THE INTEGER-VALUED AUTOREGRESSIVE (INAR(p)) MODEL

By Du Jin-Guan

Academia Sinica, Beijing

Li Yuan

University of Petroleum, Shandong

First version received January 1989

Abstract. The integer-valued autoregressive (INAR) model with lag p dependence is discussed. The existence and ergodic property of the INAR model are proved. It is shown that the correlation structure of the INAR model is similar to that of the continuous-valued autoregressive (AR) process, and the stationary conditions of INAR and AR processes are also the same.

Keywords. Discrete time series; INAR model; conditional least-squares estimation.

1. INTRODUCTION

Much effort has recently been focused on the study of discrete time series. One particular class of models, known as INAR(1) models, which could be used to model counting processes has been proposed by Al-Osh and Alzaid (1986). The study of such models has revealed that the correlation structure of INAR(1) is similar to that of the continuous-valued AR(1) model. Thus the procedures for modelling INAR(1) can be developed by analogy with those for AR(1).

In this paper we generalize the work of Al-Osh and Alzaid (1986) to integer-valued autoregressive (INAR) models with p dependence. We show that, under certain conditions, a stationary solution of the INAR model exists. It is shown that the INAR model has the same correlation structure and stationary domain as the AR model. Procedures are also developed for estimating the parameters of the INAR model.

Let X be an non-negative integer-valued random variable and $\alpha \in [0, 1]$. According to Steutel and Van Harn (1979), the $\circ$ operator is defined as follows:

$$\alpha \circ X = \sum_{j=1}^{X} y_j$$

where the $\{y_j\}$ are independent identically distributed (i.i.d.) random variables, independent of X, with $P\{y_j = 1\} = \alpha = 1 - P\{y_j = 0\}$, and are known as counting series in this paper.

Then INAR(1) is given by

0143-9782/91/02 129–14 \$02.50/0

$$X_t = \alpha \circ X_{t-1} + \varepsilon_t \tag{1.1}$$

$t \in I = \{0, \pm 1, \pm 2, \ldots\}$, where the $\{\varepsilon_t\}$ are i.i.d. non-negative integer-valued random variables with some discrete distribution and independent of all counting series in (1.1) which are mutually independent. Here we assume that the $\{\varepsilon_t\}$ are i.i.d. in order to make X_{t-1} independent of the counting series, which is stronger than the assumption made by Al Osh and Alzaid which we believe should be i.i.d.

One natural extension of (1.1) is

$$X_t = \alpha_1 \circ X_{t-1} + \ldots + \alpha_p \circ X_{t-p} + \varepsilon_t \tag{1.2}$$

where the assumptions for $\{\varepsilon_t\}$ are the same as above, all counting series are mutually independent and $\alpha_j \in [0, 1]$ $(j = 1, 2, \ldots, p)$. The $\{\varepsilon_t\}$ are independent of all counting series.

In the following section we shall consider under what conditions on α_1, $\alpha_2, \ldots, \alpha_p$, the stationary solution of (1.2) exists.

2. MAIN THEOREM

THEOREM 2.1. *Let the $\{\varepsilon_t\}$ be i.i.d. non-negative integer-valued random variables with $E\varepsilon_t = \mu$, $\operatorname{var}(\varepsilon_t) = \sigma^2$ and $\alpha_i \in [0, 1]$ $(i = 1, 2, \ldots, p)$. If the roots of*

$$\lambda^p - \alpha_1 \lambda^{p-1} - \ldots - \alpha_{p-1}\lambda - \alpha_p = 0$$

are inside the unit circle, then there exists a unique stationary non-negative integer-valued random series $\{X_t\}$ satisfying

$$\begin{gathered} X_t = \alpha_1 \circ X_{t-1} + \ldots + \alpha_p \circ X_{t-p} + \varepsilon_t \\ \operatorname{cov}(X_s, \varepsilon_t) = 0 \qquad (s < t). \end{gathered} \tag{2.1}$$

We need the following lemmas to prove Theorem 2.1.

LEMMA 2.1.

$$E(\alpha \circ X) = \alpha EX \tag{2.2}$$

$$E(\alpha \circ X)^2 = \alpha(1 - \alpha)EX + \alpha^2 EX^2;$$

$$E(\alpha \circ X - \alpha \circ Y)^2 = \alpha(1 - \alpha)E|X - Y| + \alpha^2 E(X - Y)^2 \tag{2.3}$$

where the counting series in $\alpha \circ X$ and $\alpha \circ Y$ are the same;

$$E\{(\alpha \circ X)(\beta \circ Y)\} = \alpha\beta EXY \tag{2.4}$$

where

$$\alpha \circ X = \sum_{j=1}^{X} y_j \qquad \beta \circ Y = \sum_{j=1}^{Y} y_j^*,$$

$\{y_j\}$ is independent of $\{y_j^*\}$ and $\{X, Y\}$ is independent of $\{y_j\}$, $\{y_j^*\}$.

Lemma 2.1 is easy to prove by the definition of the $\circ$ operator and conditional expectation.

