Ann. Inst. Statist. Math.
32 (1980), Part A, 81-94

ON PREDICTION OF INTEGRATED MOVING AVERAGE PROCESSES

YOSHIHIRO YAJIMA

(Received Mar. 8, 1979; revised Aug. 1, 1979)

Abstract

We shall consider the asymptotic properties of predictors with estimated coefficients for IMA processes and how to determine the order of predictors to minimize the error of prediction. For this purpose, the effect of the initial values on predictors is also considered.

1. Introduction

The Box-Jenkins ARIMA (p, d, q) model [6] for a nonstationary process is considered to be very useful to analyze many time series. However seemingly their definition involves some ambiguities. Recently Huzii has given a theoretical definition of ARIMA processes for prediction and estimation, still holding Box-Jenkins' essential idea, and considered their properties*.

In this paper, we shall consider the asymptotic properties of the error of prediction for an ARIMA process, especially, an IMA process, according to Huzii's definition, when the true coefficients of predictors are replaced by some consistent estimates. Several authors discussed the properties of a predictor with estimated coefficients in view of the asymptotic error when it is adopted for the prediction of another process having the same statistical structure as that of the process employed for estimation, but being independent of this one. (Cf. Bhansali [4], Bloomfield [5], Box and Jenkins [6], Grenander and Rosenblatt [7], Yamamoto [15].)

However we shall treat the case that estimators are constructed by using past realizations of the process itself whose future value is desired to be predicted. Further we shall consider how to determine the order of a predictor to minimize the error of prediction. For this purpose, we shall also discuss the effect of the initial values of the process on predictors.

* Unpublished paper. Huzii, M. (1979): "On an ARIMA process and estimations of parameters for prediction," *Res. Rep. on Inf. Science*, Ser. B-65.

Our main results are as follows. Firstly the initial values do not influence on predictors up to the order $T^{-1/2}$ where T is the number of observations. Next, if we neglect the initial values, we should increase the order k of the predictor so that $\lim_{T\to\infty} T(\max_j |h_j|^{2k})=0$ to minimize the error of prediction where h_j is the characteristic root of MA processes.

2. Definition of IMA processes

In this section, we shall outline the definition of Integrated Moving Average processes (hereafter abbreviated as IMA processes). Let x_t, t being integer, be a real-valued stochastic process for $t\geqq -d+1$, where d is a non-negative integer. We shall denote the Borel field generated by $\{x_t;\ n_1\leqq t\leqq n_2\}$ by $\mathcal{B}_{n_1}^{n_2}(x)$. Let z_t, $t\geqq 1$, be a real-valued $\mathcal{B}_{-d+1}^{\infty}(x)$-measurable stochastic process. For a $\mathcal{B}_{-d+1}^{\infty}(x)$-measurable random variable z, we shall express, for simplicity, the conditional expectation $\mathrm{E}(z/x_{-d+1}, x_{-d+2},\cdots, x_0)$ as $\tilde{\mathrm{E}}(z)$. And for a $\mathcal{B}_{-d+1}^{\infty}(x)$-measurable set M, we also express the conditional probability $\mathrm{P}(M/x_{-d+1}, x_{-d+2},\cdots, x_0)$ as $\tilde{\mathrm{P}}(M)$.

DEFINITION 1. We shall call z_t a weakly stationary process if $\tilde{\mathrm{E}}z_t^2<\infty$, $\tilde{\mathrm{E}}z_t=0$, and $\tilde{\mathrm{E}}z_t z_s=R_{t-s}$ for any $t, s\geqq 1$.

DEFINITION 2. A stochastic process x_t defined for $t\geqq -d+1$, is called an integrated weakly stationary process of order d for $t\geqq t_0$, t_0 being a positive integer, if $y_t=\nabla^d x_t$ is a weakly stationary process for $t\geqq t_0$ where $\nabla x_t=x_t-x_{t-1}$ with $\nabla^d x_t=\nabla(\nabla^{d-1}x_t)$.

In general, an integrated weakly stationary process x_t is expressed as

$$x_t=\sum_{j=0}^{d-1} m_j t^j+S^d y_t\ ,$$

for $t\geqq -d+1$, where $y_t=\nabla^d x_t$ and m_j is a $\mathcal{B}_{-d+1}^{0}(x)$-measurable function. Further Sy_t is defined by

$$Sy_t=\begin{cases}\sum_{l=1}^{t} y_l\ , & t>0\ ,\\ 0\ , & t\leqq 0\ ,\end{cases}$$

with $S^d y_t=S(S^{d-1}y_t)$.

