Chin. Ann. of Math.
6B (1) 1985

ESTIMATION OF THE ORDER OF ARMA MODEL BY LINEAR PROCEDURES

WANG SHOUREN (王寿仁)* CHEN ZHAOGUO (陈兆国)*

Abstract

Given an ARMA (p_0, q_0) sequence $x(t)$, a linear procedure is used to estimate (p_0, q_0) and the strong consistency of the estimate is proved.

First fit the data of size T by a long $AR(P(T))$ model to obtain an estimate of $\varepsilon(t)$, denoting it by $\hat{\varepsilon}(t)$, where $P(T)\cong(\ln T)^{1+\delta}$, $\delta>0$. Then for each p, we minimize $\frac{1}{T}\sum_{t=1}^{T}\Big(\sum_{j=0}^{p} a_{pj}x(t-j)-\sum_{j=1}^{p}b_{pj}\hat{\varepsilon}(t-j)\Big)^2$. Let $\hat{a}_{pj}$, $\hat{b}_{pj}$ be the minimizing coefficients which can be obtained by Whittle's recursive procedure. As p increases to some s, consistently greater than $r_0=\max(p_0, q_0)$, we take the second estimate of $\varepsilon(t)$ as

$$\tilde{\varepsilon}(t)=\sum_{j=0}^{s}\hat{a}_{sj}x(t-j)-\sum_{j=1}^{s}\hat{b}_{sj}\tilde{\varepsilon}(t-j).$$

For every (p, q), put

$$\tilde{\sigma}^2_{pq}=\min_{a_{pj}, b_{qj}}\frac{1}{T}\sum_{t=1}^{T}\Big(\sum_{0}^{p}a_{pj}x(t-j)-\sum_{1}^{q}b_{qj}\tilde{\varepsilon}(t-j)\Big)^2,$$

then the minimization of

$$\mathrm{BIC}(p, q)=\ln\tilde{\sigma}^2_{pq}+(p+q)\frac{\ln T}{T}$$

will offer a strongly consistent estimate of (p_0, q_0).

§ 1. Introduction

Let $X(t)$ be an ergodic stationary time series generated by an ARMA(p_0, q_0) model

$$\sum_{j=0}^{p_0}a_{0j}x(t-j)=\sum_{j=0}^{q_0}b_{0j}\varepsilon(t-j), \tag{1.1}$$

where $a_{00}=b_{00}=1$, $a_{0p_0}\neq 0$, $b_{0q_0}\neq 0$,

$$E\varepsilon(t)=0,\ E\varepsilon(s)\varepsilon(t)=\sigma^2\delta_{s,t},\ E\varepsilon(t)^4<\infty. \tag{1.2}$$

Put

$$a_0(z)=\sum_{j=0}^{p_0}a_{0j}z^j,\quad b_0(z)=\sum_{j=0}^{q_0}b_{0j}z^j.$$

We assume that $a(z)$ and $b(z)$ are prime and

$$a_0(z)\neq 0,\ b_0(z)\neq 0,\ |z|\leqslant 1. \tag{1.3}$$

In order to estimate the order (p_0, q_0) and the coefficients a_{0j}, b_{0j} from an

Manuscript received November 15, 1982.

* Institute of Applied Mathematics, Academia Sinica, Bejing, China.

observations $x_1, x_2, \cdots, x_T$, it is usually to fit the data by $ARMA\ (p, q)$ model for each (p, q) and calculate the maximum likelihood estimate $\tilde{\sigma}^2_{pq}$ of σ^2. The criteria used to get the best estimate of (p_0, q_0) is choosing $(\tilde{p}, \tilde{q})$ which minimizes

$$\ln \tilde{\sigma}^2_{pq} + (p+q)A(T)/T, \tag{1.4}$$

where $A(T) \to \infty$, $A(T)/T \to 0$ as $T \to \infty$. Under certain conditions, Hannan[1] has shown that $(\tilde{p}, \tilde{q}) \to (p_0, q_0)$ a. s. as $T \to \infty$, if

$$\liminf_{T\to\infty} \frac{A(T)}{2 \ln \ln T} > 0, \quad \frac{A(T)}{T} \to 0. \tag{1.5}$$

One particular function $A(T)$ which deserves special attention is $A(T) = \ln T$.

To compute the ML estimate $\tilde{\sigma}^2_{pq}$ for each (p, q), non-linear procedure is used. Recently, Hannan and Rissanen[2] proposed a way to avoid all non-linear procedures in the estimation of the order. They used a long $AR(P(T))$, with $P(T) = (\ln T)^{1+\delta}$, to fit the data, and used this $AR(P(T))$ to get an estimate of $s(t)$, denoted by $\hat{s}(t)$. Suppose

$$\hat{\sigma}^2_{pq} = \min_{a_{pj}, b_{qj}} \frac{1}{T} \sum_{t=1}^{T} \left(\sum_{j=0}^{p} a_{pj} x(t-j) - \sum_{j=1}^{q} b_{qj} \hat{s}(t-j) \right)^2,$$

then the estimate of (p_0, q_0) is $(\hat{p}, \hat{q})$ which minimizes

$$BIC(p, q) = \ln \hat{\sigma}^2_{pq} + (p+q) \ln T/T.$$

But as it was pointed out by the authors in the correction which was published in the same journal lately, that $\ln T$ has to be altered to $(\ln T)^{1+\delta}$, $\delta > 0$, if one wants theoretical result $(\hat{p}, \hat{q}) \to (p_0, q_0)$ a. s. as in Theorem 3 in [2]. Although $\delta = 0$ and $\delta > 0$ (we can take δ as small as we like) make no difference in practice, one might prefer not to change it, because BIC is such a typical criterion. We are going to deal with this purpose.

In this paper, we suggest a procedure of two steps estimation of $s(t)$. First we fit the data with $AR(P(T))$ as it was done in [2]. The autoregressive coefficients can be estimated from Yule-Walker equation

$$(\hat{a}_{T1} \cdots \hat{a}_{TP(T)}) \begin{bmatrix} C_x(0) & \cdots\cdots & C_x(P(T)-1) \\ \vdots & \ddots & \vdots \\ C_x(P(T)-1) & \cdots\cdots & C_x(0) \end{bmatrix} = -(C_x(1) \cdots C_x(P(T))), \tag{1.6}$$

where $C_x(l) = \frac{1}{T} \sum_{1}^{T} \hat{x}(t) \hat{x}(t+l)$, $\hat{x}(t) = x(t)$, $1 \leqslant t \leqslant T$; $\hat{x}(t) = 0$, otherwise. Then the first estimate of $s(t)$ is given by

$$\hat{s}(t) = \begin{cases} \hat{x}(t) + \hat{a}_{T1} \hat{x}(t-1) + \cdots + \hat{a}_{TP(T)} \hat{x}(t-P(T)), & 1 \leqslant t \leqslant T, \\ 0, & \text{otherwise}. \end{cases} \tag{1.7}$$

In the following of this paper, we denote $Q_T = (\ln \ln T/T)^{1/2}$ and $O(Q_T)$ means a random sequence satisfying $\limsup_{T\to\infty} O(Q_T)/Q_T < \infty$, a. s.

It was shown that (c. f. [4], Theorem 6)

$$\sup_{1\leq j\leq P(T)} |\hat{\alpha}_{Tj}-\alpha_j| = O(Q_T), \tag{1.8}$$

where $\sum_0^\infty \alpha_j z^j = a_0(z) b_0(z)^{-1} = k_0(z)^{-1} = \alpha_0(z)$.

Secondly, we are going to estimate $r_0 = \max(p_0, q_0)$. For each p, we take an estimate of σ^2 as

$$\hat{\sigma}_p^2 = \min \frac{1}{T}\sum_{t=1}^{T}\left(\sum_{j=0}^{p} a_{pj}\hat{x}(t-j) - \sum_{j=1}^{p} b_{pj}\hat{\varepsilon}(t-j)\right)^2, \tag{1.9}$$

where the minimization is over a_{pj} and b_{pj} with $a_{p_0}=1$. The corresponding minimizing coefficients satisfy the following equation

$$(\hat{a}_{p1}\hat{b}_{p1}\cdots\hat{a}_{pp}\hat{b}_{pp})\begin{bmatrix} C(0) & \cdots\cdots & C(-p+1) \\ \vdots & \ddots & \vdots \\ C(p-1) & \cdots\cdots & C(0)\end{bmatrix}$$

$$= -(1,\ 0)(C(-1)\cdots C(-p)), \tag{1.10}$$

where

$$C(l) = \begin{bmatrix} C_x(l) & -C_{x\hat{\varepsilon}}(l) \\ -C_{\hat{\varepsilon}x}(l) & C_{\hat{\varepsilon}}(l)\end{bmatrix} = \begin{bmatrix} \frac{1}{T}\sum_1^T \hat{x}(t)\hat{x}(t+l) & -\frac{1}{T}\sum_1^T \hat{x}(t)\hat{\varepsilon}(t+l) \\ -\frac{1}{T}\sum_1^T \hat{\varepsilon}(t)\hat{x}(t+l) & \frac{1}{T}\sum_1^T \hat{\varepsilon}(t)\hat{\varepsilon}(t+l)\end{bmatrix}, \tag{1.11}$$

$$C(l)' = C(-l). \tag{1.12}$$

Let equation (1.10) be embedded in the following matrix equation

$$\hat{B}_p C_p = -\zeta_p, \tag{1.13}$$

where $\hat{B}_p = (\hat{B}_{p1}\cdots\hat{B}_{pp})$, each $\hat{B}_{pj}$ is 2×2 matrix with $(\hat{a}_{pj}\hat{b}_{pj})$ as its first row,

