This is to certify that the

dissertation entitled

ON PERIODIC AUTOREGRESSION: MAXIMUM ENTROPY MODELING AND PARAMETER ESTIMATION

presented by

Hao Zhang

has been accepted towards fulfillment of the requirements for

Doctor degree in Statistics

V. Mandrekar V. Mandrekar

Major professor

Date June 10, 1995

ON PERIODIC AUTOREGRESSION: MAXIMUM ENTROPY MODELING AND PARAMETER ESTIMATION

By

Hao Zhang

A DISSERTATION

Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree of

DOCTOR OF PHILOSOPHY

Department of Statistics and Probability

1995

ABSTRACT

ON PERIODIC AUTOREGRESSION: MAXIMUM ENTROPY MODELING AND PARAMETER ESTIMATION

By

Hao Zhang

We study a special class of periodically correlated time series for which the best linear predictor of $x(t)$ based on all past information , denoted by $\hat{x}(t)$, depends on at most p steps back, i.e., there exists an integer p such that

$$\hat{x}(t) = Proj(x(t)|x(t-1), \cdots, x(t-p)), \forall t.$$

We call such a periodically correlated process a periodic autoregression (PAR). A PAR is equivalent to the following time domain model:

$$x(t) - \sum_{j=1}^{p(t)} a(j,t)x(t-j) = \sigma(t)\epsilon(t),$$

where $p(t), \sigma(t), a(j,t)$ are all periodic in t and $\epsilon(t)$ is the innovation process.

We first show that Burg's maximum entropy principle can be generalized to periodically correlated case and the generalized principle results in a PAR.

We then study estimation problems in a PAR model. We show that the Yule-Walker equations provide consistent estimators for the coefficients $a(j,t)$. We also give a uniform convergence rate of the estimators. Finally, We generalize Akaike's Bayesian Information Criterion to give consistent estimation for the orders $p(1), p(2), \cdots, p(T)$.

To my wife and my parents

Acknowledgment

My deepest thanks go to my advisor, Dr. Mandrekar for his guidance and encouragement. His patience and thoughtfulness have made my stay much easier.

I also thank my thesis committee members, Drs. LePage, Salehi and Sledd for their time and interest.

I appreciate the support of the Department of Statistics and Probability in the last five years when I experienced growth in research, teaching and statistical consulting. Never forgotten will be the numerous discussions with many faculty members.

Finally and importantly, I thank my wife, Hong, for her support and encouragement, my parents and sisters for their long-time support.

Contents

Introduction

Time series analysis is one of the fields that have attracted interests of probabilistists, statisticians and researchers from economics, engineering, social sciences and other areas. Stationary time series has been studied extensively because of its application in many fields and the adequate mathematical tools to handle it. One of the most important class of stationary time series is ARMA models which are widely used in applications. Problems studied for ARMA models are parameter estimation, spectral estimation and prediction. AR models, as special cases of ARMA models, received more attention since stationary time series can be approximated by AR models (An, Chen and Hannan, 1982).

What makes AR models even more interesting is the application of information theory in the study of stationary time series. Burg (1967) used the idea of maximizing entropy in spectral estimation. He showed this approach results in an AR model. Burg's maximum entropy principle is better justified by Parzen (1983) and has been widely used today in spectral estimation. Akaike (1974) applied information theory to develop the well-known Akaike's Information Criterion (AIC) for order estimation in AR models. AIC tends to overestimate orders and yields inconsistent estimators (Shibata,1976). To get consistent estimators for the order in an AR model, Akaike (1977) later modified AIC to Bayesian Information Criterion (BIC). An, Chen and Hannan (1982) proved that BIC gives consistent order estimation.

Although stationary time series describe many phenomena well, there are situations when data exhibit non-stationarity. Efforts to study non- stationary processes can be found constantly in literature.

One natural generalization of stationary process is Loéve's harmonizable process. Spectral domain problems for harmonizable processes are studied by Cramér (1961). Cramér (1964) also showed that time domain problems are difficult for harmonizable processes. So far, no time domain model is discussed for general harmonizable processes.

The effort is thus concentrated on non-stationary process for which both time domain and spectral domain approaches can be applied. The major example in this direction is periodically correlated (PC) processes, which are harmonizable as shown by Gladyshev (1961). After the early work of Gudzenko (1959) and Gladyshev (1961), several authors have studied the Kolmogorov-Wiener problem (Miamee and Salehi,1980, Hurd and Mandrekar,1991).

Motivated from applications (e.g., Bloomfield *et al*, 1994, Gardner, 1986), Hurd (1989), Hurd and Gerr (1991) studied some inference problems for spectral measure of PC processes. Time domain models have been also studied for PC processes. Pagano (1978) introduced periodic autoregression (PAR). Anderson and Vecchia (1993) studied periodic ARMA model and gave the asymptotic properties of sample covariances. Adams and Goodwin (1995) studied on-line parameter estimation for periodic ARMA models.

We study PC time series systematically by first showing that the analogue of Burg's algorithm holds for PC time series. For this purpose, we introduce a slightly different definition of PAR. From this definition, we can easily show that PAR satisfies maximum entropy principle. Our definition reveals an intrinsic property of

PAR models and overcomes a technical difficulty encountered by Pagano's. These are all discussed in Chapter 2.

We then consider parameter estimation for PAR models in Chapter 3. Here, we study estimation of coefficients in PAR models. We give convergence rates for these estimates by first studying the convergence rate for sample covariances. Then we apply these rates to give consistent estimators for the order of PAR models. Our work generalizes Akaike's BIC order estimation to PAR model. Numerical results are shown at the end of Chapter 3.

Throughout this thesis, we assume all sequences of random variables have zero means. Let $\|A\|$ denote the sup-norm of a real matrix A and Z denote the set of integers.

Chapter 1

Periodic Autoregression

A sequence of real valued random variables $x = \{x(t), t \in Z\}$ is called a periodically correlated if for some integer T and any t, s,

$$Ex(t) = Ex(t+T),$$
$$Cov(x(t), x(s)) = Cov(x(t+T), x(s+T)).$$

We always assume that $Ex(t) = 0$ for all t. Then the definition is equivalent to that there exists a unitary operator $\mathbf{U}$ such that for some T,

$$x(t+T) = \mathbf{U}x(t), \forall t.$$

One chooses the minimum T as the period of the PC sequence. The best linear predictor of $x(t)$, given $x(s), s < t$, is the projection of $x(t)$ onto the closed space spanned by $x(s), s < t$ (Kolmogorov, 1941). Hereafter, we denote this projection by

$$Proj(x(t)|x(s), s < t).$$

From the definition of PC sequences, we see that if

$$Proj(x(t)|x(s), s < t) = \sum_{j=1}^{\infty} c_j(t)x(t-j),$$

then

$$Proj(x(t+T)|x(s), s < t+T) = \sum_{j=1}^{\infty} c_j(t)x(t+T-j).$$

In other words, the coefficients $c_j(t)$ are periodic in t for every $j \geq 1$. Whereas for general non-stationary sequences, nothing can be said about $c_j(t)$.

From the point view of prediction, the simplest class of PC sequences are those for which prediction depends only on finite history. This leads to the following definition.

Definition 1.0.1 *A PC sequence is called a periodic autoregression (PAR for abbreviation) if there exists an integer $p > 0$, such that for any t*

$$\hat{x}(t) = Proj(x(t)|x(t-1), x(t-2), \cdots, x(t-p)).$$

We do not consider the case $\hat{x}(t)$ is zero which is less significant in view of prediction. Then for each t, there is a smallest positive integer $p(t)$ satisfying

$$\hat{x}(t) = Proj(x(t)|x(t-1), x(t-2), \cdots, x(t-p(t))).$$

We will call $p(t)$ the order of $x(t)$. We will show in Proposition 1.0.1 that $p(t)$ is periodic, then we can simply say that $x(t)$ has order $p(1), p(2), \cdots, p(T)$. We can write

$$\hat{x}(t) = \sum_{j=1}^{p(t)} a(j,t)x(t-j),$$

or equivalently

$$x(t) - \sum_{j=1}^{p(t)} a(j,t)x(t-j) = \epsilon(t), \tag{1.1}$$

where $\epsilon(t) = x(t) - \hat{x}(t)$. Thus $\epsilon(t)$ is orthogonal to $x(s)$ for all $s < t$.

Let

$$\sigma^2(t) = E|x(t) - \hat{x}(t)|^2.$$

If $\sigma^2(t) > 0$ for every t, we say that $x(t)$ is non-deterministic. In this case, we can write (1.1) as

$$x(t) - \sum_{j=1}^{p(t)} a(j,t)x(t-j) = \sigma(t)\epsilon(t), \tag{1.2}$$

where $\{\epsilon(t)\}$ is a the innovation sequence, i.e.,

$$\epsilon(t) = \frac{x(t) - \hat{x}(t)}{\sigma(t)}.$$

The next proposition says all parameters are periodic in t.

Proposition 1.0.1 *If $x(t)$ is a non-deterministic PAR, then the functions $p(\cdot)$,$a(\cdot,\cdot)$ and $\sigma^2(t)$ are unique and for any t*

$$\begin{aligned}
p(t) &= p(t+T) \\
\sigma^2(t) &= \sigma^2(t+T) \\
a(j,t) &= a(j,t+T)
\end{aligned}$$

Proof. $p(.)$ and $\sigma^2(.)$ are obviously unique by definition. Since x is non-deterministic, it follows that $x(t-1), ..., x(t-p(t))$ are linearly independent and thus $a(\cdot,\cdot)$ is unique. Notice that $x(t)$ is periodically correlated if and only if there exists a unitary operator $\mathbf{U}$ such that

$$\mathbf{U}x(t) = x(t+T).$$

It is very easy to justify

$$\mathbf{U}Proj(y|\mathcal{S}) = Proj(\mathbf{U}y|\mathbf{U}\mathcal{S})$$

for any y and (closed) subspace $\mathcal{S}$, where $\mathbf{U}\mathcal{S} = \{\mathbf{U}x : x \in \mathcal{S}\}$.

It follows that

$$\begin{aligned}
\mathbf{U}Proj(x(t)|x(s), s < t) &= Proj(x(t+T)|x(s+T), s < t) \\
\mathbf{U}Proj(x(t)|x(s), t-p \le s \le t) &= Proj(x(t+T)|x(s+T), t-p \le s \le t).
\end{aligned}$$

We conclude from the above two equations that

$$p(t) = p(t+T), \quad \sigma^2(t) = \sigma^2(t+T)$$

Since

$$\hat{x}(t) = \sum_{j=1}^{p(t)} a(j,t)x(t-j),$$

then

$$\mathbf{U}\hat{x}(t) = \sum_{j=1}^{p(t)} a(j,t)x(t-j+T).$$

On the other hand,

$$\mathbf{U}\hat{x}(t) = \hat{x}(t+T) = \sum_{j=1}^{p(t+T)} a(j,t+T)x(t+T-j).$$

From the facts that $p(t) = p(t+T)$ and $x(s)$, for $t+T-p(t) \leq s \leq t+T-1$ are linearly independent, we obtain

$$a(j,t+T) = a(j,t), \forall j, t.$$

QED.

We see that for non-deterministic PC sequences, our definition is equivalent to (1.2) with periodic parameters, which was used to define periodic autoregression by Pagano (1978). The definition here overcomes a technical difficulty. (See the remark at the end of this section.) We are ready to give a characterization of PAR model now. Let

$$\vec{x}(t) = (x(1+tT), x(2+tT), \cdots, x(T+tT)\ )', \ \forall\ t,$$

$$R(t,s) = Ex(t)x(s).$$

Theorem 1.0.1 *Let x be a PC sequence. Then x is a non-deterministic PAR is equivalent to the followings.*

(i) *$x(t)$ satisfies the following time domain model*

$$x(t) - \sum_{j=1}^{p(t)} a(j,t)x(t-j) = \sigma(t)\epsilon(t), \tag{1.3}$$

where $\epsilon(t)$ is the innovation process and $\sigma(t) > 0$.