Next let u_t and w_s be real-valued stochastic processes. We shall denote the Hilbert space generated by $\{u_t;\ t_1\leqq t\leqq t_2\}$ by $L^2\{u; t_1, t_2\}$ and the projection of w_s on $L^2\{u; t_1, t_2\}$ by $P_{L^2\{u;t_1,t_2\}}w_s$.

Let us put $y_t=\nabla^d x_t$ for $t\geqq 1$ and assume $\tilde{\mathrm{E}}y_t=0$. Then let a_t be the white noise process of y_t with respect to $\tilde{\mathrm{E}}$, so that $\tilde{\mathrm{E}}\,a_t^2=\sigma_a^2$ and $L^2\{y;1,t\}=L^2\{a;1,t\}$ holds for any t.

DEFINITION 3. We shall call y_t a Moving Average (MA) process of order q if y_t can be expressed as

$$y_t=a_t-\sum_{j=1}^{q}\theta_j a_{t-j}\ ,$$

with constants θ_j, $1\leqq j\leqq q$ for $t\geqq q+1$ where $\theta(z)=1-\sum_{j=1}^{q}\theta_j z^j=0$ has the roots outside the unit disc.

We can find that a MA process of order q is weakly stationary for $t\geqq q+1$, but not for $1\leqq t\leqq q$.

DEFINITION 4. We shall call x_t an IMA process of order (d,q) if $y_t=\nabla^d x_t$ is a MA process of order q.

Thus an IMA process of order (d,q) is an integrated weakly stationary process for $t\geqq q+1$.

3. Prediction of IMA processes ($d=q=1$)

In this section, we shall consider the asymptotic properties of the error of the prediction when the coefficients of predictors are replaced by their estimates. Hereafter all the processes are always assumed to be Gaussian. To make the idea of the proofs clear, we shall at first consider the case $d=q=1$. In the next section, we show that for any d and q the corresponding results still hold. Thus by the previous definition x_t is expressed as

$$x_t=m+\sum_{l=1}^{t}y_l\ ,\qquad t\geqq 1\ ,$$

where $y_t=a_t-\theta a_{t-1}$, $t\geqq 2$, with $|\theta|<1$, while y_t and m are independently distributed.

Let us consider the prediction of x_{T+1} given the observations $\{x_1, x_2,\cdots,x_T\}$. Because $x_{T+1}=x_T+y_{T+1}$ and $L^2\{x;1,T\}=L^2\{x_1,y_2,\cdots,y_T\}$, this problem reduces to the prediction of y_{T+1} based on $\{x_1,y_2,\cdots,y_T\}$. Let $\sum_{i=1}^{k}\beta_{i,k}y_{T+1-i}$ and $\sum_{i=1}^{T}\alpha_{i,T}z_{T+1-i}$ be the projection of y_{T+1} on $L^2\{y;T+1-k,T\}$ and $L^2\{x_1,y_2,\cdots,y_T\}$ respectively where $z_1=x_1$ and $z_i=y_i$, $i\geqq 2$. Further let $\hat{\alpha}_{i,T}^T$ and $\hat{\beta}_{i,k}^T$ be generic symbols for estimators of $\alpha_{i,T}$ and $\beta_{i,k}$ respectively. For simplicity, hereafter the superscript T of estimators is suppressed.

As soon as we try to construct predictors with some estimated coefficients, two problems naturally arise. First, can we neglect the information contained in x_1? If we can, the problem reduces to an ordinary prediction of the stationary process y_t. Seemingly Box and Jenkins [6] adopted this method. Intuitively, it seems natural because it is practically difficult to get good estimators of m^2, $\tilde{\mathrm{E}}\, y_1 y_2$ and $\tilde{\mathrm{E}}\, y_1^2$ needed for $\hat{\alpha}_{i,T}$.

Next, even if we agree to neglect x_1, we have to determine the order k of the predictor $\sum_{i=1}^{k} \hat{\beta}_{i,k} y_{T+1-i}$ to minimize the error of the prediction. If k is large, the number of coefficients which must be estimated is also large and, consequently, the error of estimation may accumulate. On the other hand, the small order k may also lead to a wrong predictor. Hence we have to select k so as to balance these difficulties as $T \to \infty$.