Let

$$A = \begin{bmatrix} \alpha_1 & \alpha_2 & \dots & \alpha_{p-1} & \alpha_p \\ 1 & 0 & \dots & 0 & 0 \\ \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & \dots & 1 & 0 \end{bmatrix} \tag{2.5}$$

$$X = (X_1, X_2, \dots, X_p)\tau.$$

We then use the notation

$$A \circ X = \left(\sum_{j=1}^{p} (\alpha_j \circ X_j), X_1, \dots, X_{p-1}\right). \tag{2.6}$$

LEMMA 2.2.

$$E(A \circ X) = AEX$$

$$E(A \circ X)(A \circ X)\tau = AE(XX\tau)A\tau + C \tag{2.7}$$

where

$$C = \begin{bmatrix} \Sigma_{j=1}^{p} \alpha_j(1-\alpha_j)EX_j & 0 \\ 0 & 0 \end{bmatrix}_{p\times p}.$$

Let $A = (a_{ij})_{m\times n}$. (X, Y are $n \times 1$ columns. If $a_{ij} \geqslant 0$, for $1 \leqslant i \leqslant m$, $1 \leqslant j \leqslant n$, then we write A as $A \geqslant 0$; if $X - Y \geqslant 0$, we write $X \geqslant Y$.

LEMMA 2.3. *If $X \geqslant Y$, $A \geqslant 0$, then $AX \geqslant AY$.*

LEMMA 2.4. *If $X_n \xrightarrow{L^2} X$, $X_n \xrightarrow{\text{a.e}} Y$, then $X = Y$ almost everywhere (a.e.).*

LEMMA 2.5. *If $X_n \xrightarrow{L^2} X$, then $\alpha \circ X_n \xrightarrow{L^2} \alpha \circ X$ where the counting series in $\alpha \circ X_n$ and $\alpha \circ X$ are the same.*

The proofs of Lemmas 2.2–2.5 are trivial, and so we omit them here. We also need the concept of the Kronecker square of a matrix in order to prove Theorem 2.1. As an illustrative example consider a 3×3 matrix. According to Bellman (1974), the Kronecker square of A, written as $A_{[2]}$, is defined as follows

$$A_{[2]} =$$

$$\begin{bmatrix} a_{11}^2 & a_{12}^2 & a_{13}^2 & 2a_{11}a_{12} & 2a_{11}a_{13} & 2a_{12}a_{13} \\ a_{21}^2 & a_{22}^2 & a_{23}^2 & 2a_{21}a_{22} & 2a_{21}a_{23} & 2a_{22}a_{23} \\ a_{31}^2 & a_{32}^2 & a_{33}^2 & 2a_{31}a_{32} & 2a_{31}a_{33} & 2a_{32}a_{33} \\ a_{11}a_{21} & a_{11}a_{22} & a_{13}a_{23} & a_{11}a_{22}+a_{12}a_{21} & a_{11}a_{23}+a_{13}a_{21} & a_{12}a_{23}+a_{13}a_{22} \\ a_{11}a_{31} & a_{12}a_{32} & a_{13}a_{33} & a_{11}a_{32}+a_{12}a_{31} & a_{11}a_{33}+a_{13}a_{31} & a_{12}a_{33}+a_{13}a_{32} \\ a_{21}a_{31} & a_{22}a_{32} & a_{23}a_{33} & a_{21}a_{32}+a_{22}a_{31} & a_{21}a_{33}+a_{23}a_{31} & a_{22}a_{33}+a_{23}a_{32} \end{bmatrix}$$

$A_{[2]}$ for a $p \times p$ matrix A can be defined in same way. Note that the dimensionality of $A_{[2]}$ is $\{p + p(p-1)/2\} \times \{p + p(p-1)/2\}$. The following lemma shows the relationship between eigenvalues of A and $A_{[2]}$. Its proof is given by Bellman (1974).

LEMMA 2.6. *If a matrix A has $\lambda_1, \lambda_2, \ldots, \lambda_p$ as eigenvalues, then the eigenvalues of $A_{[2]}$ are $\lambda_1, \lambda_2, \ldots, \lambda_p, \lambda_1\lambda_2, \ldots, \lambda_1\lambda_p, \lambda_2\lambda_3, \ldots, \lambda_{p-1}\lambda_p$.*

The proof of Theorem 2.1 is rather too long, and so we give it in the Appendix. Theorem 2.1 shows us that the stationary conditions of INAR(p) and AR(p) are the same.

3. ERGODICITY AND MOMENTS

Let $\{Y(t)\}$ be all the counting series in $\alpha_1 \circ X_{t-1} + \ldots + \alpha_p \circ X_{t-p}$ in (1.2). Obviously, $Y(t)$ is a p-dimensional time series. From (1.2), we obtain

$$\mathcal{F}(X_t, X_{t-1}, \ldots) \subset \mathcal{F}[\varepsilon_t, Y(t), \varepsilon_{t-1}, Y(t-1), \ldots]$$

and so

$$\bigcap_{t=0}^{-\infty} \mathcal{F}(X_t, X_{t-1}, \ldots) \subset \bigcap_{t=0}^{-\infty} \mathcal{F}[\varepsilon_t, Y(t), \varepsilon_{t-1}, Y(t-1), \ldots]. \tag{3.1}$$

The right-hand side of (3.1) is the tail of a σ field independent random series $\{\varepsilon_t, Y(t)\}$; the probability of an event in it is 0 or 1 and so the tail of the σ field of $\{X_t\}$ contains only the measurable sets with probability 0 or 1. We know from Wang (1982) that $\{X_t\}$ is ergodic.

Let $\hat{\gamma}_k = (1/N)\Sigma_{t=1}^{N-k}(X_t - \bar{X})(X_{t+k} - \bar{X})$ be the sample covariance of $\{X_t\}$ and $\hat{\rho}_k = \hat{\gamma}_k/\hat{\gamma}_0$ be the sample correlation coefficient of $\{X_t\}$. Because of the ergodic and stationary properties of $\{X_t\}$, $\hat{\alpha}_k$ and $\hat{\rho}_k$ are strongly consistent (see also Wang, 1982). Then we have the following theorem.