$$\zeta_p = (C(-1)\cdots C(-p))$$

and C_p is the blockwise Toeplitz matrix in the left hand side of (1.10). In order to compute $\hat{B}_p$ for $p=1, 2, \cdots$, there is a recursive procedure due to Whittle[3] (also see [2])

$$\hat{B}_{pp} = -\hat{\Delta}_{p-1}\hat{g}_{p-1}^{-1},\quad \hat{\beta}_{pp} = -\hat{\Delta}'_{p-1}\hat{G}_{p-1}^{-1},\quad \hat{\Delta}_p = \sum_0^p \hat{B}_{pj}C(j-p-1), \tag{1.14}$$

$$\hat{G}_p = (I_2 - \hat{B}_{pp}\hat{\beta}_{pp})\hat{G}_{p-1},\quad \hat{g}_p = (I_2 - \hat{\beta}_{pp}\hat{B}_{pp})\hat{g}_{p-1},\quad \hat{G}_0 = \hat{g}_0 = C(0), \tag{1.15}$$

$$\begin{aligned} &\hat{B}_{pj} = \hat{B}_{p-1,j} + \hat{B}_{pp}\hat{\beta}_{p-1,p-j},\quad j=1, 2, \cdots, p-1, \\ &\hat{\beta}_{pj} = \hat{\beta}_{p-1,j} + \hat{\beta}_{pp}\hat{B}_{p-1,p-j},\quad j=1, 2, \cdots, p-1, \\ &\hat{B}_{p0} = \hat{\beta}_{p0} = I_2. \end{aligned} \tag{1.16}$$

The (1, 1) th element of

$$\hat{G}_p = \sum_0^p\sum_0^p \hat{B}_{pj}C(j-l)\hat{B}'_{pl} = (I_2,\ \hat{B}_p)C_{p+1}(I_2,\ \hat{B}_p)' = C(0) - \sum_1^p\sum_1^p \hat{B}_{pj}C(j-l)\hat{B}'_{pl}$$

$$= C(0) + \sum_1^p \hat{B}_{pj}C(j)$$

is $\hat{\sigma}_p^2$.

Because the criterion of BIC can not furnish a consistent estimate of the true order, in this case, we suggest another criterion, that is monitoring $\det \hat{G}_p$. Let r be the first p, such that

$$\det \hat{G}_p < \left(\frac{\ln T}{T}\right)^{1/2} \hat{\sigma}_T^4. \tag{1.17}$$

Then we take r as the estimate of r_0, where $\hat{\sigma}_T^2$ is the estimate of σ^2 by $AR(P(T))$ and we know that $\hat{\sigma}_T^2 \to \sigma^2$ a. s. as $T\to\infty$(see the proof of Theorem 5 in [4]). We add the factor $\hat{\sigma}_T^4$ here only for a technical reason, it does not influence the mathematical theory. In Section 2, we shall show that as $T\to\infty$, $r\to r_0$ a. s. also $\hat{a}_{rj}=a_{0j}+O(Q_T)$, $\hat{b}_{rj}=b_{0j}+O(Q_T)$ (because for sufficiently large T, $\hat{a}_{rj}=\hat{a}_{r_0j}$, $\hat{b}_{rj}=\hat{b}_{r_0j}$ and we shall show that $\hat{a}_{r_0j}=a_{0j}+O(Q_T)$, $\hat{b}_{r_0j}=b_{0j}+O(Q_T)$).

With this r, we use an $ARMA(r, r)$ to estimate $\varepsilon(t)$ by

$$\tilde{\varepsilon}(t)=\begin{cases}\sum_{j=0}^{r}\hat{a}_{rj}\hat{x}(t-j)-\sum_{j=1}^{r}\hat{b}_{rj}\tilde{\varepsilon}(t-j), & 1\leqslant t\leqslant T;\\ 0, & \text{otherwise.}\end{cases} \tag{1.18}$$

Put

$$\begin{aligned} C_{x\tilde{\varepsilon}}(l)&=\frac{1}{T}\sum_{1}^{T}\hat{x}(t)\tilde{\varepsilon}(t+l),\\ C_{\tilde{\varepsilon}x}(l)&=\frac{1}{T}\sum_{1}^{T}\tilde{\varepsilon}(t)\hat{x}(t+l),\\ C_{\tilde{\varepsilon}}(l)&=\frac{1}{T}\sum_{1}^{T}\tilde{\varepsilon}(t)\tilde{\varepsilon}(t+l). \end{aligned} \tag{1.19}$$

For each (p, q), consider another estimate of σ^2,

$$\begin{aligned}\hat{\sigma}_{pq}^2&=\min_{a_{pj},b_{qj}}\frac{1}{T}\sum_{t=1}^{T}\left(\sum_{0}^{p}a_{pj}\hat{x}(t-j)-\sum_{1}^{q}b_{qj}\tilde{\varepsilon}(t-j)\right)^2\\ &=\min_{a_{pj},b_{qj}}(1\quad a_{p1}\cdots a_{pp}\quad b_{q1}\cdots b_{qq}).\\ &\begin{bmatrix} C_x(0) & \cdots & C_x(-p) & -C_{x\tilde{\varepsilon}}(-1) & \cdots & C_{x\tilde{\varepsilon}}(-q)\\ \vdots & \ddots & \vdots & \cdots & \cdots & \cdots\\ C_x(p) & \cdots & C_x(0) & -C_{x\tilde{\varepsilon}}(p-1) & \cdots & -C_{x\tilde{\varepsilon}}(p-q)\\ -C_{\tilde{\varepsilon}x}(1) & \cdots & -C_{\tilde{\varepsilon}x}(1-p) & C_{\tilde{\varepsilon}}(0) & \cdots & C_{\tilde{\varepsilon}}(1-q)\\ \cdots & \cdots & \cdots & \vdots & \ddots & \vdots\\ -C_{\tilde{\varepsilon}x}(q) & \cdots & -C_{\tilde{\varepsilon}x}(q-p) & C_{\tilde{\varepsilon}}(q-1) & \cdots & C_{\tilde{\varepsilon}}(0) \end{bmatrix}\begin{bmatrix}1\\ a_{p1}\\ \vdots\\ a_{pp}\\ b_{q1}\\ \cdots\\ b_{qq}\end{bmatrix}.\end{aligned} \tag{1.20}$$

The minimum is reached at $(\tilde{a}'_p, \tilde{b}'_q)=(\tilde{a}_{p1}, \cdots, \tilde{a}_{pp}, \tilde{b}_{q1}, \cdots, \tilde{b}_{qq})$. Now criterion BIC can be used. Let $(\tilde{p}, \tilde{q})$ minimizes $\ln\tilde{\sigma}_{pq}^2+(p+q)\ln T/T$ for $0\leqslant p, q\leqslant r$, then $(\tilde{p}, \tilde{q})$ shall furnish a consistent estimate for (p_0, q_0). We shall give the proof in Section 2.

Once $(\tilde{p}, \tilde{q})$ are obtained, we can use any available algorithms to calculate asymptotic efficient estimates of a_{0j} and b_{0j}, one of which was mentioned in [2].

When T is very large, the method discribed above seems to be quite safe for estimating r_0. But it is plausible to suspect that in most cases, when T can not be so large, the underestimation of r_0 would often occur and hence the estimate of (p_0, q_0) are usually smaller than the true value.

Because $(\hat{a}_{rj}, \hat{b}_{rj})$ are used only in (1.18) to give an estimate of $\varepsilon(t)=b_0(z)^{-1}\times a_0(z)x(t)$ (here, z means backward shift operator), one would prefer an overestimate of r_0 rather than this underestimate, if the new estimate of $\varepsilon(t)$ is reasonable. It was pointed out and has been used in many cases by Hannan (see [1, 2]), that for fixed P, Q with $P\geqslant p_0$, $Q\geqslant q_0$, there are ways to choose $\hat{a}_{Pj}(j=1, \cdots, P)$ and $\hat{b}_{Qj}(j=1, \cdots, Q)$, such that $\hat{b}_Q(z)^{-1}\hat{a}_P(z)=\left(\sum_0^Q \hat{b}_{Qj}z^j\right)^{-1}\left(\sum_0^P \hat{a}_{Pj}z^j\right)$ converges in a certain sense to $b_0(z)^{-1}a_0(z)$ as $T\to\infty$, though $\hat{a}_P(z)\to a_0(z)$ and $\hat{b}_Q(z)\to b_0(z)$ do not hold. In Section 3, we shall adopt this idea to carry out Whittle's recursion until R, which can be much larger than r_0 such that $\det G_R>0$ does not hold, or "all zeros of $\hat{a}_R(z)$ and $\hat{b}_R(z)$ are outside and keep away from the unit circle by a fixed distance" does not hold. This will be done in Section 3. Further, we shall prove that when $R>s\geqslant r$, all the coefficients of $\hat{\chi}_S(z)=\hat{a}_S(z)b_0(z)-\hat{b}_S(z)a_0(z)$ are $O(Q_T)$, so that we can make $\hat{b}_S(z)^{-1}\hat{a}_S(z)-b_0(z)^{-1}a_0(z)$ small enough, which is needed in proving the main result.

There is a possibility that R may be too large, say, R goes beyond $O(\ln T)$, so in practice, we combine both criteria. First by monitoring $\det G$ to get r, the consistent estimate of $r_0=\max(p_0, q_0)$, then we continue the recursion further for several steps, say to the s th step (in each step, $\det \hat{G}_p>0$ and $\hat{a}_p(z)$, $\hat{b}_p(z)$ have their zeros outside and keep away from the unit circle by a fixed distance), s is consistently greater than r_0 and is usually greater than r_0 in practice even if T is not very large. Obviously, it is much more plausible than the traditional saying that "choose P, Q big enough such that $P\geqslant p_0$, $Q\geqslant q_0$".

