \leq r_{21}$ for $r_{21} > 0$), which is not a (relocated) closed convex cone. In that case, the GORIC should be interpret with care for two reasons. First, the GORIC is not (yet) defined for these types of restrictions. Second, the level probabilities are now no longer invariant for β^0 and σ^2. In the software, we use $\beta^0 = 0$. As a consequence, H_m: $\beta_l = 0$ is examined in determining the penalty.

A.4. Save and close

When you have modified `Input.txt` and `Data.txt` (such that it applies to your data), you should save and close them.

A.5. Run `GORIC.exe`

When `GORIC.exe` is completed, the output file `Output.txt` will be created in the folder you are working in.

`Output.txt`

The output is given in `Output.txt` and will look as follows (in case of the MANOVA example):

```
This program is free.  However, when results obtained with this program are
published, please refer to:

Rebecca M. Kuiper, Herbert Hoijtink, and Mervyn J. Silvapulle (2011).
An Akaike-type Information Criterion for Model Selection under Inequality
Constraints.  Biometrika, 98 (2), 495-501.

Rebecca M. Kuiper, Herbert Hoijtink, and Mervyn J. Silvapulle (2012).
Generalization of the Order-Restricted Information Criterion for
Multivariate Normal Linear Models.
Journal of Statistical Planning and Inference, 142, 2454-2463

Rebecca M. Kuiper and Herbert Hoijtink (2013).
A Fortran 90 Program for the Generalization of the Order-Restricted
Information Criterion.
Journal of Statistical Software.

Notably, the latter is included in this software.


- - Summary of observed data - -


- Number of observations (N) -

N = 40


- Sigma estimated from the data -

h,   estimated Sigma
1    10.79750  -0.85750 -0.07000
2    -0.85750 226.75750 21.00500
3    -0.07000  21.00500 24.67500


- Order-restricted betas -

Note that the first 4 parameters belong to the first dependent variable,
..., and the last 4 to the last dependent variable.
```

```
Group number:      1     2     3     4      5      6      7      8     9
10     11     12
Sample betas:  22.70 22.80 23.70 27.30  99.30 108.40 100.90 112.90 61.90
63.80 60.20 52.90

Hypothesis 1   24.13 24.13 24.13 24.13 105.38 105.38 105.38 105.38 59.70
59.70 59.70 59.70
Hypothesis 2   24.13 24.13 24.13 24.13 105.37 105.37 105.37 105.37 63.00
63.00 60.64 52.16
Hypothesis 3   22.70 22.80 23.70 27.30  99.30 108.40 100.90 112.90 61.90
63.80 60.20 52.90



- - GORIC - -

m  log likelihood  penalty   GORIC* GORIC weight** rel.evidence pref.hyp.***
1         -406.54     4.00  821.09          0.00                  6254.99
2         -396.85     7.48  808.66          0.07                    12.52
3         -388.80    13.00  803.61          0.93                     1.00


According to the Generalized Order-Restricted Information Criterion,
out of the set of hypotheses the preferred one is number 3,
which is the unconstrained model, that is, the model without restrictions on
the parameters.


*   The value of the Generalized Order-Restricted Information Criterion
(GORIC) = -2 * log likelihood + 2 * penalty.
**  The GORIC weight is the relative likelihood / the weight of evidence of
Hypothesis m given the data and the set of hypotheses.
*** The relative evidence for the preferred hypothesis compared to
Hypothesis m reflects how many times the preferred hypothesis is more
likely than Hypothesis m.  Thus, it gives insight into the strength of the
preferred hypothesis.
```

Number of observations (N): See Section 2.1.

Sigma estimated from the data: In the software, Σ is estimated by $\hat{\Sigma}$ (Equation 6), the MLE of Σ. Bear in mind that Σ is only estimated when $t > 1$.
For more details see Section 4.

Order-restricted betas: The order-restricted βs ($\tilde{\beta}^m_{h,j}$) can be found in Equation 7, see also Section A.1. Note that the subscripts are 1 to 12 in the software, where $\tilde{\beta}^m_1$ corresponds to $\tilde{\beta}^m_{1,1}$, $\tilde{\beta}^m_2$ to $\tilde{\beta}^m_{1,2}, \ldots, \tilde{\beta}^m_5$ to $\tilde{\beta}^m_{2,1}, \ldots,$ and $\tilde{\beta}^m_{12}$ to $\tilde{\beta}^m_{3,4}$.

GORIC: The expression of the GORIC is displayed in Equation 4.
The model/hypothesis with the lowest GORIC value is the preferred one: Hypothesis "number 3", that is, H_u: $\beta_1, \beta_2, \beta_3, \beta_4, \beta_5, \beta_6, \beta_7, \beta_8, \beta_9, \beta_{10}, \beta_{11}, \beta_{12}$.

GORIC weight: The expression of the GORIC weight is displayed in Equation 9.

Relative evidence preferred hypothesis: The relative evidence for the preferred hypothesis compared to Hypothesis m gives an intuition about the strength of the hypothesis. For more details see Section 6.

Affiliation:

Rebecca M. Kuiper and Herbert Hoijtink
Department of Methodology and Statistics
Utrecht University
P.O. Box 80.140
3508 TC Utrecht, The Netherlands
E-mail: R.M.Kuiper@uu.nl
URL: http://www.uu.nl/staff/RMKuiper/

Journal of Statistical Software http://www.jstatsoft.org/
published by the American Statistical Association http://www.amstat.org/
Volume 54, Issue 8 *Submitted:* 2011-05-18
August 2013 *Accepted:* 2013-02-20