行政院國家科學委員會專題研究計畫成果報告

****
*   *
*   一些模式選擇方法在選擇兩個非巢式常態線性模式時的界限圖形   *
*   *
****

計畫類別：個別型計畫

計畫編號：NSC 89-2118-M-041-002

執行期限：88年8月1日至89年7月31日

主持人　：陳青浩

執行單位：嘉南藥理科技大學 醫務管理系

中華民國89年10月15日

行政院國家科學委員會專題研究計畫成果報告

一些模式選擇方法在選擇兩個非巢式常態線性模式時的界限圖形

計畫編號：NSC 89-2118-M-041-002

執行期限：88 年 8 月 1 日至 89 年 7 月 31 日

主持人：陳青浩　嘉南藥理科技大學 醫務管理系

一、中文摘要

關鍵詞：非巢式， 正交分解， 典形分析。

本計劃探討一些常見的模式選擇方法在選擇兩個候選的非巢式常態線性模式時，根據各自的理論探討觀察值向量落在去除重疊空間部份後之聯合模式參數空間上時選擇模式的界限範圍。在本計劃中採用 Efron (1984) 所提出的一種特殊正交對稱的座標軸系方式，並藉由典形分析(canonical analysis) 的技巧，將此參數空間予以分割成許多正交之平面及直線後引用這種特殊正交對稱的座標軸系來定位後再分別探討各選擇方法選擇模式時的界限範圍。

Abstract

In this research, I applied some commonly used model selection methods to select among two candidate nonnested normal linear models, after removed the common parameter space, the selection boundaries based on the position of the observation vector located in the combined parameter space for each different theorems were explored. The special symmetric coordinate system suggested in Efron(1984) was used in this research, by using the technique of canonical analysis, the combined parameter space was decomposed into several orthogonal planes and lines , and the selecting boundaries for different model selection methods were discussed.

Keywords： nonnested , orthogonal decomposition, canonical analysis.

二、緣由與目的

最早關於非巢式常態線性模式選擇方法的研究與討論是由 Hotelling (1940) 所提出當兩個模式僅各有一個變數不同時檢定兩個模式一樣適當的證明， 他建議移去兩個模式共同之變數後， 在簡化後的二維聯合模式參數平面(combined model parameter space)上， 如果樣本觀察值向量位於兩個模式參數向量之間且距離相等則表示此兩個模式一樣適當。

Efron (1984)則延續 Hotelling 的觀念，探討樣本觀察值向量與兩個模式參數空間之距離， 進而發展出一種非巢式常態線性模式的選擇方法。 他並舉例應用該方法在簡化後的二維聯合模式參數平面上畫出選擇模式的界限範圍。

Fraser 和 Gebotys (1987) 也曾引用類似的觀念在簡化後的聯合模式參數空間上探討樣本觀察值向量與兩個模式參數空間之間的角度，以角度來作為模式選擇的判斷。

本計劃挑選近年來一些常見的方法或法則，在兩個候選的非巢式常態線性模式中，經由觀察值向量落在兩個模式參數空間之位置來決定選擇的模式，透過選擇模式界限範圍的圖形顯示， 對這些方法或法則的選擇傾向及機率有一個基本的概念，

並提供對於這些不同的選擇方法或法則彼此之間一個比較的途徑。當兩個候選的模式具相同的維數時， 以上的圖示方式是很容易完成的；但當此參數空間為三維時，由於各種理論或多或少有選擇最簡模式的傾向，也各自對選擇具較高維數參數空間的模式有不同的懲罰權重， 所以選擇模式時的界限也就各有不同。

三、研究結果

Let the true normal linear model be given by

$$\mathbf{Y} = \boldsymbol{\theta} + \boldsymbol{\varepsilon}, \quad \boldsymbol{\varepsilon} \sim \mathbf{N}\left(0, \sigma^2 \mathbf{I}\right),$$

and let the two competing nonnested normal linear models be referred to as:

Model A ： $\mathbf{Y} = \theta_A + \varepsilon, \ \theta_A \in \Theta_A$.

and

Model B ： $\mathbf{Y} = \theta_B + \varepsilon, \ \theta_B \in \Theta_B$.

