*Biometrika* (2011), **98**, 2, *pp.* 495–501
doi: 10.1093/biomet/asr002
Advance Access publication 22 April 2011

*Printed in Great Britain*

# An Akaike-type information criterion for model selection under inequality constraints

BY R. M. KUIPER, H. HOIJTINK

*Department of Methodology and Statistics, Utrecht University, PO Box 80140, 3508 TC, Utrecht, The Netherlands*

r.m.kuiper@uu.nl    h.hoijtink@uu.nl

AND M. J. SILVAPULLE

*Department of Econometrics and Business Statistics, Monash University, Caulfield East, 3145, Australia*

mervyn.silvapulle@monash.edu

SUMMARY

The Akaike information criterion for model selection presupposes that the parameter space is not subject to order restrictions or inequality constraints. Anraku (1999) proposed a modified version of this criterion, called the order-restricted information criterion, for model selection in the one-way analysis of variance model when the population means are monotonic. We propose a generalization of this to the case when the population means may be restricted by a mixture of linear equality and inequality constraints. If the model has no inequality constraints, then the generalized order-restricted information criterion coincides with the Akaike information criterion. Thus, the former extends the applicability of the latter to model selection in multi-way analysis of variance models when some models may have inequality constraints while others may not. Simulation shows that the information criterion proposed in this paper performs well in selecting the correct model.

*Some key words*: Akaike information criterion; Analysis of variance; Constrained inference; Model selection; Order-restricted information criterion; Order restriction.

## 1. INTRODUCTION

The Akaike information criterion (Akaike, 1973) is among the most widely used criteria for model selection. This criterion assumes that the parameter space of the model is not restricted by inequality constraints of the form $\theta_1 \leqslant \theta_2$, where $\theta_1$ and $\theta_2$ are unknown parameters. In this note, we propose an Akaike-type information criterion for the analysis of variance model when the treatment means $\{\theta_1, \ldots, \theta_p\}$ are assumed to satisfy a mixture of linear equality and inequality constraints.

To illustrate the essentials, let us consider the analysis of variance model

$$y_{ij} \sim N(\theta_i, \sigma^2) \quad (i = 1, \ldots, p, \ j = 1, \ldots, n_i), \tag{1}$$

with independent observations from $p$ normal populations, and let $\theta = (\theta_1, \ldots, \theta_p)^{\mathrm{T}}$. This setting is general enough to incorporate factorial designs. For model (1), AIC $= -2\{$maximum loglikelihood $-$ number of parameters$\}$ when $\theta$ is not subject to inequality constraints. However, in many studies, prior information such as that the new treatment is at least as good as the old treatment, which may take the form $\theta_1 \leqslant \theta_2$, is available. In such cases, the Akaike information criterion is not suitable for model selection. When $\theta$ satisfies the simple order $\theta_1 \leqslant \cdots \leqslant \theta_p$, Anraku (1999) proposed the order-restricted information criterion

$$\textsc{oric} = -2\left\{\text{maximum log likelihood} - 1 - \sum_{i=1}^{p} i w_i\right\}, \tag{2}$$

where the constants $\{w_0, \ldots, w_p\}$ are the so-called level probabilities for the simple order $\theta_1 \leqslant \cdots \leqslant \theta_p$. In this note, we propose an extension of (2), called the generalized order-restricted information criterion, to the case when the parameter $\theta$ may be restricted by $R\theta \geqslant 0$ where $R$ is any matrix of known constants.

The form $R\theta \geqslant 0$ is general enough to accommodate practically any linear inequality constraints encountered in practice. Some examples to which the generalized order-restricted information criterion is applicable include the simple order, the tree order $\theta_1 \leqslant \theta_2, \ldots, \theta_1 \leqslant \theta_p$, and the matrix order (Silvapulle & Sen, 2005, pp. 43, 296). By contrast, (2) is applicable only when $\theta_1 \leqslant \cdots \leqslant \theta_p$; thus, for example, it is not applicable for the tree order or the matrix order, even after transformation of the model.