THEOREM 3.1. *$\hat{\gamma}_k$ and $\hat{\rho}_k$ are strongly consistent.*

Let $\boldsymbol{X}_t = (X_t, X_{t-1}, \ldots, X_{t-p+1})\tau$; then (1.2) can be written as

$$\boldsymbol{X}_t = A \circ \boldsymbol{X}_{t-1} + \boldsymbol{\varepsilon}_t \tag{3.2}$$

with

$$A = \begin{bmatrix} \alpha_1 & \alpha_2 & \ldots & \alpha_{p-1} & \alpha_p \\ 1 & 0 & \ldots & 0 & 0 \\ \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & \ldots & 1 & 0 \end{bmatrix}$$

$$\boldsymbol{\varepsilon}_t = \begin{bmatrix} \varepsilon_t \\ 0 \\ \vdots \\ 0 \\ 0 \end{bmatrix}.$$

If we write

$$\gamma_k = E(X_t - EX_t)(X_{t-k} - EX_{t-k})^\tau$$

then

$$\begin{aligned} \gamma_k &= EX_t X_{t-k}^\tau - EX_t EX_{t-k}^\tau \\ &= AEX_{t-1}X_{t-k}^\tau + E\boldsymbol{\varepsilon}_t EX_{t-k}^\tau - EX_t EX_{t-k}^\tau \\ &= A\gamma_{k-1} + (A - I)EX_t EX_t^\tau + E\boldsymbol{\varepsilon}_t EX_t^\tau \\ &= A\gamma_{k-1} \qquad (k > 0) \end{aligned} \tag{3.3}$$

since $EX_t = AEX_{t-1} + E\boldsymbol{\varepsilon}_t$. It is easily seen from (3.3) that

$$\gamma_k = \alpha_1\gamma_{k-1} + \alpha_2\gamma_{k-2} + \ldots + \alpha_p\gamma_{k-p} \tag{3.4}$$

or

$$\rho_k = \alpha_1\rho_{k-1} + \alpha_2\rho_{k-2} + \ldots + \alpha_p\rho_{k-p} \tag{3.5}$$

where $\gamma_k = EX_tX_{t-k}$. Equation (3.4) or (3.5) implies that the correlation structures of INAR(p) and AR(p) are same. From the above discussion we see that the INAR model is similar to the standard AR model not only in form but also in stationary domain and correlation structure, and so we can apply the theory and methods developed for the AR model to the INAR model.

4. ESTIMATION OF PARAMETERS

4.1. *Yule–Walker estimation*

For $k = 1, 2, \ldots, p$ in (3.5) we have the Yule–Walker equations

$$\Gamma\boldsymbol{\alpha} = \boldsymbol{\rho} \tag{4.1}$$

where

$$\begin{aligned} \Gamma &= [\rho_{|i-j|}]_{p\times p} \\ \boldsymbol{\alpha} &= (\alpha_1\alpha_2 \ldots \alpha_p)\tau \\ \boldsymbol{\rho} &= (\rho_1\rho_2 \ldots \rho_p)\tau. \end{aligned}$$

Replacing ρ_i with $\hat{\rho}_i$ in (4.1) yields the Yule–Walker estimate $\hat{\boldsymbol{\alpha}}$ of $\boldsymbol{\alpha}$, which satisfies

$$\hat{\Gamma}\hat{\boldsymbol{\alpha}} = \hat{\boldsymbol{\rho}}. \tag{4.2}$$

Then the estimate of μ is given by

$$\hat{\mu} = (1 - \hat{\alpha}_1 - \ldots - \hat{\alpha}_p)\bar{X} \tag{4.3}$$

and the estimate of σ^2

$$\hat{\sigma}^2 = \frac{1}{N - P} \sum_{t=p+1}^{N} (\varepsilon_t - \bar{\varepsilon}_N)^2 \tag{4.4}$$

where

$$\hat{\varepsilon}_t = X_t - \hat{\alpha}_1 X_{t-1} - \ldots - \hat{\alpha}_p X_{t-p}$$

$$\bar{\varepsilon}_N = \frac{1}{N - P} \sum_{t=p+1}^{N} \hat{\varepsilon}_t. \tag{4.5}$$

THEOREM 4.1. *$\hat{\alpha}_1, \hat{\alpha}_2, \ldots, \hat{\alpha}_p, \hat{\mu}$ and $\hat{\sigma}^2$ are strongly consistent.*

Theorem 4.1 can be proved with the aid of the consistency of $\hat{\rho}_i$.