To make the comparison easy, we formulate errors as follows,

$$(1) \quad \tilde{\mathrm{E}}\left\{y_{T+1} - \sum_{i=1}^{T} \hat{\alpha}_{i,T} z_{T+1-i}\right\}^2$$
$$= \tilde{\mathrm{E}}\left\{y_{T+1} - \sum_{i=1}^{T} \beta_{i,T} y_{T+1-i}\right\}^2 + \tilde{\mathrm{E}}\left\{\sum_{i=1}^{T} (\hat{\alpha}_{i,T} - \beta_{i,T}) z_{T+1-i} + \beta_{T,T} m\right\}^2,$$

$$(2) \quad \mathrm{E}\left\{y_{T+1} - \sum_{i=1}^{k} \hat{\beta}_{i,k} y_{T+1-i}\right\}^2$$
$$= \tilde{\mathrm{E}}\left\{y_{T+1} - \sum_{i=1}^{T} \beta_{i,T} y_{T+1-i}\right\}^2 + \tilde{\mathrm{E}}\left\{\sum_{i=1}^{T} (\beta_{i,T} - \beta_{i,T-1}) y_{T+1-i}\right\}^2$$
$$+ \tilde{\mathrm{E}}\left\{\sum_{i=1}^{k} (\hat{\beta}_{i,k} - \beta_{i,k}) y_{T+1-i} + \sum_{i=1}^{T-1} (\beta_{i,k} - \beta_{i,T-1}) y_{T+1-i}\right\}^2,$$

where $\hat{\alpha}_{i,T}$ and $\hat{\beta}_{i,k}$ are made of z_i, $1 \leqq i \leqq T$, and y_i, $2 \leqq i \leqq T$, respectively. In principle the choice of estimators are arbitrary. However, to make the discussion manageable we restrict $\hat{\alpha}_{i,T}$, $\hat{\beta}_{i,k}$ to some type given explicitly later. As the first term, which is the theoretical error and will be discussed later, is in common in (1) and (2), we have only to compare other terms. The second terms of (1) and (2) represent the influence of x_1 on the predictors stated above. While the third one of (2) implies the error which occurs when $\sum_{i=1}^{k} \hat{\beta}_{i,k} y_{T+1-i}$ is used as a predictor. From now on we consider asymptotic distributions and variances of these terms. First let us take the third one of (2), the most laborious part.

Of course, there are many estimators of $\hat{\beta}_{i,k}$. But in this paper we construct $\hat{\beta}_{i,k}$ by using $\hat{\rho}_T$ which is an estimator of $\rho = \tilde{\mathrm{E}}\, y_t y_{t+1} / \tilde{\mathrm{E}}\, y_t^2$. Because we usually obtain $\beta_{i,k}$ solving $P = AB$ where $A = (\rho_{|i-j|})$, $P = {}^T(\rho, 0, \cdots, 0)$, and $B = {}^T(\beta_{1,k}, \beta_{2,k}, \cdots, \beta_{k,k})$. For MA processes of first order $\beta_{i,k}$ and θ are continuous functions of ρ. Thus, if we set $\hat{\rho}_T = \left\{\sum_{t=2}^{T-1} y_t y_{t+1}/(T-2)\right\} \Big/ \left\{\sum_{t=2}^{T} y_t^2/(T-1)\right\}$ as an estimator of ρ, $\hat{\theta}_T=\theta(\hat{\rho}_T)$ and $\hat{\beta}_{i,k}=\beta_{i,k}(\hat{\rho}_T)$ are consistent estimators of θ and $\beta_{i,k}$ respectively. Now we shall prepare a few lemmata for later use.