To check $\hat{a}_S(z)=\sum_0^S \hat{a}_{Sj}z^j$ has all its zeros outside $|z|=1+\delta$, $\delta>0$, is equivalent to check all the zeros of $z^S+\hat{a}_{S1}z^{S-1}+\cdots+\hat{a}_{SS}$ are inside $|z|=(1+\delta)^{-1}$. Put $w=(1+\delta)z$, then it is equivalent to check $w^S+\hat{a}_{S1}(1+\delta)w^{S-1}+\cdots+\hat{a}_{SS}(1+\delta)^S$ has all its zeros inside $|w|=1$ and any criterion of Routh-Hurwitz type can be used. Similarly for $\hat{b}_S(z)=\sum_0^S \hat{b}_{Sj}z^j$.

§ 2. Monitoring Criterion

Instead of (1.13), we consider of the same type equation, but with C_p, ζ_p being substituted by Γ_p, γ_p respectively (they are estimated by C_p, ζ_p) and the unknown matrix being $B_p=(B_{p1}, \cdots, B_{pp})$, that is

$$B_p\Gamma_p=-\gamma_p, \tag{2.1}$$

where

$$\Gamma_p=\begin{bmatrix}\gamma(0) & \cdots\cdots & \gamma(-p+1)\\ \vdots & \ddots & \vdots\\ \gamma(p-1) & \cdots\cdots & \gamma(0)\end{bmatrix},\quad \gamma_p=(\gamma(-1)\cdots\gamma(-p)),$$

$$\gamma(l)=\begin{bmatrix} Ex(t)x(t+l) & -Ex(t)\varepsilon(t+l) \\ -E\varepsilon(t)x(t+l) & E\varepsilon(t)\varepsilon(t+l) \end{bmatrix}=\begin{bmatrix} \gamma_x(l) & -\gamma_{x\varepsilon}(l) \\ -\gamma_{\varepsilon x}(l) & \gamma_\varepsilon(l) \end{bmatrix}.$$

By condition (1.3)

$$x(t)=\sum_0^\infty k_j\varepsilon(t-j),\quad \varepsilon(t)=\sum_0^\infty \alpha_j x(t-j), \tag{2.2}$$

where the coefficients k_j and α_j are determined by

$$\alpha_0(z)=a_0(z)b_0(z)^{-1}=\sum_0^\infty \alpha_j z^j,\quad k_0(z)=a_0(z)^{-1}b_0(z)=\sum_0^\infty k_j z^j,$$

with $\alpha_0=k_0=1$ and $|\alpha_j|$, $|k_j|$ decrease to zero exponentially. Thus we have

$$\gamma_x(l)=\sigma^2\sum_0^\infty k_jk_{j+l}=\gamma_x(-l),\ l\geqslant 0;$$

$$\gamma_{\varepsilon x}(l)=\sigma^2k_l=\sum_0^\infty \alpha_j\gamma_x(j+l)=\gamma_{x\varepsilon}(-l),\ l\geqslant 0;$$

$$\gamma_{x\varepsilon}(l)=\sigma^2,\ l=0;\gamma_{x\varepsilon}(l)=0,\ l>0;$$

$$\gamma_\varepsilon(l)=\sigma^2\delta_{l,0},$$

and

$$\gamma(l)=\begin{bmatrix} \gamma_x(l) & 0 \\ -\sigma^2k_l & 0 \end{bmatrix},\ \gamma(-l)=\gamma(l)',\ l>0. \tag{2.3}$$

For any p, in order to obtain B_p by solving (2.1), which is similar to (1.13), the recursive formula (1.14)—(1.16) are applicable if $C(j)$ are replaced by $\gamma(j)$ and all the hats therein are erased. But it requires some discussion.

Lemma 2.1. *Γ_p is of full rank for $p\leqslant r_0$ and not for $p>r_0$.*

Proof For any $p>r_0$, we can take a non-zero $2p$-vector $\lambda'=(1\ 1\ a_1\ b_1\cdots a_{p-1}\ b_{p-1})$ with $a_j=a_{0j}$ for $j\leqslant p_0$, $b_j=b_{0j}$ for $j\leqslant q_0$ and zero otherwise, then $\lambda'\Gamma_p=0$ by (1.1). Thus Γ_p is degenerate. Now suppose $p\leqslant r_0$ and there is a non-zero $2p$-vector $\lambda'=(\lambda_0\ \mu_0\ \lambda_1\mu_1\cdots\lambda_{p-1}\ \mu_{p-1})$ such that $\lambda'\Gamma_p=0$, then $\lambda'\Gamma_p\lambda=0$, hence it must be

$$\sum_0^{p-1}\lambda_j x(t-j)-\sum_0^{p-1}\mu_j\varepsilon(t-j)=0. \tag{2.4}$$

We can assume that one of λ_0 and μ_0 is not zero, otherwise replacing p by $p-1$ in (2.4). Multiply both side of (2.4) by $\varepsilon(t)$ and take their expectations we obtain $\lambda_0\gamma_{x\varepsilon}(0)=\mu_0\gamma_\varepsilon(0)$ or $\lambda_0=\mu_0$. Thus we can put λ' in the form $(1\ 1\ \lambda_1\ \mu_1\cdots\lambda_{p-1}\ \mu_{p-1})$ and (2.4) implies that $x(t)$ is generated by an $ARMA(p-1,\ p-1)$ with $p_0>p-1$ or $q_0>p-1$, this contradicts (1.2). The lemma is proven.

Corollary. *$G_p>0$ for $p\leqslant r_0-1$.*

Proof Since (2.3), the second row of γ_p is zero, hence by the lemma and from (2.1), the second row of B_p must be zero. For any $(\lambda_1\ \lambda_2)\neq 0$, it must be $(\lambda_1\ \lambda_2)\cdot(I_2\ B_p)\neq 0$, so

$$(\lambda_1\ \lambda_2)G_p(\lambda_1\ \lambda_2)'=(\lambda_1\ \lambda_2)(I_2\ B_p)\Gamma_{p+1}(I_2\ B_p)'(\lambda_1\ \lambda_2)'>0$$

by the lemma and $G_p=\sum_0^p\sum_0^p B_{pj}\gamma(j-l)B'_{pl}$. Hence $G_p>0$.

It is easy to show that the eigenvalues of $I_2-\beta_{pp}B_{pp}$ and $I_2-B_{pp}\beta_{pp}$ are the same, from (1.15) and by induction one sees that det $g_p\neq 0$ for $p\leqslant r_0-1$.

How about G_{r_0}? Since the second row of γ_p is always zero for any p, thus for $p\leqslant r_0$, the solution B_p of (2.1) must have its second row being zero by Lemma 2.1. But for $p>r_0$, though Γ_p is degenerate, we can also choose the second row of B_p to be zero. For $p=r_0$, the first row of B_{r_0} must be $(a_{01}b_{01}\cdots a_{0r_0}b_{0r_0})$, $a_{0j}=0$, $j>p_0$; $b_{0j}=0$, $j>q_0$. Hence

$$G_{r_0}=\sum_{j=0}^{r_0}B_{r_0j}\gamma(j)=\begin{bmatrix}\gamma_{\varepsilon x}(0) & -\gamma_{x\varepsilon}(0)\\ -\gamma_{\varepsilon x}(0) & \gamma_{\varepsilon}(0)\end{bmatrix}=\begin{bmatrix}\sigma^2 & -\sigma^2\\ -\sigma^2 & \sigma^2\end{bmatrix}, \tag{2.5}$$

which is degenerate, so does g_{r_0}. It seems that the recurssive formula (1.14)—(1.16) can not be carried on beyond the r_0 th step. Nevertheless, we can do a little more. Consider $\Delta_{r_0}=\sum_0^{r_0}B_{r_0j}\gamma(j-r_0-1)$, its second row is zero since the second row of $\gamma(-l)$ is zero for $l>0$. The first row of Δ_{r_0} is

$$\left(\sum_0^{r_0}a_{0j}\gamma_x(j-r_0-1),\quad -\sum_0^{r_0}a_{0j}\gamma_{\varepsilon x}(r_0+1-j)\right)=(0,\ 0),$$

because, from (1.1), we always have

$$\sum_0^{r_0}a_{0j}\gamma_x(j-l)=0,\ \sum_0^{r_0}a_{0j}\gamma_{x\varepsilon}(j-l)=0,\ l>q_0. \tag{2.6}$$

If for $p=r_0+1$, we put (1.14) in the form

$$B_{r_0+1,r_0+1}g_{r_0}=-\Delta_{r_0}=0,\ \beta_{r_0+1,r_0+1}G_{r_0}=-\Delta'_{r_0}=0,$$

then, since G_{r_0} and g_{r_0} are degenerate, the solution B_{r_0+1,r_0+1} and β_{r_0+1,r_0+1} are quite arbitrary, but we can choose them to be zero. If we take this particular solution and do the same hereafter, then the same argument as above shows the recursion can be carried on for $p=r_0+1$, $r_0+2\cdots$ and give the solution $a_{pj}=a_{r_0j}$, $b_{pj}=b_{r_0j}$ and $a_{pj}=b_{pj}=0$, $j>r_0$, because of (1.16) and $B_{pp}=\beta_{pp}=0$.

