An Information Criteria for Order-restricted Inference

Nan Lin[a], Tianqing Liu[2,b], and Baoxue Zhang[1,2,b]

[a] *Department of Mathematics, Washington University in Saint Louis, Saint Louis, MO 63130, U.S.A.*

[b] *Key Laboratory for Applied Statistics of MOE and School of Mathematics and Statistics, Northeast Normal University, Changchun 130024 Jilin Province, P. R. China*

Abstract: A general information criterion with a general penalty which depends on the size of samples is developed for nested and non-nested models in the context of inequality constraints. The true parameters may be defined by a specified parametric model, or a set of specified estimating functions. When the true parameters are defined by estimating functions, we use the empirical likelihood approach to construct information criterion. The consistency of the relevant detection procedures are also established. A Monte Carlo study indicates that our new criterion is effective, compared to Anraku' (1999) information criterion, for detecting the configuration of normal means satisfying the simple order restriction.

Keywords: AIC; BIC; Empirical likelihood; Changepoint; No-observed-adverse-effect level; Simple order restriction; Umbrella Order restriction; Consistency.

1 Introduction

Many types of problems are concerned with identifying meaningful structure in real world situations. Structure involving orderings and inequalities is often useful since it is easy to interpret, understand, and explain. When experiments conditions have an inherent ordering, making use of ordering information can improve inference.

In bioassays of dose-effect experiments, the no-observed-adverse-effect level L_0 is used to decide the highest dose level at which the adverse effect is acceptably small. In order to determine L_0, Kikuchi, Yanagawa & Nishiyama (1993) and Yanagawa, Kikuchi & Brown (1994) suggested several multiple comparison procedures, and also derived

[1]E-mail: bxzhang@nenu.edu.cn

[2]The authors are partly supported by NSF of China (No.10871037).

a method based on Akaike's information criterion, AIC (Akaike, 1973). In Kikuchi *et al.* (1993), independent random samples corresponding to k dose levels are supposed to come from k normal populations with an unknown common variance σ^2 and means $\theta_1 \leq \theta_2 \leq \cdots \leq \theta_k$. In their paper, L_0 may be defined as a changepoint of mean parameters θ_{m_1} $(1 \leq m_1 \leq k)$ such that

$$\theta_1 = \cdots = \theta_{m_1} < \theta_{m_1+1} \leq \cdots \leq \theta_k.$$

Kikuchi *et al.* (1993) and Yanagawa *et al.* (1994) also examined the probabilities of correct decision of the AIC method and other testing procedures by simulation. However, as far as we know, in classical statistics, hardly any model selection procedure is available for nested or non-nested models in the context of inequality constraints. The problem is that most model selection criterion such AIC and BIC have a penalty for the number of parameters, and in the inequality constrained models, this number is not clear.

Recently, Anraku (1999) proposed a better theoretical information criterion for parameters under a simple order restriction by employing a more favorable penalty number. Peng *et al.* (2002) proposed an information criterion for detecting the multiplicity of the largest parameter as well as the configuration of true parameters with simple order restriction. They also established strong consistency of their procedures.

In this paper, a general information criterion with a general penalty which depends on the size of samples is developed for nested or non-nested models in the context of inequality constraints. The true parameters may be defined by a specified parametric model, or a set of specified estimating functions. When the true parameters are defined by estimating functions, we use the empirical likelihood approach to construct information criterion. The consistency of the relevant detection procedures are also established.

2 An Information Criteria for Order-restricted Inference

Suppose we have independent random samples from each of k populations specified by scalar-valued, unknown parameters $\theta_1, \cdots, \theta_k$ satisfying some unknown order restriction. Our concern is to seek distinct parameters among $\theta_1, \cdots, \theta_k$ based on the data. The following definition help us in constructing information criterion.

Definition: Let $M(\boldsymbol{\theta})$ be a saturated model with parameter $\boldsymbol{\theta} \in \boldsymbol{\Theta} \subseteq \mathcal{R}^k$, where $\boldsymbol{\Theta}$ is the parameter space. Let $\boldsymbol{\Theta}_0 = \{\boldsymbol{\theta} : \theta_1 = \cdots = \theta_k\}$ and $\boldsymbol{\Theta}_t$ be a measurable set such

that $\mathbf{\Theta}_0 \subseteq \mathbf{\Theta}_t \subseteq \mathbf{\Theta}$. Furthermore, let $M(\mathbf{\Theta}_t) =: M(\boldsymbol{\theta}), \quad \boldsymbol{\theta} \in \mathbf{\Theta}_t$, which means $M(\mathbf{\Theta}_t)$ is a sub-model with parameter space $\mathbf{\Theta}_t$. Define *model family* as follows:

$$\mathscr{F} = \{M(\mathbf{\Theta}_t) : \mathbf{\Theta}_0 \subseteq \mathbf{\Theta}_1 \subseteq \mathbf{\Theta}_2 \subseteq \cdots \subseteq \mathbf{\Theta}_T\}$$

where T is finite. It is easily seen that $M(\mathbf{\Theta}_t)$ is nested in $M(\mathbf{\Theta})$ for all t and when $t_1 < t_2$, $M(\mathbf{\Theta}_{t_1})$ is nested in $M(\mathbf{\Theta}_{t_2})$.

For example, the classic information criteria consider model $M(\boldsymbol{\theta})$, $\boldsymbol{\theta} = (\theta_1, \cdots, \theta_T)^\tau$, $\boldsymbol{\theta} \in R^T$ with $(\theta_1, \cdots, \theta_t)^\tau \in \mathbf{\Theta}_t = R^t$, that is $(\theta_{t+1}, \cdots, \theta_T)^\tau$ are fixed, it is easy to have $\mathscr{F} = \{M(\mathbf{\Theta}_t)\}_{t=1}^T$ is a model family. Under order restriction, Let

$$M_1 : \theta_1 = \theta_2 = \theta_3, \quad M_2 : \theta_1 = \theta_2 \leq \theta_3, \quad M_3 : \theta_1 \leq \theta_2 \leq \theta_3, \quad M_4 : \theta_1 \perp \theta_2 \perp \theta_3,$$

where $\perp$ means the independence of two parameters. Then, $\mathscr{F} = \{M_t\}_{t=1}^4$ is a model family.