(ii) *The Yule-Walker equations hold, i.e.,*

$$R(t-k,t) - \sum_{j=1}^{p(t)} a(j,t)R(t-k,t-j) = \sigma^2(t)\delta_{k,0} \tag{1.4}$$

for any $k \geq 0$, $\forall t$.

(iii) *$\vec{x}(t)$ is T dimensional $AR(p)$ model for*

$$p = \max_{1 \leq t \leq T} [\frac{p(t)-t}{T}] + 1.$$

(where $[a]$ denotes the integer part of a.) Namely, there exist matrixes A_j and G and a T dimensional time series $\vec{\epsilon}(t)$ such that

$$\vec{x}(t) - \sum_{j=1}^{p} A_j \vec{x}(t-j) = \vec{\epsilon}(t) \quad \text{and} \tag{1.5}$$

$$Cov(\vec{\epsilon}(t), \vec{\epsilon}(s)) = G\delta_{t,s}, \quad E\vec{\epsilon}(t)\vec{x}(s)' = 0, \forall s < t. \tag{1.6}$$

Proof. It is obvious that (i) and (ii) are equivalent and are necessary and sufficient conditions for x to be PAR by definition 1.

Let us prove (i)$\Rightarrow$(iii). Suppose x is PAR. Then x has the time domain expression given by (1.3) which can be written in vector form

$$B_0\vec{x}(t) - \sum_{j=1}^{p} B_j \vec{x}(t-j) = D\vec{\epsilon}(t) \tag{1.7}$$

where p is defined in the theorem and D=diag$(\sigma(1), ..., \sigma(T))$, $B_j = (b_j(k,l))_{k,l=1}^{T}$ for

$$\begin{aligned} b_0(k,l) &= \delta_{k,l} - a(k-l,l)\chi_{\{k>l\}} \\ b_j(k,l) &= a(Tj+k-l,k)\chi_{\{Tj+k-l \leq p(k)\}}, \quad 1 \leq j \leq p. \end{aligned}$$

Let $\vec{\epsilon}(t) = (\epsilon(1+tT), \epsilon(2+tT), \cdots, \epsilon(T+tT))'$. Then it is obvious that $\vec{\epsilon}(t) \perp \vec{x}(s)$ for any $s < t$. Let

$$A_j = B_0^{-1} B_j, \quad \vec{\varepsilon}(t) = B_0^{-1} D \vec{\epsilon}.$$

Then $\vec{\varepsilon}(t)$ satisfies (1.6) with $G = B_0^{-1} D^2 (B_0^{-1})'$ and (1.5) follows from (1.7).

(iii)$\Rightarrow$ (i): Since G is positive definite, it has the following Cholesky decomposition

$$G = LHL'$$

where L is lower triangular with all diagonal entries being 1 and H is diagonal and non-singular. Multiplying (1.5) by L^{-1}, we get

$$L^{-1}\vec{x}(t) - \sum_{j=1}^{p} L^{-1} A_j \vec{x}(t-j) = L^{-1}\vec{\varepsilon}.$$

Since L^{-1} is also lower triangular and Var$(L^{-1}\vec{\varepsilon}) = H$ is a diagonal matrix, the scalar form of the above equation will yield (i).

QED.

Corollary 1.0.1 *A nondeterministic PAR is purely nondeterministic.*

Proof. Let x be a nondeterministic PAR. Then the corresponding multiple sequence $\vec{x}$ is a stationary AR. It is well known multiple AR is purely nondeterministic. In [19], and [22], it is proved that x is purely nondeterministic if and only if $\vec{x}$ is so. Thus the proof is completed.

QED.

Remark. A stationary AR(p) model is defined in literature (Hannan, 1973) as a second order stationary sequence satisfying

$$x(t) - \sum_{j=1}^{p} a_j x(t-j) = \sigma\epsilon(t),$$

for some constants $\sigma > 0, a_j$ such that

$$|1 - \sum_{j=1}^{p} a_j z^j| \neq 0, \text{ for } |z| \leq 1.$$

and a white noise $\{\epsilon(t)\}$. The constraint for the coefficients is a necessary and sufficient condition for the existence of a solution of stationary sequence (see, *e.g.*, Hannan, 1970). Analogously, we need to know constraints on the coefficients $a(j,t)$ to guarantee a PC solution of (1.3) exists. We can give the constraint in two equivalent ways. We note that (1.3) with periodic parameters has a non-deterministic PC solution if and only if $\{p(t), \sigma^2(t), a(j,t), j = 1, \cdots, p(t), t = 1, 2, \cdots, T\}$ uniquely determines $R(t,s)$ for $|t-s| \leq p(t)$ such that

$$R(t,s) = R(s,t) = R(t+T, s+T)$$

and for any $t = 1, 2, \cdots, T$, the matrix

$$\Gamma_t = (R(t-j, t-k))_{j,k=0}^{p(t)} \tag{1.8}$$

is positive definite. The necessity is obvious and sufficiency follows from Theorem 2.2.1 in the next chapter.

We also see from Theorem 1.0.1 that (1.3) has a PC solution if and only if

$$\det(I - \sum_{j=1}^{p} A_j z^j) \neq 0, \forall\, |z| \leq 1. \tag{1.9}$$

Indeed, (1.9) implies there is a stationary solution of (1.5)(Hannan, 1970, page 326). The corresponding scalar sequence must be a PAR satisfying (1.3) by Theorem 1.0.1 It must be non-deterministic since $\sigma(t)$ is positive. We use the technique of Whittle (1963) to show the other way. Suppose that (1.3) has a PC solution. Then the corresponding vector-valued stationary sequence satisfies (1.5) and (1.6).

Define the pT dimensional random vector

$$Y_t = \begin{pmatrix} \vec{x}(t) \\ \vec{x}(t-1) \\ \vdots \\ \vec{x}(t-p+1) \end{pmatrix}$$

Project Y_t such that

$$Y_t = PY_{t-1} + Z_t, \quad Z_t \perp Y_{t-1}$$

and P is the projection matrix. We see from (1.5) that

$$P = \begin{pmatrix} A_1 & A_2 & \cdots & A_{p-1} & A_p \\ I & \cdot & \cdots & \cdot & \cdot \\ \cdot & I & \cdots & \cdot & \cdot \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ \cdot & \cdot & \cdot & I & \cdot \end{pmatrix}$$

$$Z_t = (\vec{\varepsilon}(t)', 0, ..., 0)'.$$

Let λ be an eigenvalue of P and ξ be the corresponding left eigenvector. Then

$$\xi P = \lambda \xi. \tag{1.10}$$

Observe that $Cov(Y_t) = Cov(Y_{t-1})$ is positive definite , then

$$\begin{aligned} E|\xi Z_t|^2 &= E|\xi Y_t|^2 - E|\xi P Y_{t-1}|^2 \\ &= \xi Var(Y_t)\xi'(1 - |\lambda|^2) \geq 0. \end{aligned}$$

it follows that $|\lambda| \leq 1$ with equality if and only if $\xi Z_t = 0$. Because of the special form of P, (1.10) implies that ξ and ξ_0 must be 0 together where ξ_0 is the first T entries of ξ. Since $Var(\vec{\varepsilon}_t)$ is non-singular, for non-zero ξ,

$$\xi Z_t = \xi_0 \vec{\varepsilon}_t \neq 0,$$

So all the eigenvalues of P have modules less than 1. Now suppose z satisfies

$$\det(I - \sum_{j=1}^{p} A_j z^{-j}) = 0.$$

It suffices to prove z^{-1} is an eigenvalue of P. There exists an row vector $\xi_1 \in R^T$ such that

$$\xi_1(I - \sum_{j=1}^{p} A_j z^{-j}) = 0. \tag{1.11}$$

Let

$$\xi_j = z\xi_{j-1} - \xi_1 A_{j-1}, \quad j = 2, ..., p. \tag{1.12}$$

Set

$$\xi = (\xi_1, ..., \xi_p).$$

Notice (1.11) and (1.12) imply immediately

$$\xi P = z\xi.$$

(1.9) now follows.

Chapter 2

Maximum Entropy Modeling of PC Time Series

2.1 Introduction

The entropy of a random vector in $\mathcal{R}^n$ with probability density function $f(x)$ is defined as

$$H(X) = -E \ln f(X) = -\int_{\mathcal{R}^n} f(x) \ln f(x) dx.$$

Burg (1967) developed a maximum entropy approach for spectral estimation of stationary time series which has been widely used since then. Burg's approach can be stated in the following way. Suppose $p+1$ autocovariances $R(0), R(1), \cdots, R(p)$ of a stationary sequence are known (usually estimated from observations). Instead of taking $R(n)$ to be 0 for all n greater than p, as in windowed spectral estimation, we extrapolate $R(n)$ for $n > p$ in such a way that maximizes the entropy

$$H(x(t), x(t-1), \cdots, x(t-n)),$$

for all $n > p$.

It turns out that the only such extrapolation is given by Yule-Walker equations, thus this maximum entropy method results in an AR model.

We consider here the same question for PC sequence. Suppose for each $t = 1, 2, \cdots, T$, we know the covariance matrix of

$$(x(t), x(t-1), \cdots, x(t-p_t))$$

for some integers $p_t > 0$. Because the time series is PC, we do not require that the p_t's are the same. We will extrapolate the covariance function in such a way that maximizes the entropy

$$H(x(t), x(t-1), \cdots, x(t-s))$$

for all $s < t$. Problems we will consider are

(1). whether there is a PC solution to this maximizing problem and

(2). the properties of such PC sequences which maximize the entropies.

We will prove in the next section that there is a unique Gaussian PAR sequences which maximizes the entropies.

2.2 Maximum Entropy Modeling of PC Time Series

To avoid the ambiguity of saying part of covariances of a sequence is known without knowing such sequence exists, we state the problem in a more mathematical way. Let $p_1, p_2, \cdots, p_T$ be positive integers and $r(.,.)$ be defined on the set

$$\Delta = \cup_{t=1}^{T}\{(u,v) \in Z \times Z : t - p_t \leq u, v \leq t\}.$$

We assume that

$$\Gamma_t = (r(t-j, t-k))_{j,k=0}^{p_t}$$

is positive definite for all $t = 1, 2, \cdots, T$ and

$$r(s,t) = r(s+T, t+T) = r(t,s) \tag{2.1}$$

whenever $(s,t), (t,s)$ and $(s+T, t+T)$ are in Δ.

These assumptions are seen necessary for r to be a covariance function of a PC time series. Let $\mathcal{K}$ be the set of all PC time series with period T whose covariances are $r(t,s)$ for $(t,s) \in \Delta$. The next theorem says $\mathcal{K}$ is not empty.

Theorem 2.2.1 *There is a non-deterministic Gaussian PAR in $\mathcal{K}$.*

Proof. Since Γ_t is positive definite, the equations

$$r(t-k,t) - \sum_{j=1}^{p_t} a(j,t) r(t-k, t-j) = \delta_{k,0}\sigma^2(t), \text{ for } k = 0, 1, \cdots, p_t, \tag{2.2}$$

have unique solution $a(1,t), a(2,t), \cdots, a(p_t,t), \sigma^2(t)$ and $\sigma^2(t) > 0$. These Yule-Walker equations actually provide a way to extend $r(t,s)$ to be a covariance function of a PC time series. But we will adapt a statistical approach here.