It is this property that makes the solution $\hat{B}_{pj}$ and $\hat{\beta}_{pj}$ unstable for $p>r_0$ in (1.14)—(1.16), because a very small deviation of $\hat{\Delta}_p$ from zero would make the elements of $\hat{B}_{p+1,p+1}$ and $\hat{\beta}_{p+1,p+1}$ very large, since det$\hat{G}_p$ and det$\hat{g}_p$ are near zero. This discussion suggests us by monitoring det$\hat{G}_p$ to estimate r_0, we denote this estimate by r as in Section 1.

Because of (2.3) and the following lemma, we can also put zeros on the second column of $C(l)$ for $l>0$, instead of $-C_{x\hat{\varepsilon}}(l)$ and $C_{\hat{\varepsilon}}(l)$, it gives the same asymptotic result.

Lemma 2.2. *Let $x(t)$ satisfy* (1.1), (1.2) *and* (1.3), *then for a fixed l, we have*

$$C_x(l)-\gamma_x(l)=O(Q_T); \tag{2.7a}$$

$$C_{\hat{\varepsilon}x}(l)-\gamma_{\varepsilon x}(l)=O(Q_T),\ l\geqslant 0; \tag{2.7b}$$

$$C_{x\hat{\varepsilon}}(l)=O(Q_T),\ l>0; \tag{2.7c}$$

$$C_{\hat{\varepsilon}}(l)-\gamma_{\varepsilon}(l)=O(Q_T). \tag{2.7d}$$

Proof (2.9a) was proved in [4] (Theorem 2). It was also proved in Theorem 6 of the same paper that

$$\sup_{1\leqslant j\leqslant P(T)}|\hat{\alpha}_{Tj}-\alpha_j|=O(Q_T) \quad \text{a. s.}$$

Put $\Delta\hat{\alpha}_{Tj}=\hat{\alpha}_{Tj}-\alpha_j$, $\Delta C_x(l)=C_x(l)-\gamma_x(l)$ and so on, we have

$$C_{\hat{\varepsilon}x}(l)=\frac{1}{T}\sum_1^T\hat{\varepsilon}(t)\hat{x}(t+l)=\frac{1}{T}\sum_1^T\Big(\sum_{j=0}^{P(T)}\hat{\alpha}_{Tj}\hat{x}(t-j)\Big)\hat{x}(t+l)=\sum_{j=0}^{P(T)}\hat{\alpha}_{Tj}C_x(l+j)$$
$$=\sum_{j=0}^{P(T)}(\alpha_j+\Delta\hat{\alpha}_{Tj})(\gamma_x(l+j)+\Delta C_x(l+j))$$
$$=\sum_{j=0}^{P(T)}\alpha_j\gamma_x(l+j)+O\Big(\sup_{1\leqslant j\leqslant P(T)}|\Delta\hat{\alpha}_{Tj}|\sum_{j=0}^{P(T)}|\gamma_x(l+j)|$$
$$+\sup_{0\leqslant j\leqslant P(T)}|\Delta C_x(l+j)|\sum_{j=0}^{P(T)}|\alpha_j|+P(T)\sup_{0\leqslant j\leqslant P(T)}|\Delta C_x(l+j)|\sup_{1\leqslant j\leqslant P(T)}|\Delta\alpha_{Tj}|\Big).$$

Since $|\alpha_j|$, $|\gamma_x(j)|$ converge to zero exponentially as $j\to\infty$, let $T\to\infty$, the first term on the right side converges to $\sum_0^\infty\alpha_j\gamma_x(l+j)=\gamma_{\varepsilon x}(l)$; while the second and third terms are $O(Q_T)$ and the fourth term is $P(T)O(Q_T^2)=O(Q_T)$. Thus (2.7b) is proved. Similarly we can prove (2.7c) and (2.7d).

Now we are going to prove the strong consistency of r.

Theorem 2.1. *Let $x(t)$ satisfy* (1.1), (1.2) *and* (1.3), *and r is decided by* (1.17), *then $r\to r_0$ a. s. as $T\to\infty$.*

Proof Since $G_p>0$ for $p=0, 1, \cdots, r_0-1$ and $G_{r_0}=\begin{bmatrix}\sigma^2 & -\sigma^2\\ -\sigma^2 & \sigma^2\end{bmatrix}$, we need only to prove that

$$|\det\hat{G}_p-\det G_p|=O(Q_T),\ p=0, 1, \cdots, r_0. \tag{2.8}$$

From Lemma 2.2, the elements of $C_p-\Gamma_p$ and $\zeta_p-\gamma_p$ are $O(Q_T)$ and $\Gamma_p>0$ for $p=0, 1, \cdots, r_0$, then by the similar technique as in [4] (Theorem 4), we can deduce that the elements of $\hat{B}_p-B_p$ are $O(Q_T)$ (Because now there are only fixed number of p, we need not use the maximum therein).

Now

$$\hat{G}_p=\sum_0^p\sum_0^p\hat{B}_{pj}C(j-l)\hat{B}'_{pl}=\sum_0^p\sum_0^p(B_{pj}+\Delta\hat{B}_{pj})(\gamma(j-l)+\Delta C(j-l))(B_{pl}+\Delta\hat{B}_{pl})'.$$

The elements in $\hat{G}_p$ differ from the corresponding elements in

$$G_p=\sum_0^p\sum_0^p B_{pj}\gamma(j-l)B'_{pl}$$

by $O(Q_T)$, hence (2.8) follows. But we know that $\hat{\sigma}_T^2\to\sigma^2$ a. s.; thus for large T, we have

$$\det\hat{G}_p=\det G_p+O(Q_T)>(\ln T/T)^{1/2}\hat{\sigma}_T^4,\ p=0, 1, \cdots, r_0-1;$$
$$\det\hat{G}_{r_0}=O(Q_T)<(\ln T/T)^{1/2}\hat{\sigma}_T^4,$$

which establishes the theorem.

Because Γ_{r_0} is of full rank and $C(j)-\gamma(j)=O(Q_T)$ for $j=0, \pm 1, \cdots, \pm r_0$, then as it was proved in Therorem 2.1, $\Delta\hat{B}_{r_0}=\hat{B}_{r_0}-B_{r_0}=O(Q_T)$. Noticing (1.3) we know that for sufficiently large T, $\hat{b}_{r_0}(z)$ has all its zeros outside and keep away from the unit circle by a fixed distance. Put $\alpha_0(z)=a_0(z)/b_0(z)$, $\hat{\alpha}_{r_0}(z)=\hat{a}_{r_0}(z)/\hat{b}_{r_0}(z)$, then

$$\begin{aligned}\Delta\hat{\alpha}_{r_0}(z)&=\hat{\alpha}_{r_0}(z)-\alpha_0(z)=(\hat{a}_{r_0}(z)b_0(z)-\hat{b}_{r_0}(z)a_0(z))/b_0(z)\hat{b}_{r_0}(z)\\&=(\Delta\hat{a}_{r_0}(z)b_0(z)-\Delta\hat{b}_{r_0}(z)a_0(z))/b_0(z)\hat{b}_{r_0}(z).\end{aligned}\tag{2.9}$$

Then one can see that the coefficient of z^j in the expansion of $\Delta\hat{\alpha}_{r_0}(z)$, that is $\Delta\hat{\alpha}_j$, is dominated by $\rho^j O(Q_T)$, where $0<\rho<1$, ρ depends on the location of zeros of $b_0(z)$.

We may write (1.18) in the form

$$\tilde{\varepsilon}(t)=\hat{a}_{r_0}(z)\hat{x}(t)-(\hat{b}_{r_0}(z)-1)\tilde{\varepsilon}(t),\ 1\leqslant t\leqslant T;\ \tilde{\varepsilon}(t)=0 \text{ otherwise},\tag{2.10}$$

here z is the backward shift operator. We can also write

$$\tilde{\varepsilon}(t)=\hat{\alpha}_{r_0}(z)\hat{x}(t),\ 1\leqslant t\leqslant T;\ \tilde{\varepsilon}(t)=0, \text{ otherwise}.\tag{2.11}$$

Denote $\hat{\alpha}_{r_0}(z)=\sum_0^\infty \hat{\alpha}_j z^j$ and consider

$$\begin{aligned}C_{\tilde{\varepsilon}x}(l)&=\frac{1}{T}\sum_1^T\tilde{\varepsilon}(t)\hat{x}(t+l)=\frac{1}{T}\sum_1^T\Big(\sum_{j=0}^\infty\hat{\alpha}_j\hat{x}(t-j)\Big)\hat{x}(t+l)\\&=\sum_{j=0}^\infty\hat{\alpha}_jC_x(j+l)\quad(\text{notice that } C_x(l)=0 \text{ for } |l|\geqslant T)\\&=\sum_{j=0}^\infty\alpha_jC_x(j+l)+\sum_{j=0}^\infty\Delta\hat{\alpha}_jC_x(j+l)\\&=\sum_{j=0}^\infty\alpha_j\gamma_x(j+l)+\sum_{j=0}^\infty\alpha_j\Delta C_x(j+l)+\sum_{j=0}^\infty\Delta\hat{\alpha}_jC_x(j+l).\end{aligned}$$

In the following, m_1, m_2 and so on denote some constants which depend on $a_0(z)$, $b_0(z)$, then we have $\alpha_j=m_1e^{-m_2j}$. Take $J=[\ln T/m_2]$, since

$$\max_{0\leqslant l\leqslant T-1}\Delta C_x(l)=O(\ln T/T)^{1/2}$$

(see [4], Theorem 3), then

$$\begin{aligned}\Big|\sum_{j=J+1}^\infty\alpha_j\Delta C_x(j+l)\Big|&\leqslant\Big|\sum_{J+1}^{T-l-1}\alpha_j\Delta C_x(j+l)\Big|+\Big|\sum_{T-l}^\infty\alpha_j\gamma_x(j+l)\Big|\\&\leqslant 0(\ln T/T)^{1/2}m_3e^{-\ln T}+m_4e^{-T}=O(T^{-1}).\end{aligned}$$