$\Theta_A \cap \Theta_B \neq \Theta_A$ and $\Theta_A \cap \Theta_B \neq \Theta_B$. Since Θ_A and Θ_B may overlap, After using orthogonal decomposition to remove the overlap subspace, model A becomes:

$$Y_{A+B} = \eta_A + \varepsilon, \ \eta_A \in L_A;$$

model B becomes :

$$Y_{A+B} = \eta_B + \varepsilon, \ \eta_B \in L_B,$$

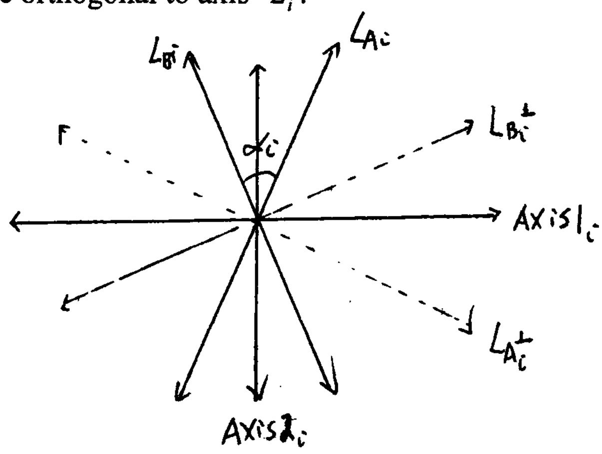
here L_A and L_B represent the model parameter spaces after the overlap subspace has been removed from model A and model B respectively, L_A and L_B only intersects at the origin point, i.e., $L_A \cap L_B = \{0\}$; y_{A+B} represents the projection of the sample observation vector $\mathbf{y}$ onto the combined parameter space $L_A \oplus L_B$.

Let $\mathbf{y}_{A+B} = \mathbf{y}_{A^\perp} + \mathbf{y}_A = \mathbf{y}_{B^\perp} + \mathbf{y}_B$

And

$$\eta = \eta_{A^\perp} + \eta_A = \eta_{B^\perp} + \eta_B$$

is the orthogonal decompositions where $\mathbf{y}_K$ and η_K denote the projections of $\mathbf{y}_{A+B}$ and η, respectively, onto L_K, for $K \in \{A, B, A^\perp, B^\perp\}$. The p_E dimensional orthogonal complement of $\Theta_A \oplus \Theta_B$ in R^n, denoted by Θ_E, is the portion of the error space common to both models. The projection of $\mathbf{y}$ onto Θ_E will be denoted by $\mathbf{y}_E$. The symmetric coordinate system for $L_A \oplus L_B$, suggested in Efron (1984), is adopted below. Let d_A and d_B represent the dimensions of L_A and L_B respectively, assume that the models are labeled so that $d_A \leq d_B$. Decompose L_A into d_A orthogonal one dimensional spaces, L_{A1}, L_{A2}, …, L_{Ad_A} and a similar decomposition of L_B into d_B orthogonal one dimensional spaces, L_{B1}, L_{B2}, …, L_{Bd_B}. Note that L_{Ai} is orthogonal to L_{Bj} for $i \neq j$ and, L_{Ai} and L_{Bi} is the ith pair of the canonical variables, and α_i denotes the smaller angle between L_{Ai} and L_{Bi}. The symmetric coordinate system will be defined in stages sarting with the $L_{Ai} \oplus L_{Bi}$ planes. Within such a plane, the axes of the symmetric coordinate system will be called axis 1_i and axis 2_i. Axis 2_i is the line which bisects the angle α_i between L_{Ai} and L_{Bi}, while axis 1_i is chosen to be orthogonal to axis 2_i.



$L_{Ai} \oplus L_{Bi}$ plane

Assume that the axes are labeled so that the positive direction on L_{Ai} is the first quadrant of the symmetric coordinate. Let the coordinates $(y_{Ai}, y_{Ai^\perp})$ and $(y_{Bi}, y_{Bi^\perp})$ denote the projection $\mathbf{y}_{Ai+Bi} = P(\mathbf{y}_{A+B} : L_{Ai} \oplus L_{Bi})$ with respect to to $L_{Ai} \ominus L_{Ai^\perp}$ and $L_{Bi} \ominus L_{Bi^\perp}$, respectively, and the symmetric coordinates (y_{1i}, y_{2i}) of $\mathbf{y}_{Ai+Bi}$ satisfy the linear relationships

