Generalization of the order-restricted information criterion for multivariate normal linear models

R. M. Kuiper^{a,1,*}, H. Hoijtink^a, M. J. Silvapulle^b

^a*Department of Methods and Statistics, Utrecht University, the Netherlands*

^b*Department of Econometrics and Business Statistics, Monash University, Australia*

Abstract

The generalized order-restricted information criterion (GORIC) is a model selection criterion which can, up to now, solely be applied to the analysis of variance models and, so far, only evaluate restrictions of the form $R\theta \leq 0$, where θ is a vector of k group means and R a $c_m \times k$ matrix. In this paper, we generalize the GORIC in two ways: (i) such that it can be applied to t -variate normal linear models and (ii) such that it can evaluate a more general form of order restrictions: $R\theta \leq r$, where θ is a vector of length tk , r a vector of length c_m , and R a $c_m \times tk$ matrix of full rank (when $r \neq 0$). At the end, we illustrate that the GORIC is easy to implement in a multivariate regression model.

Keywords: Inequality constrained inference, Model selection, Normal linear model, Order-restricted information criterion, Order restriction.

This is a preprint; when citing, please reference the published version:

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. DOI: <https://doi.org/10.1016/j.jspi.2012.03.007>

This preprint is shared under a CC-BY-NC-ND license; the published article is Copyright © 2012 Elsevier B.V. All rights reserved.

*Corresponding address: Department of Methods and Statistics, Utrecht University, P.O. Box 80.140, 3508 TC, Utrecht, the Netherlands. Tel.: +31 30 253 1571. E-mail address: R.M.Kuiper@uu.nl.

Email addresses: R.M.Kuiper@uu.nl (R. M. Kuiper), mervyn.silvapulle@monash.edu (M. J. Silvapulle)

URL: <http://staff.fss.uu.nl/RMKuiper> (R. M. Kuiper)

1. Introduction

Anraku (1999) proposes the order-restricted information criterion (ORIC), which is used for models of the form $y_{ij} = \theta_j + \epsilon_{ij}$, where y_{ij} is observation i (with $i = 1, \dots, N_j$) for group j (with $j = 1, \dots, k$), θ_j the mean of group j , and ϵ_{ij} the error term, which follows a normal distribution with mean 0 and variance σ^2 . This model selection criterion can only be applied to hypotheses which can be written as simple order restrictions: $H_m : \theta_1 \leq \dots \leq \theta_k$, where “ $\leq$ ” may be replaced by “ $=$ ”. Furthermore, Hughes & King (2003) derive the so-called one-sided AIC which is “applicable in problems where the signs of some or all the parameters are known or can be inferred on basis of a priori information”. In addition, Kuiper et al. (2011) propose a generalization of the order-restricted information criterion, GORIC. This model selection criterion can be used in analysis of variance (ANOVA) models for the evaluation of a more general form of order restrictions, namely closed convex cones: $R\theta \leq 0$, where θ is a vector of the k group means and R a $c_m \times k$ matrix.

In this paper, we generalize the GORIC of Kuiper et al. (2011) even further. First, its use is extended to t -variate normal linear models (with k predictors which might include a constant). Second, we prove that the GORIC can be applied to relocated closed convex cones: hypotheses of the type $H_m : R\theta \leq r \neq 0$ when R is of full rank, with θ a vector of length tk , R a $c_m \times k$ matrix, and r a vector of length c_m . It should be stressed that the matrix of full rank may be obtained by discarding redundant restrictions. Note that a hypothesis including the restrictions $\theta_l \geq r_{11}$ and $\theta_l \leq r_{12}$ is not a relocated closed convex cone for $r_{11} \neq r_{12}$, with θ_l the l th element of θ for $l = 1, \dots, tk$, since in that case R is not of full rank and there are no redundant restrictions. The same remains valid for $\theta_l \geq r_{11}, \theta_{l'} \geq r_{12}, \theta_l + \theta_{l'} \leq r_{13}$ for $l \neq l'$. For $H_m : \theta_l \leq r_{11}, \theta_{l'} \leq r_{12}, \theta_l + \theta_{l'} \leq r_{13}$, R is also not of full rank. However, when $r_{11} + r_{12} \geq r_{13}$, there is a redundant restriction, namely $\theta_l + \theta_{l'} \leq r_{13}$. When discarding it, R is of full rank. Hence, in that case $H_m : \theta_l \leq r_{11}, \theta_{l'} \leq r_{12}, \theta_l + \theta_{l'} \leq r_{13}$ equals $H_m : \theta_l \leq r_{11}, \theta_{l'} \leq r_{12}$ which is a relocated closed convex cone. Notably, equality restrictions can be a part of the hypothesis, that is, $H_m : R_1\theta \leq r_1, R_2\theta = r_2$ can also be evaluated, when $[R'_1, R'_2]'$ is of full rank, with R_1 a $c_{m1} \times tk$ matrix, R_2 a $c_{m2} \times tk$, r_1 a vector of

length c_{m1} , and r_2 of c_{m2} .

The derivation of the GORIC is done in two steps. In the next section, Section 2, we give the main part of the derivation of the GORIC for simple multivariate models of the form $y_i = \theta + \epsilon_i$, where y_i (for $i = 1, \dots, N$) is the vector of the k dependent variables for observation i (i.e., $y_i = [y_{i1}, \dots, y_{ik}]'$), θ the vector of the k group means, and ϵ_i the vector of the k error terms for observation i . This forms a good starting point for the derivation of the GORIC for t -variate normal linear models, which is discussed in Section 3. The demonstration that the GORIC can be applied to relocated closed convex cones is given in Section 4. Like in Kuiper et al. (2011), the GORIC incorporates a likelihood part and a penalty part. In Section 5, we elaborate on the penalty part of the GORIC. Note that in Kuiper et al. (2011), where the ANOVA model is inspected, the penalty depends on only one unknown positive scalar (i.e., the variance σ^2), that is, all the other terms are known. In this paper, where simple multivariate models and t -variate normal linear models are examined, we have to assume that the residual covariance matrix is equal to an unknown positive constant (σ^2) times a known scale matrix. We discuss briefly in Section 6 what consequences there are if this scale matrix is not known but estimated from the data. We end, in Section 7, with an illustration of the GORIC in a multivariate regression model.