On the other hand, we have

$$\Big|\sum_{j=0}^J\alpha_j\Delta C_x(j+l)\Big|=O(Q_T),$$

because $\sum_0^\infty|\alpha_j|<\infty$ and $\max_{0\leqslant j\leqslant \ln T/m_2}\Delta C_x(j)=O(Q_T)$ (also see [4], Theorem 2). Thus

$$\sum_{j=0}^\infty\alpha_j\Delta C_x(j+l)=O(Q_T).$$

Because $\Delta\hat{\alpha}_j=\rho^jO(Q_T)$, $C_x(j)=\gamma_x(j)+\Delta\ C_x(j)$ is obviously bounded a. s., thus $C_{\tilde{\varepsilon}x}(l)=\gamma_{\varepsilon x}(l)+O(Q_T)$. Similarly we can prove $C_{x\tilde{\varepsilon}}(l)=\gamma_{x\tilde{\varepsilon}}(l)+O(Q_T)$ and $C_{\tilde{\varepsilon}}(l)=\gamma_\varepsilon(l)+O(Q_T)$. Thus we have the following lemma:

Lemma 2.2.′ *Under the same conditions as in Lemma 2.2, $\tilde{\varepsilon}(t)$ is defined by (2.10) or (2.11), then (2.7b)—(2.7d) still hold if $C_{\hat{\varepsilon}x}(l)$, $C_{x\hat{\varepsilon}}(l)$ and $C_{\hat{\varepsilon}}(l)$ are substituted by $C_{\tilde{\varepsilon}x}(l)$, $C_{x\tilde{\varepsilon}}(l)$ and $C_{\tilde{\varepsilon}}(l)$ respectively.*

Theorem 2.2. *Suppose $x(t)$ satisfy (1.1), (1.2) and (1.3), $\tilde{\varepsilon}(t)$ is obtained by (2.10) or (2.11) and form $\tilde{\sigma}^2_{pq}$ as (3.17), if $(\tilde{p}, \tilde{q})$ minimize (3.19) for $0\leqslant p\leqslant r$, $0<q\leqslant r$, then $\tilde{p}\to p_0$, $\tilde{q}\to q_0$ a. s. as $T\to\infty$.*

Because this theorem is the special case of Theorem 3.1 in the next section, we omit the simpler proof (it is similar to Theorem 3.1, using Lemma 2.2′, fomula (2.9) and so on).

§ 3. The Main Theorem

As mentioned just before Lemma 2.2, the second row of $C(-l)$, $l>0$, can be replaced by zeros without effecting the asymptotic results. From now on we denote

$$C(0)=\begin{pmatrix} C_x(0) & -C_{x\hat{\varepsilon}}(0) \\ -C_{\hat{\varepsilon}x}(0) & C_{\hat{\varepsilon}}(0) \end{pmatrix}>0,\quad C(-l)=\begin{pmatrix} C_x(-l) & -C_{x\hat{\varepsilon}}(-l) \\ 0 & 0 \end{pmatrix},\ l>0,$$

and

$$C_l=\begin{pmatrix} C(0) & \cdots\cdots & C(-l+1) \\ \vdots & \ddots & \vdots \\ C(l-1) & \cdots\cdots & C(0) \end{pmatrix}. \tag{3.1}$$

In order to prove the main result we require the following lemmas.

Lemma 3.1 *Suppose that $C_{p+1}>0$ and*

$$(\hat{B}_{p1}, \cdots, \hat{B}_{pp})C_p=-(C(-1), \cdots, C(-p))\ (\text{c. f. } (1.13)).$$

For $t>p$, let us define

$$\bar{C}(-t)=-\sum_1^p B_{pj}C(j-t),\quad \bar{C}(t)=\bar{C}(-t)'.$$

Then $\{\bar{C}(t),\ t=0,\ \pm1,\ \cdots\}$ is a positive definite sequence of matrices.

Proof Construct $\bar{C}_l$ from $\bar{C}(t)$ as C_l from $C(t)$.

First we prove that $\bar{C}_{p+2}>0$. We need only toprove that $\det\bar{C}_{p+2}>0$.

By the definition of $C(-p-1)$ we have

$$(\hat{B}_{p1}\cdots\hat{B}_{pp}0)\bar{C}_{p+1}=-(\bar{C}(-1)\cdots\bar{C}(-p-1)),$$

thus

$$\det\bar{C}_{p+2}=\det\bar{C}_{p+1}\det\left\{\bar{C}(0)-(\bar{C}(-1)\cdots\bar{C}(-p-1))\bar{C}^{-1}_{p+1}\begin{pmatrix}\bar{C}(1)\\ \vdots\\ \bar{C}(p+1)\end{pmatrix}\right\}$$

$$=\det\bar{C}_{p+1}\det\left\{\bar{C}(0)+(\hat{B}_{p1}\cdots\hat{B}_{pp}0)\begin{pmatrix}\bar{C}(1)\\ \vdots\\ \bar{C}(p+1)\end{pmatrix}\right\}$$

$$=\det \bar{C}_{p+1} \det\left\{\bar{C}(0)+(\hat{B}_{p1}\cdots\hat{B}_{pp})\begin{pmatrix}\bar{C}(1)\\ \vdots \\ \bar{C}(p)\end{pmatrix}\right\}$$

$$=\det \bar{C}_{p+1} \det\left\{\bar{C}(0)-(\bar{C}(-1)\cdots\bar{C}(-p))\bar{C}_p^{-1}\begin{pmatrix}\bar{C}(1)\\ \vdots \\ \bar{C}(p)\end{pmatrix}\right\}$$

$$=(\det \bar{C}_{p+1})^2/\det \bar{C}_p)>0.$$

Again, obviously,

$$\det\begin{pmatrix} C_\varepsilon(0) & & 0 \\ & \ddots & \\ 0 & & C_{p+1}\end{pmatrix}>0, \text{ so } \bar{C}_{p+2}>0.$$

Using the same procedure we can prove that $\bar{C}_{p+3}>0$, $\cdots$and the lemma follows.

Lemma 3.2. *Let* $\hat{B}_p(z)=\sum_0^p \hat{B}_{pj}z^j$, $\hat{B}_{p0}=I_2$. *Then under the conditions of Lemma 3.1,* $\det \hat{B}_p(e^{i\lambda})\neq 0$, $-\pi\leqslant\lambda\leqslant\pi$.

Proof By the definition of $\hat{B}_{pj}$ and

$$\bar{C}(t)'=\bar{C}(-t)=-\sum_1^p \hat{B}_{pj}\bar{C}(j-t)$$

for $t>p$, we have

$$\sum_0^p \hat{B}_{pj}\bar{C}(j-l)=0 \text{ for } l\geqslant 1.$$

Now, we treat $\bar{C}(t)$, $\hat{B}_{pj}$ as constants, because $\{\bar{C}(t)\}$ is positive definite, so there exists a vector stationary series $y(t)$ with $\bar{C}(t)$ as its covariance and

$$\bar{C}(t)=\int_{-\pi}^{\pi} e^{-it\lambda}dF(\lambda), \tag{3.2}$$

$$y(t)=\int_{-\pi}^{\pi} e^{-it\lambda}d\zeta(\lambda), \tag{3.3}$$

where $E\, d\zeta(\lambda)d\zeta(\mu)^*=\delta_{\lambda,\mu}dF(\lambda)$.

Put $\xi(t)=\sum_0^p \hat{B}_{pj}y(t-j)=\int_{-\pi}^{\pi}e^{-it\lambda}\sum_0^p \hat{B}_{pj}e^{ij\lambda}d\zeta(\lambda)$, $t=0, \pm 1, \cdots$, $\xi(t)$ is stationary, so it may be written as

$$\xi(t)=\int_{-\pi}^{\pi} e^{-it\lambda}d\zeta_\xi(\lambda) \tag{3.4}$$

and by the uniqueness of spectral representation we have

$$\hat{B}_p(e^{i\lambda})d\zeta(\lambda)=d\zeta_\xi(\lambda), \tag{3.5}$$

and

$$\hat{B}_p(e^{i\lambda})dF(\lambda)\hat{B}_p(e^{i\lambda})^*=E\, d\zeta_\xi(\lambda)d\zeta_\xi(\lambda)^*. \tag{3.6}$$

But

$$E\xi(t+l)y(t)'=\sum_0^p \hat{B}_{pj}\bar{C}(j-l)=0 \text{ for } l>0,$$

thus
$$E\xi(t+l)\xi(t)'=E\xi(t+l)\Big(\sum_0^p \hat{B}_{pj}y(t-j)\Big)'=0,\ l>0;$$
$$E\xi(t-l)\xi(t)'=(E\xi(t)\xi(t-l)')'=0,\ l>0.$$

From the above we see that $\xi(t)$ has spectral density $\frac{1}{2\pi}E\xi(t)\xi(t)'$ and from (3.6) we have

$$\hat{B}_p(e^{i\lambda})\frac{dF(\lambda)}{d\lambda}\hat{B}_p(e^{i\lambda})^*=\frac{1}{2\pi}E\xi(t)\xi(t)'=\frac{1}{2\pi}E\Big(\sum_0^p\hat{B}_{pj}y(t-j)\Big)\Big(\sum_0^p\hat{B}_{pj}y(t-j)\Big)'$$
$$=\frac{1}{2\pi}\sum_0^p\sum_0^p\hat{B}_{pj}C(j-l)\hat{B}'_{pl}=\frac{1}{2\pi}\hat{G}_p.$$

Now $\hat{G}_p$ is of full rank, so $\hat{B}_p(e^{i\lambda})$ must be of full rank for every $\lambda\in[-\pi,\pi]$. This proves the lemma.