Let $i \in \{1, 2, \cdots, T\}$ be such that

$$i - p_i \le t - p_t \text{ for } \forall t = 1, 2, \cdots, T.$$

Then there are Gaussian random variables $x(t)$ of 0 mean, for $i - p_i \le t \le i - 1$, such that

$$Ex(t)x(s) = r(t,s), \text{ for } i - p_i \le t, s \le i - 1.$$

Let $\epsilon(t), t \ge i$ be a sequence of i.i.d standard normal random variables and also independent of $\{x(t), i - p_i \le t \le i - 1\}$. Define, for $t \ge i$,

$$x(t) = \sum_{j=1}^{p(t)} a(j,t) x(t-j) + \sigma(t)\epsilon(t)$$

where $p(t), a(j,t)$ and $\sigma(t)$ are the periodic version of $p_t, a(j,t)$ and $\sigma(t)$ respectively.

This definition together with (2.2) yield

$$Ex(t)x(s) = r(t,s), \text{ for } (t,s) \in \Delta.$$

We now show that $x(t)$ is PC, *i.e.*,

$$Ex(t+T)x(s+T) = Ex(t)x(s), \tag{2.3}$$

for $\forall t, s \geq i - p_i$. We observe that (2.3) is true for $i - p_i \leq t, s \leq 0$, because of (2.1). Assume it is true for $i - p_i \leq t, s \leq n$. Replacing $x(n+1)$ by the definition, we have for $t < n+1$,

$$\begin{aligned} Ex(t)x(n+1) &= \sum_{j=1}^{p(n+1)} Ex(t)x(n+1-j) \\ &= \sum_{j=1}^{p(n+1)} Ex(t+T)x(n+1+T) \quad \text{(by the induction assumption)} \\ &= Ex(t+T)x(n+1+T) \end{aligned}$$

Similarly, we can prove

$$Ex^2(n+1) = Ex^2(n+1+T).$$

Thus (2.3) is true for $t, s \leq n+1$. So we have proved (2.3).

Let

$$r(t,s) = Ex(t)x(s)$$

for all $t, s \geq i - p_i$. Then $r(t,s)$ still satisfies (2.1). Now we extend $r(t,s)$ to Z^2 by

$$r(t - mT, s - mT) = r(t,s), \forall m \geq 0.$$

For any $m \leq n$, the matrix

$$\{r(t,s) : m \leq t, s \leq n\}$$

is positive definite because of the periodicity of $r(t,s)$ and the fact that $\{x(t), t \geq 0\}$ are linearly independent. So there is a Gaussian sequence with $r(t,s)$ as covariance function by Kolmogorov's Theorem. This sequence must be a PAR by Theorem 2 and non-deterministic since $\sigma^2(t) > 0$.

QED.

This Gaussian PAR must be unique in distribution. It might have orders $p(t) \leq p_t$ for $t = 1, 2, \cdots, T$ because $a(j, p_t)$ might be zero. But the orders are uniquely determined by Γ_t's.

The next theorem says that it is the one that maximizes entropy.

Theorem 2.2.2 *Let $x(t)$ be a Gaussian PAR in $\mathcal{K}$, then for any $s \leq t$,*

$$H(x(t), x(t-1), ..., x(s)) = \sup_{Y \in \mathcal{K}} H(y(t), y(t-1), ..., y(s)) \tag{2.4}$$

where the supremum is taken over all sequences Y in $\mathcal{K}$ for which the entropies in (2.4) can be defined.

Conversely, if a PC sequence $y(t)$ in $\mathcal{K}$ satisfies (2.4), *then $y(t)$ is a Gaussian PAR.*

Remark. The problem we considered here is more general than assuming $\mathbf{R}_0, \mathbf{R}_1, \cdots, \mathbf{R}_p$, the covariance of a vector-valued stationary sequence, are known. Actually, the later is a special case of our problem here. Thus Theorem 3 contains the maximum entropy method for stationary vector-valued sequence as a special case. There is a practical consideration why we assume $r(t,s)$ is known in the set Δ instead of a square area $\{1-q \leq t, s \leq T\}$ for some $q > 0$. To approximate a PC sequence using a PAR, we might choose different orders $p(1), p(2), \cdots, p(T)$. Given finite observations of a PC sequence, we should just estimate $r(t,s)$ for $(t,s) \in \Delta$ and extrapolate it through Yule-Walker equations since smaller $|t-s|$ tends to give better estimate of $r(t,s)$.

To prove this theorem, we need some basic properties of entropy. The following two lemmas are known.(*e.g.*, for Lemma 1, Choi, 1983, Parzen, 1983; for Lemma 2, Gallager, 1965, Kullback, 1978.)

Lemma 2.2.1 *For any random vector Y, let X be a normal random vector having the same covariance matrix as Y. Then*

$$H(Y) \leq H(X)$$

The equality holds if and only if Y has normal distribution.

For two random vectors with joint probability function $f(x,y)$, the conditional entropy of X given Y is defined as, provided it exists,

$$H(X|Y) = -\int \ln f(x|y) f(x,y) dx dy.$$

where $f(x|y)$ is the conditional probability density of X given Y. Thus

$$H(X|Y) = H(X,Y) - H(Y). \tag{2.5}$$

We use $H(X|Y,Z)$ to denote the conditional entropy of X given Y and Z. $H(X|Y)$ can be interpreted as the remaining uncertainty of X given Y is observed. Then the following lemma is clear intuitively based on this interpretation.

Lemma 2.2.2 *For any three random vectors X, Y, Z with joint probability density function $f(x,y,z)$,*

$$H(X|Y,Z) \leq H(X|Y),$$

with equality if and only if X and Y are independent conditionally on Z, i.e.,

$$f(x|y,z) = f(x|y), \ \textit{a.e.}$$

Proof of Theorem 2.2.2. We know from Lemma 1 that we should look for a maximizer of entropies among Gaussian sequences. Let $y(t)$ be a Gaussian sequence in $\mathcal{K}$. Using (2.5) repeatedly, we get for any $s < t$,

$$\begin{aligned}
&H(y(t), y(t-1), \cdots, y(s)) \\
&= H(y(s)) + \sum_{u=s+1}^{t} H(y(u)|y(u-1), \cdots, y(s)).
\end{aligned}$$

We see that for those terms for which $u \geq s \geq u - p(u)$, the conditional entropies are known. (since y is Gaussian and the covariance matrix is known.)

Thus, to find a maximizer is to maximize, for $s < u - p(u)$,

$$\begin{aligned}
&H(y(u)|y(u-1), \cdots, y(s)) \\
&\leq H(y(u)|y(u-1), \cdots, y(u-p(u))) \quad \text{(by Lemma 2)} \qquad (2.6) \\
&= H(x(u)|x(u-1), \cdots, x(y-p(u)))
\end{aligned}$$

The last equality is true because $(x(u), x(u-1), \cdots, x(y-p(u)))$ and $(y(u)|y(u-1), \cdots, y(u-p(u)))$ have the same Gaussian distribution. We see this upper bound is reached by the Gaussian PAR $x(t)$. So $x(t)$ maximizes the entropy.

Conversely, if a Gaussian sequence y in $\mathcal{K}$ maximizes all the entropies, then the equality in (2.6) must hold. Then, following Lemma 2,

$$\begin{aligned}
&f(y(u)|y(u-1), \cdots, y(s)) \\
&= f(y(u)|y(u-1), \cdots, y(u-p(u))).
\end{aligned}$$

In terms of conditional expectation, it means

$$\begin{aligned}
&E(y(u)|y(u-1), \cdots, y(s)) \\
&= E(y(u)|y(u-1), \cdots, y(u-p(u))).
\end{aligned}$$

Since y has Gaussian distribution, the conditional expectation is the projection onto corresponding space. Then the last equality is exactly

$$\begin{aligned}
& Proj(y(u)|y(u-1),\cdots,y(s)) \\
&= Proj(y(u)|y(u-1),\cdots,y(u-p(u))).
\end{aligned}$$

so y is a PAR.

QED

Finally, we note that MEM picks up the most random or the most unpredictable time series. It is much clearer to state it terms of prediction. Let $\sigma^2(t;x)$ be the mean square of prediction error defined by

$$\sigma^2(t;x) = E|x(t) - Proj(x(t)|x(s), s < t)|^2$$

for a sequence $\{x(t), t \in Z\}$ in $\mathcal{K}$.

Theorem 2.2.3 *$x \in \mathcal{K}$ is a PAR if and only if*

$$\sigma^2(t;x) \geq \sigma^2(t;y), \forall y \in \mathcal{K}.$$

Proof. Let $x(t)$ be the PAR in $\mathcal{K}$ and $y(t)$ be any sequence in $\mathcal{K}$, we have for any $t = 1, 2, \cdots, T$,

$$\begin{aligned}
\sigma^2(t;y) &\leq E|y(t) - \sum_{j=1}^{p(t)} a(j,t)y(t-j)|^2 \qquad (2.7) \\
&= r(t,t) - \sum_{j=1}^{p(t)} a(j,t)r(t,t-j) = \sigma^2(t;x)
\end{aligned}$$

Conversely, if $y(t)$ satisfies

$$\sigma^2(t;y) = \sup_{\xi \in \mathcal{K}} \sigma^2(t;\xi),$$

The equality must hold in (2.7). It follows that

$$\sum_{j=1}^{p(t)} a(j, L(t))y(t-j) = Proj(y(t)|y(s), s < t).$$

So $y(t)$ is a PAR.

QED.

Chapter 3

Parameter Estimation in PAR Model

In this Chapter, we consider parameter estimation in PAR. We assume, throughout this Chapter, that $x(1), x(2), \cdots, x(N)$ are observations from a non-deterministic PAR $\{x(t) : t = 0, \pm 1, \pm 2, \cdots\}$.

Define the sample covariance by

$$R_N(t,s) = [NT^{-1}]^{-1} \sum_{j \in D(t,s)} x(t+jT)x(s+jT) \tag{3.1}$$

for $t = 1, 2, \cdots, T$, $s = 0, 1, \cdots, N-t-1$, where $D(t,s) = \{j : \max(t,s) + jT \leq N\}$. $R_N(.,.)$ can be extended by

$$R_N(t,s) = R_N(s,t), \quad R_N(t+kT, s+kT) = R_N(t,s)$$

Then $R_N(t,s)$ serves as estimator of $R(t,s)$. In Theorem 3.2.1, we give the uniform convergence rate of these estimators. The solution of Yule-Walker equation with R replaced by R_N estimates regression coefficients. Uniform convergence rate of them is given in Theorem 3.3.1. In the last part of this Chapter, we generalize

Akaike's BIC for stationary AR model to PAR model to get consistent estimator of order $p(1), p(2), \cdots, p(T)$.

The following assumption is made throughout this Chapter. Since we will consider properties of sample covariances, we assume the moments of order more than 4 exist and are bounded. For two sequences of real numbers a_n and b_n, $a_n \approx b_n$ means $a_n = O(b_n)$ and $b_n = O(a_n)$.

Assumption 3.0.1 *$\{\epsilon(t) : t \in Z\}$ is the innovation process of $\{x(t) : t \in Z\}$ and*

$$\begin{aligned} &E(\epsilon(t) \mid \mathcal{F}_{t-1}) = 0, \\ &E(\epsilon^2(t) \mid \mathcal{F}_{t-1}) = 1, \quad a.s \\ &\sum_{t=1}^{n} V(\epsilon^2(tT|\mathcal{F}_{tT-T}) \approx n, a.s \end{aligned}$$

where $\quad \mathcal{F}_t = \sigma\{\epsilon(s) : s \le t\}$*. Suppose also that for some $\delta > 0$,*

$$\sup_{t \in Z} E|\epsilon(t)|^{4+\delta} < \infty$$

3.1 Preliminary Results on Martingale Differences

In this section, we give some results of sample covariances of a martingale difference. We need this to derive results for PAR model.