Suppose we have $\mathbb{s}$ model families

$$\mathscr{F}_s = \{M(\mathbf{\Theta}_{st}) : \mathbf{\Theta}_0 \subseteq \mathbf{\Theta}_{s1} \subseteq \mathbf{\Theta}_{s2} \subseteq \cdots \subseteq \mathbf{\Theta}_{s\mathbb{t}_s}\}, \quad s = 1, 2, \cdots, \mathbb{s}. \tag{2.1}$$

For convenience, let $M(s,t)$ denote the model $M(\mathbf{\Theta}_{st})$. In this paper, we consider information criteria as follows:

$$c_n(s,t) = -2\log(L_n(x,s,t))/n + p(s,t) \cdot \varphi(n)/n, \tag{2.2}$$

where $p(s,t)$ is the dimension of model $M(s,t)$. $L_n(x,s,t)$ is the maximum likelihood based on the model $M(s,t)$, the model is selected that corresponds to

$$(\hat{s}, \hat{t}) = \arg \min_{s \leq \mathbb{s}, t \leq \mathbb{t}_s} c_n(s,t). \tag{2.3}$$

Theorem 2.1 *Let $\mathbb{s}$ and $\mathbb{t}_s$ be some finite positive integers. Suppose model families $\{\mathscr{F}_s\}_{s=1}^{\mathbb{s}}$ subject to $\mathscr{F}_i \bigcap \mathscr{F}_j = \emptyset$, $\forall i \neq j$. Let (s_0, t_0) be the correct model family and the correct number of parameters with $s_0 \leq \mathbb{s}$ and $t_0 \leq \mathbb{t}_s$, which means model $M(s_0, t_0)$ is the correct model. Assume that $0 < p(s,t) < \mathbb{p}$ for $s = 1, \cdots, \mathbb{s}$, $t = 1, \cdots, \mathbb{t}$, where $\mathbb{p}$ is a finite constant, and for fixed s, $p(s,t_1) < p(s,t_2)$ if $t_1 < t_2$. Furthermore, assume $\mathscr{F}_{s_0}$ is the largest model family that contains true model $M(s_0,t_0)$. Let $L_n(s,t)$ be the maximum likelihood based on model $M(s,t)$. For $t > t_0$ and $M(s_0,t) \in \mathscr{F}_{s_0}$, assume that*

$$2[\log(L_n(s_0,t)) - \log(L_n(s_0,t_0))] \xrightarrow{d} W(s_0,t_0,t) \tag{2.4}$$

where $\xrightarrow{d}$ denotes the convergence in distribution, $W(s_0,t_0,t)$ is a non-degenerate distribution. Furthermore, we assume

$$\lim_{n\to\infty} \varphi(n)/n = 0, \qquad \lim_{n\to\infty} \varphi(n) = \infty. \tag{2.5}$$

Then,

$$\lim_{n\to\infty} P[\hat{s} = s_0, \hat{t} = t_0] = 1. \tag{2.6}$$

3 Order-restricted Information Criteria Based on Parametric Likelihood

In this section, suppose we specify a probability model $p(x|\boldsymbol{\theta})$ for data x. Let $L_n(x, \boldsymbol{\theta})$ be the likelihood under the model $M(\boldsymbol{\theta})$, where $\boldsymbol{\theta} \in \boldsymbol{\Theta}$. Denote $\hat{\boldsymbol{\theta}}_\lambda$ as the MLE of $\boldsymbol{\theta}$ under model $M(\boldsymbol{\Theta}_\lambda)$ that is

$$\hat{\boldsymbol{\theta}}_\lambda(x) = \arg \sup_{\boldsymbol{\theta} \in \boldsymbol{\Theta}_\lambda} L_n(x, \boldsymbol{\theta}), \tag{3.1}$$

where $\boldsymbol{\Theta}_0 = \{\boldsymbol{\theta} : \theta_1 = \cdots = \theta_k\}$. Furthermore, let

$$\mathcal{A}_n(y, \lambda) = 2\left\{\log(L_n(y, \hat{\boldsymbol{\theta}}_\lambda(y))) - \log(L_n(y, \hat{\boldsymbol{\theta}}_0(x)))\right\} \tag{3.2}$$

Then, the dimension for the model $M(\boldsymbol{\Theta}_\lambda)$ is defined as

$$p_w(\lambda) = \int \mathcal{A}_n(y, \lambda) p(y|\hat{\boldsymbol{\theta}}_0(x)) dy \tag{3.3}$$

It is easy to verify that the model dimension $p_w(\lambda)$ satisfy the conditions of Theorem 2.1. Then, the information criteria (2.2) reduces to

$$ORIC_w(\lambda) = -2l(\lambda) + p_w(\lambda) \cdot \varphi(n), \tag{3.4}$$

where $l(\lambda)$ is the maximum log-likelihood under the λ-th candidate model. According to Theorem 2.1, (3.4) is a consistent criterion.

Let $\nu_1(\lambda)$ and $\nu_2(\lambda)$ are the number of $\perp$ and $\{\geq, \leq\}$ specified in the $\lambda-$th candidate model. Then, we define the model dimension as follows

$$p(\lambda) = \nu_1(\lambda) + \sum_{i=0}^{\nu_2(\lambda)} \frac{1}{1+i}. \tag{3.5}$$

It is easy to verify that the model dimension $p(\lambda)$ satisfy the conditions of Theorem 2.1. Then, the information criteria (2.2) reduces to

$$ORIC(\lambda) = -2l(\lambda) + p(\lambda) \cdot \varphi(n), \tag{3.6}$$

where $l(\lambda)$ is the maximum log-likelihood under the λ-th candidate model. According to Theorem 2.1, (3.6) is a consistent criterion.

Under order restrictions, we choose

$$\varphi_1(n) = 2$$

$$\varphi_2(n) = a \cdot \log(\log(n))$$

$$\varphi_3(n) = b \cdot \log(n)$$

$$\varphi_4(n) = c \cdot \sqrt{n}$$

where, a, b and c is constants. Generally, we can take $a = 2/\log(\log(10k))$, $b = 2/\log(10k)$, and $c = 2/\sqrt{10k}$.