From the above proof we see that the spectral density of $y(t)$ is
$$f(\lambda)=\frac{1}{2\pi}\hat{B}_p(e^{i\lambda})^{-1}\hat{G}_p(\hat{B}_p(e^{i\lambda})^{-1})^*. \tag{3.7}$$

And by Lemma 3.2, $\hat{a}_p(e^{i\lambda})\neq 0$, for $\lambda\in[-\pi,\pi]$, we have
$$\hat{B}_p(e^{i\lambda})=\begin{bmatrix}\hat{a}_p(e^{i\lambda}) & \hat{b}_p(e^{i\lambda})-1\\ 0 & 1\end{bmatrix},\quad \hat{B}_p(e^{i\lambda})^{-1}=\begin{bmatrix}\hat{a}_p(e^{i\lambda})^{-1} & (1-\hat{b}_p(e^{i\lambda}))/\hat{a}_p(e^{i\lambda})\\ 0 & 1\end{bmatrix}.$$

Lemma 3.3. *Under the same conditions as above, all zeros of $\hat{a}_p(z)$ are outside the unit circle (or, all zeros of $\det\hat{B}_p(z)$ are outside the unit circle).*

Proof From Lemma 3.2 and (3.5) we have
$$y(t)=\int_{-\pi}^{\pi}e^{-it\lambda}\hat{B}_p(e^{i\lambda})^{-1}d\zeta_\xi(\lambda)=\int_{-\pi}^{\pi}e^{-it\lambda}\hat{a}_p(e^{i\lambda})^{-1}\begin{bmatrix}1 & 1-\hat{b}_p(e^{i\lambda})\\ 0 & \hat{a}_p(e^{i\lambda})\end{bmatrix}d\zeta_\xi(\lambda).$$

For $l>0$,
$$0=Ey(t)\xi(t+l)'=\frac{1}{2\pi}\int_{-\pi}^{\pi}e^{il\lambda}\hat{a}_p(e^{i\lambda})^{-1}\begin{bmatrix}1 & 1-\hat{b}_p(e^{i\lambda})\\ 0 & \hat{a}_p(e^{i\lambda})\end{bmatrix}\hat{G}_p d\lambda.$$

Since $\hat{G}_p>0$, thus we can blotout$\hat{G}_p$, considering the (1, 1) th element of the above equality, we obtain
$$\int_{-\pi}^{\pi}e^{il\lambda}\hat{a}_p(e^{i\lambda})^{-1}d\lambda=0,\ l>0.$$

From this, we immediately infer that $\hat{a}_p(z)^{-1}$ is analytic in $|z|\leq 1$ and hence $\hat{a}_p(z)$ has all its zeros outside the unit circle.

Lemma 3.3, for the scalar case, is well-known, but in the case of two dimensions, we have no available result and it seems to us that putting $C_{\hat{\varepsilon}x}(-l)=C_{\hat{\varepsilon}}(\pm l)=0$, $l>0$, makes the proof easy.

Lemma 3.4. *Suppose that*
$$C_{\varepsilon x}(l)=\gamma_{\varepsilon x}(l)+O(Q_T),\ C_{x\hat{\varepsilon}}(l)=\gamma_{x\varepsilon}(l)+O(Q_T),\ C_{\hat{\varepsilon}}(l)=\gamma_\varepsilon(l)+O(Q_T),$$
$$|l|\leq\max(p,q) \tag{3.8}$$
hold and suppose that $\tilde{a}_{pj}$, $\tilde{b}_{qj}$ are the coefficients which are a priori bounded and at

which the minimum in (1.20) *is reached and* $p \geqslant p_0$, $q \geqslant q_0$. *Put*

$$\tilde{\chi}_{pq}(z) = \tilde{a}_p(z) b_0(z) - \tilde{b}_q(z) a_0(z),$$

this is a polynomial of order $n_0 = \max(p+q_0,\ q+p_0)$ *with constant term* 0. *Then the coefficients of* $\tilde{\chi}_{pq}(z)$ *are* $O(Q_T)$.

Proof Put

$$a_p(z) = \sum_0^p a_{pj} z^j,\quad b_q(z) = \sum_0^q b_{qj} z^j,\quad \chi_{pq}(z) = a_p(z) b_0(z) - b_q(z) a_0(z) = \sum^{n_0} \chi_j z^j,$$

$$a = (a_{p1}, \cdots, a_{pp})',\quad a_0 = (a_{01}, \cdots, a_{0p})',\quad (a_{0j} = 0,\ j > p_0);$$

$$b = (b_{q1}, \cdots, b_{qq})',\quad b_0 = (b_{01}, \cdots, b_{0q})',\quad (b_{0j} = 0,\ j > q_0);$$

$$\chi = (\chi_1, \cdots, \chi_{n_0})',\quad \psi = (\psi_1, \cdots, \psi_{n-n_0}),$$

where $n = p+q$. We use Hannan's transformation (see [1])

$$\begin{bmatrix} \chi \\ \psi \end{bmatrix} = G \begin{bmatrix} a - a_0 \\ b - b_0 \end{bmatrix} = \begin{bmatrix} G_1 \\ G_2 \end{bmatrix} \begin{bmatrix} a - a_0 \\ b - b_0 \end{bmatrix}, \tag{3.9}$$

G_1 is of rank n_0, G_2 is of rank $n - n_0$, since

$$G_1 = \begin{bmatrix} \overbrace{\begin{matrix} 1 & & & & 0 \\ b_{01} & 1 & & & \\ \ddots & b_{01} & \ddots & & \\ & \ddots & \ddots & 1 & \\ b_{0q_0} & & \ddots & b_{01} & \\ & b_{0q_0} & & \ddots & \\ 0 & & \ddots & & b_{0q_0} \end{matrix}}^{p} & \overbrace{\begin{matrix} -1 & & & & 0 \\ -a_{01} & -1 & & & \\ \ddots & -a_{01} & \ddots & -1 & \\ & \ddots & \ddots & -a_0 & \\ -a_{0p_0} & & \ddots & & \\ & -a_{0p_0} & & \ddots & \\ 0 & & \ddots & & -a_{0p_0} \end{matrix}}^{q} \end{bmatrix}_{n_0 \times n}$$

$$G_2 = \begin{pmatrix} \overbrace{\begin{matrix} 1 & a_{01} & \ddots & a_{0p_0} & 0 \\ 0 & \ddots\ 1 & a_{01} & \ddots & a_{0p_0} \end{matrix}}^{p} & \overbrace{\begin{matrix} 1 & b_{01} & \ddots & b_{0q_0} & 0 \\ 0 & \ddots\ 1 & b_{01} & \ddots & b_{0q_0} \end{matrix}}^{q} \end{pmatrix}_{(n-n_0) \times n}.$$

One can see that every row of G_2 is orthogonal to any row of G_1, hence the linear space $\{G_1\}$ spanned by the rows of G_1 is orthogonal to the linear space $\{G_2\}$ spanned by the rows of G_2.

Put $G^{-1} = H = (H_1 H_2)$, where H_1 is $n \times n_0$ matrix and H_2 is $n \times (n - n_0)$ matrix. From $GH = I_n$ one can see that $\{H_1\} = \{G_1\}$, $\{H_2\} = \{G_2\}$ and hence $\{H_1\} \perp \{H_2\}$, where $\{H_i\}$ is the linear space spanned by the columns of H_i, $i = 1, 2$.

From the assumption of the model, it is obvious that $h \in \{H_2\} = \{G_2\}$ iff

$$(x(t), \cdots, x(t-p+1),\ \varepsilon(t), \cdots, \varepsilon(t-q+1))h = 0. \tag{3.10}$$

Denote

$$\Gamma^*_{pq}=\begin{bmatrix} \gamma_x(1) & \gamma_x(0) & \cdots & \gamma_x(1-p) & -\gamma_{xs}(0) & \cdots & -\gamma_{xs}(1-q) \\ \vdots & \vdots & \ddots & \vdots & \vdots & \cdots & \cdots \\ \gamma_x(p) & \gamma_x(p-1) & \cdots & \gamma_x(0) & -\gamma_{xs}(p-1) & \cdots & \gamma_{xs}(p-q) \\ -\gamma_{sx}(1) & -\gamma_{sx}(0) & \cdots & -\gamma_{sx}(1-p) & \gamma_s(0) & \cdots & \gamma_s(1-q) \\ \cdots & \cdots & \cdots & \cdots & \vdots & \ddots & \cdots \\ -\gamma_{sx}(q) & -\gamma_{sx}(q-1) & \cdots & -\gamma_{sx}(q-p) & \gamma_s(q-1) & \cdots & \gamma_s(0) \end{bmatrix}_{n\times(n+1)}$$

and Γ_{pq} is obtained by erasing the first column of Γ^*_{pq}. Let

$$y_j=(x(t)\cdots x(t-p+1)\quad -\varepsilon(t)\cdots-\varepsilon(t-q+1))h_j,\ Y=(y_1,\ \cdots,\ y_{n_0})',\ \Gamma_Y=EYY',$$

where h_j is the jth column of H_1. Then from (3.10), we have

$$H'\Gamma_{pq}H=E\{H_1\ H_2)'(x(t)\cdots x(t-p+1)\quad -\varepsilon(t)\cdots-\varepsilon(t-q+1))(x(t)\cdots x(t-p+1)\quad \varepsilon(t)\cdots\varepsilon(t-q+1))'(H_1\ H_2)\}$$

$$=\begin{bmatrix}\Gamma_Y & 0\\ 0 & 0\end{bmatrix}. \tag{3.11}$$

Γ_y is non-degenerate, because $(y_1,\ \cdots,\ y_n)$ is linearly independent. Otherwise there is a $\lambda=(\lambda_1,\ \cdots\lambda_{n_0})'$ such that

$$0=Y'\lambda=(x(t)\cdots x(t-p+1)\quad -\varepsilon(t)\cdots-\varepsilon(t-q+1))H_1\lambda.$$

From (3.10), $H_1\lambda\in\{H_2\}$, but $\{H_2\}\perp\{H_1\}$, thus, there must be $H_1\lambda=0$ or $\lambda=0$, since rank $H_1=n_0$.