2. Derivation of the GORIC

2.1. Preliminaries

The GORIC is based on the Kullback–Leibler discrepancy (Kullback & Leibler, 1951):

$$E_{g(y|\xi)} \{-2 \log f(y|\xi^m)\} = -2 \int_{-\infty}^{\infty} \log \{f(y|\xi^m)\} g(y|\xi) dy,$$

where $g(y|\xi)$ is the true generating model, with ξ the true parameter, and $f(y|\xi^m)$ a candidate model or hypothesis, that is, a statistical model to approximate $g(y|\xi)$, with ξ^m the parameter corresponding to Hypothesis H_m . The preferred hypothesis is the one which renders the lowest Kullback–Leibler discrepancy (parameterized by $\xi^m = \hat{\xi}^m$). Therefore, the GORIC, like the ORIC of Anraku (1999), is an estimate of the Kullback–Leibler discrepancy or rather of minus two times the expected log-likelihood.

Let the data be $y = [y_1, \dots, y_N] \in \mathbb{R}^{k \times N}$, with $y_i = [y_{i1}, \dots, y_{ik}]' \in \mathbb{R}^{k \times 1}$. It is assumed that the data y_i are normally and independently distributed with mean θ and covariance matrix V :

$$y_i \sim \mathcal{N}_k(\theta, V) \text{ for } i = 1, \dots, N, \quad (1)$$

where $\theta \in \mathbb{R}^{k \times 1}$ and V a $k \times k$ positive definite matrix. Models of this type are used in multivariate one-sided testing and repeated measures analysis (without between-subject factors).

The log-likelihood of y is written as

$$\begin{aligned} \log f(y|\theta, V) &= \sum_{i=1}^N \left[-\frac{k}{2} \log\{2\pi\} - \frac{1}{2} \log |V| - \frac{1}{2} (y_i - \theta)' V^{-1} (y_i - \theta) \right] \\ &= -\frac{Nk}{2} \log\{2\pi\} - \frac{N}{2} \log |V| - \frac{1}{2} \sum_{i=1}^N [(y_i - \theta)' V^{-1} (y_i - \theta)]. \end{aligned}$$

Based on the premise of normality, we postulate that the true density $g(y|\xi)$ is a normal distribution with mean θ and covariance matrix V :

$$g(y|\xi) = f(y|\theta, V).$$

Let the order-restricted maximum likelihood estimators of Hypothesis H_m be denoted by $\tilde{\theta}^m$ and $\tilde{V}^m$. They are obtained by

$$\min_{\theta \in H_m, V} \sum_{i=1}^N [(y_i - \theta)' V^{-1} (y_i - \theta)],$$

which leads to

$$\tilde{\theta}^m = \arg \min_{\theta \in H_m} \sum_{i=1}^N \left[(y_i - \theta)' (\tilde{V}^m)^{-1} (y_i - \theta) \right], \quad (2)$$

$$\tilde{V}^m = N^{-1} \sum_{i=1}^N \left[(y_i - \tilde{\theta}^m)(y_i - \tilde{\theta}^m)' \right]. \quad (3)$$

Since $\tilde{\theta}^m$ depends on $\tilde{V}^m$ and $\tilde{V}^m$ on $\tilde{\theta}^m$, iterations are required to calculate them. One could, for example, first set $\tilde{\theta}^m$ equal to the vector of groups means ($\bar{y}$). Based on these

values, one can iterate between (2) and (3) until convergence is reached. To calculate values of $\tilde{\theta}^m$, one could employ a quadratic program algorithm like the IMSL subroutine QPROG (Visual Numerics, 2003, pp. 1307–1310) in Fortran 90.

The statistical model to approximate $f(y|\theta, V)$ corresponding to Hypothesis H_m is

$$f(y|\xi^m) = f(y|\tilde{\theta}^m, \tilde{V}^m).$$

For fixed $\tilde{\theta}^m$ and $\tilde{V}^m$, the expected log-likelihood at $(\tilde{\theta}^m, \tilde{V}^m)$, where the expectation is taken with respect to $f(y|\theta, V)$, is

$$\begin{aligned} & E_{f(y|\theta, V)} \left\{ \log f(y|\tilde{\theta}^m, \tilde{V}^m) \right\} \\ &= \int_{-\infty}^{\infty} \log \left\{ f(y|\tilde{\theta}^m, \tilde{V}^m) \right\} f(y|\theta, V) dy \\ &= -\frac{Nk}{2} \log\{2\pi\} - \frac{N}{2} \log |\tilde{V}^m| - \frac{1}{2} E_{f(y|\theta, V)} \left\{ \sum_{i=1}^N \left[(y_i - \tilde{\theta}^m)' (\tilde{V}^m)^{-1} (y_i - \tilde{\theta}^m) \right] \right\}. \end{aligned} \quad (4)$$

The GORIC is minus two times an approximation of (4). We explain next that, for Hypothesis H_m , the GORIC is written as

$$\text{GORIC}_m = -2 \log f(y|\tilde{\theta}^m, \tilde{V}^m) + 2 PT_m.$$

Since (4) depends on the unknown θ and V , it is ideally estimated by $\log f(y|\tilde{\theta}^m, \tilde{V}^m)$. However, this is not a good estimator, hence a bias results. To adjust for this, the GORIC comprises a likelihood part and a penalty part (denoted by PT_m), where the latter is the infimum of the expectation of the bias. To derive an expression for the penalty term, we need the following definitions. Let

$$\begin{aligned} V &= \sigma^2 U, \\ \tilde{V}^m &= \tilde{\sigma}_m^2 U, \\ \tilde{\sigma}_m^2 &= (Nk)^{-1} \sum_{i=1}^N \left[(y_i - \tilde{\theta}^m)' U^{-1} (y_i - \tilde{\theta}^m) \right], \end{aligned} \quad (5)$$

where U is known. In Section 6, we will return to the issue of U being unknown.

The expectation of the bias between $\log f(y|\tilde{\theta}^m, \tilde{V}^m)$ and (4) with respect to $f(\tilde{\theta}^m, \tilde{V}^m|\theta, V)$ (which is for ease of notation denoted by E) is

$$\begin{aligned} B^m(\theta, V) &= E \left\{ \log f(y|\tilde{\theta}^m, \tilde{V}^m) - E_{f(y|\theta, V)} \left\{ \log f(y|\tilde{\theta}^m, \tilde{V}^m) \right\} \right\} \\ &= E \left\{ -\frac{1}{2} \sum_{i=1}^N \left[(y_i - \tilde{\theta}^m)' (\tilde{V}^m)^{-1} (y_i - \tilde{\theta}^m) \right] + \frac{1}{2} \left[Nk \frac{\sigma^2}{\tilde{\sigma}_m^2} + N(\tilde{\theta}^m - \theta)' (\tilde{V}^m)^{-1} (\tilde{\theta}^m - \theta) \right] \right\} \quad (6) \\ &= -\frac{Nk}{2} + \frac{Nk}{2} E \left\{ \frac{\sigma^2}{\tilde{\sigma}_m^2} \right\} + \frac{1}{2} E \left\{ N(\tilde{\theta}^m - \theta)' (\tilde{V}^m)^{-1} (\tilde{\theta}^m - \theta) \right\}, \quad (7) \end{aligned}$$