## 2. The generalized order-restricted information criterion

### 2·1. *Preliminaries*

Consider the analysis of variance model (1) with $\theta$ subject to $R\theta \geqslant 0$, where $R$ is a $r \times p$ matrix. Let $n = n_1 + \cdots + n_p$ be the total number of observations. It follows from (1) that the loglikelihood is

$$\ell(\theta, \sigma) = -2^{-1} n \log(2\pi\sigma^2) - 2^{-1} \sum_{i=1}^{k} \sum_{j=1}^{n_i} \left\{\frac{y_{ij} - \theta_i}{\sigma}\right\}^2 .$$

Let $\ell_0(\theta^*, \sigma^*)$ denote the expected loglikelihood function $E\{\ell(\theta^*, \sigma^*) \mid \theta, \sigma\}$ at an arbitrary point $(\theta^*, \sigma^*)$ in the parameter space, where the expectation is evaluated at the true value $(\theta, \sigma)$. Then

$$\ell_0(\theta^*, \sigma^*) = -\frac{1}{2} n\{\log(2\pi) + \log \sigma^{*2}\} - \frac{1}{2}\left\{n\left(\frac{\sigma}{\sigma^*}\right)^2 + \sum_{i=1}^{p} n_i \left(\frac{\theta_i - \theta_i^*}{\sigma^*}\right)^2\right\}.$$

Let $(\tilde{\theta}, \tilde{\sigma})$ denote the maximum likelihood estimator of $(\theta, \sigma)$ under equality and/or inequality constraints, if there are any. The objective of an information criterion-based approach is to choose the model for which $\ell_0(\tilde{\theta}, \tilde{\sigma})$, the expected loglikelihood at the maximum likelihood estimator $(\tilde{\theta}, \tilde{\sigma})$, is maximized. However, $\ell_0(\cdot)$ depends on the unknown population distribution, and therefore, the standard method is to use an estimator of $\ell_0(\tilde{\theta}, \tilde{\sigma})$. A natural estimator of this is the maximum loglikelihood, $\ell(\tilde{\theta}, \tilde{\sigma})$. However, this is not a good estimator because its bias $B(\theta, \sigma) = E\{\ell(\tilde{\theta}, \tilde{\sigma}) - \ell_0(\tilde{\theta}, \tilde{\sigma}) \mid \theta, \sigma\}$ does not reduce to zero as $n \to \infty$. Details of these derivations for the case when there are no inequality constraints are well known (Claeskens & Hjort, 2008; Hurvich & Tsai, 1989; McQuarrie & Tsai, 1998; Gourieroux & Monfort, 1995, § 22.3.2). If $\theta$ is restricted by $R\theta \geqslant 0$, then

$$B(\theta, \sigma) = -\frac{n}{2} + \frac{n}{2} E\left\{\left(\frac{\sigma}{\tilde{\sigma}}\right)^2 ; \theta, \sigma\right\} + \frac{1}{2} E\left\{\sum_{i=1}^{p} n_i \frac{(\tilde{\theta}_i - \theta_i)^2}{\tilde{\sigma}^2} ; \theta, \sigma\right\}.$$

Let us temporarily suppose that there are no constraints on $\theta$, and let $(\hat{\theta}, \hat{\sigma})$ denote the unconstrained maximum likelihood estimator of $(\theta, \sigma)$ Then the bias $E\{\ell(\hat{\theta}, \hat{\sigma}) - \ell_0(\hat{\theta}, \hat{\sigma}); \theta, \sigma\}$ in estimating $\ell_0(\hat{\theta}, \hat{\sigma})$ by $\ell(\hat{\theta}, \hat{\sigma})$ is $p + o(1)$. Therefore, an asymptotically unbiased estimator of $\ell_0(\hat{\theta}, \hat{\sigma})$ is $\{\ell(\hat{\theta}, \hat{\sigma}) - p\}$, which is proportional to AIC.

In the inequality constrained case, suppose that $\theta$ is subject to $R\theta \geqslant 0$. Now, $B(\theta, \sigma)$ is no longer constant up to a term of order $o(1)$, and therefore, the bias cannot be removed by subtracting a constant. For this setting, Anraku (1999) proposed the order-restricted information criterion $-2\{\ell(\tilde{\theta}, \tilde{\sigma}) - \inf_{\theta,\sigma} B(\theta, \sigma)\}$, which resembles the AIC and is most favourable to the parametric model. Because $B(\theta, \sigma)$ depends on the particular inequality constraints, a challenge is to find simple and practical ways of computing $\inf_{\theta,\sigma} B(\theta, \sigma)$ for different inequality constraints.