We will give a law of the iterated logarithm for martingale difference and apply it to give the convergence rate of covariance of martingale difference. For stationary and ergodic martingale difference, a law of the iterated logarithm has been given in literature. (See, for example,[30]). But we assume higher moments instead of stationarity and ergodicity since we believe that the assumption of higher moments being bounded is less restrictive in time series than the assumption of stationarity and ergodicity.

Theorem 3.1.1 *Let $\{Y_n, \mathcal{F}_n, n \geq 1\}$ be a martingale difference such that for some $\delta > 0,\ M > 0$ and for any n,*

$$E|Y_n|^{2+\delta} \leq M$$

Let $s_n^2 = \sum_{i=1}^n E\{Y_i^2|\mathcal{F}_{i-1}\}$. Suppose also that

$$\liminf_{n\to\infty} \frac{s_n^2}{n} > 0, \quad a.s.$$

Then almost surely

$$\limsup_{n\to\infty} \frac{\sum_{i=1}^n Y_n}{\sqrt{2s_n^2 \ln\ln s_n^2}} = 1$$

Proof. Stout [29] proved that for martingale difference Y_n, the following law of the iterated logarithm holds

$$\limsup_{n\to\infty} \frac{\sum_{i=1}^n Y_n}{\sqrt{2s_n^2 \ln\ln s_n^2}} = 1$$

if $s_n^2 = \sum_{i=1}^n E(Y_n^2|\mathcal{F}_{i-1}) \to \infty$ and there exists a sequence K_n which is $\mathcal{F}_{n-1}$ measurable and goes to zero such that

$$\sum_{i=1}^{\infty} v_n^{-2} E(Y_n^2 \chi\{Y_n^2 > v_n^2\}) < \infty, \tag{3.2}$$

where

$$v_n^2 = \frac{K_n s_n^2}{\ln\ln s_n^2}.$$

In our theorem, it is obvious that $s_n^2 \to \infty$. Take

$$K_n = \frac{1}{\ln\ln s_n^2}.$$

We only need to check (3.2). Since

$$\begin{aligned} & E(Y_n^2 \chi\{Y_n^2 > v_n^2\}) \\ & \quad \leq (E|Y_n|^{2+p}) v_n^{-p} \leq M \cdot s_n^{-p}, \end{aligned}$$

we see that the sum in (3.2) is bounded by

$$M \sum_{n=1}^{\infty} s_n^{-2-p}$$

which is finite almost surely because $\liminf s_n^2/n > 0$. So the theorem follows.

QED.

We first state a lemma which is needed to prove our next theorem.

Lemma 3.1.1 *Let $\{Y_n, \mathcal{F}_n, n \geq 1\}$ be a supermartingale difference with $EY_1 = 0$ and for some $K \in (0,1/2]$,*

$$Y_n \leq K \frac{s_n}{\sqrt{2 \ln \ln s_n^2}} \qquad a.s$$

where $s_n^2 = \sum_{i=1}^n E\{Y_i^2 | \mathcal{F}_{i-1}\}$. Suppose for some constants $b \geq 9$, almost surely,

$$s_n^2 \leq b^2, \forall n.$$

Then for any $0 < \delta < 2$,

$$P(\sup_{n\geq 1} \sum_{i=1}^n Y_i > \delta\{2b^2 \ln \ln b^2\}^{1/2}) \leq \exp(-\beta \ln \ln b^2)$$

where $\beta = \delta^2(1 - \frac{\delta K}{2})$.

Proof. Let

$$c = Kb/\sqrt{2 \ln \ln b^2}, \quad \lambda = \delta b^{-1} \sqrt{2 \ln \ln b^2}.$$

Then

$$\lambda c = \delta K \leq 1.$$

Since $x(x \ln \ln x)^{-1/2}$ is increasing for $x \geq 9$, we have

$$Y_n \leq c, \forall n.$$

It follows [31] (Lemma 5.4.1 on page 299) that

$$T_n = \exp(\lambda \sum_{i=1}^{n} Y_i) \exp(-\frac{\lambda^2}{2}(1 + \frac{\lambda c}{2}) s_n^2).$$

is a super-martingale w.r.t $\mathcal{F}_n$ and $ET_1 \leq 1$. Thus for any $\alpha > 0$,

$$P(\sup_n T_n > \alpha) \leq 1/\alpha.$$

Then

$$\begin{aligned}
&P(\sup_n \sum_{i=1}^{n} Y_i > \delta(2b^2 \ln \ln b^2)^{1/2}) \\
&= P(\sup_n \sum_{i=1}^{n} Y_i > \lambda b^2) \\
&\leq P(\sup_n T_n > \exp(\lambda^2 b^2 - \frac{2}{b^2}(1 + \frac{\lambda c}{2}) b^2) \\
&\leq \exp(-\lambda^2 b^2 (1/2 - \lambda c/4)) = \exp(-\beta \ln \ln b^2).
\end{aligned}$$

QED.

Theorem 3.1.2 *If* $\{\epsilon(t), \mathcal{F}_t, t \geq 1\}$ *satisfies Assumption* 3.0.1, *then for any positive real number* d *and integer* T,

$$\limsup_{n \to \infty} \frac{|\sum_{s=1}^{n} \epsilon^2(sT) - n|}{\sqrt{2n \ln \ln n}} < \infty, \tag{3.3}$$

$$\limsup_{n \to \infty} \frac{\max_{0<t<d \ln n} |\sum_{s=1}^{n} \epsilon(sT)\epsilon(sT+t)|}{\sqrt{2n \ln \ln n}} \leq \sqrt{2}. \tag{3.4}$$

Proof. Let

$$Y(s) = \epsilon^2(sT) - 1.$$

Notice that Assumption 3.0.1 implies $\{Y(s), \mathcal{F}_{sT}, s \geq 1\}$ is a martingale difference and

$$\sup_t E|Y(t)|^{2+\delta/2} < \infty,$$

Applying Theorem 3.1.1 to $\{Y(s), \mathcal{F}_{sT}, s \geq 1\}$ and the fact

$$\sum_{s=1}^{n} E(Y^2(s)|\mathcal{F}_{sT-T}) \approx n,$$

$$\limsup_{n\to\infty} \frac{\sum_{s=1}^{n} Y(s)}{\sqrt{2n \ln\ln n}} < \infty. \tag{3.5}$$

(3.3) is proved now. To prove (3.4), we need first truncate $\epsilon(s)$. Let

$$\begin{aligned}
\lambda(s) &= \left(\frac{s}{\ln\ln s}\right)^{1/2} \frac{1}{\ln s} \\
\xi(s) &= \epsilon(s)\chi_{\{|\epsilon(s)|<\sqrt{\lambda(s)}\}} \\
\eta(s) &= \epsilon(s)\chi_{\{|\epsilon(s)|\geq\sqrt{\lambda(s)}\}}
\end{aligned}$$

For a fixed $t > 0$, let

$$\begin{aligned}
\mathcal{F}'_s &= \mathcal{F}_{sT+t\vee T} \\
Y_t(s) &= \xi(sT)\xi(sT+t) - E\{\xi(sT)\xi(sT+t)|\mathcal{F}'_{s-1}\}
\end{aligned}$$

Then $\{Y_t(s), \mathcal{F}'_s, s \geq 1\}$ is a martingale difference. We will finish the proof by proving the following,

$$\max_{0<t<d\ln n} \sum_{s=1}^{n} |\epsilon(sT)\epsilon(sT+t) - Y_t(s)| = o(\sqrt{n}), \tag{3.6}$$

$$\limsup_{n\to\infty} \max_{0<t<d\ln n} \frac{|\sum_{s=1}^{n} Y_t(s)|}{\sqrt{2n\ln\ln n}} \leq \sqrt{2}. \tag{3.7}$$

To prove (3.6), we first note that

$$E|\eta(s)|^2 \le E|\epsilon(s)|^{4+\delta}\lambda(s)^{-1-\delta/2}.$$

It follows

$$\sum_{s=1}^{\infty} \frac{E|\eta(s)|^2}{\sqrt{s}} < \infty. \tag{3.8}$$

Then Kronecker's lemma implies

$$\sum_{s=1}^{n} \eta(s)^2 = o(\sqrt{n}) \quad a.s. \tag{3.9}$$

and

$$\sum_{s=1}^{n} E(\eta^2(s)|\mathcal{F}'_{s-1}) = o(\sqrt{n}), a.s. \tag{3.10}$$

We will show first

$$\begin{aligned} &\max_{1\le t\le d\ln n} \sum_{s=3}^{n} |\epsilon(sT)\epsilon(sT+t) - \xi(sT)\xi(sT+t)| \\ &= \max_{1\le t\le d\ln n} \sum_{s=3}^{n} |\eta(sT)\eta(sT+t) + \eta(sT)\xi(sT+t) + \xi(sT)\eta(sT+t)| \\ &= o(\sqrt{n}) \end{aligned} \tag{3.11}$$

Applying Holder's inequality and (3.9), we get

$$\begin{aligned} &\max_{1\le t\le d\ln n} \sum_{s=3}^{n} |\eta(sT)\eta(sT+t)| \\ &\le \sum_{s=1}^{nT+d\ln n} \eta^2(s) = o(\sqrt{nT + d\ln n}) = o(\sqrt{n}) \end{aligned} \tag{3.12}$$

Since $\sqrt{\ln\ln s}\cdot \ln s$ is increasing and

$$(s+t)^{1/4} \le s^{1/4} + t^{1/4},$$

then

$$\lambda^{1/2}(s+t) \le \lambda^{1/2}(s) + \lambda^{1/2}(t). \tag{3.13}$$

Because $\xi(s) \le \lambda^{1/2}(s)$, (3.13) implies

$$\sum_{s=3}^{n} |\eta(sT)\xi(sT+t)| \le \sum_{s=3}^{nT} |\eta(s)\xi(s+t)|$$
$$\le \sum_{s=3}^{nT} |\eta(s)|(\lambda^{1/2}(s) + \lambda^{1/2}(t)) \tag{3.14}$$

Notice that by (3.8)

$$\sum_{s=3}^{\infty} \frac{|\eta(s)|\lambda^{1/2}(s)}{\sqrt{s}} \le \sum_{s=3}^{\infty} \frac{|\eta(s)|^2}{\sqrt{s}} < \infty, \text{a.s.}$$

Then Borel-Cantelli lemma implies

$$\sum_{s=3}^{nT} |\eta(s)| = o(\sqrt{\frac{nT}{\lambda(nT)}}).$$

Applying it to (3.14), we get

$$\max_{1 \le t \le d \ln n} \sum_{s=3}^{n} |\eta(sT)\xi(sT+t)| = o(\sqrt{n}) \tag{3.15}$$

Similarly, we can prove that

$$\max_{1 \le t \le d \ln n} \sum_{s=3}^{n} |\xi(sT)\eta(sT+t)| = o(\sqrt{n}) \tag{3.16}$$

Now, (3.11) follows from (3.12), (3.15) and (3.16).