Now we denote the quadratic form of RHS in (1.20) by $\sigma^2_{pq}(a,\ b)=\sigma^2(\chi,\ \psi)$. Its minimum is reached at

$$\begin{bmatrix}\tilde a\\ \tilde b\end{bmatrix}\quad\text{or}\quad\begin{bmatrix}\tilde\chi\\ \tilde\psi\end{bmatrix}=G\begin{bmatrix}\tilde a-a_0\\ \tilde b-b_0\end{bmatrix},$$

then we have

$$0=\frac{\partial\sigma^2(\tilde\chi,\ \tilde\psi)}{\partial\chi}=\frac{\partial\sigma^2(0,\ \tilde\psi)}{\partial\chi}+\frac{\partial^2\sigma^2(\chi^*,\ \tilde\psi)}{\partial\chi\,\partial\chi'}\tilde\chi, \tag{3.12}$$

where χ^* is between 0 and $\tilde\chi$, and the first derivative is an n-vector evaluated at $\begin{pmatrix}\tilde\chi\\ \tilde\psi\end{pmatrix}\left(\text{or}\begin{pmatrix}0\\ \tilde\psi\end{pmatrix}'\right)$, the second derivative is an $n\times n$ matrix evaluated at $\begin{pmatrix}\chi^*\\ \tilde\psi\end{pmatrix}$, but actually it does not depend on $\begin{pmatrix}\chi^*\\ \tilde\psi\end{pmatrix}$, since $\sigma^2(\chi,\ \psi)$ is a quadratic form.

Denote by C^*_{pq} the matrix which is obtained by erasing the first row of the matrix of RHS in (1.20), denote by C_{pq} the matrix which is obtained by erasing the first column of C^*_{pq}. From (3.8) we have

$$C^*_{pq}=\Gamma^*_{pq}+(O(Q_T))_{n\times(n-1)},\ C_{pq}=\Gamma_{pq}+(O(Q_T))_{n\times n}. \tag{3.13}$$

Suppose $\begin{pmatrix}\bar a\\ \bar b\end{pmatrix}$ is the vector such that $\begin{pmatrix}0\\ \tilde\psi\end{pmatrix}=G\begin{pmatrix}\bar a-a_0\\ \bar b-b_0\end{pmatrix}$, then

$$\frac{\partial\sigma^2(0,\tilde{\psi})}{\partial\chi}=\frac{\partial(\bar{a}',\bar{b}')}{\partial\chi}\frac{\partial\sigma^2_{pq}(\bar{a},\bar{b})}{\partial\begin{pmatrix}a\\b\end{pmatrix}}=2(I_{n_0}0)H'C^*_{pq}\begin{pmatrix}1\\\bar{a}\\\bar{b}\end{pmatrix}. \tag{3.14}$$

From (1.1), $x(t)+\cdots+a_{0p}x(t-p)-\varepsilon(t)-\cdots-b_{0q}\varepsilon(t-q)=0$ (noticing that $a_{0j}=0$, $j>p_0$; $b_{0j}=0$, $j>q_0$). Multiply both sides of the equation by $x(t-j)$, $j=1,\cdots,p$ and $\varepsilon(t-j)$, $j=1,\cdots,q$, then take expectations, we obtain $\Gamma^*_{pq}\begin{pmatrix}1\\a_0\\b_0\end{pmatrix}=0$ and from (3.13), $2(I_{n_0}\ \ 0)\,H'C^*_{qp}\begin{pmatrix}1\\a_0\\b_0\end{pmatrix}=(O(Q_T))_{n_0\times1}$. Combining this with (3.14), we obtain

$$\begin{aligned}\frac{\partial\sigma^2(0,\tilde{\psi})}{\partial\chi}&=2(I_{n_0}\ \ 0)H'C_{pq}\begin{pmatrix}\bar{a}-a_0\\\bar{b}-b_0\end{pmatrix}+(O(Q_T))_{n_0\times1}\\&=2(I_{n_0}\ \ 0)H'C_{pq}H\begin{pmatrix}0\\\tilde{\psi}\end{pmatrix}+(O(Q_T))_{n_0\times1}\\&=2(I_{n_0}\ \ 0)H'\Gamma_{pq}H\begin{pmatrix}0\\\tilde{\psi}\end{pmatrix}+(O(Q_T))_{n_0\times1}\\&=2(I_{n_0}\ \ 0)\begin{pmatrix}\Gamma_Y&0\\0&0\end{pmatrix}\begin{pmatrix}0\\\tilde{\psi}\end{pmatrix}+(O(Q_T))_{n_0\times1}=(O(Q_T))_{n_0\times1}.\end{aligned} \tag{3.15}$$

Here the third equality is assured by (3.13) and "$\tilde{a}$, $\tilde{b}$ are a priori bounded" (so does $\tilde{\psi}$).

Again $\dfrac{\partial\sigma^2(\chi,\psi)}{\partial\chi\,\partial\chi'}=2(I_{n_0}\ 0)H'C_{pq}H\begin{pmatrix}I_{n_0}\\0\end{pmatrix}=2\Gamma_Y+(O(Q_T))_{n_0\times n_0}$, then from (3.12), (3.15) we have

$$\tilde{\chi}=-\Gamma_Y^{-1}(O(Q_T))_{n_0\times1}=(O(Q_T))_{n_0\times1}.$$

This establishes the lemma.

Lemma 3.4.' *Let $q=p$; $C_{x\hat{\varepsilon}}(j)$, $C_{\hat{\varepsilon}x}(j)$, $C_{\hat{\varepsilon}}(j)$ in (1.20) are replaced by $C_{x\tilde{\varepsilon}}(j)$, $C_{\tilde{\varepsilon}x}(j)$, $C_{\tilde{\varepsilon}}(j)$ respectively, the minimizing values $\hat{a}_{pj}$, $\hat{b}_{pj}$ are a priori bounded and correspondingly put $\hat{\chi}_p(z)=\hat{a}_p(z)b_0(z)-\hat{b}_p(z)a_0(z)$. Then the coefficients of $\hat{\chi}_p(z)$ are $O(Q_T)$.*

The proof is the same as above.

Lemma 3.5. *Suppose $R>p\geqslant r_0$, $\hat{a}_{pj}$, $\hat{b}_{pj}$ are a priori bounded and T sufficiently large. Then $\hat{a}_p(z)$ has all its zeros outside and keeping away from the unit circle by a fixed distance if and only if $\hat{b}_p(z)$ has the same properties.*

Proof Since

$$\hat{b}_p(z)=\frac{\hat{a}_p(z)b_0(z)}{a_0(z)}-\frac{\hat{\chi}_p(z)}{a_0(z)},$$

where the coefficients of $\hat{\chi}_p(z)$ are $O(Q_T)$, and by assumption that $a_0(z)$ and $b_0(z)$ have their zeros outside unit circle, so if the zeros of $\hat{a}_p(z)$ has the properties mentioned in the lemma, then $\hat{b}_p(z)$ has the same properties. The converse is true by considering $\hat{a}_p(z) = \hat{\chi}_p(z)/b_0(z) + \hat{b}_p(z)a_0(z)/b_0(z)$.

Now we carry on Whittle's recursion (1.14)—(1.16) when $\hat{G}_p > 0$ remains true. From Lemmas 3.1—3.3, we know that $\hat{a}_p(z)$ have all their zeros outside the unit circle, they are, say, $\hat{z}_1, \cdots, \hat{z}_p$. Then $\hat{a}_p(z) = (1-\hat{z}_1^{-1}z)\cdots(1-\hat{z}_p^{-1}z)$. But $|\hat{z}_j^{-1}| < 1$, thus all the coefficients of $\hat{a}_p(z)$ are a priori bounded, say, by $\max\limits_{1\le j\le p}\binom{p}{j}$. In addition to that, we also check the zeros of $\hat{a}_p(z)$ to assure that they keep away from the unit circle by a fixed distance. As it was mentioned in the end of Section 1, we carry the recursion to s-th step, for every p, $0 \le p \le s$. By Lemmas 3.4′, 3.5, $\hat{b}_p(z)$ has all its zeros outside and keep away from the unit circle by a fixed distance and all its coefficients are bounded by $\max\limits_{0\le j\le s}\binom{s}{j}$.

In the following, put $\tilde{\varepsilon}(t)$ as in (1.18) with r replaced by s, then

$$\tilde{\varepsilon}(t) = \hat{\alpha}_s(z)\hat{x}(t),\ 1 \le t \le T; \tilde{\varepsilon}(t) = 0,\ \text{otherwise}.\ \hat{\alpha}_s(z) = \hat{a}_s(z)/\hat{b}_s(z). \tag{3.16}$$

Again, as (2.9), we have

$$\Delta\hat{\alpha}_s(z) = \hat{\alpha}_s(z) - \alpha_0(z) = (\hat{a}_s(z)b_0(z) - \hat{b}_s(z)a_0(z))/b_0(z)\hat{b}_s(z) = \hat{\chi}_p(z)/b_0(z)\hat{b}_s(z).$$

By Lemma 3.4′ and the properties of zeros of $b_0(z)$ and $\hat{b}_s(z)$, one can see that the coefficients $\Delta\alpha_j$ of z^j in the expansion of $\Delta\alpha_s(z)$ are dominated by $\rho^j O(Q_T)$. Then the same proof as in Lemma 2.2′ leads to (3.8).