For the simple order restriction $\theta_1 \leqslant \cdots \leqslant \theta_p$, Anraku (1999) showed that $\inf_{\theta,\sigma} B(\theta, \sigma)$ has the closed form $(1 + \sum_{i=1}^{p} i w_i)$, which in turn led to (2). A main contribution of this note is to provide a similar simple closed form for $\inf_{\theta,\sigma} B(\theta, \sigma)$ when $\theta$ is restricted by $R\theta \geqslant 0$.

### 2·2. *A closed form for the penalty term* $\inf_{\theta,\sigma} B(\theta, \sigma)$

Let $W = \text{diag}\{n_1^{-1}, \ldots, n_p^{-1}\}$ be the diagonal matrix with the $i$th diagonal being $n_i^{-1}$ $(i = 1, \ldots, p)$. Let $\mathcal{C} = \{\theta^* : R\theta^* \geqslant 0\}$, $X \sim N(0, W)$ and $\tilde{X} = \arg\min_x\{(X - x)^{\mathrm{T}} W^{-1}(X - x) : x \in \mathcal{C}\}$. Now, let $\{w_i(p, W, \mathcal{C}), i = 0, \ldots, p\}$ be the nonnegative constants that are uniquely defined and appear in the chi-bar square distribution, $\text{pr}(\tilde{X}^{\mathrm{T}} W^{-1} \tilde{X} \leqslant c) = \sum_{i=0}^{p} w_i(p, W, \mathcal{C})\text{pr}(\chi_i^2 \leqslant c)$. These constants, also known as chi-bar square weights, arise naturally in constrained statistical inference, where their computation has been studied in detail. For details and references, see § 3.5 in Silvapulle & Sen (2005) and Silvapulle (1996).

The crucial result to extend order-restricted information criterion to accommodate more general order restrictions is the following.

Proposition 1. *Consider the normal theory analysis of variance model* (1). *Let $\mathcal{C}$ be a closed convex cone in $R^p$ or $\mathcal{C} = R^p$. Let $\theta \in \mathcal{C}$ and $\sigma > 0$. Then* $1 + \sum_{i=1}^{p} i w_i(p, W, \mathcal{C}) + O(n^{-1}) \leqslant B(\theta, \sigma) \leqslant (1 + p) + O(n^{-1})$, *where the lower bound is reached if and only if $\theta$ lies in the largest linear space contained in $\mathcal{C}$.*

This result is applicable when the constraints are of the form $R\theta \geqslant 0$ because $\mathcal{C}$ can then be taken to be $\{\theta \in R^p : R\theta \geqslant 0\}$. In view of the greatest lower bound for $B(\theta, \sigma)$ in Proposition 1, and the form $-2\{\ell(\tilde{\theta}, \tilde{\sigma}) - \inf_{\theta,\sigma} B(\theta, \sigma)\}$ for the information criterion, we define the generalized order-restricted information criterion

$$\text{GORIC} = -2\left\{\ell(\tilde{\theta}, \tilde{\sigma}) - 1 - \sum_{i=1}^{p} i w_i(p, W, \mathcal{C})\right\}. \tag{3}$$

We propose that the model for which this is the minimum be chosen. For the special case of simple order, (3) reduces to (2). Suppose that there are no inequality constraints on $\theta$. Then, Proposition 1 is applicable with $\mathcal{C} = R^p$. With this choice, we have $w_p(p, W, \mathcal{C}) = 1$, $w_i(p, W, \mathcal{C}) = 0$ for $i < p$, and thus, the generalized order-restricted information criterion reduces to AIC.

The approach proposed in this paper shares a consistency property with the traditional Akaike information criterion approach. To establish this, let us consider the two models $H_a : \theta \in \mathcal{C}_a$ and $H_b : \theta \in \mathcal{C}_b$, where $\mathcal{C}_a$ and $\mathcal{C}_b$ are closed convex cones and are not equal. Suppose that the true parameter $\theta$ lies in $\mathcal{C}_a$ and not in $\mathcal{C}_b$. Then, by mimicking the arguments in Anraku (1999) for his Theorem 4, we have $n^{-1}(\text{GORIC}^a - \text{GORIC}^b)/(-2) = c + o_p(1)$, where $c = E[\log\{f(y; \theta, \sigma)\} \mid \theta, \sigma] - \log\{f(y; \theta^b, \sigma^b)\} > 0$, and $(\theta^b, \sigma^b)$ is the probability limit of the maximum likelihood estimator of $(\theta, \sigma)$ under model $\mathcal{C}_b$. Hence, the correct model will be chosen with probability going to 1 for $n \to \infty$.