Using the same approach, we can prove that

$$\max_{1 \le t \le d \ln n} \sum_{s=3}^{n} |E(\epsilon(sT)\epsilon(sT+t) - \xi(sT)\xi(sT+t)|\mathcal{F}'_{s-1})| = o(\sqrt{n}). \tag{3.17}$$

(3.6) follows (3.14) and (3.17). Next, we will use the exponential inequality in lemma 3.1.1 to get (3.7). Let us now investigate sum of the conditional variance of $Y_t(s)$.

$$\max_{1 \le t \le d \ln n} \sum_{s=3}^{n} |E\{\xi(sT)\xi(sT+t)|\mathcal{F}'_{s-1}\}|^2$$

$$
\begin{aligned}
&\leq \max_{1\leq t\leq d\ln n} \sum_{s=3}^{n} |\xi(sT)^2 |E\{\eta(sT+t)|\mathcal{F}'_{s-1}\}|^2 \\
&\leq \lambda(nT) \sum_{s=3}^{nT+d\ln n} E(\eta^2(s)|\mathcal{F}'_{s-1}) \\
&= o(\sqrt{n})o(\sqrt{nT+d\ln n}) = o(n) \qquad (3.18)
\end{aligned}
$$

Here we have used (3.10). Since

$$
\begin{aligned}
E(\epsilon^2(sT+t)|\mathcal{F}_{sT+t-1}) &= 1, \quad \text{a.s and} \mathcal{F}'_{s-1} \subset \mathcal{F}_{sT+t-1}, \quad \text{then} \\
E(\epsilon^2(sT+t)|\mathcal{F}'_{s-1}) &= 1.
\end{aligned}
$$

Consequently,

$$
\begin{aligned}
&\sum_{i=3}^{n} E(\xi^2(sT)\xi^2(sT+t)|\mathcal{F}'_{s-1}) \\
&= \sum_{s=3}^{n} \xi^2(sT)(1 - E(\eta^2(sT+t)|\mathcal{F}'_{s-1})) \qquad (3.19)
\end{aligned}
$$

Applying (3.10) and the fact that

$$\xi^2(sT) \leq \lambda(sT),$$

we get

$$
\begin{aligned}
&\max_{1\leq t\leq d\ln n} \sum_{s=3}^{n} \xi^2(sT)E(\eta^2(sT=t)|\mathcal{F}'_{s-1}) \\
&\leq \lambda(nT)\sum_{s=3}^{n} E(\eta^2(sT+t)|\mathcal{F}'_{s-1}) \\
&\leq \sqrt{\frac{nT}{\ln\ln nT}} \cdot o(\sqrt{n}) = o(n) \qquad (3.20)
\end{aligned}
$$

Notice also that (3.3) and (3.9) imply

$$
\begin{aligned}
\sum_{s=1}^{n} \xi^2(sT) &= \sum_{s=1}^{n} (\epsilon^2(sT) - \eta^2(sT)) \\
&= n + o(n) \qquad (3.21)
\end{aligned}
$$

Then (3.19), (3.20) and (3.21) yield that uniformly in $1 \le t \le d \ln n$,

$$\sum_{s=1}^{n} E\{\xi^2(sT)\xi^2(sT+t)|\mathcal{F}'_{s-1}\} = n + o(n). \tag{3.22}$$

It follows from (3.18) and (3.22) that uniformly in $t \le d \ln n$,

$$s_t^2(n) = \sum_{s=1}^{n} Var\{Y_t(s)|\mathcal{F}'_{s-1}\} = n + o(n). \tag{3.23}$$

Notice that from the definition of $Y_t(n)$,

$$\begin{aligned} |Y_t(n)| &\le 2\lambda(nT)\lambda(nT+t) \\ &= K_t(n)\frac{s_t(n)}{\sqrt{\ln\ln s_t^2(n)}} \end{aligned} \tag{3.24}$$

where

$$K_t(n) = 2\left(\frac{nT}{\ln\ln nT} \cdot \frac{nT+t}{\ln\ln(nT+t)}\right)^{1/4}\left(\frac{s_t(n)}{\sqrt{\ln\ln s_t^2(n)}}\right)^{-1}\frac{1}{(\ln nT)\ln(nT+t)}.$$

Then by (3.23)

$$\lim_{n\to\infty} K_t(n) = 0. \tag{3.25}$$

Choose $0 < K < \frac{\sqrt{2}}{3}$, let

$$\tilde{Y}_t(s) = Y_t(s)\chi_{\{K_t(s)\le K\}}.$$

Then

$$\limsup_{n\to\infty} \frac{\max_t \sum_{s=1}^n Y_t(s)}{(2n\ln\ln n)^{1/2}} = \limsup_{n\to\infty} \frac{\max_t \sum_{s=1}^n \tilde{Y}_t(s)}{(2n\ln\ln n)^{1/2}}. \tag{3.26}$$

We will prove that

$$\limsup_{n\to\infty} \frac{\max_t \sum_{s=1}^n \tilde{Y}_t(s)}{(2n\ln\ln n)^{1/2}} \le \gamma(K), d \tag{3.27}$$

where $\gamma(K) \in (0, \frac{4}{9K^2})$ is the unique root of $\beta(x,K) - 2 = 0$ and $\beta(x,K) = x^2(1 - \frac{Kx}{2})$. Since $\beta(x,K)$ is decreasing in K, we see that $\gamma(K)$ decreases to $\sqrt{2}$ as K goes to 0. Then (3.26) and (3.27) yield

$$\limsup_{n\to\infty} \frac{\max_t \sum_{s=1}^n Y_t(s)}{(2n\ln\ln n)^{1/2}} \le \sqrt{2}. \tag{3.28}$$

Apply this results to the martingale difference $\{-Y_t(s), \mathcal{F}'_s, s \geq 1\}$, we get

$$\liminf_{n\to\infty} \frac{\min_t \sum_{s=1}^n Y_t(s)}{(2n \ln\ln n)^{1/2}} \geq -\sqrt{2}. \tag{3.29}$$

Now (3.7) follows (3.28) and(3.29). So it is enough to prove (3.27). To reach that goal, let $p > 1$ and define the stopping times

$$\tau_{m,t} = \inf\{n \geq 1, s_t^2(n+1) \geq p^{2m}\}$$

Let

$$\tau_m = \inf\{n \geq 1, \max_{1\leq t\leq d\ln n} s_t^2(n+1) \geq p^{2m}\}$$

Then $\tau_m \leq \tau_{m,t}$ for any t. (Notice that τ_m is not a stopping time.) Then for $0 < \delta < \infty$,

$$\begin{aligned} &P(\max_{3\leq t\leq d\ln n} \sum_{s=1}^n \tilde{Y}_t(s) > \delta\sqrt{2n\ln\ln n}, i.o) \\ &\leq P(\sup_{n\leq \tau_{m+1}} \max_{3\leq t\leq d\ln n} \sum_{s=1}^n \tilde{Y}_t(s) > \delta\sqrt{(\tau_m+1)\ln\ln(\tau_m+1)}, \text{i.o in} m) \end{aligned} \tag{3.30}$$

Since we have proved

$$\lim_{n\to\infty} n^{-1} \max_{1\leq t\leq d\ln n} s_t^2(n) = 1,$$

then almost surely for sufficiently large m,

$$p^{2m-1} < \tau_m < p^{2m+1}$$

Using this inequality and the fact that

$$\sqrt{p^{2m-1}\ln\ln p^{2m-1}} > p^{-2}\sqrt{p^{2m+2}\ln\ln p^{2m+2}},$$

for sufficiently large m, the probability in (3.30) is less than

$$P(\{\sup_{n\leq \tau_{m+1}} \max_{3\leq t\leq d\ln n} \sum_{s=1}^n \tilde{Y}_t(s) > \delta p^{-2}\sqrt{p^{2m+2}\ln\ln p^{2m-1}}, \text{i.o in } m)\}) \tag{3.31}$$

It is easy to verify for a fixed t, the martingale difference $\{\tilde{Y}_t(s)\chi_{\{s \le \tau_{m+1,t}\}}, \mathcal{F}'_s, s \ge 1\}$ satisfies conditions in Lemma 3.1.1 with this $b^2 = p^{2m+2}$.

From Lemma 3.1.1 and the fact $\tau_m \le \tau_{m,t}$ for any t, we have

$$
\begin{aligned}
&P(\sup_{n \le \tau_{m+1}} \sum_{s=1}^{n} \tilde{Y}_t(s) > \delta p^{-2}\sqrt{b^2 \ln\ln b^2}) \\
&\le P(\sup_{n \le \tau_{m+1,t}} \sum_{s=1}^{n} \tilde{Y}_t(s) > \delta p^{-2}\sqrt{b^2 \ln\ln b^2}) \\
&\le \exp(-\beta \ln\ln p^{2m-1})) = ((2m-1)\ln p)^{-\beta}, \quad \forall 1 \le t \le d\ln n
\end{aligned}
$$

where $\beta = \beta(\delta p^{-2}, K)$. Since $\beta(x, K)$ is increasing in $x \in (0, 4/(3K))$, then $\beta > 2$ for $\delta > p^2\gamma(K)$. It follows then for such a δ,

$$
\begin{aligned}
&\sum_{m=1}^{\infty} \sum_{t=3}^{d\ln p^{2m+3}} P(\{\sup_{n} \sum_{s=1}^{n\wedge\tau_{m+1}} \tilde{Y}_t(s) > \delta\sqrt{p^{2m-1}\ln\ln p^{2m-1}}\}) \\
&\le \sum_{m=1}^{\infty} d(2m+3)(\ln p)((2m-1)\ln p)^{-\beta} < \infty.
\end{aligned} \tag{3.32}
$$

(3.31), (3.32) and Borel-Cantelli lemma imply

$$
P(\sup_{n \le \tau_{m+1}} \max_{3 \le t \le d\ln n} \sum_{s=1}^{n} \tilde{Y}_t(s) > \delta\sqrt{(\tau_m+1)\ln\ln(\tau_m+1)}, \text{i.o in} m)) = 0 \tag{3.33}
$$

for $\delta > p^2\gamma(K)$. Since $p > 1$ is arbitrary, let p goes to 1,then (3.33) is true for $\delta > \gamma(K)$.

Then we have proved

$$
P(\max_{1 \le t \le d\ln n} \sum_{s=1}^{n} \tilde{Y}_t(s) > \gamma(K)\sqrt{2n\ln\ln n}, \quad i.o) = 0.
$$

(3.27) follows now. The proof is completed.

QED.

3.2 Convergence Rate of Sample Covariances

In this section, we prove the following theorem.

Theorem 3.2.1 *If $\{x(t) : t \in Z\}$ is PAR and Assumption 3.0.1 holds, then for any constant $d > 0$, almost surely,*

$$\sup_{|t-s|<d\ln N} | R_N(t,s) - R(t,s) | = O(\sqrt{\frac{\ln\ln N}{N}})$$

where R_N is defined by (3.1). $R(t,s)$ is the autocovariance function of x.

We will need some lemmas to prove the theorem. Clearly, $R_N(t,s)$ can be linearly expressed by the sample covariances of the corresponding martingale difference $\epsilon(t)$. So, we first investigate the Wold coefficients for a PAR model. It is well known that the corresponding multivariate stationary AR model $\vec{x}$ has representation

$$\vec{x}(t) = \sum_{j=0}^{\infty} C_j \vec{\epsilon}(t-j) \tag{3.34}$$

where

$$\vec{\epsilon}(t) = (\epsilon(1+tT), \epsilon(2+tT), \cdots, \epsilon(T+tT))',$$
$$\vec{x}(t) = (x(1+tT), x(2+tT), \cdots, x(T+tT))'.$$

If we write (3.34) for each component, we get

$$x(t) = \sum_{j=0}^{\infty} c(j,t)\epsilon(t-j) \tag{3.35}$$

and obviously $c(j,t)$ is periodic in t. We call $c(j,t)$ the Wold coefficients of the PAR $x(t)$.

The following fact about PAR is analogous to a well known one for stationary AR model.

Lemma 3.2.1 *There exists constants $\gamma_1 > 0$ and $\gamma_2 > 0$ such that for any j, t*

$$|c(j,t)| \leq \gamma_1 exp(-\gamma_2 j)$$

Proof. It is well known that the Wold coefficients C_n of a stationary multivariate process go to zero at an exponential rate, i.e., there exists positive constants α and β such that

$$\|C_n\| \leq \alpha \exp(-n\beta),$$

where the norm $\|C_n\|$ is the maximum of entries in C_n. Observe that for any $0 \leq m < T$ and $j = nT + m$, $c(j,t)$ is an element of the matrix C_n for any $t = 1, 2, \cdots, T$. Hence

$$| \; c(j,t) \; | \leq \|C_n\| \leq \alpha \exp(-n\beta) \leq \alpha \exp(-j\beta/T).$$

QED.