$$\begin{pmatrix} y_{Ai} \\ y_{Ai^\perp} \end{pmatrix} = \begin{pmatrix} \sin\frac{\alpha_i}{2} & \cos\frac{\alpha_i}{2} \\ -\cos\frac{\alpha_i}{2} & \sin\frac{\alpha_i}{2} \end{pmatrix} \begin{pmatrix} y_{1i} \\ y_{2i} \end{pmatrix} \quad \text{............(1)}$$

and

$$\begin{pmatrix} y_{Bi} \\ y_{Bi^\perp} \end{pmatrix} = \begin{pmatrix} -\sin\frac{\alpha_i}{2} & \cos\frac{\alpha_i}{2} \\ \cos\frac{\alpha_i}{2} & \sin\frac{\alpha_i}{2} \end{pmatrix} \begin{pmatrix} y_{1i} \\ y_{2i} \end{pmatrix} \quad \text{..............(2)}$$

The symmetric coordinates $\eta_{11}, \cdots, \eta_{1dA}, \eta_{21}, \cdots, \eta_{2dB}$ of η are defined in the same fashion.

To select among two candidate models, the ratio of the squared lengths of the projection of $\mathbf{y}$ onto the space corresponding to the violation of the two models are computed, the selection rule can be written as :

select model A if

$$\|\mathbf{y}_{A^\perp}\|^2 - c_1\|\mathbf{y}_{B^\perp}\|^2 + c_2\|\mathbf{y}_E\|^2 < 0, \text{............(3)}$$

where c_1 and c_2 are suitable constants.

Twelve commonly used model selection methods : the maximum likelihood method (ML), the mean square error approach (MSE), Mallow's (1973) C_p criterion, Efron's (1984) $\hat{\delta}$ criterion, Akaike's (1973) AIC criterion, Schwarz's (1978) and Leonard's(1982) BIC criterion, Hannan's and Quinn's (1979) HQIC criterion, Hurvich's and Tsai's (1991) CAIC criterion, Aitkin's (1991) POBJ and POBN criteria, Shibata's (1980) SIC criterion and Laud's and Ibrahim's (1995) L_m criterion were applied and their appropriate values of $\log c_1$ and c_2 are provided in the following table.

Table 1

Criterion	$\log c_1$	c_2
ML	0	0
MSE	$\log\left(\frac{n-p_A}{n-p_B}\right)$	$1-c_1$
POBN	$\frac{p_B-p_A}{n}\log 2$	$1-c_1$
POBJ	$\frac{2}{n}\log\left(\frac{\Gamma\left(\frac{2n-p_A}{2}\right)\Gamma\left(\frac{n-p_B}{2}\right)}{\Gamma\left(\frac{2n-p_B}{2}\right)\Gamma\left(\frac{n-p_A}{2}\right)}\right)$	$1-c_1$
AIC	$\frac{2(p_B-p_A)}{n}$	$1-c_1$
CAIC	$\frac{2(p_B-p_A)(n-1)}{(n-p_A-2)(n-p_B-2)}$	$1-c_1$
BIC	$\frac{(p_B-p_A)}{n}\log n$	$1-c_1$
SIC	$2\log\left(\frac{n+2p_B+2}{n+2p_A+2}\right)$	$1-c_1$
HQIC	$\frac{2(p_B-p_A)}{n}\log\log n$	$1-c_1$
C_P	0	$\frac{2(p_A-p_B)}{p_E}$
$\hat{\delta}$	0	$\frac{2(p_A-p_B)}{p_E-2}$
L_m	$\log\left(\frac{n-p_A-2}{n-p_B-2}\right)$	0

The above definitions and the relationships of equations (1) and (2) yield

$$\|\mathbf{y}_{A^\perp}\|^2 = \sum_{i=1}^{d_A}\left(-\cos\frac{\alpha_i}{2}y_{1i} + \sin\frac{\alpha_i}{2}y_{2i}\right)^2 + \sum_{i=d_A+1}^{d_B} y_{2i}^2 \text{......(4)}$$

$$\|\mathbf{y}_{B^\perp}\|^2 = \sum_{i=1}^{d_A}\left(\cos\frac{\alpha_i}{2}y_{1i} + \sin\frac{\alpha_i}{2}y_{2i}\right)^2 \text{...............(5)}$$