4 Order-restricted Information Criteria Based on Empirical Likelihood

Suppose we have k independent random samples. For the ith sample, let $x_{i1}, \cdots, x_{in_i}$ be $i.i.d.$ observations from a d_i-variate distribution F_i, that there is a 1-dimensional parameter θ_i associated with F_i and information about θ_i and F_i is available in the form of $Eg_i(x_i, \theta_i) = 0$. Let $n = \sum_{i=1}^{k} n_i$, then the log-empirical likelihood is given by

$$l_n(F_1, \cdots, F_k) = \sum_{i=1}^{k} \sum_{j=1}^{n_i} \log p_{ij} \tag{4.1}$$

Obviously, we should maximize $l(F_1, \cdots, F_k)$ under constraints given by

$$\sum_{j=1}^{n_i} p_{ij} = 1 \qquad \sum_{j=1}^{n_i} p_{ij} g_i(x_{ij}, \theta_i) = 0$$

for $i = 1, \cdots, k$. Then, it is easy to have

$$p_{ij}(\theta_i) = \frac{1}{n_i\{1 + \lambda_i^\tau g_i(x_{ij}, \theta_i)\}}$$

and for fixed θ_i, λ_i is a Lagrange multiplier satisfying

$$\sum_{j=1}^{n_i} p_{ij}(\theta_i) g_i(x_{ij}, \theta_i) = 0$$

for $i = 1, \cdots, k$. The corresponding profile log-empirical likelihood is given by

$$l_E(\boldsymbol{\theta}) = \sum_{i=1}^{k} \sum_{j=1}^{n_i} \log\{p_{ij}(\theta_i)\} \tag{4.2}$$

where $\boldsymbol{\theta} = (\theta_1^\tau, \cdots, \theta_k^\tau)^\tau$. Denote $\hat{\boldsymbol{\theta}}_\lambda$ as the maximum empirical likelihood estimate (MELE) of $\boldsymbol{\theta}$ under model $M(\boldsymbol{\Theta}_\lambda)$ that is

$$\hat{\boldsymbol{\theta}}_\lambda = \arg \sup_{\boldsymbol{\theta} \in \boldsymbol{\Theta}_\lambda} l_E(\boldsymbol{\theta}). \tag{4.3}$$

Then, we define the empirical-likelihood-based information criteria as

$$ORIC_{EL}(\lambda) = -2l_E(\hat{\boldsymbol{\theta}}_\lambda) + p(\lambda) \cdot \varphi(n), \tag{4.4}$$

where $p(\lambda)$ is defined in (3.5) . According to Theorem 2.1, (4.4) is a consistent criterion.

5 A simulation study

To examine the performance of the order-restricted information criterion, we executed Monte Carlo simulations for $k = 4$. We generated normal variates with means $\mu_1, \mu_2, \mu_3, \mu_4$ and variance 1. The following configurations of parameters were selected for $\boldsymbol{\mu} = (\mu_1, \mu_2, \mu_3, \mu_4)$:

$H_0: \ \mu 1 = \mu 2 = \mu 3 = \mu 4, \qquad \boldsymbol{\mu} = (0, 0, 0, 0);$
$K_1: \ \mu 1 = \mu 2 = \mu 3 \perp \mu 4, \qquad \boldsymbol{\mu} = (0, 0, 0, -0.5);$
$K_2: \ \mu 1 = \mu 2 \perp \mu 3 = \mu 4, \qquad \boldsymbol{\mu} = (0, 0, -0.5, -0.5);$
$K_3: \ \mu 1 \perp \mu 2 = \mu 3 = \mu 4, \qquad \boldsymbol{\mu} = (0, -0.5, -0.5, -0.5);$
$K_4: \ \mu 1 = \mu 2 \perp \mu 3 \perp \mu 4, \qquad \boldsymbol{\mu} = (0, 0, -0.5, -1);$
$K_5: \ \mu 1 \perp \mu 2 = \mu 3 \perp \mu 4, \qquad \boldsymbol{\mu} = (0, 0.5, 0.5, 0);$
$K_6: \ \mu 1 \perp \mu 2 \perp \mu 3 = \mu 4, \qquad \boldsymbol{\mu} = (0, 0.5, 0, 0);$
$K_7: \ \mu 1 \perp \mu 2 \perp \mu 3 \perp \mu 4, \qquad \boldsymbol{\mu} = (0, 0.5, 0, 0.5);$
$M_1: \ \mu 1 = \mu 2 = \mu 3 \leq \mu 4, \qquad \boldsymbol{\mu} = (0, 0, 0, 0.5);$
$M_2: \ \mu 1 = \mu 2 \leq \mu 3 = \mu 4, \qquad \boldsymbol{\mu} = (0, 0, 0.5, 0.5);$
$M_3: \ \mu 1 \leq \mu 2 = \mu 3 = \mu 4, \qquad \boldsymbol{\mu} = (0, 0.5, 0.5, 0.5);$
$M_4: \ \mu 1 = \mu 2 \leq \mu 3 \leq \mu 4, \qquad \boldsymbol{\mu} = (0, 0, 0.5, 1);$
$M_5: \ \mu 1 \leq \mu 2 = \mu 3 \leq \mu 4, \qquad \boldsymbol{\mu} = (0, 0.5, 0.5, 1);$
$M_6: \ \mu 1 \leq \mu 2 \leq \mu 3 = \mu 4, \qquad \boldsymbol{\mu} = (0, 0.5, 1, 1);$
$M_7: \ \mu 1 \leq \mu 2 \leq \mu 3 \leq \mu 4, \qquad \boldsymbol{\mu} = (0, 0.5, 1, 1.5).$

We examined two types of comparison. First we compared the performance of our order-restricted criterion with that of Anraku(1999) for detecting the true configuration of the parameters. In this simulation study, models $H_0,\ K_1,\ K_2, \cdots,\ K_7,\ M_1,\ M_2, \cdots,\ M_7$ are jointly considered. Here we assume the variance is known to be 1. First, the penalty numbers corresponding to $H_0,\ K_1,\ K_2, \cdots,\ K_7,\ M_1,\ M_2, \cdots,\ M_7$ are given by (3.3) for Figure 1 and Figure 2. Second, the penalty numbers corresponding to $H_0,\ K_1,\ K_2, \cdots,\ K_7,\ M_1,\ M_2, \cdots,\ M_7$ are given by (3.5) for Figure 3 and Figure 4. Furthermore, in order to provide some clues for how to select the $\varphi(n)$ under moderate-size sample, or even small sample, we choose $\varphi(n) = (2/\log(\log(40))) \cdot \log(\log(n)),\ (2/\log(40)) \cdot \log(n),\ (2/\sqrt{40}) \cdot \sqrt{n}$ for our three case. The sample sizes for each population are taken equally, ranging form 10 to 200 with step 10.