where to obtain the second part in (6) we used

$$\begin{aligned} E_{f(y|\theta, V)} \left\{ \sum_{i=1}^N \left[(y_i - \tilde{\theta}^m)' (\tilde{V}^m)^{-1} (y_i - \tilde{\theta}^m) \right] \right\} &= \\ E_{f(y|\theta, V)} \left\{ \sum_{i=1}^N \left[(y_i - \theta)' (\tilde{V}^m)^{-1} (y_i - \theta) \right] \right\} &+ N(\tilde{\theta}^m - \theta)' (\tilde{V}^m)^{-1} (\tilde{\theta}^m - \theta), \end{aligned}$$

and

$$\begin{aligned} E_{f(y|\theta, V)} \left\{ \sum_{i=1}^N \left[(y_i - \theta)' (\tilde{V}^m)^{-1} (y_i - \theta) \right] \right\} &= \text{trace} \left\{ (\tilde{V}^m)^{-1} V N \right\} \\ &= Nk \frac{\sigma^2}{\tilde{\sigma}_m^2}, \end{aligned}$$

and to obtain the first part in (7) we used

$$E \left\{ \sum_{i=1}^N \left[(y_i - \tilde{\theta}^m)' (\tilde{V}^m)^{-1} (y_i - \tilde{\theta}^m) \right] \right\} = Nk.$$

It holds true that $B^m(\theta, V) \geq B^m(\theta^0, V)$ for all $\theta^0 \in \mathcal{C}_0 = \{\theta \in \mathbb{R}^k | \theta_1 = \dots = \theta_k\}$ and all $\theta \in \mathcal{C}_m$ and that $B^m(\theta^0, V)$ has the same value for all $\theta^0 \in \mathcal{C}_0$ (Anraku, 1999; Robertson et al., 1988, pp. 101–102). Hence,

$$PT_m = \inf_{\theta \in \mathcal{C}_m} B^m(\theta, V) = \inf_{\theta \in \mathcal{C}_0} B^m(\theta, V) = B^m(\theta^0, V). \quad (8)$$

$B^m(\theta^0, V)$ is calculated by (7), where the expectation is now with respect to $f(\tilde{\theta}^m, \tilde{V}^m|\theta \in \mathcal{C}_0, V)$ (for brevity, denoted by E) and θ is replaced by θ_0 , which yields

$$B^m(\theta^0, V) = -\frac{Nk}{2} + \frac{Nk}{2} E \left\{ \frac{\sigma^2}{\tilde{\sigma}_m^2} \right\} + \frac{1}{2} E \left\{ N(\tilde{\theta}^m - \theta^0)' (\tilde{V}^m)^{-1} (\tilde{\theta}^m - \theta^0) \right\}. \quad (9)$$

In the sequel, we use, without loss of generalization, $H_0 : \theta = \theta^0$ in lieu of $H_0 : \theta \in \mathcal{C}_0 = \{\theta \in \mathbb{R}^k | \theta_1 = \dots = \theta_k\}$.

To obtain $B^m(\theta^0, V)$ in (9), one requires to determine

$$E \left\{ \frac{\sigma^2}{\tilde{\sigma}_m^2} \right\} \approx \frac{1}{E \left\{ \frac{\tilde{\sigma}_m^2}{\sigma^2} \right\}} + \frac{\text{var} \left(\frac{\tilde{\sigma}_m^2}{\sigma^2} \right)}{\left[E \left\{ \frac{\tilde{\sigma}_m^2}{\sigma^2} \right\} \right]^3} \quad (10)$$

and

$$E \left\{ N(\tilde{\theta}^m - \theta^0)' \left(\tilde{V}^m \right)^{-1} (\tilde{\theta}^m - \theta^0) \right\} = E \left\{ \frac{N(\tilde{\theta}^m - \theta^0)' U^{-1} (\tilde{\theta}^m - \theta^0)}{\sigma^2} \middle/ \frac{\tilde{\sigma}_m^2}{\sigma^2} \right\}, \quad (11)$$

where (10) is based on a second order Taylor expansion of $1/x$ around $E \{x\}$, with $x = \tilde{\sigma}_m^2/\sigma^2$. We first need to rewrite $\tilde{\sigma}_m^2/\sigma^2$ before we can obtain its null distribution and expectation. It can be shown, using Theorem 1 in Appendix A, that

$$\begin{aligned} \frac{\tilde{\sigma}_m^2}{\sigma^2} &= \frac{1}{Nk} \frac{\sum_{i=1}^N [(y_i - \tilde{\theta}^m)' U^{-1} (y_i - \tilde{\theta}^m)]}{\sigma^2} \\ &= \frac{1}{Nk} \left[\frac{\sum_{i=1}^N [(y_i - \theta^0)' U^{-1} (y_i - \theta^0)]}{\sigma^2} - \frac{N(\tilde{\theta}^m - \theta^0)' U^{-1} (\tilde{\theta}^m - \theta^0)}{\sigma^2} \right]. \end{aligned} \quad (12)$$

The first term in brackets in (12) has (assuming that H_0 is true) a chi-square distribution with Nk degrees of freedom (i.e., χ_{Nk}^2) and, therefore, has an expectation of $E \{\chi_{Nk}^2\} = Nk$. Consequently, we only require the null distribution and expectation of the second term in brackets in (12) to compute $E \{\tilde{\sigma}_m^2/\sigma^2\}$. The expression for $\text{var}(\tilde{\sigma}_m^2/\sigma^2)$ in (10) is written down in Appendix B. In addition, Appendix B demonstrates that (10) can be written as

$$E \left\{ \frac{\sigma^2}{\tilde{\sigma}_m^2} \right\} = 1 + \frac{2}{Nk} + \frac{1}{Nk} E \left\{ \frac{N(\tilde{\theta}^m - \theta^0)' U^{-1} (\tilde{\theta}^m - \theta^0)}{\sigma^2} \right\} + O((Nk)^{-2}). \quad (13)$$

Note that the part in the expectation equals the second term in brackets in (12). Moreover, the two terms in the second expectation in (11) are equal to the second term in brackets in (12) and to (12), respectively. As a consequence, to calculate (11) and (13), we need the null distribution of the second term in brackets in (12).