As mentioned in Section 1, we use $\tilde{\varepsilon}(t)$ defined in (3.16) to estimate σ^2, that is, for each (p, q), $0 \le p$, $q \le s$, we calculate

$$\tilde{\sigma}^2_{pq} = \min_{a_{pj}, b_{qj}} \frac{1}{T}\sum_1^T\Big(\sum_{j=0}^p a_{pj}\hat{x}(t-j) - \sum_{j=1}^q b_{qj}\tilde{\varepsilon}(t-j)\Big)^2, \tag{3.17}$$

under the bounded condition

$$\max_{p,q,j}\{|a_{pj}|,\ |b_{qj}|\} \le \max_{0\le j\le s}\binom{s}{j}. \tag{3.18}$$

This restriction is reasonable because, first, we require that

$$\tilde{a}_p(z) \ne 0,\ \tilde{b}_p(z) \ne 0,\ |z| \le 1,$$

where $\tilde{a}_p(z) = \sum\limits_0^p \tilde{a}_{pj}z^j$, $\tilde{b}_p(z) = \sum\limits_0^q \tilde{b}_{qj}z^j$ and $\tilde{a}_{pj}$, $\tilde{b}_{qj}$ minimize (3.17); secondly, from (3.8) and the discussion in Section 2, one can see that $\tilde{a}_{p_0}(z) \to a_0(z)$, $\tilde{b}_{q_0}(z) \to b_0(z)$, a. s., so (3.18) always holds for $p = p_0$, $q = q_0$ (for large T), thus (3.18) do not rule out the true order. We can abandon those (p, q), $0 \le p$, $q \le s$, for which some $\tilde{a}_{pj}$ or $\tilde{b}_{qj}$ do not satisfy (3.18). Thus Lemma 3.4 holds. If $(\tilde{p}, \tilde{q})$ minimizes

$$BIC(p, q) = \ln\tilde{\sigma}^2_{pq} + (p+q)\ln T/T, \tag{3.19}$$

then $(\tilde{p}, \tilde{q})$ is our estimate of (p_0, q_0). We are now going to prove the main theorem in the following.

Theorem 3.1. *Let $x(t)$ satisfy (1.1), (1.2) and (1.3), and we derive $(\tilde{p}, \tilde{q})$ as above, then $\tilde{p} \to p_0$, $\tilde{q} \to q_0$ a. s. as $T \to \infty$.*

Proof As in the proof of Theorem 3 in [2], for $p<p_0$ or $q<q_0$,

$$\liminf_{T\to\infty}\left\{\tilde{\sigma}^2_{pq}+(p+q)\frac{\ln T}{T}-\tilde{\sigma}^2_{p_0q_0}-(p_0+q_0)\frac{\ln T}{T}\right\}>0,$$

so we must have $\tilde{p}\geqslant p_0$, $\tilde{q}\geqslant q_0$. Now for $p\geqslant p_0$, $q\geqslant q_0$

$$\begin{aligned}\tilde{\sigma}^2_{pq}&=\frac{1}{T}\sum_1^T[\tilde{a}_p(z)\hat{x}(t)-(\tilde{b}_q(z)-1)\tilde{\varepsilon}(t)]^2\\&=\frac{1}{T}\sum_1^T[\{\tilde{a}_p(z)-(\tilde{b}_q(z)-1)\hat{\alpha}_s(z)\}\hat{x}(t)]^2\\&=\frac{1}{T}\sum_1^T[\{(\tilde{a}_p(z)-\tilde{b}_q(z)\alpha_0(z))+\alpha_0(z)-(\tilde{b}_q(z)-1)(\hat{\alpha}_s(z)-\alpha_0(z))\}\hat{x}(t)]^2\\&=\frac{1}{T}\sum_1^T\left[\alpha_0(z)\hat{x}(t)+\frac{\tilde{\chi}_{pq}(z)\hat{b}_s(z)-(\tilde{b}_q(z)-1)\hat{\chi}_s(z)}{b_0(z)\hat{b}_s(z)}\hat{x}(t)\right]^2.\end{aligned}\tag{3.20}$$

From Lemmas 3.4, 3.4′ and $\hat{b}_s(z)$, $\tilde{b}_q(z)$ has a priori bound, one can see that the numerator of the second term (polymomial) above has all its coefficients equaling to $O(Q_T)$, and the coefficient of z^0 is zero by the definition of $\tilde{\chi}_{pq}(z)$ and $\hat{\chi}_s(z)$ or $\tilde{b}_q(z)$. Again because of (1.3) and all zeros of $\hat{b}_s(z)$ are outside and keep away from the unit circle by fixed distance, the second term can be written as $\{O(Q_T)\times\sum_{j=0}^{\infty}\phi_j\hat{x}(t-1-j)\}$, where ϕ_j (random sequence) satisfies $|\phi_j|\leqslant\rho^j$ uniformly in T, $0<\rho<1$. Denote $\alpha_1(z)\hat{x}(t)=\bar{\varepsilon}(t)$, then for a fixed constant $m>0$, we have

$$\frac{1}{T}\sum_1^T\bar{\varepsilon}(t)\hat{x}(t-l)=\begin{cases}O(Q_T) & \text{uniformly in } l,\ 0<l\leqslant m\ln T,\\ O\left(\frac{\ln T}{T}\right)^{1/2} & \text{uniformly in, } 0<l.\end{cases}\tag{3.21}$$

In fact, the left hand side of (3.21) is

$$\begin{aligned}&\frac{1}{T}\sum_{t=1}^T\sum_{j=0}^{\infty}\alpha_j\hat{x}(t-j)\hat{x}(t-l)=\sum_{j=0}^{\infty}\alpha_j\frac{1}{T}\sum_{t=1}^T\hat{x}(t-j)\hat{x}(t-l)=\sum_{j=0}^{\infty}\alpha_jC_x(j-l)\\&=\sum_{j=0}^{\infty}\alpha_j\gamma_x(j-l)+\sum_{j=0}^{\infty}\alpha_j\triangle C_x(j-l)=\sum_{j=0}^{\infty}\alpha_j\Delta C_x(j-l).\end{aligned}$$

Here we notice that $\sum_{j=0}^{\infty}\alpha_j\gamma_x(j-l)=\gamma_{x\varepsilon}(l)=0$. Using the same technique as in Lemma 2.2′, (3.21) follows. The uniformity can be seen from the proof of that lemma.

Thus

$$\begin{aligned}\tilde{\sigma}^2_{pq}&=\frac{1}{T}\sum_1^T\left(\bar{\varepsilon}(t)+O(Q_T)\sum_{j=0}^{\infty}\phi_j\hat{x}(t-1-j)\right)^2\\&=\frac{1}{T}\sum_1^T\bar{\varepsilon}(t)^2+O(Q_T)\sum_{j=0}^{\infty}\phi_j\frac{1}{T}\sum_1^T\bar{\varepsilon}(t)\hat{x}(t-1-j)\\&\quad+O(Q_T^2)\sum_0^{\infty}\sum_0^{\infty}\phi_j\phi_l\frac{1}{T}\sum_{t=1}^T\hat{x}(t-1-j)\hat{x}(t-1-l).\end{aligned}\tag{3.22}$$

Because $|\phi_j|<\rho^j$, by a suitable choice of m, we can make

$$\sum_{j=m\ln T}^{\infty}|\phi_j|\leqslant T^{-1},$$

thus from (3.21), one can see that

$$\sum_{j=0}^{\infty}\phi_j\frac{1}{T}\sum_{1}^{T}\bar{\varepsilon}(t)\hat{x}(t-1-j)=\sum_{j=0}^{m\ln T-1}\cdots+\sum_{m\ln T}^{\infty}\cdots=O(Q_T),$$

and the second term of RHS in (3.21) is $O(Q_T^2)$. The third term of RHS in (3.22) is also $O(Q_T^2)$ by noticing that $|C_x(l)|\leqslant C_x(0)\to\gamma_x(0)$ a. s. and $|\phi_j|$ decreases exponencially. Thus we have, at last,

$$BIC(p,\ q)=\frac{1}{T}\sum_{1}^{T}\bar{\varepsilon}(t)^2+O\left(\frac{\ln\ln T}{T}\right)+\frac{(p+q)\ln T}{T},\ s\geqslant p\geqslant p_0,\ s\geqslant q\geqslant q_0,$$

which is minimized at $p=p_0$, $q=q_0$ for sufficient large T. That establishes the theorem.

Acknowledgement

We are grateful to Professor E. J. Hannan. Beased on his work and the stimulating discussion which was carried on when one of us was visiting Australian National University, this work was done.

References

[1] Hannan, E. J., The estimation of the order of an ARMA process. *Ann. Statist*, **8**(1980), 1071—1081.

[2] Hannan, E. J. and Rissanen, J., Recursive estimation of mixed autoregressive moving average order, Biometrika, **69**(1982), 81—94.

[3] Whittle, P., On the fitting of multivariate autroregression, and the approximate canonical factorization of a spectral density matrix, Aiometrika, **50**(1963), 129—134.

[4] An Hongzhi Chen Zhaoguo and Hannan, E. J., Autocorrelation, autoregression and autoregressive approximation, *Ann. Statist*, **10**(1982), 926—936.