Figure 1 and Figure 2 show the frequency of error detection by these four methods based on 1000 simulations for true models $H_0,\ K_1,\ K_2, \cdots,\ K_7,\ M_1,\ M_2, \cdots,\ M_7$, respectively. We observe that our order information criterion work much better than Anraku(1999)'s order information criterion when $\varphi(n) = (2/\log(\log(40))) \cdot \log(\log(n))$ or $\varphi(n) = (2/\log(40)) \cdot \log(n)$. And generally, under the restriction that $\lim_{n\to\infty} \varphi(n)/n =$

0, frequencies of error detection for criteria with $\varphi(n)$ increasing to ∞ more rapidly converge to 0 more quickly than those with "slower" $\varphi(n)$. But the "slower" $\varphi(n)$ is more robust than the "faster" $\varphi(n)$, see K_6 and K_7 of Figure 2. When we used the penalty numbers (3.5), from Figure 3 and Figure 4, we can draw the same conclusions.

Secondly we compared four procedures for detecting the first changepoint (see (6) of Anraku(1999)). Models $H_0,\ M_1,\ M_2, \cdots,\ M_7$ are considered. Here we assumed the variances of normal variates are common and unknown, and the penalty numbers are given by (3.3). And $\varphi(n)$ is choose as above. The sample sizes for each population are taken equally, ranging form 10 to 200 with step 10. Figure 5 shows the frequency of error detection by these four methods based on 1000 simulations for true models $H_0,\ M_1,\ M_2, \cdots,\ M_7$, respectively. We also observe that our order information criterion work much better than Anraku(1999)'s order information criterion when $\varphi(n) = (2/\log(\log(40))) \cdot \log(\log(n))$ or $\varphi(n) = (2/\log(40)) \cdot \log(n)$. And the "slower" $\varphi(n)$ is more robust than the "faster" $\varphi(n)$, see M_3, M_6 and M_7 of Figure 5.

For empirical likelihood inference, we generated $X_1 \sim N(\mu_1, 1),\ X_2 \sim N(\mu_2, 2),\ X_3 \sim t(6)+\mu_3,\ X_4 \sim \chi^2(1)-1+\mu_4$ Figure 6 and Figure 7 shows the frequency of error detection by these four methods based on 1000 simulations for true models $H,\ K_1,\ K_2, \cdots,\ K_7$, $M_1,\ M_2, \cdots,\ M_7$, respectively. We observe that our order information criterion works well when $\varphi(n) = (2/\log(\log(40))) \cdot \log(\log(n))$ or $\varphi(n) = (2/\log(40)) \cdot \log(n)$. And generally, under the restriction that $\lim_{n\to\infty} \varphi(n)/n = 0$, frequencies of error detection for criteria with $\varphi(n)$ increasing to ∞ more rapidly converge to 0 more quickly than those with "slower" $\varphi(n)$. But the "slower" $\varphi(n)$ is more robust than the "faster" $\varphi(n)$, see K_4, K_5, K_6 and K_7 of Figure 6 and Figure 7.

6 Comments and Conclusions

As is well known, the calculation of $P(i, k, w)$ may be difficult except for special weights under various order restrictions. Our information criterion avoids this computation problem.

7 Appendix

Proof of Theorem 2.1. When $s \neq s_0$, $t \leq \mathbb{t}_s$ then the model $M(s,t)$ with $p(s,t)$ parameters is misspecified, so that

$$plim_{n\to\infty} \log(L_n(s,t))/n < plim_{n\to\infty} \log(L_n(s_0,t_0))/n, \tag{A.1}$$

where *plim* denotes convergence in probability. Hence, it follows from (A.1) and $\lim_{n\to\infty} \varphi(n)/n = 0$ that

$$\begin{aligned}
& \lim_{n\to\infty} P[c_n(s_0,t_0) \geq c_n(s,t)] \\
= & \lim_{n\to\infty} P[-2\log(L_n(s_0,t_0)) + p(s_0,t_0)\varphi(n)/n \geq -2\log(L_n(s,t)) + p(s,t)\varphi(n)/n] \\
= & \lim_{n\to\infty} P[\log(L_n(s_0,t_0))/n - \log(L_n(s,t))/n \leq 0.5(p(s_0,t_0) - p(s,t))\varphi(n)/n] \\
\leq & \lim_{n\to\infty} P[\log(L_n(s_0,t_0))/n - \log(L_n(s,t))/n \leq \mathbb{p}\varphi(n)/n] \\
= & 0
\end{aligned} \tag{A.2}$$

so that

$$\begin{aligned}
& \lim_{n\to\infty} P[\hat{t} \leq t_m, \hat{s} \neq s] \\
\leq & \lim_{n\to\infty} P[c_n(s_0,t_0) \geq c_n(s,t) \ for\ some\ t \leq \mathbb{t}_s, s \neq s_0] \\
\leq & \sum_{s\neq s_0} \sum_{t<t_0} \lim_{n\to\infty} P[c_n(s_0,t_0) \geq c_n(s,t)] \\
= & 0
\end{aligned} \tag{A.3}$$

When $s = s_0$, if $t < t_0$ then the model with $p(s_0,t)$ parameters is misspecified, so that

$$plim_{n\to\infty} \log(L_n(s_0,t))/n < plim_{n\to\infty} \log(L_n(s_0,t_0))/n \tag{A.4}$$