2.2. The null distribution and expectation of $N(\tilde{\theta}^m - \theta^0)'U^{-1}(\tilde{\theta}^m - \theta^0)/\sigma^2$

To obtain the null distribution of

$$\frac{N(\tilde{\theta}^m - \theta^0)'U^{-1}(\tilde{\theta}^m - \theta^0)}{\sigma^2} = (\tilde{\theta}^m - \theta^0)'[V/N]^{-1}(\tilde{\theta}^m - \theta^0),$$

we require the following expression:

$$\begin{aligned} \bar{\chi}^2(V/N, \mathcal{C}_m) &= \min_{\theta=\theta^0} \sum_{i=1}^N [(y_i - \theta)'V^{-1}(y_i - \theta)] - \min_{\theta \in \mathcal{C}_m} \sum_{i=1}^N [(y_i - \theta)'V^{-1}(y_i - \theta)] \\ &= \sum_{i=1}^N [(y_i - \theta^0)'V^{-1}(y_i - \theta^0)] - \sum_{i=1}^N [(y_i - \tilde{\theta}^m)'V^{-1}(y_i - \tilde{\theta}^m)] \\ &= N(\tilde{\theta}^m - \theta^0)'V^{-1}(\tilde{\theta}^m - \theta^0) \\ &= (\tilde{\theta}^m - \theta^0)'[V/N]^{-1}(\tilde{\theta}^m - \theta^0), \end{aligned} \quad (14)$$

where the one but last line is obtained by using Theorem 1 in Appendix A. According to Silvapulle & Sen (2005, pp. 75–77) and Robertson et al. (1988, pp. 70), the null distribution of (14) is given by

$$Pr(\bar{\chi}^2(V/N, \mathcal{C}_m) \leq x) = \sum_{j=0}^k w_j(k, V/N, \mathcal{C}_m) Pr(\chi_j^2 \leq x),$$

where $w_j(k, V/N, \mathcal{C}_m)$ is the level probability / chi-bar-square weight for Hypothesis H_m . A level probability $w_j(k, V/N, \mathcal{C}_m)$ is the probability that $\tilde{\theta}^m$ has j levels or, rather, the probability that the parameter space in accordance with the active constraints in $\mathcal{C}_m$ is of dimension j . An elaboration on the level probabilities is given in Section 5. Employing $E\{\chi_j^2\} = j$, gives

$$E\left\{\frac{N(\tilde{\theta}^m - \theta^0)'U^{-1}(\tilde{\theta}^m - \theta^0)}{\sigma^2}\right\} = \sum_{j=1}^k w_j(k, V/N, \mathcal{C}_m)j. \quad (15)$$

2.3. The GORIC

Equations (13) and (15) amount to

$$E\left\{\frac{\sigma^2}{\bar{\sigma}_m^2}\right\} = 1 + \frac{2}{Nk} + \frac{1}{Nk} \sum_{j=1}^k w_j(k, V/N, \mathcal{C}_m)j + O((Nk)^{-2}). \quad (16)$$

From (15) it follows that, if j level sets are given, the second expression in brackets in (12) and the term in brackets in (12) are conditionally independent and are distributed as chi-square distributions with j and $Nk - j$ degrees of freedom, respectively (see Anraku, 1999; Robertson et al., 1988, pp. 69–74). Therefore,

$$\frac{\frac{N(\tilde{\theta}^m - \theta^0)'U^{-1}(\tilde{\theta}^m - \theta^0)}{\sigma^2} / j}{Nk \frac{\tilde{\sigma}_m^2}{\sigma^2} / (Nk - j)} = \frac{Nk - j}{Nk j} \left[\frac{N(\tilde{\theta}^m - \theta^0)'U^{-1}(\tilde{\theta}^m - \theta^0)}{\tilde{\sigma}_m^2} \right] \quad (17)$$

has an F distribution with $(j, Nk - j)$ degrees of freedom for $\theta^0 \in \mathcal{C}_0$. Hence, when j level sets are given, the limit of j times (17) has a chi-square distribution with j degrees of freedom for $(Nk - j) \rightarrow \infty$. Consequently, (11) can be written as

$$E \left\{ \frac{N(\tilde{\theta}^m - \theta^0)'U^{-1}(\tilde{\theta}^m - \theta^0)}{\tilde{\sigma}_m^2} \right\} = \sum_{j=1}^k w_j(k, V/N, \mathcal{C}_m) j + O((Nk)^{-1}). \quad (18)$$

From (16) and (18), it follows that (9) can be rewritten as

$$B^m(\theta^0, V) = 1 + \sum_{j=1}^k w_j(k, V/N, \mathcal{C}_m) j + O((Nk)^{-1}).$$

Thus, the GORIC is calculated by

$$\begin{aligned} \text{GORIC}_m &= -2 \log f(y|\tilde{\theta}^m, \tilde{V}^m) + 2 PT_m, \text{ where} \\ PT_m &= 1 + \sum_{j=1}^k w_j(k, V/N, \mathcal{C}_m) j. \end{aligned} \quad (19)$$

Now we have obtained the GORIC for a special type of models and not yet for normal linear models in general. Examples of this special type of models are models for multivariate one-sided testing and repeated measures analysis without between-subject factors. In the next section, we will derive the GORIC that can be applied to univariate normal linear models and multivariate normal linear models.

3. The GORIC for t -variate normal linear models

The derivation of the GORIC in the previous sections elucidates the one that is feasible for t -variate normal linear models:

$$\left[\begin{array}{c} y_1 \\ \vdots \\ y_t \end{array} \right] \middle| X \sim \mathcal{N}_{tN} \left([I_t \otimes X] \begin{bmatrix} \beta_1 \\ \vdots \\ \beta_t \end{bmatrix}, V = \Sigma \otimes U \right), \quad (20)$$

where $y = \begin{bmatrix} y_1 \\ \vdots \\ y_t \end{bmatrix} \in \mathbb{R}^{tN \times 1}$, $I_t \otimes X = \text{diag}(X, \dots, X) \in \mathbb{R}^{tN \times tk}$, $\beta = \begin{bmatrix} \beta_1 \\ \vdots \\ \beta_t \end{bmatrix} \in \mathbb{R}^{tk \times 1}$, $V \in \mathbb{R}^{tN \times tN}$, $\Sigma \in \mathbb{R}^{t \times t}$, and $U \in \mathbb{R}^{N \times N}$, with $y_h \in \mathbb{R}^{N \times 1}$, $X \in \mathbb{R}^{N \times k}$, and $\beta_h \in \mathbb{R}^{k \times 1}$ for $h = 1, \dots, t$. When $t = 1$, hence for univariate normal linear models, V reduces to $\sigma^2 U$. In addition, for regression models it holds true that $V = \Sigma \otimes I_N$, with I_N the $N \times N$ identity matrix.