4.2. *Conditional least-squares estimation*

Let

$$F_t = \mathscr{F}(X_1, X_2, \ldots, X_t)$$

$$\boldsymbol{\beta} = (\mu, \alpha_1, \ldots, \alpha_p)\tau \triangleq (\beta_1, \beta_2, \ldots \beta_{p+1})\tau$$

$$g(\boldsymbol{\beta}, F_t) = E(X_t | F_{t-1})$$

$$= \mu + \alpha_1 X_{t-1} + \ldots + \alpha_p X_{t-p} \qquad (t > p) \tag{4.6}$$

$$Q_N(\boldsymbol{\beta}) = \sum_{t=p+1}^{N} \{X_t - g(\boldsymbol{\beta}, F_t)\}^2. \tag{4.7}$$

Select the $\hat{\boldsymbol{\beta}}$ making $Q_N(\boldsymbol{\beta})$ minimum as the estimator of $\boldsymbol{\beta}$, i.e.

$$Q_N(\hat{\boldsymbol{\beta}}) = \min Q_N(\boldsymbol{\beta}). \tag{4.8}$$

Then $\hat{\boldsymbol{\beta}}$ is called the conditional least-squares (CLS) estimation of $\boldsymbol{\beta}$ (Klimko and Nelson, 1978). $\hat{\boldsymbol{\beta}}$ can be solved from

$$\frac{\partial Q_N}{\partial \mu} = 0 \tag{4.9}$$

$$\frac{\partial Q_N}{\partial \alpha_l} = 0 \qquad (l = 1, 2, \ldots, p), \tag{4.10}$$

which admits

$$\hat{\mu}^* = \bar{X}_N^{(0)} - \sum_{j=1}^{p} \hat{\alpha}_j^* \bar{X}_N^{(j)} \tag{4.11}$$

$$\hat{\Gamma}^* \hat{\boldsymbol{\alpha}}^* = \hat{\boldsymbol{\rho}}^* \tag{4.12}$$

with $\hat{\boldsymbol{\beta}} = (\hat{\mu}^*, \hat{\boldsymbol{\alpha}}^{\tau *})\tau$, where

$$\bar{X}_N^{(j)} = \frac{1}{N-P} \sum_{t=p+1}^{N} X_{t-j}$$

$$\hat{\gamma}_{k-j}^* = \frac{1}{N-P} \sum_{t=p+1}^{N} (X_{t-j} - \bar{X}_N^{(j)})(X_{t-1} - \bar{X}_N^{(k)})$$

$$\hat{\rho}_{k-j}^* = \frac{\hat{\gamma}_{k-j}^*}{\hat{\gamma}_0^*}$$

$$\hat{\Gamma}^* = [\hat{\rho}_{|i-j|}^*]_{p\times p}$$

$$\hat{\boldsymbol{\rho}} = [\hat{\rho}_1^* \hat{\rho}_2^* \ldots \hat{\rho}_p^*]\tau.$$

It is easily seen from (4.2), (4.3), (4.11) and (4.12) that, when N is large enough, $\hat{\Gamma}^* - \hat{\Gamma}$, $\hat{\boldsymbol{\rho}}^* - \hat{\boldsymbol{\rho}}$ and $\bar{X}_N^{(j)} - \bar{X}$ are nearly zero, so that $\hat{\boldsymbol{\alpha}}^*$ and $\hat{\boldsymbol{\alpha}}$ are very close, and so are $\hat{\mu}^*$ and $\hat{\mu}$, which shows us that the CLS estimates in (4.11) and (4.12) are very close to those in (4.2) and (4.3).

It can easily be seen that g in (4.6), $\partial g/\partial\beta_i$ and $\partial^2 g/\partial\beta_i\partial\beta_j$ satisfy all the regularity conditions proposed by Klimko and Nelson (1978). Consequently, the CLS estimation in (4.11) and (4.12) is strongly consistent. Define

$$g_t = g(\boldsymbol{\beta}, F_t) \qquad (t \geqslant p+1)$$

$$U_p(\boldsymbol{\beta}) = X_{p+1} - \alpha_1 X_p - \ldots - \alpha_p X_1 - \mu. \tag{4.13}$$

Then we know from Klimko and Nelson (1978) that the CLS estimation is asymptotically normal with the asymptotic variance $V^{-1}WV^{-1}$, where

$$W = \left[\mathrm{E}U_p^2(\boldsymbol{\beta}) \frac{\partial g_{p+1}}{\partial\beta_i} \frac{\partial g_{p+1}}{\partial\beta_j}\right]_{(p+1)\times(p+1)} \tag{4.14}$$

$$V = \left[E \frac{\partial g_{p+1}}{\partial\beta_i} \frac{\partial g_{p+1}}{\partial\beta_j}\right]_{(p+1)\times(p+1)}. \tag{4.15}$$

For $i \geqslant 2$, $j \geqslant 2$,

$$V_{ij} = \mathrm{E} \frac{\partial g_{p+1}}{\partial\beta_i} \frac{\partial g_{p+1}}{\partial\beta_j} = EX_{t-i}X_{t-j} = \gamma_{i-j} + \mu_x^2$$

where $\mu_x = EX_t$, $V_{1j} = EX_{t-j} = \mu_x$, so that

$$V = \begin{bmatrix} 1 & \mu_x \mathbf{1}\tau \\ \mu_x \mathbf{1} & G + \mu_x \mathbf{1}\mathbf{1}\tau \end{bmatrix} \tag{4.16}$$

where $\mathbf{1} = (1, 1, \ldots, 1)\tau_{p\times 1}$ and $G = [\gamma_{i-j}\tau]_{p\times p}$. From (4.16) we have

$$V^{-1} = \begin{bmatrix} 1 + \mu_x^2 \mathbf{1}\tau G^{-1}\mathbf{1} & -\mu_x \mathbf{1}\tau G^{-1} \\ -\mu_x G^{-1}\mathbf{1} & G^{-1} \end{bmatrix} \tag{4.17}$$