LEMMA 1. *Let* $\theta(z)=\prod_{j=1}^{q}(1-h_j z)$ *be the characteristic polynomial of a MA process of order* q *where* h_j *are distinct each other and located inside the unit disc. And let its reciprocal be* $\theta^{-1}(z)=\sum_{j=1}^{q} M_j/(1-h_j z)$. *Then* $\beta_{i,k}$ *is expressed as*

$$\beta_{i,k}=\beta_i-\gamma_{i,k}\ ,$$

where $\beta_i=-\sum_{j=1}^{q} M_j h_j^i$, $\gamma_{i,k}=\sum_{j=1}^{q} A_{j,k} h_j^i+\sum_{j=1}^{q} B_{j,k} h_j^{-i}$, *and* $A_{j,k}$, $B_{j,k}$ *are the solutions of*

$$0=\sum_{j=1}^{q}\left(\sum_{l=1}^{q+i}\rho_{l-i}h_j^l\right)A_{j,k}+\sum_{j=1}^{q}\left(\sum_{l=1}^{q+i}\rho_{l-i}h_j^{-l}\right)B_{j,k}\ ,\qquad 1\leqq i\leqq q\ ,$$

$$-\sum_{j=1}^{i}\rho_{q+j-i}\beta_{k+j}=\sum_{j=1}^{q}\left(\sum_{l=1}^{2q+1-i}\rho_{q+1-l}h_j^{k-2q-1+l+i}\right)A_{j,k}$$
$$+\sum_{j=1}^{q}\left(\sum_{l=1}^{2q+1-i}\rho_{q+1-l}h_j^{-(k-2q-1+l+i)}\right)B_{j,k}\qquad 1\leqq i\leqq q\ ,$$

with $\rho_i=\tilde{\mathrm{E}}\, y_t y_{t+i}/\tilde{\mathrm{E}}\, y_t^2$.

PROOF. If we define $w_t=\varepsilon_t-\sum_{j=1}^{q}\theta_j\varepsilon_{t-j}$, $t=0, \pm 1, \pm 2, \cdots$, where ε_t is an orthogonal sequence with mean zero and variance σ_ε^2, then w_t, $t=0, \pm 1, \cdots$, and y_t, $t\geqq q+1$, have the same correlation structure with respect to E and $\tilde{\mathrm{E}}$ respectively. So

$$\sum_{i=1}^{\infty}\beta_i w_{T+1-i}=P_{L^2\{w;\,-\infty,T\}}w_{T+1}\ ,$$

$$\sum_{i=1}^{\infty}\beta_{i,k} w_{T+1-i}=P_{L^2\{w;\,T+1-k,T\}}w_{T+1}\ ,$$

where $\beta_{i,k}=0$ for $i>k$. As $\sum_{i=1}^{\infty}\gamma_{i,k}w_{T+1-i}$ is orthogonal to w_{T+1-i}, $1\leqq i\leqq k$,

$$\sum_{j=0}^{2q}\rho_{q-j}\gamma_{q+i-j,k}=0\ ,\qquad 1\leqq i\leqq k\ , \tag{3}$$

where $\gamma_{i,k}=0$, $i\leqq 0$, and $\gamma_{i,k}=\beta_i$, $i>k$. The characteristic equation of (3) $\sum_{j=0}^{2q}\rho_{q-j}z^j=0$ has the roots h_j and h_j^{-1}. Hence for $k\geqq 2q+1$ $\beta_{i,k}$ is expressed as asserted. These solutions also satisfy (3) even if $k<2q+1$. Thus the proof is completed.

Remark 1. Mentz [9] gave the inverse matrix of $A=(\rho_{|i-j|})$ in a similar way as Lemma 1.

Hereafter it is assumed that k is a function of T such that $0\leqq k\leqq T-1$ and $\lim_{T\to\infty} k=\infty$. And it should be remarked that the limit in probability or law, which appear in the results, is taken with respect to $\tilde{\mathrm{P}}$.

LEMMA 2. *If y_t is a MA process of first order, then the following three sums converge to zero in probability as $T\to\infty$,*

(i) $\sum_{i=1}^{k}(\hat{\theta}_T^i-\theta^i)y_{T+1-i}$,

(ii) $T^{1/2}\sum_{i=1}^{k}(\hat{\theta}_T^{2k-i+2}-\theta^{2k-i+2})y_{T+1-i}$,

(iii) $T^{1/2}\sum_{i=1}^{k}(\hat{\theta}_T^{2k+i+2}-\theta^{2k+i+2})y_{T+1-i}$,

where $\hat{\theta}_T$ is the estimator of θ with $|\theta|<1$ defined previously.