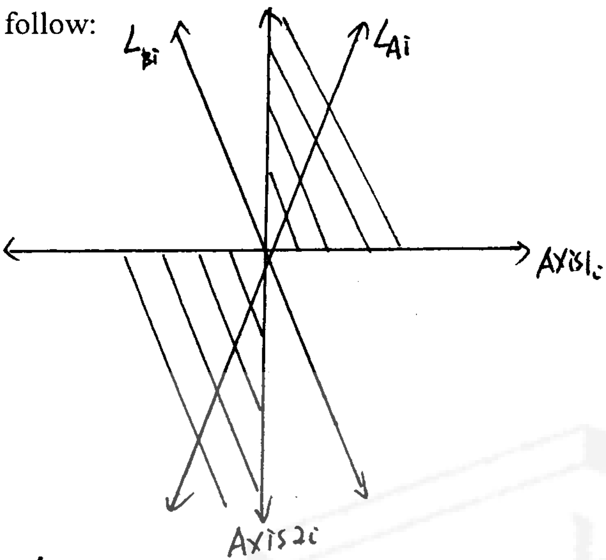
When $d_A = d_B$, all of the selection rules are equivalent, their values of c_1=1 and c_2=0 in table 1, the selection rule of equation (3) becomes select model A if

$$\|\mathbf{y}_{A^\perp}\|^2 - \|\mathbf{y}_{B^\perp}\|^2 < 0. \text{.........................(6)}$$

By the equations (4) and (5), equation (6) is equivalent to

$$\sum_{i=1}^{d_A} 2\sin\alpha_i y_{1i} y_{2i} > 0. \text{.......................(7)}$$

This implies for $i \ni \{1, \cdots, d_A\}$, on each $L_{Ai} \oplus L_{Bi}$ plane, the selection boundaries for all selection rules are the same, if $\mathbf{y}_{Ai+Bi}$ locates on the first or third quadrants on the plane, then select model A, they are as follow:



$L_{Ai} \oplus L_{Bi}$ plane

Consider the case when $d_A = 1$ and $d_B = 2$, since $c_1 \geq 1$, $c_2 \leq 0$ and $\|\mathbf{y}_E\|^2 > 0$, let y_1, y_2 and y_3 denote the coordinates of $\mathbf{y}_{A+B}$ with respect to the coordinate system $\{axis1_i, axis2_i, L_{B2}\}$ on $L_A \oplus L_B$, in terms of the symmetric coordinate system, the selection rule is

$$y_3^2 + (y_1, y_2)\mathbf{Q}\begin{pmatrix} y_1 \\ y_2 \end{pmatrix} < -c_2\|\mathbf{y}_E\|^2 \quad \ldots\ldots\ldots\ldots(8)$$

where

$$\mathbf{Q} = \begin{pmatrix} (1-c_1)\cos^2\frac{\alpha_1}{2} & -(1-c_1)\sin\frac{\alpha_1}{2}\cos\frac{\alpha_1}{2} \\ -(1-c_1)\sin\frac{\alpha_1}{2}\cos\frac{\alpha_1}{2} & (1-c_1)\sin^2\frac{\alpha_1}{2} \end{pmatrix} \ldots.(9)$$

and α_1 is the angle between L_{A1} and L_{B1}. The eigenvector-eigenvalue pairs for $\mathbf{Q}$ are

$$\mathbf{d}_1 = \begin{pmatrix} \cos\beta \\ \sin\beta \end{pmatrix}, \lambda_1 = \frac{1}{2}[(1-c_1)(1+\cos\alpha_1) - (1+c_1)\sin\alpha_1\tan\beta]$$

and

$$\mathbf{d}_2 = \begin{pmatrix} -\sin\beta \\ \cos\beta \end{pmatrix}, \lambda_2 = \frac{1}{2}[(1-c_1)(1-\cos\alpha_1) + (1+c_1)\sin\alpha_1\tan\beta]$$

where $(c_1+1)\cot(2\beta) = (c_1-1)\cot\alpha_1$ and $\lambda_1 < 0 < \lambda_2$.

The following table shows the value of β for each criterion.