The order-restricted maximum likelihood estimators, $\tilde{\beta}^m$ and $\tilde{V}^m$, are obtained by

$$\arg \min_{\beta \in H_m, V} (y - [I_t \otimes X]\beta)' V^{-1} (y - [I_t \otimes X]\beta).$$

From this, it follows that

$$\tilde{\beta}^m = \arg \min_{\beta \in H_m} (y - [I_t \otimes X]\beta)' \left(\tilde{V}^m \right)^{-1} (y - [I_t \otimes X]\beta), \quad (21)$$

$$\tilde{V}^m = (y - [I_t \otimes X]\tilde{\beta}^m)(y - [I_t \otimes X]\tilde{\beta}^m)'. \quad (22)$$

Since $\tilde{\beta}^m$ depends on $\tilde{V}^m$ and $\tilde{V}^m$ on $\tilde{\beta}^m$, iterations are needed to calculate them. One could, for example, first set $\tilde{\beta}^m$ equal to $([I_t \otimes X]'[I_t \otimes X])^{-1}[I_t \otimes X]'y$. Based on these values one can iterate between (21) and (22) until convergence is reached. To calculate $\tilde{\beta}^m$ (in software), one could use a quadratic program algorithm like the IMSL subroutine QPROG (Visual Numerics, 2003, pp. 1307–1310) in Fortran 90. It should be stressed that if V is known (up to a positive constant), like in univariate regression models (where $V = \sigma^2 I_N$), no iterations are required. Namely, in univariate regression models, $\tilde{\beta}^m$ does not depend on V at all and $\tilde{V}^m = \tilde{\sigma}_m^2 I_N$. As opposed to univariate regression models, one does require to iterate

between $\tilde{\beta}^m$ and $\tilde{V}^m = \tilde{\Sigma}^m \otimes I_N$ in multivariate regression models, since $\tilde{\Sigma}^m$ is a unknown matrix; $\tilde{\Sigma}^m$ is calculated by (24) in Section 6 with $\hat{B}$ replaced by $\tilde{B}^m$, the $k \times t$ matrix where column h equals $\tilde{\beta}_h^m$.

As in Section 2, to derive an expression for the penalty term, let

$$\begin{aligned} V &= \sigma^2 S \otimes U, \\ \tilde{V}^m &= \tilde{\sigma}_m^2 S \otimes U, \\ \tilde{\sigma}_m^2 &= \left[(y - [I_t \otimes X] \tilde{\beta}^m)' [S^{-1} \otimes U^{-1}] (y - [I_t \otimes X] \tilde{\beta}^m) \right], \end{aligned}$$

where in the last line we used that $[S \otimes U]^{-1} = [S^{-1} \otimes U^{-1}]$. Initially, we assume that $S \otimes U$ is known or, rather, that both S and U are known. In Section 6, we return to the case where S (or U) is unknown.

Furthermore, let $H_0 : \beta \in \mathcal{C}_0 = \{\beta \in \mathbb{R}^{tk} | \beta_1 = \dots = \beta_{tk}\}$, that is, $H_0 : \beta = \beta^0$, with $\beta^0 \in \mathcal{C}_0$, and $H_m : \beta \in \mathcal{C}_m$. The derivation of the GORIC modified for multivariate normal linear models resembles the one in Section 2 up to (14). In those formulas, we replace θ by $[I_t \otimes X]\beta$, k by tN , and N by 1 and we use $V = \Sigma \otimes U = \sigma^2[S \otimes U]$. Then, (9) and (14) yield

$$\begin{aligned} B^m(\beta^0, V) &= -\frac{tN}{2} + \frac{tN}{2} E \left\{ \frac{\sigma^2}{\tilde{\sigma}_m^2} \right\} + \\ &\quad \frac{1}{2} E \left\{ ([I_t \otimes X] \tilde{\beta}^m - [I_t \otimes X] \beta^0)' (\tilde{V}^m)^{-1} ([I_t \otimes X] \tilde{\beta}^m - [I_t \otimes X] \beta^0) \right\}, \end{aligned}$$

and

$$\bar{\chi}^2(V, \mathcal{C}_m) = ([I_t \otimes X] \tilde{\beta}^m - [I_t \otimes X] \beta^0)' [\Sigma^{-1} \otimes U^{-1}] ([I_t \otimes X] \tilde{\beta}^m - [I_t \otimes X] \beta^0),$$

respectively. The latter can be rewritten as

$$\bar{\chi}^2(V, \mathcal{C}_m) = (\tilde{\beta}^m - \beta^0)' W^{-1} (\tilde{\beta}^m - \beta^0), \quad (23)$$

with

$$\begin{aligned} W^{-1} &= [I_t \otimes X]' [\Sigma^{-1} \otimes U^{-1}] [I_t \otimes X] \\ &= \Sigma^{-1} \otimes [X' U^{-1} X] \\ &= [\sigma^2 S]^{-1} \otimes [X' U^{-1} X], \end{aligned}$$

since $[I_t \otimes X]' = [I_t \otimes X']$ and $[A \otimes B][C \otimes D] = [AC \otimes BD]$. For univariate regression models (where $V = \sigma^2 I_N$, thus for $t = 1$ and $U = I_N$), $W = \sigma^2 [X'X]^{-1}$. Moreover, for $U = I_N$, $k = 1$, $t = k$, and X a vector of ones, the multivariate normal linear model simplifies to the model discussed in Section 2. In that case, it evidently holds true that $W = \Sigma/N$, where the $k \times k$ matrix Σ in this section equals the $k \times k$ matrix V of Section 2.

Equation (23) equals (14) with θ replaced by β , k by tk and the $k \times k$ matrix V/N by the $tk \times tk$ matrix W . Therefore, the remainder of the derivation resembles the one in Section 2 with these replacements. As a result, the null distribution of (23) is given by

$$Pr(\bar{\chi}^2(V, \mathcal{C}_m) \leq x) = \sum_{l=1}^{tk} w_l(tk, W, \mathcal{C}_m) Pr(\chi_l^2 \leq x).$$

Thus, for $H_m : \beta \in \mathcal{C}_m$ in t -variate normal linear models,

$$\begin{aligned} \text{GORIC}_m &= -2 \log f(y|[I_t \otimes X]\tilde{\beta}^m, \tilde{V}^m) + 2 PT_m, \text{ with} \\ \log f(y|[I_t \otimes X]\tilde{\beta}^m, \tilde{V}^m) &= -\frac{tN}{2} \log\{2\pi\} - \frac{1}{2} \log |\tilde{V}^m| - \\ &\quad \frac{1}{2} \left[(y - [I_t \otimes X]\tilde{\beta}^m)' (\tilde{V}^m)^{-1} (y - [I_t \otimes X]\tilde{\beta}^m) \right], \\ PT_m &= 1 + \sum_{l=1}^{tk} w_l(tk, W, \mathcal{C}_m) l. \end{aligned}$$

In Section 7, we establish how competing hypotheses for multivariate regression models can be evaluated with the GORIC. In the next section, we demonstrate that the GORIC can be used for restrictions of the form $H_m : R\theta \leq r$, with θ the parameter of interest.