PROOF. (i) A Taylor expansion gives

$$\sum_{i=1}^{k}(\hat{\theta}_T^i-\theta^i)y_{T+1-i}=(\hat{\theta}_T-\theta)\sum_{i=1}^{k}i\tilde{\theta}_{i,T}^{i-1}y_{T+1-i}\ .$$

Then we obtain, for any $\varepsilon>0$,

$$\begin{aligned}
&\mathrm{P}\left\{\left|\sum_{i=1}^{k}(\hat{\theta}_T^i-\theta^i)y_{T+1-i}\right|>\varepsilon\right\}\\
&\leqq\mathrm{P}\left\{|\hat{\theta}_T-\theta|<\delta,\ \left|\sum_{i=1}^{k}(\hat{\theta}_T^i-\theta^i)y_{T+1-i}\right|>\varepsilon\right\}+\mathrm{P}\ \{|\hat{\theta}_T-\theta|\geqq\delta\}\\
&\leqq\mathrm{P}\left\{|\hat{\theta}_T-\theta|<\delta,\ \sum_{i=1}^{k}i|\tilde{\theta}_{i,T}|^{i-1}|y_{T+1-i}|>\varepsilon/\delta\right\}+\mathrm{P}\ \{|\hat{\theta}_T-\theta|\geqq\delta\}\\
&\leqq(\delta/\varepsilon)\ \mathrm{E}\,|y_2|\sum_{i=1}^{\infty}i(|\theta|+\delta)^{i-1}+\mathrm{P}\ \{|\hat{\theta}_T-\theta|\geqq\delta\}\ .
\end{aligned}$$

Taking δ small enough and then letting $T\to\infty$, we can show that the right-hand side of the inequality can be made as small as desired.
(ii) A Taylor expansion also gives

$$T^{1/2}\sum_{i=1}^{k}(\hat{\theta}_T^{2k-i+2}-\theta^{2k-i+2})y_{T+1-i}=T^{1/2}(\hat{\theta}_T-\theta)\sum_{i=1}^{k}(2k-i+2)\tilde{\theta}_{i,T}^{2k-i+1}y_{T+1-i}\ .$$

The theorem (ii) in Section 6a.2 of Rao [11], p. 387, combined with the theorem of Anderson and Walker [2] show that the asymptotic distribution of $T^{1/2}(\hat{\theta}_T-\theta)$ is normal. So it is sufficient to show

$$p-\lim_{T\to\infty}\sum_{i=1}^{k}(2k-i+2)\tilde{\theta}_{i,T}^{2k-i+1}y_{T+1-i}=0 .$$

The absolute value of $\sum_{i=1}^{k}(2k-i+2)\tilde{\theta}_{i,T}^{2k-i+1}y_{T+1-i}$ is dominated by

$$(2k+1)(\max_i|\tilde{\theta}_{i,T}|)^k\sum_{i=1}^{k}|\tilde{\theta}_{i,T}|^{k-i+1}|y_{T+1-i}| ,$$

which is shown to go to zero in probability in a similar way as (i).
(iii) We omit the proof because it is almost the same as that of (ii).

LEMMA 3. *Let y_t be a MA process of order q, then* $\tilde{\mathrm{E}}\left\{\sum_{i=1}^{T-1}(\beta_{i,k}-\beta_{i,T-1})y_{T+1-i}\right\}^2$ *is at most of the same order as* $(\max_j|h_j|)^{2k}$ *where* $\beta_{i,k}=0$, $i>k$.

PROOF. Let us introduce the process w_t defined by $w_t=\varepsilon_t-\sum_{j=1}^{q}\theta_j\varepsilon_{t-j}$, $t=0, \pm1,\cdots$, where ε_t is an orthogonal sequence with mean zero and variance σ_a^2 with respect to E. Then

$$\begin{aligned}
&\tilde{\mathrm{E}}\left\{\sum_{i=1}^{T-1}(\beta_{i,k}-\beta_{i,T-1})y_{T+1-i}\right\}^2\\
&\quad=\mathrm{E}\left\{\sum_{i=1}^{T-1}(\beta_{i,k}-\beta_{i,T-1})w_{-i}\right\}^2\\
&\quad=\mathrm{E}\left\{\sum_{i=1}^{\infty}(\beta_i-\beta_{i,k})w_{-i}\right\}^2-\mathrm{E}\left\{\sum_{i=1}^{\infty}(\beta_i-\beta_{i,T-1})w_{-i}\right\}^2\\
&\quad=\mathrm{E}\left\{\sum_{i=1}^{\infty}\gamma_{i,k}w_{-i}\right\}^2-\mathrm{E}\left\{\sum_{i=1}^{\infty}\gamma_{i,T-1}w_{-i}\right\}^2\\
&\quad=O(\max_j|h_j|^{2k})-O(\max_j|h_j|^{2(T-1)}) .
\end{aligned}$$

The last equality is obtained by using Lemma 1.