The only computer program required to compute generalized order-restricted information criterion is a quadratic program. Since such programs are available in many mathematical and statistical computer software, computation of generalized order-restricted information criterion does not encounter any serious difficulties. The computer time required to compute the penalty term $1 + \sum_{i=1}^{p} i w_i(p, W, \mathcal{C})$ in generalized order-restricted information criterion does not depend on the number of observations in the sample, but only on the dimension of $\theta$ and the nature of inequality constraints on $\theta$. In most practical settings, the computation of generalized order-restricted information criterion would take only a matter of seconds.

## 3. An example

Zelano et al. (1972) conducted an experiment to evaluate the effect of exercise on the age $y$ at which a child starts to walk. The data are available in Silvapulle & Sen (2005, p. 34). Each of the four groups received a different walking exercise. The first group received a seven-week walking exercise for twelve minutes a day beginning at the age of one week. The second group received a daily exercise, but not a daily walking exercise. The third group did not receive any exercises, and serves as control group. The fourth

Table 1. *Estimates of the penalty term $\inf_{(\theta,\sigma)} B(\theta,\sigma)$, the loglikelihood $\ell(\tilde{\theta},\tilde{\sigma}^2)$ and generalized order-restricted information criterion*

| Hypothesis | $\inf_{(\theta,\sigma)} B(\theta,\sigma)$ | $\ell(\tilde{\theta},\tilde{\sigma}^2)$ | GORIC |
|---|---|---|---|
| $H_0: \theta_1 = \theta_2 = \theta_3 = \theta_4$ | 2·00 | −43·36 | 90·73 |
| $H_1: \theta_1 \leqslant \theta_2 \leqslant \theta_4 \leqslant \theta_3$ | 3·10 | −40·01 | 86·23 |
| $H_2: \theta_1 \leqslant \theta_3, \theta_2 \leqslant \theta_3, \theta_1 \leqslant \theta_4, \theta_2 \leqslant \theta_4$ | 3·61 | −40·01 | 87·25 |
| $H_u$: No restrictions on the parameters | 5·00 | −40·01 | 90·03 |

GORIC, generalized order-restricted information criterion.

group also did not receive any exercise, but they were checked weekly for progress. The model used here is (1), with $p = 4$, $n_1 = n_2 = n_4 = 6$, $n_3 = 5$, and $\theta_i$ the mean age in months at which a child starts to walk $(i = 1, \ldots, 4)$.

Because the effect of the exercises is not completely understood, several different possible hypotheses are of interest. One possible hypothesis is $H_1$ in Table 1, that the mean age decreases with increasing intensity of exercise. Another is $H_2$ in Table 1, that Treatments 1 and 2 are at least as good as Treatments 3 and 4, but no ordering is suggested between Treatments 1 and 2, or between Treatments 3 and 4.

The hypotheses $H_0$, $H_1$, $H_2$ and $H_u$ in Table 1 have different inequality constraints on $\theta$. Consequently, the generalized order-restricted information criterion has different values for these hypotheses. Table 1 suggests that, in terms of the generalized order-restricted information criterion, model $H_1$ fits better than the other three. The traditional approach based on AIC is unable to provide such a discrimination between these models. Because the order restriction in $H_2$ cannot be expressed as a simple order, the method in Anraku (1999) is inadequate to compare $H_2$ with the other models in Table 1. In this sense, the generalized order-restricted information criterion extends the applicability of the order-restricted information criterion to more general linear order restrictions.

## 4. Simulation

A simulation study was carried out to evaluate the performance of the generalized order-restricted information criterion, using the design of a real data example. Berzonsky et al. (2003) studied the effects of parthenogenesis on wheat embryo formation in the presence and in the absence of maize pollination. This experiment was conducted as a balanced $4 \times 2$ factorial design in a glass house. The response variable $y$ is a measure of embryo formation. Factor A is genotype with four levels, and Factor B is maize pollination with two levels. Berzonsky et al. (2003) studied the two-way analysis of variance model, $Y_{ijk} = \mu_{ij} + \eta_{ijk}$ $(i = 1, \ldots, 4, j = 1, 2, k = 1, \ldots, 20)$. They also discussed possible orderings of the cell mean parameters based on prior knowledge about the relationship among the cell means. The main one is stated below as $H_1$. One use of a well-fitting model in the context of this study is prediction of embryo formation under different experimental conditions.