4. Restrictions of the form $H_m : R\theta \leq r$

Let θ be the parameter of interest. Until now, we have focussed on $H_m : \theta \in \mathcal{C}_m$, that is, on closed convex cones. A special case (according to Silvapulle & Sen, 2005) is $H_m : \theta \in \mathcal{C}_m = \{\theta \in \mathbb{R}^k : R\theta \leq 0\}$, where R is a $c_m \times k$ matrix. When R is of full rank (after discarding the redundant restrictions),

$$\{\theta \in \mathbb{R}^k : R\theta \leq r\} = \{\theta \in \mathbb{R}^k : R\theta - r \leq 0\} = \{\theta \in \mathbb{R}^k : R\theta^* \leq 0\},$$

where $\theta^* = \theta - q$ and $Rq = r$, with $q \in \mathbb{R}^k$.

It should be stressed that q cannot be defined when R is not of full rank (after discarding the redundant restrictions). For example, q cannot be determined for $H_m : \theta_i \geq r_{11}, \theta_i \leq r_{12}$ when $r_{11} \neq r_{12}$. However, for $r_{11} = r_{12}$, $H_m : \theta_i \geq r_{11}, \theta_i \leq r_{12}$ simplifies to $H_m : \theta_i = r_{11}$. Then, q is defined and equals r_{11} . Consequently, the equality restrictions should be handled separately by looking at $\mathcal{C}_m = \{\theta \in \mathbb{R}^k : R_1\theta \leq r_1, R_2\theta = r_2\}$, where R_1 is a $c_{m1} \times k$ matrix, r_1 a vector of length c_{m1} , R_2 a $c_{m2} \times k$ matrix, and r_2 a vector of length c_{m2} . In that case, $\mathcal{C}_m = \{\theta \in \mathbb{R}^k : R_1\theta^* \leq 0, R_2\theta^* = 0\}$ for $[R'_1, R'_2]'q = [r'_1, r'_2]'$. Now, q exists when $[R'_1, R'_2]'$ is of full rank (after discarding the redundant restrictions).

Since $\{\theta \in \mathbb{R}^k : R\theta \leq 0\}$ is a closed convex cone, $\{\theta \in \mathbb{R}^k : R\theta^* \leq 0\} = \{\theta \in \mathbb{R}^k : R\theta \leq r\}$ is too for the shifted data: both y_i and θ are shifted by q . But, only when R is of full rank (when discarding the redundant restrictions). Hence, $\{\theta \in \mathbb{R}^k : R\theta \leq r\}$ is a closed convex cone with a relocated base, that is, the origin $(0, 0)$ is no longer the base of the cone. Therefore, we will refer to this cone as a relocated closed convex cone. Previous result implies that the GORIC for $H_m : \theta \in \mathcal{C}_m = \{\theta \in \mathbb{R}^k : R\theta \leq r\}$ has the same expression as the GORIC for $H_m : \theta \in \mathcal{C}_m = \{\theta \in \mathbb{R}^k : R\theta \leq 0\}$ denoted in (19). Evidently, the same remains valid for $H_m : \theta \in \mathcal{C}_m = \{\theta \in \mathbb{R}^k : R_1\theta \leq r_1, R_2\theta = r_2\}$ when $[R'_1, R'_2]'$ is of full rank (after discarding the redundant restrictions).

The analogue remains true for t -variate normal linear models, where β is the parameter of interest, since

$$\{\beta \in \mathbb{R}^{tk} : R\beta \leq r\} = \{\beta \in \mathbb{R}^{tk} : R\beta - r \leq 0\} = \{\beta \in \mathbb{R}^{tk} : R\beta^* \leq 0\},$$

where $\beta^* = \beta - q$ and $Rq = r$, with $q \in \mathbb{R}^{tk}$. Here, y is shifted by $[I_t \otimes X]q$ and β by q .

In the next section, we elaborate on the level probabilities for all discussed types of models. Furthermore, we give an interpretation of the penalty term.

5. Level probabilities

A level probability $w_l(L, W, \mathcal{C}_m)$ is the probability that there are l levels among the L order-restricted maximum likelihood estimators (see also Anraku (1999); Silvapulle & Sen

(2005, pp. 77–83); Robertson et al. (1988, p. 69)). In other words, it is the probability that the parameter space in accordance with the active constraints in $\mathcal{C}_m$ is of dimension j . Note that the parameters ν (i.e., $\nu = \theta$ in Section 2 and $\nu = \beta$ in Section 3) emanate from the null distribution which is the normal distribution with mean ν^0 and covariance matrix W .

Anraku (1999) laid out on page 149 that in general the calculation of the level probabilities is difficult (see also Robertson et al., 1988, pp. 74–86). However, Silvapulle & Sen (2005, pp. 78–81) point out that it can be done computationally convenient via simulation. To implement this in software, one requires a quadratic program algorithm. For example, one can use the IMSL subroutine QPROG (Visual Numerics, 2003, pp. 1307–1310) in Fortran 90.

In case the restrictions are of the type $H_m : R\nu \leq r$, with $r \neq 0$, the data should be shifted accordingly, as explained in Section 4. Notably, the calculation of the level probabilities does not involve simulation of the data y , but only the parameters ν , and furthermore q does not need to be determined. Therefore, we can just simulate the shifted parameters ν^* and all simulation steps remain valid.

The level probabilities are invariant for the values of ν^0 and σ^2 (Silvapulle & Sen (2005, pp. 86) and Robertson et al. (1988, pp. 69)). However, there is one exception, namely when the restrictions are of the type $\nu \leq r$, including $r = 0$. In case the hypothesis of interest contains this type of restriction, one must use $\nu^0 = 0$. This results in level probabilities that are invariant for the value of σ^2 . Observe that setting ν^0 equal to 0 yields the same result as for $\nu^0 \neq 0$ with $\sigma^2 \rightarrow \infty$.

The penalty term

$$PT_m = 1 + \sum_{l=1}^L w_l(L, W^*, \mathcal{C}_m) l$$

can be seen as the expected dimension of the parameters. In other words, the expected dimension of ν plus 1 because of the unknown variance term σ^2 . When there are no restrictions, the penalty is equal to the number of distinct parameter values. Hence, in that case, the GORIC reduces to the AIC.