For $2 \leqslant i \leqslant p+1$, $2 \leqslant j \leqslant p+1$,

$$W_{ij} \triangleq EU_p^2(\boldsymbol{\beta}) \frac{\partial g_{p+1}}{\partial \beta_i} \frac{\partial g_{p+1}}{\partial \beta_j}$$

which, by a conditioning argument, reduces to

$$\begin{aligned} W_{ij} &= \sum_{k=1}^{p} \alpha_k(1 - \alpha_k) EX_{p-k+1}X_{p+2-i}X_{p+2-j} + \sigma^2 EX_{p+2-i}X_{p+2-j} \\ &= \sum_{k=1}^{p} \alpha_k(1 - \alpha_k) EX_{p-k+1}X_{p+2-i}X_{p+2-j} + \sigma^2 V_{ij} \end{aligned} \tag{4.18}$$

Similarly we can find

$$\begin{aligned} W_{1j} &\triangleq \mathrm{E}U_p^2(\boldsymbol{\beta}) \frac{\partial g_{p+1}}{\partial \beta_1} \frac{\partial g_{p+1}}{\partial \beta_j} \\ &= \sum_{k=1}^{p} \alpha_k(1 - \alpha_k) EX_{p-k+1}X_{p+1-j} + \sigma^2 V_{1j} \end{aligned} \tag{4.19}$$

From (4.18) and (4.19) we have

$$W_{ij} = \sum_{k=1}^{p} \alpha_k(1 - \alpha_k) EX_{p-k+1}X_{p+2-i}X_{p+2-j} + \sigma^2 V_{ij} \tag{4.20}$$

for $1 \leq i \leq p + 1$, $1 \leq j \leq p + 1$ with the temporary notational convention that $X_{p+1} = 1$. Then we obtain the following theorem.

THEOREM 4.2. *$\hat{\beta} = (\mu^*, \hat{\boldsymbol{\alpha}}^{\tau *})\tau$ is strongly consistent. If $E\varepsilon_t^3 < \infty$, then*

$$(N - P)^{1/2}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) \rightarrow \mathrm{MVN}(0, V^{-1}WV^{-1}) \tag{4.21}$$

5. PREDICTION

Let $F_N = \mathcal{F}(X_1, \ldots, X_N)$ be the σ field generated by $X_1, X_2, \ldots, X_N$. Then the minimum variance predictor $X_N(1)$ of X_{N+1} is

$$\hat{X}_N(1) = E(X_{N+1}|F_N) = \alpha_1 X_N + \ldots + \alpha_p X_{N+1-p} + \mu$$

Similarly we have

$$\hat{X}_N(m) = E(X_{N+m}|F_N) = \sum_{j=1}^{p} E(\alpha_j \circ X_{N+m-j}|F_N) + \mu$$

We know (Yan, 1985) that

$$\begin{aligned} E(\alpha_j \circ X_{N+m-j}|F_N) &= E\{E(\alpha_j \circ X_{N+m-j}|X_{N+m-j}, F_N)|F_N\} \\ &= \alpha_j E(X_{N+m-j}|F_N) = \alpha_j \hat{X}_N(m - j) \end{aligned}$$

so that

$$\hat{X}_N(m) = \sum \alpha_j \hat{X}_N(m - j) + \mu \tag{5.1}$$

It can be seen from (5.1) that the prediction formula of INAR(p) is the same as that of AR(p).

ACKNOWLEDGEMENT

The authors would like to thank the referee for a very careful reading of the paper and for his valuable comments which led to several improvements. They would also like to thank him for corrections to the English.

APPENDIX: PROOF OF THEOREM 2.1

Let

$$S_{n,t} = \begin{cases} 0 & n < 0 \\ \varepsilon_t & n = 0 \\ \alpha_1 \circ S_{n-1,t-1} + \ldots + \alpha_p \circ S_{n-p,t-p} + \varepsilon_t & n > 0 \end{cases} \tag{A1}$$

$$\boldsymbol{S}_{n,t} = (S_{n,t}, S_{n-1,t-1}, \ldots, S_{n-p+1,t-p+1})\tau.$$

Then (A1) can be written as

$$\boldsymbol{S}_{n,t} = A \circ \boldsymbol{S}_{n-1,t-1} + \boldsymbol{\varepsilon}_t \tag{A2}$$

with

$$A = \begin{bmatrix} \alpha_1 & \alpha_2 & \ldots & \alpha_{p-1} & \alpha_p \\ 1 & 0 & \ldots & 0 & 0 \\ \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & \ldots & 1 & 0 \end{bmatrix}$$

$$\boldsymbol{\varepsilon}_t = \begin{bmatrix} \varepsilon_t \\ 0 \\ \vdots \\ 0 \\ 0 \end{bmatrix}.$$

Denote $\boldsymbol{\mu} = E\boldsymbol{\varepsilon}_t$, $\Sigma_1 = E\boldsymbol{\varepsilon}_t\boldsymbol{\varepsilon}_t^{\tau}$. Write $L^2 = \{X, EX^2 < \infty\}$. Define the scalar product on L^2 as $(X, Y) = EXY$. Then L^2 is a Hilbert space.