Hence, it follows from (A.4) and $\lim_{n\to\infty} \varphi(n)/n = 0$ that

$$\begin{aligned}
& \lim_{n\to\infty} P[c_n(s_0,t_0) \geq c_n(s_0,t)] \\
= & \lim_{n\to\infty} P[-2\log(L_n(s_0,t_0))/n + p(s_0,t_0)\varphi(n)/n \geq -2\log(L_n(s_0,t))/n + p(s_0,t)\varphi(n)/n] \\
= & \lim_{n\to\infty} P[\log(L_n(s_0,t_0))/n - \log(L_n(s_0,t))/n \leq 0.5(p(s_0,t_0) - p(s_0,t))\varphi(n)/n] \\
= & 0
\end{aligned} \tag{A.5}$$

so that

$$\begin{aligned}
& \lim_{n\to\infty} P[\hat{t} < t_0, \hat{s} = s_0] \\
= & \lim_{n\to\infty} P[\hat{t} < t_0 | \hat{s} = s_0] P[\hat{s} = s_0] \\
\leq & \lim_{n\to\infty} P[c_n(s_0,t_0) \geq c_n(s_0,t) \ for\ some\ t < t_0 | \hat{s} = s_0] P[\hat{s} = s_0] \\
\leq & \sum_{t<t_0} \lim_{n\to\infty} P[c_n(s_0,t_0) \geq c_n(s_0,t) | \hat{s} = s_0] P[\hat{s} = s_0] \\
= & 0
\end{aligned} \tag{A.6}$$

For $t > t_0$ it is follows form the assumption that

$$2[\log(L_n(s_0,t)) - \log(L_n(s_0,t_0))] \xrightarrow{d} W \tag{A.7}$$

where $\xrightarrow{d}$ denotes the convergence in distribution. Therefore, by $\lim_{n\to\infty}\varphi(n) = \infty$ and (A.7) implies that

$$plim_{n\to\infty} 2[\log(L_n(s_0,t)) - \log(L_n(s_0,t_0))]/\varphi(n) = 0$$

hence

$$\begin{aligned} & plim_{n\to\infty} n(c_n(s_0,t_0) - c_n(s_0,t))/\varphi(n) \\ = \; & plim_{n\to\infty} 2[\log(L_n(s_0,t)) - \log(L_n(s_0,t_0))]/\varphi(n) + p(s_0,t_0) - p(s_0,t) \\ = \; & p(s_0,t_0) - p(s_0,t) < 0 \end{aligned}$$

so that

$$\lim_{n\to\infty} P[c_n(s_0,t_0) \geq c_n(s_0,t))] = 0$$

This implies that

$$\lim_{n\to\infty} P[\hat{t} > t_0, \hat{s} = s_0] = \lim_{n\to\infty} P[\hat{t} > t_0 | \hat{s} = s_0] P[\hat{s} = s_0] = 0.$$

Thus, we have

$$\lim_{n\to\infty} P[\hat{s} = s_0, \hat{t} = t_0] = 1 \tag{A.8}$$

References

[1] Anraku, K. (1999), "An information criterion for parameters under a simple order restriction." *Biometrika,* 86:141-152.

[2] Barlow, R.E., Bartholomew, D.J., Bremner, J.M. and Brunk, H.D. (1972), "Statistical Inference under Order Restrictions." New York: Wiley.

[3] Barmi H.E. (1996), "Empirical likelihood ratio test for or against a set of inequality constraints" *Journal of Statistical Planning and Inference,* 55:191-204.

[4] Bartholomew, D.J. (1959), "A test of homogeneity for ordered alternatives." *Biometrika,* 46:36-48.

[5] Bohrer, R. and Chow, W. (1978), "Weights for one-sided multivariate inference." *Appl. Statist,* 27:100-104.

[6] Childs, D.R. (1967), "Reduction of the multivariate normal integral to characteristic form." *Biometrika,* 54:293-300.

[7] Konishi, S. and Kitagawa, G. (1996), "Generalised information criteria in model selection" *Biometrika,* 83:875-890.

[8] Kullback, S. and Leibler, R.A. (1951), "On information and sufficiency." *Ann.Math. Statist,* 22:79-86.

[9] LEE, C.I.C., Robertson, T. and Wright, F.T. (1993), "Bounds on distributions arising in order restricted inferences with restricted weights." *Biometrika,* 80:405-416.

[10] Perlman, M. D. (1969), "One-sided problems in multivariate analysis." *Ann.Math. Statist,* 40:549-567.

[11] Qin, J. and J. Lawless (1955), "empirical likelihood and tests of parameters subject to constraints." *Canad. J. Statist,* 23:145-149.

[12] Rao, C.R. (1973), "Linear Statistical Inference and its Applications," 2nd ed. New York: Wiley.

[13] Robertson, T. and Wright, F.T. (1982), "Bounds on mixtures of distributions arising in order restricted inference." *Ann. Statist,* 10:302-306.

[14] Robertson, T. and Wright, F.T. and Dykstra, R.L. (1988), "Order Restricted Statistical Inference." New York: Wiley.

[15] Williams, D.A. (1971), "A test for differences between treatment means when several dose levels are compared with a zero dose control" *Biometrics,* 27:103-117.

[16] Zhao, L. and Peng, L. (2002), "Model selection under order restriction." *Statistics & Probability Letters,* 57:301-306.