6. U or S unknown

Until now, we have assumed that $W = \sigma^2 W^*$ and that W^* is known in the calculation of the penalty term. As mentioned in Section 5, the level probabilities are invariant of positive constants (like σ^2), which implies that $w_l(L, W^*, \mathcal{C}_m) = w_l(L, W, \mathcal{C}_m)$. Consequently, if W^* is known, one can use W^* (in the simulation steps); otherwise, one should use an estimate of W instead of W^* . For the model described in Section 2 in (1), where $W = V/N = \sigma^2 U/N$, this implies that when W^* or rather U is unknown, W or rather V should be estimated from the data; for instance by the maximum likelihood estimator (m.l.e.) of V :

$$\hat{V} = N^{-1} \sum_{i=1}^N [(y_i - \bar{y})(y_i - \bar{y})'],$$

with $\bar{y}$ a vector of the sample means of y . The analogue remains valid for t -variate normal linear models described in Section 3, where $W = \sigma^2 S \otimes [X'U^{-1}X]^{-1}$. For example, in a t -variate regression model (where $U = I_N$), S or rather $\Sigma = \sigma^2 S$ can be estimated by the m.l.e.:

$$\hat{\Sigma} = N^{-1}(Y - X\hat{B})'(Y - X\hat{B}), \quad (24)$$

with Y the $N \times t$ matrix where column h equals y_h for $h = 1, \dots, t$ and $\hat{B}$ the unrestricted maximum likelihood estimator of B , which is a $k \times t$ matrix where column h equals β_h . Importantly, for univariate regression models (where $U = I_N$ is known and $\Sigma = \sigma^2$), no estimation is required, since level probabilities are invariant for positive scalars. Nevertheless, this is not the case for multivariate regression models (where $U = I_N$ is known), because S is unknown.

It should be stressed that $E\{\hat{V}\} = \frac{N-1}{N}V$. Because the level probabilities are invariant of positive constants, $w_l(L, \hat{V}, \mathcal{C}_m) \rightarrow w_l(L, V, \mathcal{C}_m)$ for $N \rightarrow \infty$. Analogously, $w_l(L, \hat{\Sigma}, \mathcal{C}_m) \rightarrow w_l(L, \Sigma, \mathcal{C}_m)$ for $N \rightarrow \infty$.

If V or Σ is estimated from the data, the number of unknown distinct variance terms is $(k+1)k/2$ or $(t+1)t/2$, respectively, instead of 1. Since the restrictions are always on the ν parameters and never on the variance terms, the number of unknown variance terms is

equal for all hypotheses of interest. Thus, although the penalty should then be corrected, the correction is equal for all the hypotheses in the set.

7. The GORIC illustrated

In this section, we will illustrate the GORIC based on real data. This example shows the data of Rencher (1995), originally presented by Box & Youle (1955). The descriptive statistics of the data are given in Table 1.

Table 1: *The descriptive statistics of the dependent variables (y_h) and the predictors (x_j) from Box & Youle (1955)*

Means and standard deviations						Sample covariance		
y_h			x_j			matrix of the y_h s		
h	$\bar{y}_h$	$s.d.(y_h)$	j	$\bar{x}_j$	$s.d.(x_j)$	1	2	3
1	20.18	9.70	1	167.32	6.05	0.06	-0.03	-0.07
2	56.34	4.59	2	27.18	4.12	-0.03	0.79	-0.40
3	20.78	6.55	3	6.50	1.59	-0.07	-0.40	0.36

The data resulted from $N = 19$ designed experiments. There are $t = 3$ dependent variables, namely (1) y_1 , the percentage of unchanged starting material, (2) y_2 , the percentage converted to the desired product, and (3) y_3 , the percentage of unwanted by-product. These dependent variables are measured in experiments involving a chemical reaction in which various combinations of the temperature (x_1), the concentration (x_2), and the time (x_3) were used. These three are used as the predictors in a multivariate regression model. To compare the parameters (β), we have to standardize the dependent variables and the predictors. Consequently, the intercepts are zero and will therefore not be reported. Let $\beta_h = [\beta_{h1}, \beta_{h2}, \beta_{h3}]'$ be the vector of parameters assigned to y_h for $h = 1, 2, 3$.

We expect that the three predictors each have a negative impact on y_1 , the percentage of unchanged starting material, and a positive effect on the changed material measured by y_2 and y_3 . Stated differently, we expect that $\beta_{1j} \leq 0$ and $\beta_{hj} \geq 0$ for $h = 2, 3$ and $j = 1, 2, 3$.

One theory could be that the effects of both temperature (x_1) and concentration (x_2) are lower on the percentage converted to the desired product (y_2) than on that of unwanted by-product (y_3), which yields $\beta_{2j} \leq \beta_{3j}$ for $j = 1, 2$. Based on this, we can formulate two competing hypotheses, namely H_1 and H_2 which are stated below. Moreover, we would like to know whether H_1 or H_2 is the preferred hypothesis. Since both can be bad/weak, it is informative to include the unconstrained hypothesis H_u in which there are no restrictions on the parameters. For illustration purposes, we also include H_0 , in which the β parameters regarding the predictors are restricted to zero, in the set of hypotheses:

$$\begin{aligned}
H_0 : \quad & \beta_{hj} = 0, \text{ for } h = 1, 2, 3 \text{ and } j = 1, 2, 3, \\
H_1 : \quad & \beta_{1j} \leq 0, \beta_{hj} \geq 0, \text{ for } h = 2, 3 \text{ and } j = 1, 2, 3, \\
H_2 : \quad & \beta_{1j} \leq 0, \beta_{hj} \geq 0, \text{ for } h = 2, 3 \text{ and } j = 1, 2, 3, \text{ and} \\
& \beta_{2j} \leq \beta_{3j}, \text{ for } j = 1, 2, \\
H_u : \quad & \beta_{h1}, \beta_{h2}, \beta_{h3}, \text{ for } h = 1, 2, 3.
\end{aligned}$$

In Table 2, the order-restricted maximum likelihood estimators of β ($\tilde{\beta}^m$), the log likelihood values ($\log f(y|[I_t \otimes X]\tilde{\beta}^m, \tilde{\Sigma}^m \otimes I_{19})$), the penalty terms (PT_m), and the GORIC values are given for the four hypotheses of interest. Since the covariance matrix Σ is unknown, it is estimated from the data (which is proportional to the sample covariance in Table 1), which is used in determining the level probabilities $w_l(tk, W, \mathcal{C}_m)$. The hypothesis with the lowest GORIC value is the preferred one. Hence, it is concluded that H_2 is the preferred hypothesis.

This example illustrates that the evaluation of a set of competing hypotheses is done easily.

8. Discussion

In this paper, we extended the use of the GORIC to (i) univariate and multivariate normal linear models and (ii) relocated closed convex cones: $R\theta \leq r$, where θ is a vector of length tk , r a vector of length c_m , and R a $c_m \times tk$ matrix of full rank (when $r \neq 0$).