A1. $S_{n,t} \in L^2$

From (A2) and Lemma 2.2 we have

$$\begin{aligned} \boldsymbol{\mu}_n \triangleq E\boldsymbol{S}_{n,t} &= A\boldsymbol{S}_{n-1,t-1} + \boldsymbol{\mu} \\ &= \boldsymbol{\mu} + A\boldsymbol{\mu} + \ldots + A^n\boldsymbol{\mu} \end{aligned} \tag{A3}$$

independent of t. Then

$$ES_{n,t}S_{n,t}^{\tau} = \Sigma_1 + \boldsymbol{\mu}\boldsymbol{\mu}_{n-1}^{\tau}A\tau + A\boldsymbol{\mu}_{n-1}\boldsymbol{\mu}\tau + B_{n-1} + AES_{n-1,t-1}S_{n-1,t-1} \quad \text{(A4)}$$

where

$$B_{n-1} = \begin{bmatrix} \sum_{i=1}^{p} \alpha_t(1-\alpha_i)ES_{n-i,t-i} & 0 \\ 0 & 0 \end{bmatrix}_{p\times p} \quad \text{(A5)}$$

independent of t. Repeating (A4) n times can show that $ES_{n,t}S_{n,t}^{\tau} < \infty$, also independent of t. Therefore $S_{n,t} \in L^2$.

A2. *Existence*

$$S_{n,t} - S_{n-1,t} = \sum_{t=1}^{p} (\alpha_i \circ S_{n-i,t-i} - \alpha_i \circ S_{n-i-1,t-i})$$

$$E(S_{n,t} - S_{n-1,t})^2 = \sum_{i=1}^{p} E(\alpha_i \circ S_{n-i,t-i} - \alpha_i \circ S_{n-i-1,t-i})^2$$

$$+ 2 \sum_{1\leqslant i\leqslant j\leqslant p} (E(\alpha_i \circ S_{n-i,t-i} - \alpha_i \circ S_{n-i-1,t-i})(\alpha_j \circ S_{n-j,t-j} - \alpha_j \circ S_{n-j-1,t-j})$$

$$= \sum_{i=1}^{p} \alpha_i^2 E(S_{n-i,t-i} - S_{n-i-1,t-i})^2$$

$$+ \sum_{i=1}^{p} \alpha_i(1-\alpha_i)E|S_{n-i,t-i} - S_{n-i-1,t-i}|$$

$$+ 2 \sum_{1\leqslant i<j\leqslant p} \alpha_i\alpha_j|E(S_{n-i,t-i} - S_{n-i-1,t-i})(S_{n-j,t-j} - S_{n-j-1,t-j}) \quad \text{(A6)}$$

$$E|S_{n-i,t-i} - S_{n-i-1,t-i}| = E\left|\sum_{j=1}^{p} (\alpha_j \circ S_{n-j,t-j} - \alpha_j \circ S_{n-j-1,t-j})\right|$$

$$\leqslant \sum_{j=1}^{p} \alpha_j E|S_{n-j,t-j} - S_{n-j-1,t-j}| \quad \text{(A7)}$$

$$E(S_{n,t} - S_{n-1,t})(S_{n-j,t-j} - S_{n-j-1,t-j})$$

$$= \sum_{k=1}^{p} E(\alpha_k \circ S_{n-k,t-k} - \alpha_k \circ S_{n-k-1,t-k})(S_{n-j,t-j} - S_{n-j-1,t-j})$$

$$= \sum_{k=1}^{p} \alpha_k E(S_{n-k,t-k} - S_{n-k-1,t-k})(S_{n-j,t-j} - S_{n-j-1,t-j})$$

$$= \alpha_j E(S_{n-j,t-j} - S_{n-j-1,t-j})^2 + \sum_{k \neq j} \alpha_k E(S_{n-k,t-k} - S_{n-k-1,t-k})(S_{n-j,t-j} - S_{n-j-1,t-j}) \qquad j = 1, 2, \ldots, p-1. \tag{A8}$$

Define

$$\begin{aligned} X(n, t) = [&E(S_{n,t} - S_{n-1,t})^2, \ldots, E(S_{n-p+1,t-p+1} - S_{n-p,t-p+1})^2, \\ &E|S_{n,t} - S_{n-1,t}|, \ldots, E|S_{n-P+1,t-P+1} - S_{n-P,t-P+1}|, \\ &E(S_{n,t} - S_{n-1,t})(S_{n-1,t-1} - S_{n-2,t-1}), \ldots, \\ &E(S_{n-p+2,t-p+2} - S_{n-p+1,t-p+2})(S_{n-p+1,t-p+1} - S_{n-p,t-p+1})]\tau. \end{aligned} \tag{A9}$$

Then $X(n, t)$ is a $\{2p + p(p-1)/2\} \times 1$ vector. Obviously, $X(n, t)$ admits

$$X(n, t) \leqslant DX(n-1, t-1) \tag{A10}$$

where

$$D = \begin{bmatrix} A_1 & C & A_2 \\ 0 & A & 0 \\ A_3 & 0 & A_4 \end{bmatrix} \tag{A11}$$

with

$$C = \begin{bmatrix} \alpha_1(1-\alpha_1) & \ldots & \alpha_p(1-\alpha_p) \\ 0 & \ldots & 0 \\ \vdots & & \vdots \\ 0 & \ldots & 0 \end{bmatrix}_{p \times p} \tag{A12}$$

$$A = \begin{bmatrix} \alpha_1 & \alpha_2 & \ldots & \alpha_{p-1} & \alpha_p \\ 1 & 0 & \ldots & 0 & 0 \\ \vdots & \vdots & \ldots & & \vdots \\ 0 & 0 & \ldots & 1 & 0 \end{bmatrix}_{p \times p} \tag{A13}$$