Table 2

Criterion	β
ML, C_P, $\hat{\delta}$	$\pi/4$
MSE	$\frac{1}{2}\cot^{-1}\left(\left(\frac{p_B - p_A}{2n - p_A - p_B}\right)\cot\alpha_1\right)$
POBN	$\frac{1}{2}\cot^{-1}\left(\left(\frac{2^{\frac{(p_B-p_A)}{n}} - 1}{2^{\frac{(p_B-p_A)}{n}} + 1}\right)\cot\alpha_1\right)$
POBJ	$\frac{1}{2}\cot^{-1}\left(\frac{\left(\frac{\Gamma\left(\frac{2n-p_A}{2}\right)\Gamma\left(\frac{n-p_B}{2}\right)}{\Gamma\left(\frac{2n-p_B}{2}\right)\Gamma\left(\frac{n-p_A}{2}\right)}\right)^{\frac{2}{n}} - 1}{\left(\frac{\Gamma\left(\frac{2n-p_A}{2}\right)\Gamma\left(\frac{n-p_B}{2}\right)}{\Gamma\left(\frac{2n-p_B}{2}\right)\Gamma\left(\frac{n-p_A}{2}\right)}\right)^{\frac{2}{n}} + 1}\cot\alpha_1\right)$
AIC	$\frac{1}{2}\cot^{-1}\left(\left(\frac{e^{\frac{2(p_B-p_A)}{n}} - 1}{e^{\frac{2(p_B-p_A)}{n}} + 1}\right)\cot\alpha_1\right)$
CAIC	$\frac{1}{2}\cot^{-1}\left(\left(\frac{e^{\frac{2(p_B-p_A)(n-1)}{(n-p_A-2)(n-p_B-2)}} - 1}{e^{\frac{2(p_B-p_A)(n-1)}{(n-p_A-2)(n-p_B-2)}} + 1}\right)\cot\alpha_1\right)$
BIC	$\frac{1}{2}\cot^{-1}\left(\left(\frac{n^{\frac{(p_B-p_A)}{n}} - 1}{n^{\frac{(p_B-p_A)}{n}} + 1}\right)\cot\alpha_1\right)$
SIC	$\frac{1}{2}\cot^{-1}\left(\left(\frac{\left(\frac{n+2p_B+2}{n+2p_A+2}\right)^2 - 1}{\left(\frac{n+2p_B+2}{n+2p_A+2}\right)^2 + 1}\right)\cot\alpha_1\right)$
HQIC	$\frac{1}{2}\cot^{-1}\left(\left(\frac{(\log n)^{\frac{2(p_B-p_A)}{n}} - 1}{(\log n)^{\frac{2(p_B-p_A)}{n}} + 1}\right)\cot\alpha_1\right)$
L_m	$\frac{1}{2}\cot^{-1}\left(\left(\frac{p_B - p_A}{2n - p_A - p_B - 4}\right)\cot\alpha_1\right)$

Let $\{axis1_{new}, axis2_{new}\}$ denote the coordinate system obtained by rotating the symmetric coordinates $\{axis1_1, axis2_1\}$ for $L_{A1} \oplus L_{B1}$ through the angle β. The coordinates of $\mathbf{y}_{A_1+B_1}$ with respect to this rotated coordinate system are

$$t_1 = y_2 \sin\beta + y_1 \cos\beta \quad \text{.....................(10)}$$

and

$$t_2 = y_2 \cos\beta - y_1 \sin\beta \quad \text{.....................(11)}$$

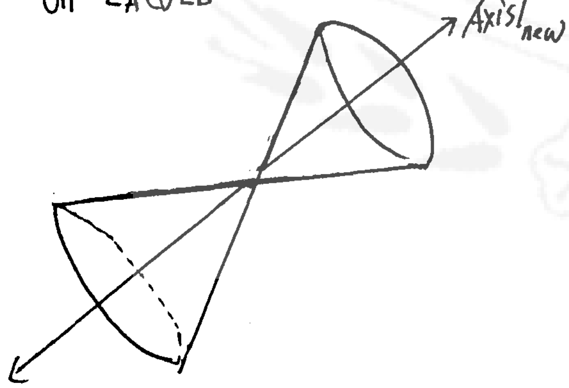
In terms of these coordinates, the selection rules of equation(8) can be writen as

$$y_3^2 + \lambda_1 t_1^2 + \lambda_2 t_2^2 < -c_2 \|\mathbf{y}_E\|^2 \quad \text{...............(12)}$$

When $c_2 = 0$, the quadratic surface of the selection boundary is an elliptic cone. The elliptical cross sections of this cone lie on planes parallel to the plane determined by $axis2_{new}$ and L_{B2}, and the centers of these ellpses lie on $axis1_{new}$, their graphs are as follow:

Elliptic cone (When $C_2=0$)

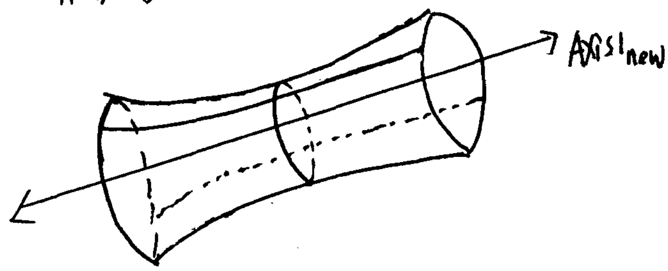
on $L_A \oplus L_B$ 3 dimensional space



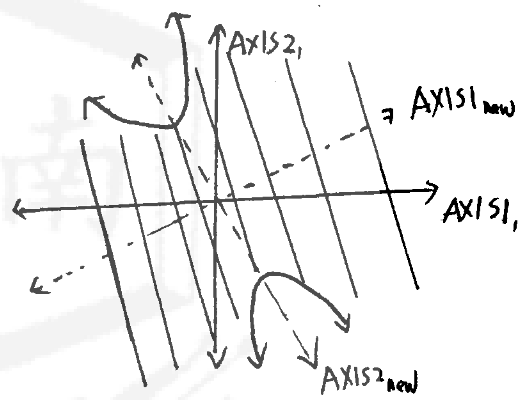
If $c_2 < 0$, the selection boundary is a hyperbolid of one sheet, on the 3 dimensional space, the new axes $\{axis1_{new}, axis2_{new}\}$ of these elliptic hyperbolid of one sheet which is asymptotic to the elliptic cone described above. In both cases, the selection region includes $axis1_{new}$.

hyperboloid of one sheet (When $C_2<0$)

on $L_A \oplus L_B$ 3 dimensional space



The selection boundary on $L_{A1} \oplus L_{B1}$ plane



四、參考文獻

[1] Aitkin , M. (1991). Posterior Bayes factors. *J. Roy. Statist. Soc. Ser. B,* **53**, 111-142.

[2] Akaike , H. (1973). Information theory and an extension of the maximum likelihood principle. In *Proc. of the Second International Symposium on Information Theory,* B.N. Petrov and F. Czaki, eds. Akademiai Kiado, Budapest, 267-281.

[3] Akaike , H. (1973). A Bayesian analysis of the minimum AIC procedure . *Ann. Inst. Statist. Math. A,* **30**, 9-14.

[4] Efron,B. (1984). Comparing nonested linear models. *J. Amer. Statist. Assoc.,* **79**, 791-803.

[5] Fraser D.A.S. and R.J. Gebotys (1987). Nonnested linear models: a conditional confidence approach. *Canad. J. Statist.*, **15**, 375-386.

[6] Hannan, E.J. and B.G. Quinn. (1979). The determination of the order of an autoregression. *J. Roy. Statist. Soc. Ser. B*, **41**, 190-195

[7] Hotelling, H. (1940). The selection of variates for use in prediction with some comments on the general problem on nuisance parameters. *Ann. Math. Statist.*, **11**, 271-283.

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

[9] Hurvich, C.M. and C.L.Tsai (1991). Bias of the corrected AIC criterion for underfitted regression and time series models. *Biometrika*, **78**, 499-509.

[10] Laud, P.W. and J.G. Ibrahim (1995). Predictive model selection. *J. Roy. Statist. Soc. Ser. B*, **57**, 247-262.

[11] Leonard, T. (1982). Comment on the paper by M. Lejeune and G.D. Faulkenberry. *J. Amer. Statist. Assoc.*, **77**, 657-658.

[12] Linhart, H. and W. Zucchini (1986). *Model Selection.* Wiley, New York.

[13] Mallows, C.L. (1973). Some comments on C_p. *Technometrics*, **15**, 661-675.

[14] Schwarz, G. (1978). Estimating the dimension of a model. *Ann. Math. Statist.*, **6**, 461-464.

[15] Shibata, R. (1980). Asymptotically efficient selection of the order of the model for estimating paramters of a linear process. *Ann. Statist.*, **8**, 147-164.