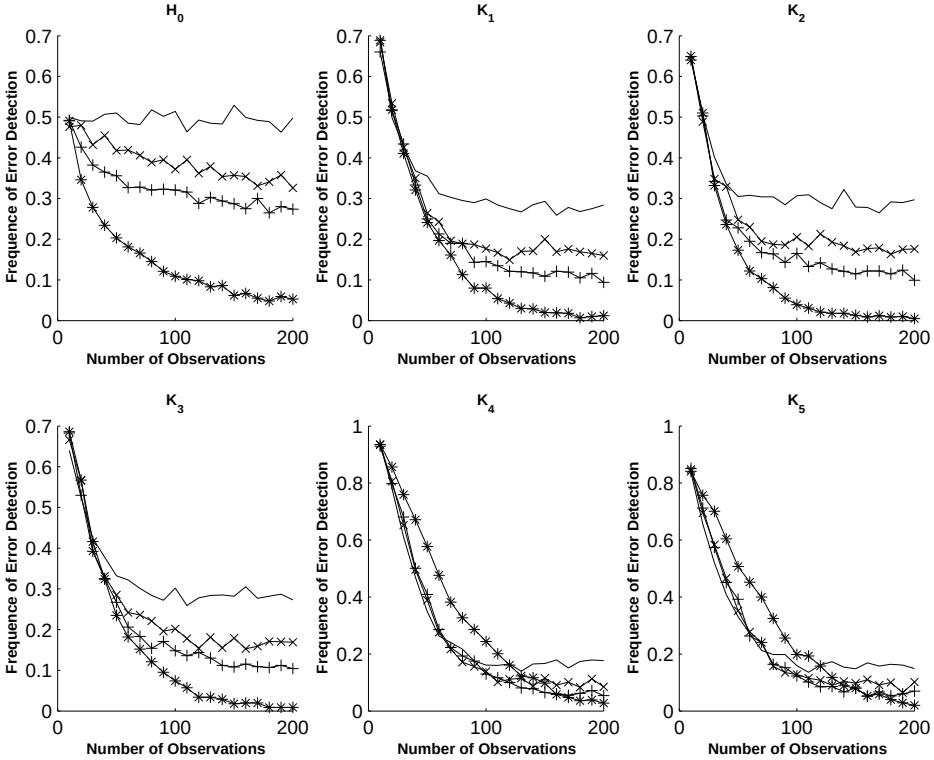

Figure 1: The — curve denote the curve for Anraku(1999)'s order information criterion. Those for our restricted criteria with the penalty numbers (3.3), and $\varphi(n)$ being choose as $(2/\log(\log(40)))\cdot\log(\log(n))$, $(2/\log(40))\cdot\log(n)$, $(2/\sqrt{40})\cdot\sqrt{n}$ are showed by —x, —+, —* curve, respectively.

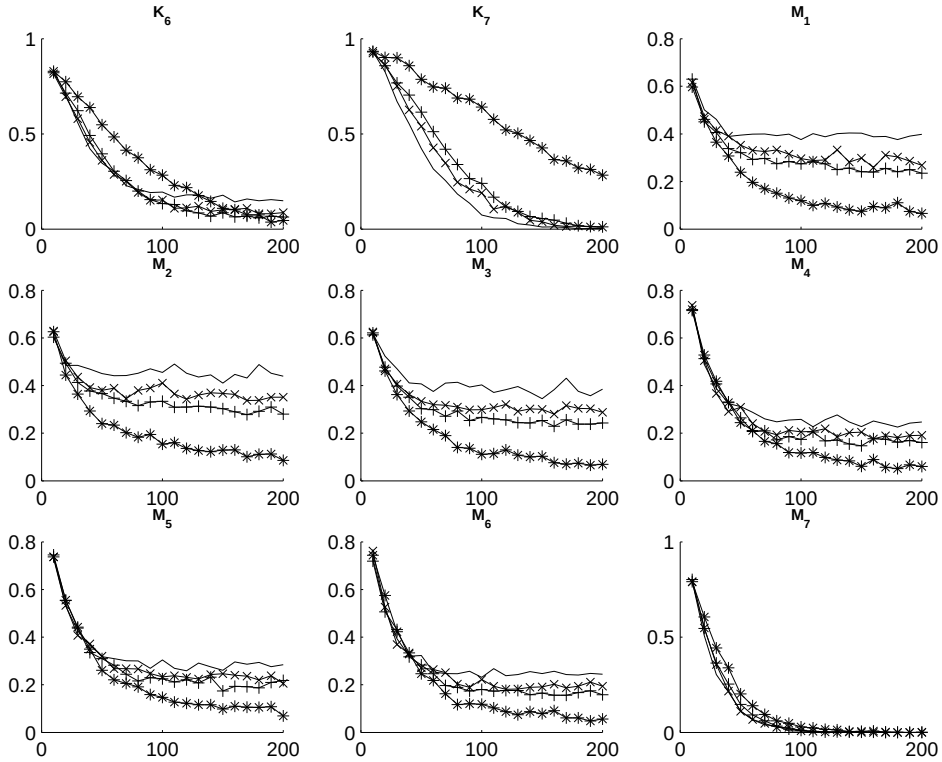

Figure 2: The — curve denote the curve for Anraku(1999)'s order information criterion.Those for our restricted criteria with the penalty numbers (3.3), and $\varphi(n)$ being choose as $(2/\log(\log(40)))\cdot\log(\log(n))$, $(2/\log(40))\cdot\log(n)$, $(2/\sqrt{40})\cdot\sqrt{n}$ are showed by —x, —+, —* curve, respectively.