Next, we consider the sample covariance of the innovation process $\epsilon(t)$. For any positive integers t, s and positive real number b, let

$$u(t,s;b) = \sum_{m=0}^{[b]} \{\epsilon(t+mT)\epsilon(s+mT) - E\epsilon(t)\epsilon(s)\} \tag{3.36}$$

where $[b]$ as before denotes the integer part of b.

Lemma 3.2.2 *Let $\{M_n\}$ be a sequence of increasing , non-negative random variables and $\{A_n\}$ be an increasing sequence of real numbers. If $A_n \to \infty$ and $E(M_n) = O(A_n)$, then*

$$M_n = o(A_n \ln A_n (\ln\ln A_n)^{1+\delta})$$

for any $\delta > 0$.

Proof. For a given $\delta > 0$, let $\lambda(n) = A_n \ln A_n (\ln\ln A_n)^{1+\delta}$. Without loss of generality, assume that

$$E(M_n) \le cA_n \tag{3.37}$$

for some constant $c > 0$.

For $j \ge 1$, let

$$n_j = \inf\{n \ge 1 : \ln A_n > j\}. \tag{3.38}$$

Then n_j increases to ∞ as $j \to \infty$.

It follows from Markov's inequality, (3.37) and (3.38) that

$$P(M_{n_j} > \lambda(n_j)) \le \frac{cA_{n_j}}{\lambda(n_j)} \le \frac{c}{j(\ln j)^{1+\delta}}$$

Since $\sum_{j=3}^{\infty} \frac{1}{j(\ln j)^{1+\delta}} < \infty$, Borel-Cantelli Lemma implies that

$$P(M_{n_j} > \lambda(n_j), i.o) = 0.$$

Now, for $n_j \le n \le n_{j+1} - 1$, we have from the monotonicity

$$\lambda(n) \ge \lambda(n_j) \quad and \quad M_n \le M_{n_{j+1}-1}$$

Then almost surely, for sufficiently large n,

$$\frac{M_n}{\lambda(n)} \le \frac{M_{n_{j+1}}}{\lambda(n_j)} \le \frac{\lambda(n_{j+1})}{\lambda(n_j)} \tag{3.39}$$

Since

$$\lim_{j\to\infty} \frac{\lambda(n_{j+1})}{\lambda(n_j)} \le \lim_{j\to\infty} \frac{\exp(j+1)(j+1)^{1+\delta}\ln(j+1)}{\exp(j)j^{1+\delta}\ln j} = e,$$

(3.39) implies

$$M_n = O(\lambda(n)) = O(A_n \ln A_n (\ln\ln A_n)^{1+\delta}).$$

Since it is true for any $\delta > 0$, O can be replaced by o. The proof is completed.

QED.

The following lemma is needed to prove our theorem and is of interests of its own.

Lemma 3.2.3 *Under Assumption 3.0.1, for any constant* $d > 0$,

$$\limsup_{n\to\infty} \max_{|t|\le d\ln n} \frac{|u(t,t;n)|}{\sqrt{2n\ln\ln n}} \quad < \quad \infty, \qquad a.s. \tag{3.40}$$

$$\limsup_{n\to\infty} \max_{|t|,|s|\le d\ln n, t\neq s} \frac{|u(t,s;n)|}{\sqrt{2n\ln\ln n}} \quad \le \quad \sqrt{2}, \qquad a.s. \tag{3.41}$$

Proof. The proof is an application of Theorem 3.1.2 and Lemma 3.2.2 with some computation. We only need to prove the lemma for $t \le s$ since $u(t,s;n)$ is symmetric in t, s.

For a fixed $|t| < d\ln n$, let n_0 be the integer such that

$$t_0 = t - n_0 T \in [1, T]$$

First notice that Theorem 3.1.2 implies that for any fixed t_0,

$$\limsup_{n\to\infty} \frac{|\sum_{m=1}^{n}(\epsilon^2(t_0+mT)-1)|}{\sqrt{2n\ln\ln n}} < \infty, \qquad a.s. \tag{3.42}$$

$$\limsup_{n\to\infty} \max_{t_0<s<d\ln n} \frac{|\sum_{m=1}^{n}\epsilon(t_0+mT)\epsilon(s+mT)|}{\sqrt{2n\ln\ln n}} \le \sqrt{2}, \qquad a.s. \tag{3.43}$$

Let $s_0 = s - n_0 T$. It is clear that $u(t,s;n)$ can be written as

$$u(t,s;n) = \sum_{m=n_0}^{n+n_0} (\epsilon(t_0+mT)\epsilon(s_0+mT) - \delta_{t,s})$$

and

$$|u(t,s;n) - \sum_{m=0}^{n} \epsilon(t_0+mT)\epsilon(s_0+mT)| \le 2M_n, \tag{3.44}$$

where M_n denotes the the maximum of $|u(t,s;i)|$ over $|t|, |s|, i \le d\ln n$. Since $\{u(t,s,;i), \mathcal{F}_{\max(t,s)+iT-1}, i \ge 1\}$ is a L^2 martingale under Assumption 3.0.1 and

$$\begin{aligned} &E|u(t,s;i)|^2 \\ &= \sum_{m=0}^{i} E|\epsilon(t+mT)\epsilon(s+mT) - E\epsilon(t)\epsilon(s)|^2 \le c(i+1). \end{aligned} \tag{3.45}$$

for $c = 1 + \sup_t E|\epsilon(t)|^4$. It is evident that

$$\begin{aligned} E(M_n^2) &\leq E\sum_{t,s,i} |u(t,s;i)|^2 \\ &\leq (1 + 2d\ln n)^3 \cdot c(\ln n + 1). \end{aligned}$$

Then $E(M_n) = O(A_n)$ for $A_n = (\ln n)^2$. Lemma 3.2.2 implies

$$\begin{aligned} M_n &= o(A_n \ln A_n (\ln\ln A_n)^2) \\ &= o(n^\alpha), \quad \forall \alpha > 0. \end{aligned}$$

The lemma now follows.

QED.

Proof of Theorem 3.2.1 Without loss of generality, assume $N = nT$. Then by (3.1)

$$R_N(t,s) = n^{-1} \sum_{m=1}^{n-\max(t,s)/T} x(t+mT)x(s+mT) \tag{3.46}$$

for $t = 1, \cdots, T., and \quad s = 0, 1, \cdots, N-t-1$.

Since both R_N and R are symmetric and periodic, we only need to prove the theorem for $t = 1, .., T.$ and $t \leq s \leq t + d\ln n$, and

Let $Q_n = \sqrt{\frac{\ln\ln n}{n}}$. Notice that from (3.35) and the orthogonality of $\{\epsilon(t)\}$,

$$R(t,s) = \sum_{j,k=0}^{\infty} c(j,t)c(k,s)\delta_{t-j,s-k} \tag{3.47}$$

Then it follows from (3.46) and (3.47) that

$$\begin{aligned} R_N(t,s) - R(t,s) = n^{-1} \sum_{j,k=0}^{\infty} c(j,t)c(k,s) \\ \times \sum_{m=0}^{n-s/T} [\epsilon(t+mT-j)\epsilon(s+mT-k) - \delta_{t-j,s-k}] \\ + \frac{n - [n - s/T]}{n} \sum_{j,k=0}^{\infty} c(j,t)c(k,s)\delta_{t-j,s-k} \end{aligned}$$

The sum in the second term is finite by Lemma 3.2.1, thus the second term is obviously $O(n^{-1}\ln n)$ uniformly in $s \le d\ln$.

Denote the first term by $W_n(t,s)$. Then

$$W_n(t,s) = n^{-1}\sum_{j,k=0}^{\infty} c(j,t)c(k,s)u(t-j,s-k;n-s/T) \tag{3.48}$$

where $u(t,s;x)$ is defined by (3.36).

Next, truncate the sum in (3.48) at $j,k \le d\ln n$. Denote by $Z_n(t,s)$ the truncated sum, i.e.

$$Z_n(t,s) = n^{-1}\sum_{j,k=0}^{d\ln n} c(j,t)c(k,s)u(t-j,s-k;n-s/T) \tag{3.49}$$

Then it follows Lemma 3.2.3

$$\max_{0<s-t\le d\ln n}\max_{0\le j,k\le d\ln n}|u(t-j,s-k;n-s/T)| = O(\sqrt{n\ln\ln n}).$$

Consequently

$$\max_{0\le s-t\le d\ln n}|Z_n(t,s)| \le n^{-1}O(\sqrt{n\ln\ln n})\max_{t,s\in Z}\sum_{j,k=0}^{d\ln n}|c(j,t)c(k,s)| = O(Q_n) \tag{3.50}$$

So it is sufficient to show

$$\max_{0\le s-t\le d\ln n}|W_n(t,s)-Z_n(t,s)| = O(Q_n). \tag{3.51}$$

The left hand side is dominated by $I_{1,n}+I_{2,n}$, where

$$I_{1,n} = \max_{0\le s-t\le d\ln n} n^{-1}|\sum_{j=d\ln n}^{\infty}\sum_{k=0}^{\infty} c(j,t)c(k,s)u(t-j,s-k;n-s/T)|$$
$$I_{2,n} = \max_{0\le s-t\le d\ln n} n^{-1}|\sum_{0\le j<d\ln n}\sum_{k>d\ln n} c(j,t)c(k,s)u(t-j,s-k;n-s/T)|$$

Applying Markov's inequality, it is easy to see

$$P(\max_{0\le s-t\le d\ln n} n^{-1}|\sum_{j=d\ln n}^{\infty}\sum_{k=0}^{\infty} c(j,t)c(k,s)u(t-j,s-k;n-s/T)| > Q_n)$$

$$\leq Q_n^{-2} n^{-2} \sum_{0 \leq s-t \leq d \ln n} E|\sum_{j=d\ln n}^{\infty} \sum_{k=0}^{\infty} c(j,t)c(k,s)u(t-j, s-k; n-s/T)|^2 \tag{3.52}$$

We have proved in (3.45)

$$E|u(t,s;i)|^2 \leq c(i+1),$$

then Holder's inequality implies for $\forall t_i, n_j, i = 1, \cdots, 4, j = 1, 2,$

$$E|u(t_1, t_2; n_1)u(t_3, t_4; n_2)| \leq c\sqrt{(n_1+1)(n_2+1)}.$$

This together with Lemma 3.2.1 implies that the expectation in (3.52) is dominated by

$$(n+1)c|\sum_{j=d\ln n}^{\infty} c(j,t)|^2| \sum_{k=0}^{\infty} c(k,s)|^2 \leq \frac{c\gamma_1^4}{(1-exp(-\gamma_2))^4}(n+1)n^{-2\gamma_2 d}.$$

where γ_1, γ_2 are the same as in proposition 2.

Then (3.52) is bounded by $(d\ln n)^2 Q_n^{-2} c n^{-1-2\gamma_2 d} = O(n^{-2})$. It follows that

$$\sum_n P(I_{1,n} > Q_n) < \infty.$$

Borel-Cantelli lemma implies

$$I_{1,n} = O(Q_n) \quad a.s$$

Similarly, we can prove that

$$I_{2,n} = O(Q_n) \quad a.s.$$

(3.51) is established now and the proof is finished.

QED.

3.3 Convergence Rate of Coefficients

Solution of Yule-Walker equations provides estimators for the coefficients. If the orders $p(1), p(2), \cdots, p(T)$ are known, then we will have no difficulties to show, using the results in last section, that these estimators are consistent and have the same convergence rate as the sample covariances. Since the orders are unknown, we need a little extra work and notations get complicated. To make our statements clearer, we will define random inner product which will simplify our statements.