$$A_1 = \begin{bmatrix} \alpha_1^2 & \alpha_2^3 & \ldots & \alpha_{p-1}^2 & \alpha_p^2 \\ 1 & 0 & \ldots & 0 & 0 \\ \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & \ldots & 1 & 0 \end{bmatrix}_{p \times p} \tag{A14}$$

$$A_2 = \begin{bmatrix} 2\alpha_1\alpha_2 \ldots 2\alpha_1\alpha_p 2\alpha_2\alpha_3 \ldots 2\alpha_{p-1}\alpha_p \\ 0 \end{bmatrix}_{p \times p(p-1)/2} \tag{A15}$$

$$A_3 = \begin{bmatrix} \alpha_1 & 0 & \cdots & 0 & 0 \\ 0 & \alpha_2 & \cdots & 0 & 0 \\ \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & \cdots & \alpha_{p-1} & 0 \\ & & 0 & & \end{bmatrix}_{p(p-1)/2\times p} \qquad \text{(A16)}$$

$A_4 =$

$$\begin{array}{cc}
\begin{array}{ccccccccccccccccccc}
 & & p-1 & & & & & p-2 & & & & & p-3 & & & & 2 & & 1
\end{array} & \\
\left[\begin{array}{ccccccccccccccccccc}
\alpha_2 & \alpha_3 & \cdots & \alpha_{p-1} & \alpha_p & 0 & 0 & \cdots & 0 & 0 & 0 & 0 & \cdots & 0 & 0 & \cdots & 0 & 0 & 0 \\
\alpha_1 & 0 & \cdots & 0 & 0 & \alpha_3 & \alpha_4 & \cdots & \alpha_{p-1} & \alpha_p & 0 & 0 & \cdots & 0 & 0 & \cdots & 0 & 0 & 0 \\
0 & \alpha_1 & \cdots & 0 & 0 & \alpha_2 & 0 & \cdots & 0 & 0 & \alpha_4 & \alpha_5 & \cdots & \alpha_{p-1} & \alpha_p & \cdots & 0 & 0 & 0 \\
\vdots & \vdots & & \vdots & \vdots & \vdots & \vdots & & \vdots & \vdots & \vdots & \vdots & & \vdots & \vdots & & \vdots & \vdots & \vdots \\
0 & 0 & \cdots & \alpha_1 & 0 & 0 & 0 & \cdots & \alpha_2 & 0 & 0 & 0 & \cdots & \alpha_3 & 0 & \cdots & \alpha_{p-2} & 0 & \alpha_p \\
1 & 0 & \cdots & 0 & 0 & 0 & 0 & \cdots & 0 & 0 & 0 & 0 & \cdots & 0 & 0 & \cdots & 0 & 0 & 0 \\
0 & 1 & \cdots & 0 & 0 & 0 & 0 & \cdots & 0 & 0 & 0 & 0 & \cdots & 0 & 0 & \cdots & 0 & 0 & 0 \\
\vdots & \vdots & & \vdots & \vdots & \vdots & \vdots & & \vdots & \vdots & \vdots & \vdots & & \vdots & \vdots & & \vdots & \vdots & \vdots \\
0 & 0 & \cdots & 1 & 0 & 0 & 0 & \cdots & 0 & 0 & 0 & 0 & \cdots & 0 & 0 & \cdots & 0 & 0 & 0 \\
0 & 0 & \cdots & 0 & 0 & 1 & 0 & \cdots & 0 & 0 & 0 & 0 & \cdots & 0 & 0 & \cdots & 0 & 0 & 0 \\
0 & 0 & \cdots & 0 & 0 & 0 & 1 & \cdots & 0 & 0 & 0 & 0 & \cdots & 0 & 0 & \cdots & 0 & 0 & 0 \\
\vdots & \vdots & & \vdots & \vdots & \vdots & \vdots & & \vdots & \vdots & \vdots & \vdots & & \vdots & \vdots & & \vdots & \vdots & \vdots \\
0 & 0 & \cdots & 0 & 0 & 0 & 0 & \cdots & 1 & 0 & 0 & 0 & \cdots & 0 & 0 & \cdots & 0 & 0 & 0 \\
\vdots & \vdots & & \vdots & \vdots & \vdots & \vdots & & \vdots & \vdots & \vdots & \vdots & & \vdots & \vdots & & \vdots & \vdots & \vdots \\
0 & 0 & \cdots & 0 & 0 & 0 & 0 & \cdots & 0 & 0 & 0 & 0 & \cdots & 0 & 0 & \cdots & 0 & 0 & 0
\end{array}\right] &
\begin{array}{l}
\\ \\ p-1 \\ \\ \\ \\ \\ p-2 \\ \\ \\ \\ p-3 \\ \\ \\ 1
\end{array}
\end{array}$$

(A17)

The eigenvalues of D are the union of the eigenvalues of A and

$$B = \begin{bmatrix} A_1 & A_2 \\ A_3 & A_4 \end{bmatrix}.$$

The eigenpolynomial of A is $\lambda^p - \alpha_1\lambda^{p-1} - \ldots - \alpha_{p-1}\lambda - \alpha_p$, so that the eigenvalues of A are inside the unit circle because of the condition given in Theorem 2.1. It is easily seen that B is a condensed form of the Kronecker square of A, i.e. $B = A_{[2]}$. We know from Lemma 2.6 that the eigenvalues of B are inside the unit circle because the eigenvalues of A are inside the unit circle. Therefore all eigenvalues of D are inside the unit circle.