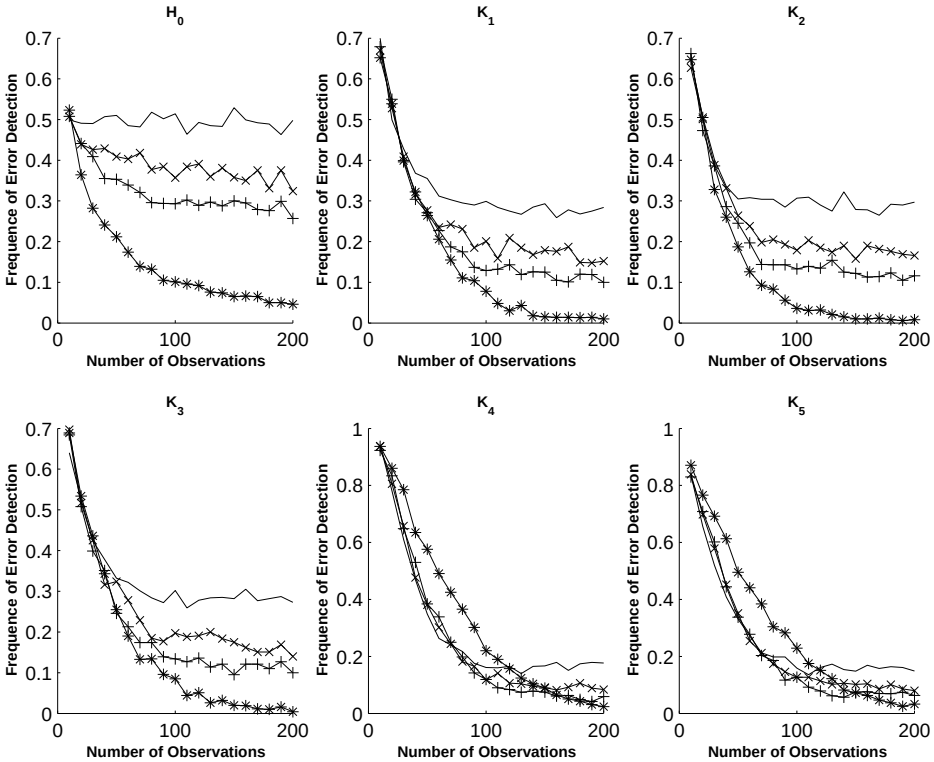

Figure 3: The — curve denote the curve for Anraku(1999)'s order information criterion. Those for our restricted criteria with the penalty numbers (3.5), and $\varphi(n)$ being choose as $(2/\log(\log(40)))\cdot\log(\log(n))$, $(2/\log(40))\cdot\log(n)$, $(2/\sqrt{40})\cdot\sqrt{n}$ are showed by —x, —+, —* curve, respectively.

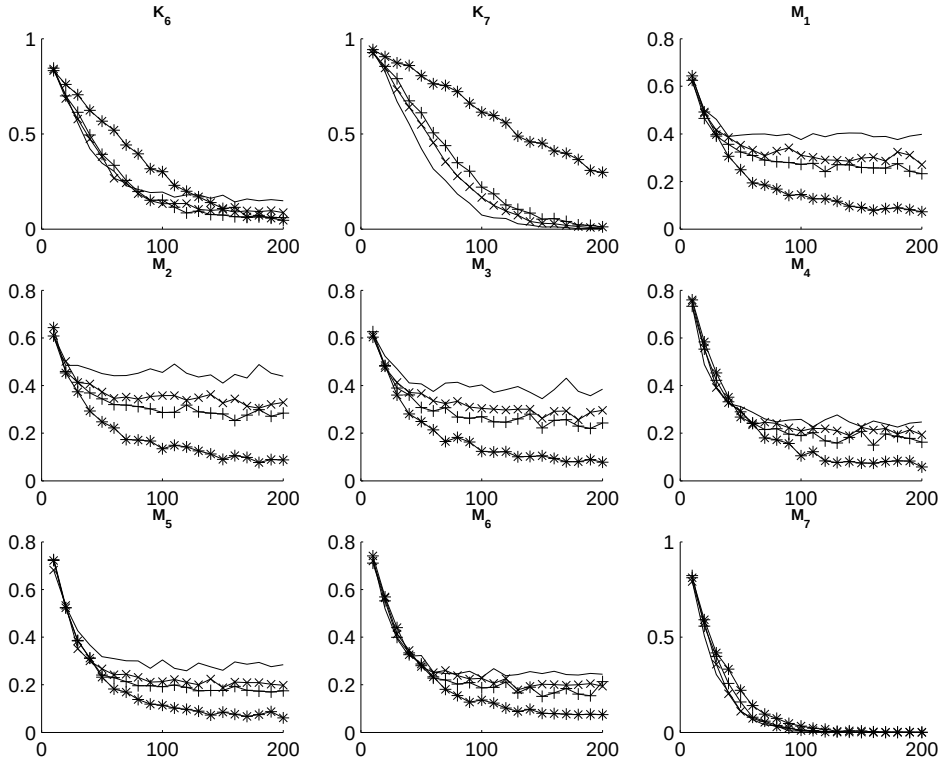

Figure 4: The — curve denote the curve for Anraku(1999)'s order information criterion. Those for our restricted criteria with the penalty numbers (3.5), and $\varphi(n)$ being choose as $(2/\log(\log(40)))\cdot\log(\log(n))$, $(2/\log(40))\cdot\log(n)$, $(2/\sqrt{40})\cdot\sqrt{n}$ are showed by —x, —+, —* curve, respectively.

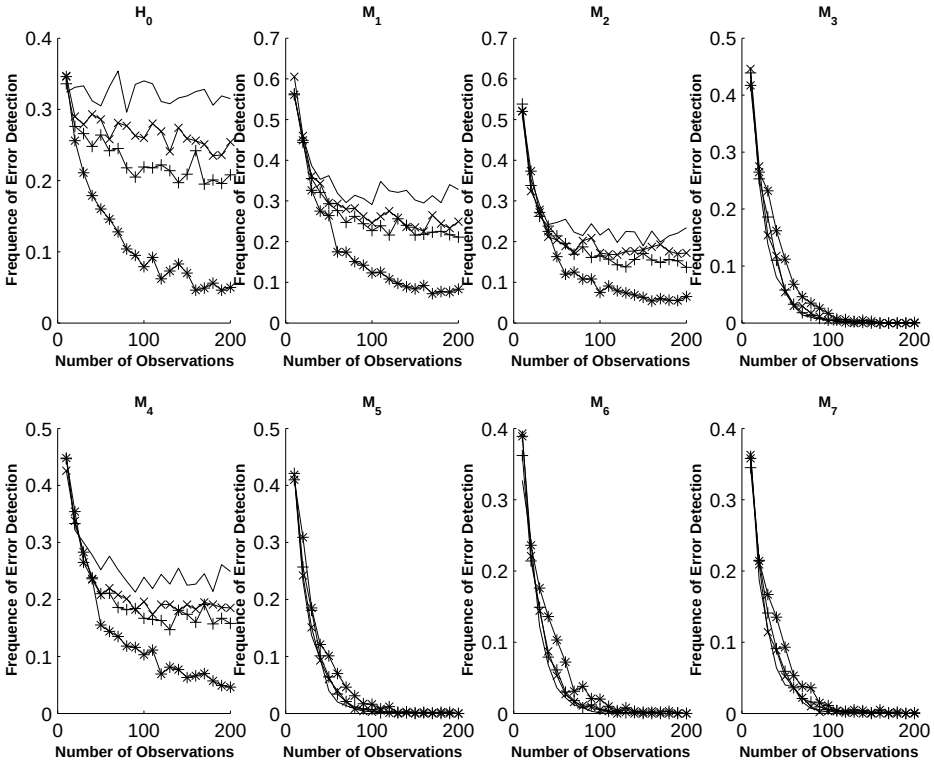

Figure 5: The — curve denote the curve for Anraku(1999)'s order information criterion. Those for our restricted criteria with the penalty numbers (3.3), and $\varphi(n)$ being choose as $(2/\log(\log(40)))\cdot\log(\log(n))$, $(2/\log(40))\cdot\log(n)$, $(2/\sqrt{40})\cdot\sqrt{n}$ are showed by —x, —+, —* curve, respectively.