For univariate regression models, evaluating hypotheses with the GORIC is simple and straightforward. In case of multivariate regression models, the same remains valid when Σ is

Table 2: GORIC of the four specified hypotheses (H_m)							
m	j	Restricted β 's ($\tilde{\beta}_{hj}^m$)			$\log f(\cdot)$	PT_m	GORIC $_m$
		$\tilde{\beta}_{1j}^m$	$\tilde{\beta}_{2j}^m$	$\tilde{\beta}_{3j}^m$			
0	1	0.00	0.00	0.00	-42.33	5.00	94.66
	2	0.00	0.00	0.00			
	3	0.00	0.00	0.00			
1	1	-0.96	0.53	0.84	-9.97	10.08	40.09
	2	-0.61	0.26	0.57			
	3	-0.37	0.36	0.28			
2	1	-0.96	0.53	0.84	-9.97	9.72	39.37
	2	-0.61	0.26	0.57			
	3	-0.37	0.36	0.28			
u	1	-0.96	0.53	0.84	-9.97	13.00	45.93
	2	-0.61	0.26	0.57			
	3	-0.37	0.36	0.28			

Note. GORIC = generalized order-restricted information criterion. Bold indicates the lowest value.

known (up to a positive constant). When Σ is unknown, one has to estimate Σ to compute the penalty term of the GORIC, but the calculation remains easy.

Acknowledgement

This study has been funded by the Netherlands Organization for Scientific Research NWO-VICI-453-05-002. Silvapulle gratefully acknowledges support from the Australian Research Council under Discovery Project.

Appendix A. Theorem 1

According to Silvapulle & Sen (2005, pp. 75), where $\theta^0 = 0$,

$$\sum_{i=1}^N [y_i - \tilde{\theta}^m]' V^{-1} \tilde{\theta}^m = 0.$$

In other words, $y_i - \tilde{\theta}^m$ and $\tilde{\theta}^m$ are V -orthogonal. Since θ^0 is a constant, this property remains valid when y and $\tilde{\theta}^m$ are both shifted by θ^0 , that is

$$\sum_{i=1}^N [(y_i - \theta^0) - (\tilde{\theta}^m - \theta^0)]' V^{-1} (\tilde{\theta}^m - \theta^0) = 0.$$

Using this property yields

$$\begin{aligned} \sum_{i=1}^N (y_i - \theta^0)' V^{-1} (y_i - \theta^0) &= \sum_{i=1}^N ([y_i - \tilde{\theta}^m] + [\tilde{\theta}^m - \theta^0])' V^{-1} ([y_i - \tilde{\theta}^m] + [\tilde{\theta}^m - \theta^0]) \\ &= \sum_{i=1}^N (y_i - \tilde{\theta}^m)' V^{-1} (y_i - \tilde{\theta}^m) + N(\tilde{\theta}^m - \theta^0)' V^{-1} (\tilde{\theta}^m - \theta^0). \end{aligned}$$

The analogue remains true for t -variate normal linear models, where θ should be replaced by $[I_t \otimes X]\beta$ and N by 1. Notably, in this case, y is shifted by $[I_t \otimes X]\beta^0$ and $\tilde{\beta}^m$ by β^0

Appendix B. The expression for $\text{var}(\tilde{\sigma}_m^2/\sigma^2)$

Let $Q = -(\tilde{\theta}^m - \theta^0)' (U/N)^{-1} (\tilde{\theta}^m - \theta^0)/\sigma^2$. From (12), it follows that

$$\begin{aligned} E \{ \tilde{\sigma}_m^2 / \sigma^2 \} &= 1 - \frac{1}{Nk} E \{ -Q \}, \text{ and} \\ \text{var} \left(\frac{\tilde{\sigma}_m^2}{\sigma^2} \right) &= \frac{1}{(Nk)^2} [\text{var}(\chi_{Nk}^2) + \text{var}(Q)], \end{aligned}$$

with $\text{var}(\chi_{Nk}^2) = 2Nk$. Because $Q = O(1)$,

$$\begin{aligned}\text{var}(Q) &= E\{Q^2\} - (E\{Q\})^2 = O(1), \text{ and} \\ \text{var}\left(\frac{\tilde{\sigma}_m^2}{\sigma^2}\right) &= \frac{1}{(Nk)^2} [2Nk + \text{var}(Q)] = \frac{1}{(Nk)^2} [2Nk + O(1)] \\ &= \frac{2}{Nk} + O((Nk)^{-2}).\end{aligned}$$

Using the power series expansion, it holds true that

$$\frac{1}{1 - \frac{1}{Nk}E\{-Q\}} = 1 + \frac{1}{Nk}E\{-Q\} + O((Nk)^{-2}).$$

Now, it follows from (10) that

$$\begin{aligned}E\left\{\frac{\sigma^2}{\tilde{\sigma}_m^2}\right\} &= \left[1 + \frac{1}{Nk}E\{-Q\} + O((Nk)^{-2})\right] + \frac{\frac{2}{Nk} + O((Nk)^{-2})}{\left[1 + \frac{1}{Nk}E\{-Q\} + O((Nk)^{-2})\right]^3} \\ &= 1 + \frac{1}{Nk}E\{-Q\} + \frac{2}{Nk} + O((Nk)^{-2}),\end{aligned}$$

which equals the expression in (13).

The analogue remains true for t -variate normal linear models, where θ should be replaced by β , Nk by tN , and U/N by $S \otimes [X'U^{-1}X]^{-1}$.

References

- Anraku, K. (1999). An information criterion for parameters under a simple order restriction. *Biometrika*, 86, 141–152.
- Box, G. E. P., & Youle, P. V. (1955). The exploration and exploitation of response surfaces: An example of the link between the fitted surface and the basic mechanism of the system. *Biometrics*, 11, 287–323.
- Hughes, A., & King, M. (2003). Model selection using AIC in the presence of one-sided information. *Journal of Statistical Planning and Inference*, 115, 397–411.
- Kuiper, R. M., Hoijsink, H., & Silvapulle, M. J. (2011). An Akaike-type information criterion for model selection under inequality constraints. *Biometrika*, 98 (2), 495–501.
- Kullback, S., & Leibler, R. A. (1951). On information and sufficiency. *Annals of Mathematical Statistics*, 22, 79–86.
- Rencher, A. C. (1995). *Methods of multivariate analysis*. New York: John Wiley and Sons.
- Robertson, T., Wright, F., & Dykstra, R. (1988). *Order restricted statistical inference*. Chichester: Wiley.
- Silvapulle, M. J., & Sen, P. K. (2005). *Constrained Statistical Inference*. New Jersey: Wiley.
- Visual Numerics (2003). *IMSL Fortran Library User's Guide MATH/LIBRARY Volume 2 of 2: Mathematical Functions in Fortran*.