Let $L^2(\Omega, F, dP)$ denote the Hilbert space of random variables with zero means and finite second moments. Then $\{x(t), t \in Z\}$ is a set in this Hilbert space and Yule-Walker equations are just normal equations of projection. We want to use this convenience of projection even when the covariances $R(t,s)$ in Yule-Walker equations are replaced by the sample covariances $R_N(t,s)$. For this purpose, we introduce random inner product. Let $\mathcal{X}$ denote the the subset $\{x(t), t \in Z\}$ of $L^2(\Omega, F, dP)$. For each integer N, let $< \cdot, \cdot >_N (\cdot)$ be a map from $\mathcal{X} \times \mathcal{X} \times \Omega$ to the set of real numbers such that

$$< x(t), x(s) >_N (\omega) = R_N(t,s)(\omega).$$

We can not yet say $< \cdot, \cdot > (\omega)$ is an inner product for a fixed ω. But for a given finite sequence of integers $t_0, t_1, \ldots, t_m$ and a fixed ω, $(R_N(t_j, t_k))_{j,k=0}^m$ is positive definite for sufficiently large N. So for such a N, $< \cdot, \cdot, >_N (\omega)$ can be regarded as an inner product on a linear space spanned by $x(t_0), \ldots, x(t_m)$ such that

$$< x(t_j), x(t_k) >_N (\omega) = R_N(t_j, t_k)(\omega).$$

We will suppress ω in the inner product and write it as $< \cdot, \cdot >_N$. The corresponding norm will be denoted by $\|.\|_N$.

For the sake of convenience and unity of notations, let $< .,. >_\infty$ and $\|.\|_\infty$ be the inner product and norm in $L^2(\Omega, F, dP)$, i.e.

$$\begin{aligned} < x(t), x(s) >_\infty &= Ex(t)x(s) \\ \|x(t)\|_\infty &= Ex^2(t). \end{aligned}$$

Then for any t, s, almost surely,

$$\lim_{N\to\infty} < x(t), x(s) >_N = < x(t), x(s) >_\infty .$$

For each $N = 1, 2, \cdots, \infty$, denote by $Proj_N[x(t_0)|x(t_1), \cdots, x(t_m)]$ the projection of $x(t_0)$ onto the subspace spanned by $x(t_1), \cdots, x(t_m)$ under the $\|.\|_N$. Let

$$Proj_N[x(t) \mid x(t-1), \cdots , x(t-p)] = \sum_{j=1}^{p} a_N(j, t; p)x(t-j). \tag{3.53}$$

$a_N(j, t; p), j = 1, \cdots, p$ are actually the solution of

$$\Gamma_N(t; p)a_N(t; p) = R_N(t; p), \tag{3.54}$$

where

$$\begin{aligned} \Gamma_N(t;p) &= (R_N(t-j, t-k))_{j,k=1}^p && (3.55) \\ R_N(t;p) &= (R_N(t-1,t), \cdots, R_N(t-p,t))' && (3.56) \\ a_N(t;p) &= (a_N(1,t;p), \cdots, a_N(p,t;p))' && (3.57) \end{aligned}$$

For $N = \infty$, the above equations are just the Yule-Walker equations we discussed in Chapter 2.

Let $l(.)$ be a periodic function from $\mathcal{N}$ to $\mathcal{N}$ with period T. $l(t)$ may depend on sample and serves as an estimator of $p(t)$. Choose a dominating function $L(N)$ from $\mathcal{N}$ to $\mathcal{N}$ and assume the following throughout the rest of this section.

Assumption 3.3.1 *$L(N)$ increases to ∞ and $L(N) = O(\ln N)$*

Theorem 3.3.1 *Let $\{x(t), t \in Z\}$ be a PAR with order $p(1), \cdots, p(T)$. Then under Assumption 3.0.1 and 3.3.1,*

$$\sup_{j,t,l} \mid a_N(j,t;l(t)) - a_\infty(j,t;l(t)) \mid = O(\sqrt{\frac{\ln \ln N}{N}})$$

where the supremum is taken over any t, $j \leq l(t)$ and any periodic function $l(.)$ with period T such that $l(t) \leq L(N)$, $\forall\ t$.

Remark. Notice that for $p \geq p(t)$, $a_\infty(j,t;p)$ are the actual regression coefficients in the PAR model, this theorem says if we choose order p greater than the true order in (3.54), then the estimator from Yule-Walker equation converges to true parameter at the rate of $\sqrt{\frac{\ln \ln N}{N}}$.

The proof of this Theorem needs the following Lemma.

Lemma 3.3.1 *Let $x(t)$ be a PAR. $\Gamma(t;q) = (R(t-j,t-k))_{j,k=0}^{q}$. Then there exists an $M > 0$ such that for any t and q, $\|\Gamma^{-1}(t;q)\| \leq M$.*

Proof. We note that for any positive definite matrix Γ, $||\Gamma||$ is less than or equal to the maximum eigen value of Γ. Since $\Gamma^{-1}(t;q)$ is positive definite, we only need to show that the eigenvalues of $\Gamma^{-1}(t;q)$ is bounded from above, or equivalently, all eigenvalues of $\Gamma(t;q)$, for any $t = 1, 2, \cdots, T$ and $q \geq 1$, are no less than a positive number λ.

Let λ_q be the minimum of all eigenvalues of $\Gamma(t;q)$ for all $t = 1, 2, \cdots, T$. Evidently,

$$\lambda_q > 0.$$

It suffices to show that

$$\lambda_{q+1} \geq \min(\lambda_q, \frac{\min \sigma^2(t)}{1 + \max_t \sum_{j=1}^{p(t)} a(j,t)^2}). \tag{3.58}$$

Let

$$\mathbf{X}_q(t) = \begin{pmatrix} x(t) \\ x(t-1) \\ \cdots \\ x(t-q) \end{pmatrix}, \forall t, q.$$

Then

$$\mathbf{X}_{q+1}(t) = \begin{pmatrix} x(t) \\ \mathbf{X}_q(t-1) \end{pmatrix}$$

For any vector $\mathbf{C}_{q+1}$ of $q+2$ dimension such that

$$||\mathbf{C}_{q+1}||^2 = 1,$$

we write it as

$$\mathbf{C}_{q+1} = \begin{pmatrix} c \\ \mathbf{C}_q \end{pmatrix},$$

where $\mathbf{C}_q$ is a $(q+1)$-dimensional vector. Using (1.2), we get

$$\mathbf{C}'_{q+1}\mathbf{X}_{q+1}(t) = cx(t) + (c\vec{a}_p(t) + \mathbf{C}_p)'\mathbf{X}_p(t),$$

where $\vec{a}_p = (a(1,t), a(2,t), \cdots, a(p(t),t), 0, \cdots, 0)$.

The orthogonality of $\epsilon(t)$ with $x(s), s < t$ together with the definition of λ_q imply

$$\begin{aligned} &\mathbf{C}'_{q+1}\Gamma(t; q+1)\mathbf{C}_{q+1} = ||\mathbf{C}'_{q+1}\mathbf{X}_{q+1}(t)||^2 \\ &= c^2\sigma^2(t) + ||(c\vec{a}_p(t) + \mathbf{C}_p)'\mathbf{X}_p(t)||^2 \\ &\geq c^2\sigma^2(t) + \lambda_q||c\vec{a}_q(t) + \mathbf{C}_q||^2 \\ &\geq c^2\sigma^2(t) + \lambda_q|c^2||\vec{a}_q(t)||^2 - ||\mathbf{C}_q||^2| \end{aligned}$$

Since $c^2 + ||\mathbf{C}_q||^2 = 1$, and

$$\inf_{0 \leq x \leq 1}(ax + \lambda|bx - 1|) = \min(\lambda, a/b), \forall a > 0, \lambda > 0, b > 1,$$

we have

$$\mathbf{C}'_{q+1}\Gamma(t;q+1)\mathbf{C}_{q+1} \geq \min(\lambda_q, \frac{\sigma^2(t)}{||\vec{a}_q(t)||^2+1}).$$

(3.58) now follows.

QED.

Proof of Theorem 3.3.1 For brevity, we omit $l(t)$ in Γ_N, R_N and a_N. Observe that

$$\begin{aligned} & I+\Gamma_\infty^{-1}(t)(\Gamma_N(t)-\Gamma_\infty(t))\ (a_\infty(t)-a_N(t)) \\ & = \Gamma_\infty^{-1}(t)[R_\infty(t)-R_N(t)+(\Gamma_N(t)-\Gamma_\infty)a_\infty(t)] \end{aligned} \tag{3.59}$$

Theorem 3.2.1 and Lemma 3.3.1 imply that the maximum absolute value of the entries of $\Gamma_\infty^{-1}(t))(\Gamma_N(t)-\Gamma_\infty(t))$ is $O(\sqrt{\frac{\ln\ln N}{N}})$. Thus

$$\begin{aligned} \| & \ \Gamma_\infty^{-1}(t)(\Gamma_N(t)-\Gamma_\infty(t)(a_\infty(t)-a_N(t))\| \\ \leq & \ O(\sqrt{\frac{\ln\ln N}{N}})l^2(t)\|a_\infty(t)-a_N(t)\| & (3.60) \\ = & \ o(1)\|a_\infty(t)-a_N(t)\| & (3.61) \end{aligned}$$

Similar argument proves RHS of (3.59) is

$$O(\sqrt{\frac{\ln\ln N}{N}})(1,1,\cdots,1)'. \tag{3.62}$$

Also notice that every $O(1)$ and $o(1)$ appeared above is uniform in t and functions l such that $l(t) \leq L(N)$. Then the Theorem follows from (3.59),(3.61) and (3.62).

3.4 BIC for Order Estimation

For stationary AR(p_0) model, Akaike(1977) first proposed to estimate p_0 by $\hat{p}$ which minimizes

$$\ln\hat{\sigma}_p^2 + p\ln N/N$$

Here $\hat{\sigma}_p^2$ is the estimate of σ^2 from the Yule-Walker equations of order p.

An-Chen-Hannan (1982) proved BIC estimator is consistent under general conditions.

In this section, we develop similar criterion for PAR models. It turns out that BIC is included as a special case. Let $x(1), x(2), \cdots, x(N)$ be observations from a PAR model with order $p(\cdot)$. Let $R_N(t,s)$ be defined by (3.1). We will follow the notations in the last section. Let

$$\sigma_N^2(t;l) = \|x(t) - Proj_N[x(t) \mid x(t-1), \cdots, x(t-l)]\|_N^2 \tag{3.63}$$

Let $q(N)$ be a sequences of positive integers such that

$$\lim_{N\to\infty} \frac{\ln\ln N}{q(N)} = 0, \quad \text{and} \quad \lim_{N\to\infty} \frac{q(N)}{N} = 0. \tag{3.64}$$

Let $\hat{p}(t) = \hat{p}_N(t)$ minimize

$$\ln \sigma_N^2(t;l) + l \cdot q(N)/N, \quad \forall 0 \le l \le L(N). \tag{3.65}$$

Then $\hat{p}(t)$ is a consistent estimator of $p(t)$ under general assumptions.

Theorem 3.4.1 *Let $x(t)$ be PAR satisfying Assumption 3.0.1 and $L(N)$ satisfy Assumption 3.3.1. Then for any t, almost surely*

$$\hat{p}(t) \to p(t), \ as N \to \infty.$$

Proof. Since $R_N(t,s) \to R(t,s)$, a.s, then

$$\| \sum_{j=1}^m c_j x(t_j) \|_N^2 \to \| \sum_{j=1}^m c_j x(t_j) \|_\infty^2. \tag{3.66}$$

for all real $c_1, \cdots, c_m$ and integers $t_1, t_2, \cdots, t_m$.