To apply the results of this paper, let us rewrite the foregoing model as $y_{ij} = \theta_i + \varepsilon_{ij}$ $(i = 1, \ldots, 8, j = 1, \ldots, 20)$. Now, let us define $H_u$ as the model with no restrictions on $\theta$, $H_0: \theta_1 = \theta_2 = \theta_3 = \theta_4, \theta_5 = \theta_6 = \theta_7 = \theta_8$, and

$$
\begin{aligned}
H_1:\ & \theta_1 \geqslant \{\theta_2, \theta_3, \theta_4\}, \theta_5 \geqslant \{\theta_6, \theta_7, \theta_8\}, \theta_1 \geqslant \theta_5, \theta_2 \geqslant \theta_6, \theta_3 \geqslant \theta_7, \theta_4 \geqslant \theta_8, \\
& \theta_1 - \theta_5 \geqslant \{\theta_2 - \theta_6, \theta_3 - \theta_7, \theta_4 - \theta_8\}.
\end{aligned}
$$

To choose suitable parameter values for the simulation, we used the effect size (Cohen, 1992) and the true hypothesis to guide us. Nine different values for the vector of population means were studied. For each value of $\theta$, estimates were obtained using 1000 independent samples. Based on these, we computed the

Table 2. *Percentage of times that different models were chosen by the generalized order-restricted information criterion*

| | Case 1: $H_0$ is true | | | Case 2: $H_1$ is true | | | Case 3: $H_u$ is true | | |
|---|---|---|---|---|---|---|---|---|---|
| *ES* | $H_0$ | $H_1$ | $H_u$ | $H_0$ | $H_1$ | $H_u$ | $H_0$ | $H_1$ | $H_u$ |
| 0·1 | 84 | 9 | 7 | 48 | 49 | 3 | 60 | 30 | 10 |
| 0·25 | 91 | 1 | 9 | 7 | 92 | 1 | 18 | 35 | 47 |
| 0·4 | 91 | 0 | 9 | 0 | 99 | 1 | 1 | 9 | 91 |

*ES*, effect size.

percentage of times that the correct model was chosen. Table 2 shows that the method introduced in this paper, selected the correct model at least 90% of the times, when the effect size was greater or equal to 0·25. When the effect size was equal to 0·1, the method selected $H_0$ more often, as expected. More simulation results and the computer program for computing generalized order-restricted information criterion are available in the online Supplementary Material.


## Acknowledgement

The first two authors were funded by the Netherlands Organization for Scientific Research, and the third author was partially supported by a grant from the Australian Research Council under a Discovery Project. The authors would like to thank the referees, the associate editor and the editor for constructive comments.


## Supplementary material

Supplementary Material available at *Biometrika* online includes simulation results and the computer program for computing the generalized order-restricted information criterion.

## Appendix

### *Proofs*

Let $W$ be a positive definite matrix of order $p \times p$. Let $\|x\|^2$ denote the squared length $x^{\mathrm{T}} W^{-1} x$, $\langle x, y\rangle$ denote the inner product $x^{\mathrm{T}} W^{-1} y$ and $P(x \mid \mathcal{C})$ denote the projection of $x$ onto $\mathcal{C}$ defined by $\arg\min_{c \in \mathcal{C}} \|x - c\|$. Thus, $P(x \mid \mathcal{C})$ is the point in $\mathcal{C}$ closest to $x$ with respect to the distance $\|\cdot\|$. For any set $A \subset R^p$, let $\rho(x, A)$ denote the distance $\inf_{a \in A} \|x - a\|$ between the point $x$ and the set $A$. Let $\mathcal{M}$ be the largest linear space contained in $\mathcal{C}$.

Lemma A1. *Let $X \in R^p$, $\theta_0 \in \mathcal{M}$, $\theta_1 \in \mathcal{C}$ and $\lambda = \theta_1 - \theta_0$. Then* (i) $\|X - P(X \mid \mathcal{C})\| \geqslant \|(X + \theta_1) - P(X + \theta_1 \mid \mathcal{C})\|$ *and* (ii) $\|P(X \mid \mathcal{C}) - \theta_0\| \leqslant \|P(X + \lambda \mid \mathcal{C}) - (\lambda + \theta_0)\|$.

The first part of the lemma follows from $\mathcal{C} \subset \mathcal{C} - \theta_1$. The second part follows from Nomakuchi (2002, Thm 2.1).