Remark 2. Let w_t be the process defined in the proof of Lemma 3. If we define $\mu_k=\mathrm{E}|w_0-P_{L^2\{w;-k,-1\}}w_0|^2$, then

$$\begin{aligned}
\mu_k-\mu_\infty&=\mathrm{E}\left\{w_0-\sum_{i=1}^{k}\beta_{i,k}w_{-i}\right\}^2-\mathrm{E}\left\{w_0-\sum_{i=1}^{\infty}\beta_iw_{-i}\right\}^2\\
&=\mathrm{E}\left\{\sum_{i=1}^{\infty}(\beta_i-\beta_{i,k})w_{-i}\right\}^2=\mathrm{E}\left\{\sum_{i=1}^{\infty}\gamma_{i,k}w_{-i}\right\}^2=O(\max_j|h_j|^{2k}) .
\end{aligned}$$

This is the explicit version for MA processes of order q of a general result given by Grenander and Szegö [8], p. 189.

Using the preceding lemmata, we have the asymptotic distribution of the third term in (2) multiplied by $T^{1/2}$.

THEOREM 1. (i) *The limit distribution as* $T\rightarrow\infty$ *of*

$$T^{1/2}\sum_{i=1}^{k}(\hat{\beta}_{i,k}-\beta_{i,k})y_{T+1-i}+T^{1/2}\sum_{i=1}^{T-1}(\beta_{i,k}-\beta_{i,T-1})y_{T+1-i}$$

is the same as the distribution of $WX+Y$ *if* $\lim_{T\to\infty}T\theta^{2k}(\neq 0)<\infty$, WX *if* $\lim_{T\to\infty}T\theta^{2k}=0$ *where* W, X, *and* Y *are random variables having the normal distributions which are limit distributions of* $T^{1/2}(\hat{\rho}_T-\rho)$, $-\theta'\sum_{i=1}^{k}i\theta^{i-1}\cdot y_{T+1-i}$, *and* $T^{1/2}\sum_{i=1}^{T-1}(\beta_{i,k}-\beta_{i,T-1})y_{T+1-i}$ *respectively. Here* θ' *means* $d\theta/d\rho$.
(ii) $\tilde{\mathrm{E}}(WX+Y)^2=\tilde{\mathrm{E}}W^2\,\tilde{\mathrm{E}}X^2+\tilde{\mathrm{E}}Y^2$.

PROOF. With $\beta_{i,k}$ defined in Lemma 1, we obtain

$$\begin{aligned}(4)\qquad & T^{1/2}\sum_{i=1}^{k}(\hat{\beta}_{i,k}-\beta_{i,k})y_{T+1-i}\\ &=-T^{1/2}\sum_{i=1}^{k}(\hat{\theta}_T^{i}-\theta^{i})y_{T+1-i}-T^{1/2}\sum_{i=1}^{k}(\hat{\gamma}_{i,k}-\gamma_{i,k})y_{T+1-i}\,.\end{aligned}$$

For $q=1$, using Lemma 1 and the relation $\rho=-\theta/(1+\theta^2)$, we have

$$\gamma_{i,k}=(\theta^{2k+i+2}-\theta^{2k-i+2})/(1-\theta^{2k+2})\,.$$

So Lemma 2 asserts that the second term of the right-hand side of (4) converges to zero in probability as $T\rightarrow\infty$. While in a similar way as in Lemma 2, the first term of (4) can be expressed as

$$\begin{aligned}-T^{1/2}\sum_{i=1}^{k}(\hat{\theta}_T^{i}-\theta^{i})y_{T+1-i}&=-T^{1/2}(\hat{\theta}_T-\theta)\sum_{i=1}^{k}i\theta^{i-1}y_{T+1-i}+o_p(1)\\ &=-T^{1/2}(\hat{\rho}_T-\rho)\theta'\sum_{i=1}^{k}i\theta^{i-1}y_{T+1-i}+o_p(1)\end{aligned}$$

where $o_p(1)$ stands for a term which converges to zero in probability as $T\rightarrow\infty$. Hence we have only to consider the asymptotic distribution of

$$-T^{1/2}(\hat{\rho}_T-\rho)\theta'\sum_{i=1}^{k}i\theta^{i-1}y_{T+1-i}+T^{1/2}\sum_{i=1}^{T-1}(\beta_{i,k}-\beta_{i,T-1})y_{T+1-i}\,.$$