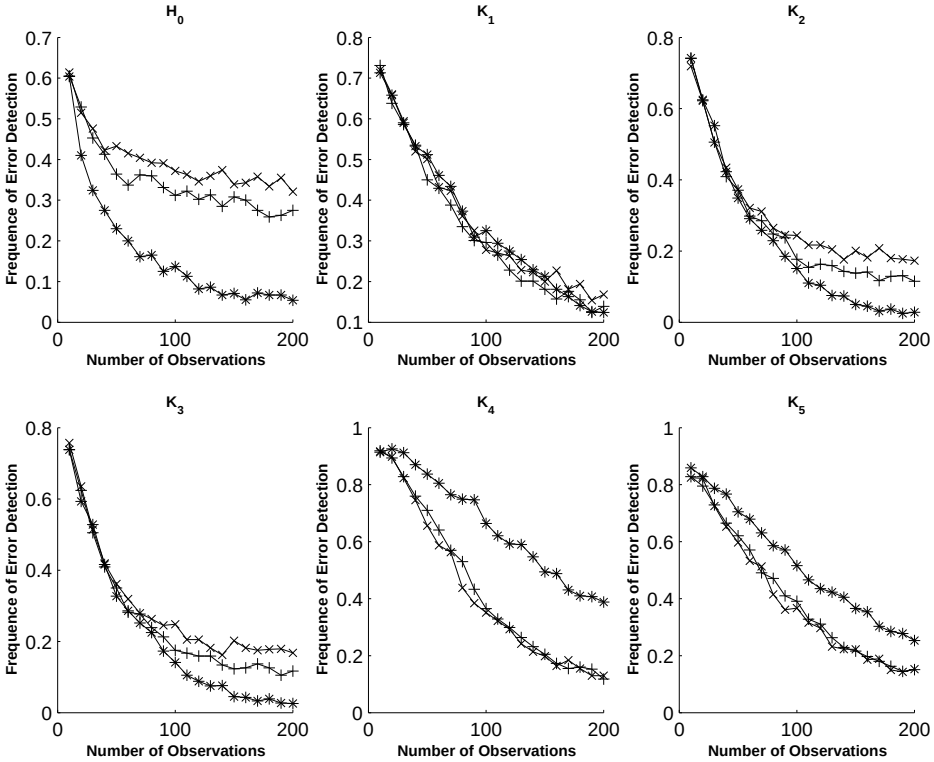

Figure 6: Those for our empirical-likelihood-based information criteria (4.4), and $\varphi(n)$ being choose as $(2/\log(\log(40)))\cdot\log(\log(n))$, $(2/\log(40))\cdot\log(n)$, $(2/\sqrt{40})\cdot\sqrt{n}$ are showed by —x, —+, —* curve, respectively.

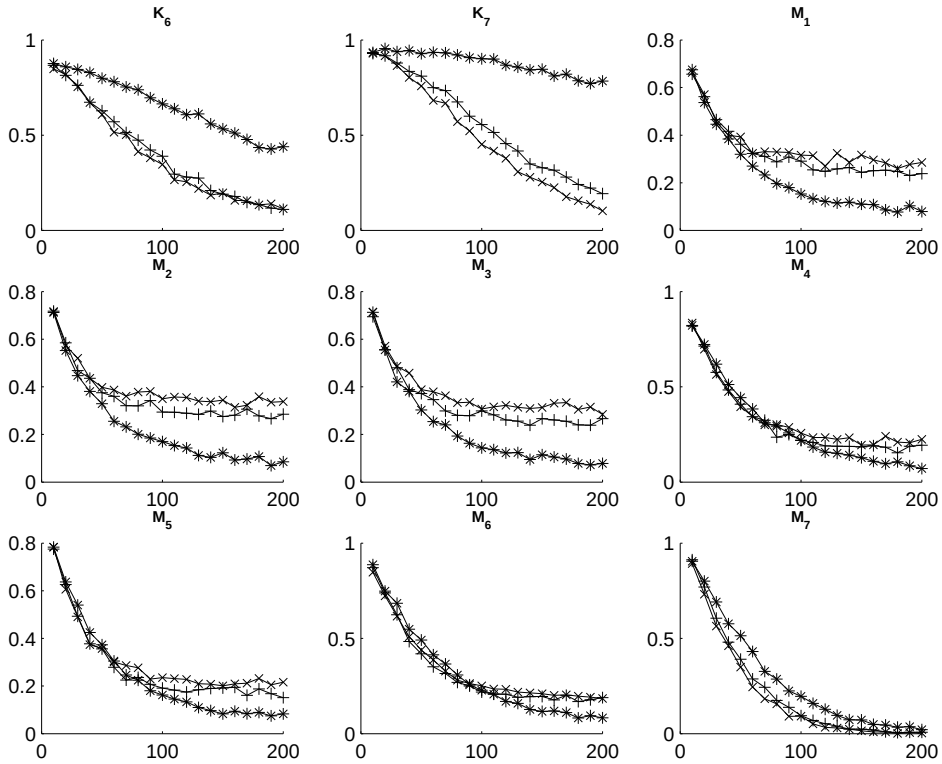

Figure 7: Those for our empirical-likelihood-based information criteria (4.4), and $\varphi(n)$ being choose as $(2/\log(\log(40)))\cdot\log(\log(n))$, $(2/\log(40))\cdot\log(n)$, $(2/\sqrt{40})\cdot\sqrt{n}$ are showed by —x, —+, —* curve, respectively.