Lemma A2. *Let $R\theta_b \geqslant 0$ and $R\theta_a = 0$. For a given vector of error terms $E$, let $Y_{aij} = \theta_{ai} + E_{ij}$ and $Y_{bij} = \theta_{bi} + E_{ij}$. Let the suffices a and b correspond to $\theta_a$ and $\theta_b$, respectively. Then,* (i) $\tilde{\sigma}_b^2 \leqslant \tilde{\sigma}_a^2$, (ii) $\|\tilde{\theta}_b - \theta_b\|^2 \geqslant \|\tilde{\theta}_a - \theta_a\|^2$ *and* (iii) $B_2(\theta_b, \sigma) \geqslant B_2(\theta_a, \sigma)$.

*Proof.* Let $\lambda = \theta_b - \theta_a$. If $R\gamma \geqslant 0$, then there exists a $\theta^*$ such that $R\theta^* \geqslant 0$ and $\gamma = \theta^* - \lambda$. Therefore, $n\tilde{\sigma}_b^2 = \min_{R\theta^* \geqslant 0} \sum_{ij} (Y_{bij} - \theta_i^*)^2 = \min_{R\theta^* \geqslant 0} \sum_{ij} \{\theta_{ai} + E_{ij} - (\theta_i^* - \lambda_i)\}^2 \leqslant \min_{R\gamma \geqslant 0} \sum_{ij} (\theta_{ai} + E_{ij} - \gamma_i)^2 = n\tilde{\sigma}_a^2$. By Lemma A2(ii), we have $\|\tilde{\theta}_b - \theta_b\|^2 = \|P(\hat{\theta}_b \mid \mathcal{C}) - \theta_b\|^2 = \|P(\hat{\theta}_a + \lambda \mid \mathcal{C}) - (\theta_a + \lambda)\|^2 \geqslant \|P(\hat{\theta}_a \mid \mathcal{C}) - \theta_a\|^2 = \|\tilde{\theta}_a - \theta_a\|^2$. Part (iii) follows from (i) and (ii) in Lemma A2. □

Lemma A3. *Suppose that $R\theta = 0$. Then $E\{\sigma^2/\tilde{\sigma}^2 \mid (\theta, \sigma)\} = 1 + n^{-1} \sum_{i=0}^{p} i w_i(p, W, \mathcal{C}) + 2n^{-1} + O(n^{-2})$.*

*Proof.* We use the lemmas from Silvapulle & Sen (2005, pp. 125–132) without further comment. Let $\{F_1, \ldots F_m\}$ be the partition of $\mathcal{C} = \{x \in R^p : Rx \geqslant 0\}$, where each $F_s$ is the relative interior of a face of $\mathcal{C}$ $(s = 1, \ldots, m)$; for a definition of face see Silvapulle & Sen (2005, p. 124). Let $S_s = \{x \in R^p : P(x \mid \mathcal{C}) \in F_s\}$ $(s = 1, \ldots, m)$. Then $\{S_1, \ldots, S_m\}$ is a partition of $R^p$, except for a set of measure zero. Let $L_s$ denote the linear space spanned by $F_s$ for $s = 1, \ldots, m$. By arguments similar to the proof of Theorem 3.4.2 in Silvapulle & Sen (2005), we have

$$\mathrm{pr}\left(\frac{\tilde{\sigma}^2}{\sigma^2} \leqslant t\right) = \sum_{i=0}^{p} \sum_{\text{over all } s \text{ with } \dim(L_s)=p-i} \mathrm{pr}(\hat{\theta} \in S_s)\, \mathrm{pr}\left(\frac{\tilde{\sigma}^2}{\sigma^2} \leqslant t \,\middle|\, \hat{\theta} \in S_s\right)$$

$$= \sum_{i=0}^{p} w_{p-i}(p, W, \mathcal{C}) \mathrm{pr}(\chi^2_{n-p+i} \leqslant nt).$$