As a special case of (3.66),we have

$$\sigma_N^2(t;l) \to \sigma_\infty^2(t;l) = \|x(t) - Proj(x(t) \mid x(t-1), \cdots, x(t-l))\|_\infty^2 \tag{3.67}$$

Suppose that

$$Proj_N[x(t) \mid x(t-1), \cdots, x(t-l)] = \sum_{j=1}^{l} a_N(j,t;l)x(t-j) \tag{3.68}$$

for $N = 1, 2, \cdots, \infty$.

It is helpful to realize that for $l \geq 2$,

$$1 - \frac{\sigma_N^2(t;l)}{\sigma_N^2(t;l-1)} = a_N^2(l,t;l) \tag{3.69}$$

In fact, (3.69) is just application of Pythagorean Theorem. In fact, let

$$\hat{x}(t) = Proj_N[x(t) \mid x(t-1), \cdots, x(t-l+1)].$$

Then

$$\begin{aligned}
&\sigma_N^2(t;l-1) - \sigma_N^2(t;l) \\
&= |x(t) - Proj_N[x(t) \mid x(t-1), \cdots, x(t-l+1)]\|_N^2 \\
&- \|x(t) - Proj_N[x(t) \mid x(t-1), \cdots, x(t-l)]\|_N^2 \\
&= \|\hat{x}(t) - Proj_N[\hat{x}(t)|x(t-1), \cdots, x(t-l+1)]\|^2 \\
&= a_N^2(l,t;l)\|x(t) - Proj_N[x(t) \mid x(t-1), \cdots, x(t-l)]\|_N^2 \\
&= a_N^2(l,t;l)\sigma_N^2(t;l)
\end{aligned}$$

Since $a_\infty^2(t,p(t);p(t))$ is positive, (3.69) implies

$$\sigma_\infty^2(t,p(t)-1) > \sigma_\infty^2(t,p(t)) \tag{3.70}$$

Thus

$$\sigma_\infty^2(t,l) \geq \sigma_\infty^2(t,p(t)-1) > \sigma_\infty^2(t,p(t)) \tag{3.71}$$

It follows that for any $l < p(t)$,

$$\lim_{N\to\infty}(\ln \sigma_N^2(t,p(t)) + p(t)\frac{q(N)}{N}) < \lim_{N\to\infty}(\ln \sigma_N^2(t,l) + l\frac{q(N)}{N}).$$

This inequality implies that asymptotically

$$\hat{p}(t) \geq p(t) \tag{3.72}$$

Using (3.69) repeatedly and applying Theorem 3.2.1, we get for $l > p(t)$,

$$\begin{aligned} & 1 - \frac{\sigma_N^2(t,l)}{\sigma_N^2(t,p(t))} \\ &= \sum_{j=p(t)+1}^{l} a_N^2(j,t;j) = (l - p(t))O(\ln\ln N/N) \end{aligned} \tag{3.73}$$

Since

$$\lim_{N\to\infty} \sigma_N^2(t;l) = \lim_{N\to\infty} \sigma_N^2(t;p(t)) = \sigma_\infty^2(t;p(t)),$$

then for sufficiently large N,

$$\ln \frac{\sigma_N^2(t,l)}{\sigma_N^2(t,p(t))} \geq -(1 - \frac{\sigma_N^2(t;l)}{\sigma_N^2(t;p(t))} \tag{3.74}$$

It follows from (3.73)-(3.74) that

$$\ln \sigma_N^2(t,l) - \ln \sigma_N^2(t,p(t)) = (l - p(t))O(\ln\ln N/N). \tag{3.75}$$

where the $O(1)$ is uniform in $l \leq L(N)$.

The assumption on $q(N)$ and (3.75) imply that

$$\min_{p(t)<l<L(N)} [\ln \sigma_N^2(t,l) - \ln \sigma_N^2(t,p(t)) + [l - p(t)]q(N)/N] > 0,$$

for sufficiently large N. So asymptotically

$$\hat{p}(t) \leq p(t). \tag{3.76}$$

Then the assertion follows from (3.72) and (3.76).

QED.

Corollary 3.4.1 *Suppose $\vec{x}$ is a multivariate $AR(p)$ model and $\hat{p}(t), t = 1, \cdots, T$, are the order estimators for the corresponding PAR x defined before. Then*

$$\hat{p} = \max_{1 \le t \le T} [\frac{\hat{p}(t) - t}{T}] + 1$$

is a consistent estimator of p.

Proof. It follows from theorem 3.4.1 and Theorem 1.0.1 (iii).

QED.

We also use simulated data to estimate order of the following model

$$x_{2n} - 0.7x_{2n-1} + .25x_{2n-2} = \epsilon_{2n}$$

$$x_{2n+1} - 0.5x_{2n} - .25x_{2n-1} = \epsilon_{2n-1}$$

where ϵ_n are i.i.d normal sequence. So $T = 2, p_1 = p_2 = 2$. We took $N = 200, q(N) = \ln N,$. Our simulated results are shown in the tables. We see that it picks up the right order.

$\hat{p}_1$	1	2	3	4	5	6	7
BIC	.05429	.02231	.04451	.06129	.08767	.10455	.12899

$\hat{p}_2$	1	2	3	4	5	6	7
BIC	.06602	-0.03244	-0.00682	0.01884	0.04428	.06567	.09199

Bibliography

[1] ADAMS G. J AND GOODWIN G. C.(1995) Parameter estimation for periodic ARMA models. *J. Time Ser. Anal.* Vol. 16, No. 2. 127-145

[2] ANDERSON, P. L. AND VECCHIA, A. V. (1993) Asymptotic results for periodic autoregressive moving average processes. *J. Time Ser. Anal.* Vol. 1, 1-18.

[3] AKAIKE, H. (1974) A new look at the statistical model identification. IEEE Trans. Automatic Control. AC-19, 716-722

[4] AKAIKE, H. (1977). On entropy maximization principle. In *Application of Statistics* , Ed. P.R. Krishnaiah, 27-41. North-holland, Amsterdam.

[5] AN, H. Z., CHEN, Z. G. AND HANNAN, E. J (1982) autocovariance, Autoregression and Autoregressive Approximation. *The Ann. of Statist.* Vol. 10. N0.3, 926-936

[6] CRAMÉR, H (1961) On some classes of non-stationary processes. *Proc. 4th Berkeley Sympos. II,* 57-77

[7] CRAMÉR, H (1964) Stochastic processes as curves in Hilbert space. *Probability Theory and Applications,* 2, 1964.

[8] BLOOMFIELD, P. HURD, H. L. AND LUND, R. B (1994) Periodic Correlation in Stratospheric Ozone Data,*J. Time Series Analysis,* Vol. 15, No. 2.

[9] BURG, J.P. (1967) Maximum entropy spectral analysis. *Proc. 33th Ann. Intern. Meeting, Soc. of Explor. Geophys.*, Oklahoma City, Oklahoma. Also reprinted in *Modern spectrum analysis,* ed., D. G. Childers. (1978) IEEE Press, New York.

[10] CHILDERS, D. G. ed. (1978) *Modern spectrum analysis.* IEEE Press, New York.

[11] CHOI, B. S. AND COVER, T. M. (1987) A proof of Burg's theorem. In C. R. Smith and G. J. Erickson (eds.), *Maximum-Entropy and Bayesian Spectral Analysis and Estimation Problems*, 75-84, D. Reidel Pub. Co., Boston.

[12] GALLAGER, R. (1968) *Information Theory and Reliable Communication,* Wiley, New York.

[13] GARDNER, W.A. (1986) *Introduction to random processes with application to signals and systems.* New York: Macmillian.

[14] GLADYSHEV, E.G. (1961) Periodically Correlated Random Sequence, Soviet Math. 2, 383-388.

[15] HANNAN, E.J. (1970) *Multiple Time Series.* Wiley, New York.

[16] GUDZENKO, L.I.(1959) On Periodically Nonstationary Precesses, *Radiotek. Elektron.* Vol. 4, no.6, 1062-1064

[17] HAYKIN, S. S. ed. (1983)*Nonlinear methods of spectral analysis.* Springer-Verlag, New York.

[18] HURD, H.L. (1989) Nonparametric time series analysis for periodically correlated processes. *IEEE Trans. Inform. Theory.* Vol.35, no. 2, 350-359

[19] HURD, H.L. AND MANDREKAR, V. (1991) Spectral Theory of Periodically and Quasi-Periodically Stationary SαS Sequences. Tech. Report No. 349. Center For Stochastic Processes, Univ. of North Carolina at Chapel Hill.

[20] KOLMOGOROV, A. N. (1941) Stationary sequences in Hilbert space. *Bull. Math. Moscow.* **2,** No. 6.

[21] KULLBACK, S. (1978) *Information Theory and Statistics.* Peter Smith, Massachusetts.

[22] MIAMEE,A.G. AND SALEHI, H.(1980) On the Prediction of Periodially Correlated Stochastic Processes. In *Multivariate Analysis–V*, P.R. Krishnaiah ed., North Holland., 167-179

[23] MORF, M., VIEIRA, A., LEE, D. T. L. AND KAILATH, T. (1978) Recursive multichannel maximum entropy spectral estimation. *IEEE Trans. Acoust. Speech, Signal Process.* **ASSP-28**, 441-454.

[24] PAGANO, M.(1978) On Periodic and Multiple Autoregressions. *The Ann. of Statist.* Vol.6. No. 6. 1310-1317.

[25] PARZEN, E. (1977) Multiple time series: determining the order of approximation autoregressive scheme. In: P. Krishnaiah, ed. *Multivariate Analysis: IV,* 283-295. North- Holland, Amsterdam.

[26] PARZEN, E. (1983) Autoregressive spectral estimation. In D. R. Brillinger and P. R. Krishnaiah ed. *Time Series in the Frequency Domain, (Handbook of Statistics 3).* 221-247. North-Holland, Amsterdam.

[27] SHIBATA, R. (1976) Selection of the order of an autoregressive model by Akaike's information criterion, *Biometrika*, **63**, 117-126.

[28] SMITH, C. R. AND ERICKSON, G. J. (1983) *Maximum-Entropy and Bayesian Spectral Analysis and Estimation Problems.* D. Reidel Pub. Co., Boston.

[29] STOUT, W.F.(1970) A Martingale Analogue of Kolmogorov's Law of the Iterated Logarithm. Z. Wahrscheinlichkeitstheorie verw. Geb. 12, 279-290.

[30] STOUT, W.F(1970)The Hartman-Wintner Law of the Iterated Logarithm for Martingales. *Ann. Math. Statist.* **41**, 2158–2160.

[31] STOUT, W.F. (1974) *Almost Sure Convergence.* Academic, New York.

[32] TIAN, C. J. (1988) A Limiting Property of Sample Autocovariance of Periodically Correlated Processes with Application to Period Detection. *Journal of Time Series Analysis,* Vol. 9, no. 4, 411-417.

[33] ULRYCH, T. J. AND BISHOP, Y. N. (1975) Maximum Entropy Spectral Analysis and Autoregressive Decomposition. *Rev. Geophysics and Space Physics* **13**, 183-200

[34] VECCHIA, A.V. (1985) Periodic autoregressive-moving average modeling with applications to water resources. *Water Resources Bulletin,* Vol 21,no 5, 721-730.

[35] VECCHIA, A. V. (1991) Testing for Periodic Autoregressions in Seasonal Time Series Data, *Biometrika*,Vol 78.

[36] WHITTLE, P. (1963) On the Fitting Of Multivariate Autoregressions, and the Approximate Canonical Factorization Of A Spectral Density Matrix. *Biometrika*, 50. 1 and 2, p129.