According to Lemma 2.4

$$X(n, t) \leqslant DX(n-1, t-1) \leqslant D^2X(n-2, t-2) \leqslant \ldots \leqslant D^nX(0, t-n)$$

and

$$X(0, t-n) = (E\varepsilon_{t-n}^2, 0, \ldots, 0, E\varepsilon_{t-n}, 0, \ldots, 0)\tau$$
$$= (\sigma^2 + \mu^2, 0, \ldots, 0, \mu, 0, \ldots, 0)\tau.$$

Let $\lambda = \max\{|v|,\ v \text{ is an eigenvalue of } D\}$. It is easily shown that, when n is large enough,

$$X(n, t) \leqslant M\lambda^n (1\ 1 \ldots 1)\tau$$

where M is constant. Therefore

$$E(S_{n,t} - S_{n-1,t})^2 \leqslant M\lambda^n$$
$$||S_{n,t} - S_{n-1,t}|| \leqslant M^{1/2}\lambda^{n/2}$$
$$||S_{n+k,t} - S_{n,t}|| \leqslant \sum_{j=1}^{k} ||S_{n+j,t} - S_{n+j-1,t}||$$
$$\leqslant \sum_{j=1}^{k} M^{1/2}\lambda^{(n+j)/2}$$
$$\leqslant M_1\lambda^{(n+1)/2} \to 0 \qquad (n \to \infty).$$

Therefore S_{n-t} converges in L^2. Let $X_t = \lim_{n\to\infty} S_{n,t}$, $n \to \infty$, in (A1). Then

$$X_t = \sum_{j=1}^{p} (\alpha_j \circ X_{t-j}) + \varepsilon_t$$

A3. *Stationarity*

From (A2), we have

$$ES_{n,t}S_{n-k,t-k}^{\tau} = \boldsymbol{\mu}\boldsymbol{\mu}_{n-k}^{\tau} + AES_{n-1,t-1}S_{n-k,t-k}^{\tau} \tag{A18}$$

Repeating (A18) k times, we know that $ES_{n,t}S_{n-k,t-k}^{\tau}$ is independent of t, since $ES_{n-k,t-k}S_{n-k,t-k}^{\tau}$ is independent of t. Because $EX_tX_{t-k}^{\tau} = \lim ES_{n,t}S_{n-k,t-k}^{\tau}$, $EX_tX_{t-k}^{\tau}$ is independent of t. Therefore $\{X_t\}$ is stationary.

A4. *X_t is non-negative integer valued*

$$E|S_{n,t} - S_{n-1,t}| \leqslant ||S_{n-t} - S_{n-1,t}|| \leqslant M^{1/2}\lambda^{n/2} \qquad (0 < \lambda < 1)$$

so that

$$\sum E|S_{n,t} - S_{n-1,t}| < \infty.$$

From Chung (1974), we obtain $\Sigma(S_{n,t} - S_{n-1,t}) < \infty$ a.e. Write

$$Y_t = \sum (S_{n,t} - S_{n-1,t}) = \lim_{n\to\infty} S_{n,t}.$$

Then from Lemma 2.4, $X_t = Y_t$ a.e. is non-negative integer valued and so is X_t.

A5. Uniqueness

Suppose that there exists another stationary series $\{Y_t\}$ satisfying (2.1). Then

$$X_t - Y_t = \sum (\alpha_i \circ X_{t-i} - \alpha_i \circ Y_{t-i})$$

It can easily be shown by similar arguments to those used previously that

$$\|X_t - Y_t\| \leq M\lambda^n.$$

Therefore

$$\|X_t - Y_t\| = 0, \text{ i.e. } X_t = Y_t.$$

A6. $cov(X_{t-k}, \varepsilon_t) = lim\, cov(S_{n-k,t-k}, \varepsilon_t)$

Since $\operatorname{cov}(S_{n-k,t-k}, \varepsilon_t) = 0$, $(k > 0)$, $\operatorname{cov}(X_{t-k}, \varepsilon_t) = 0$, i.e. $\operatorname{cov}(X_s, \varepsilon_t) = 0$, $s < t$.

REFERENCES

Al-Osh, M. A. and Alzaid, A. A. (1987) First-order integer-valued autoregressive (INAR(1)) process. *J. Time Ser. Anal.* 8(3).

An Hong-Zhi, Cheng Zhao-Guo, Du Jin-Guan and Pan Yi-Min (1982) *Time Series Analysis and Application*. Beijing: Scientific Press.

Bellman, R. (1974) *Introduction to Matrix Analysis* (2nd edn). Bombay: Tata McGraw-Hill.

Chung, K. L. (1974) *A Course in Probability Theory* (2nd edn). New York: Academic Press.

Fudan (1977) *Probability Theory*. Beijing: Academic Press.

Jacob, P. A. and Lewis, P. A. W. (1978a) Discrete time series generated by mixtures: correlation and runs properties. *J. R. Statist. Soc. Ser. B.* 40, 94–105.

—— and —— (1978b) Discrete time series generated by mixtures: asymptotic properties. *J. R. Statist. Soc. Ser. B.* 40, 222–28.

Klimko, L. A. and Nelson, P. I. (1978) On conditional least squares estimation for stochastic process. *Ann. Statist.* 6,629–42.

Steutel, F. W. and Van Harn, K. (1979) Discrete analogues of self-decomposability and stability. *Ann. Prob.* 7,893–99.

Wang Zi-Kun (1978) *Stochastic Process*. Beijing: Scientific Press.

Yan Shi-Jian (1985) *Probability Theory*. Beijing: Scientific Theory.