Now, with $N_i = (n - p + i)/2$, we have

$$E\left(\frac{\sigma^2}{\tilde{\sigma}^2}\right) = \int_0^\infty t^{-1} d\left\{\mathrm{pr}\left(\frac{\tilde{\sigma}^2}{\sigma^2}\right) \leqslant t\right\} = \int_0^\infty t^{-1} d\left\{\sum_{i=0}^{p} w_{p-i}(p, W, \mathcal{C}) \mathrm{pr}(\chi^2_{n-p+i} \leqslant nt)\right\}$$

$$= \sum_{i=0}^{p} w_{p-i}(p, W, \mathcal{C}) \int_0^\infty t^{-1}\{\Gamma(N_i)2^{N_i}\}^{-1} \exp\left(\frac{-nt}{2}\right)(nt)^{N_i-1} n\, dt$$

$$= \frac{n}{2} \sum_{i=0}^{p} w_{p-i}(p, W, \mathcal{C})(N_i - 1)^{-1}$$

$$= 1 + n^{-1}\left\{2 + \sum_{i=0}^{p} i w_i(p, W, \mathcal{C})\right\} + O(n^{-2}).$$

□

Lemma A4. *If $R\theta = 0$, then $E(\tilde{\sigma}^{-2}\|\tilde{\theta} - \theta\|^2) = \sum_{i=0}^{p} i w_i(p, W, \mathcal{C}) + O(n^{-1})$.*

*Proof.* Let $F_s$, $S_s$ and $L_s$, $s = 1, \ldots, m$, be the same as those in the proof of the previous lemma. Conditional on $\{\hat{\theta} \in S_s\}$, $\|\tilde{\theta} - \theta\|^2$ and $n\tilde{\sigma}^2$ are independent and are distributed as $\chi^2_i$ and $\chi^2_{n-i}$, respectively, where $i = \dim(L_s)$. Now,

$$\mathrm{pr}(\tilde{\sigma}^{-2}\|\tilde{\theta} - \theta\|^2 \leqslant c) = \sum_{i=0}^{p} \sum_{\text{over all } s \text{ with } \dim(L_s)=i} \mathrm{pr}(\hat{\theta} \in S_s)\mathrm{pr}(\tilde{\sigma}^{-2}\|\tilde{\theta} - \theta\|^2 \leqslant c \mid \hat{\theta} \in S_s)$$

$$= \sum_{i=0}^{p} w_i(p, W, \mathcal{C}) \mathrm{pr}\left(\frac{ni}{n-i} F_{i,n-i} \leqslant c\right).$$

Hence, $E(\tilde{\sigma}^{-2}\|\tilde{\theta} - \theta\|^2) = \sum_{i=0}^{p} i w_i(p, W, \mathcal{C})\{1 + O(n^{-1})\} = \sum_{i=0}^{p} i w_i(p, W, \mathcal{C}) + O(n^{-1})$. □

## References

Akaike, H. (1973). Information theory and an extension of the maximum likelihood principle. In *Proc. 2nd Int. Symp. Information Theory*, Ed. B. N. Petrov and F. Csàki, pp. 267–81. Budapest: Akademiai kiado.

Anraku, K. (1999). An information criterion for parameters under a simple order restriction. *Biometrika* **86**, 141–52.

Berzonsky, W. A., Kleven, S. L. & Leach, G. D. (2003). The effects of parthenogenesis on wheat embryo formation and haploid production with and without maize pollination. *Euphytica* **133**, 285–90.

Claeskens, G. & Hjort, N. (2008). Minimising average risk in regression models. *Economet. Theory* **24**, 493–527.

Cohen, J. (1992). A power primer. *Psychol. Bull.* **112**, 155–9.

Gourieroux, C. & Monfort, A. (1995). *Statistics and Econometric Models*, vol. 2. Cambridge: Cambridge University Press.

Hurvich, C. M. & Tsai, C. L. (1989). Regression and time series model selection in small samples. *Biometrika* **76**, 297–307.

McQuarrie, A. D. R. & Tsai, C. L. (1998). *Regression and Time Series Model Selection*. Singapore: World Scientific Publications.

Nomakuchi, K. (2002). A monotonicity of moments concerned with order restricted statistical inference. *Ann. Inst. Statist. Math.* **54**, 621–5.

Silvapulle, M. J. (1996). On an $F$-type statistic for testing one-sided hypotheses and computation of chi-bar-squared weights. *Statist. Prob. Lett.* **28**, 137–41.

Silvapulle, M. J. & Sen, P. K. (2005). *Constrained Statistical Inference: Inequality, Order, and Shape Restrictions*. New York: Wiley.

Zelano, P. R., Zelano, N. A. & Kolb, S. (1972). Walking in the newborn. *Science* **176**, 314–5.

[*Received March* 2010. *Revised December* 2010]