First we put

$$V_T=V_{1T}+V_{2T}=T^{1/2}\left\{\sum_{t=2}^{T-1}y_ty_{t+1}/(T-2)-\rho\sum_{t=1}^{T-1}y_t^2/(T-1)\right\}\,,$$

$$X_T=X_{1T}+X_{2T}=-\theta'\sum_{i=1}^{k}i\theta^{i-1}y_{T+1-i}\,,$$

$$Y_T=Y_{1T}+Y_{2T}=T^{1/2}\sum_{i=1}^{T-1}(\beta_{i,k}-\beta_{i,T-1})y_{T+1-i}\,,$$

where

$$V_{1T}=T^{1/2}\left\{\sum_{t=2}^{T-l-2} y_t y_{t+1}/(T-2)-\rho \sum_{t=2}^{T-l-1} y_t^2/(T-1)\right\}$$

$$X_{1T}=-\theta' \sum_{i=1}^{l} i\theta^{i-1} y_{T+1-i} ,$$

$$Y_{1T}=T^{1/2} \sum_{i=1}^{l} (\beta_{i,k}-\beta_{i,T-1}) y_{T+1-i} .$$

Further l is assumed to go to infinity with the same order as T^{δ}, $0<\delta<1$. As $T^{1/2}(\hat{\rho}_T-\rho)=V_T\Big/\left\{\sum_{t=2}^{T} y_t^2/(T-1)\right\}$ and $p-\lim_{T\to\infty}\sum_{t=2}^{T} y_t^2/(T-1)=\tilde{\mathrm{E}}\, y_t^2$, it is sufficient to show that (V_T, X_T, Y_T) converge to $(\tilde{\mathrm{E}}\, y_t^2 W, X, Y)$ jointly in law as $T\to\infty$. (Cf. Billingsley [3], Chapter 1, Corollary 1, p. 31.) Here Y is shown to be identically zero if $\lim_{T\to\infty} T\theta^{2k}=0$ by Lemma 3. The preceding assertion is immediately seen if we show that V_{2T}, X_{2T} and Y_{2T} converge to zero in probability as $T\to\infty$. As y_t is a Gaussian MA process of first order, V_{1T} and (X_{1T}, Y_{1T}) are independently distributed. Thus, if we show that V_{2T}, X_{2T} and Y_{2T} go to zero in probability as $T\to\infty$, noting that V_T and (X_T, Y_T) converge to $\tilde{\mathrm{E}}\, y_t^2 W$ and (X, Y) in law respectively as $T\to\infty$, we can conclude that (V_{1T}, X_{1T}, Y_{1T}) converge to $(\tilde{\mathrm{E}}\, y_t^2 W, X, Y)$ jointly in law as $T\to\infty$ and the assertion follows. (Cf. Billingsley [3], Chapter 1, Theorem 3.2, p. 21.) From now, we shall consider $\tilde{\mathrm{E}}\, V_{2T}^2$, $\tilde{\mathrm{E}}\, X_{2T}^2$, and $\tilde{\mathrm{E}}\, Y_{2T}^2$. Using a well known formula for higher order moments of Gaussian random variables, we have $\tilde{\mathrm{E}}\, V_{2T}^2=O(l/T)$. Next, as y_t is a MA process of first order, we have $\tilde{\mathrm{E}}\, X_{2T}^2=O(l^2\theta^{2l})$ and $\tilde{\mathrm{E}}\, Y_{2T}^2=O(T\theta^{2l})$. Clearly all these amounts go to zero as $T\to\infty$.

(ii) We find, through the proof of (i), that W and (X, Y) can be taken to be independent. Then the proof is immediate since $\tilde{\mathrm{E}}\, W=0$.

Next, we shall consider the second terms of (1) and (2). These terms represent the influence of x_1 on the predictors. Let $\omega_0=(\tilde{\mathrm{E}}\, y_1^2+m^2)/\mathrm{E}\, y_t^2$ and $\omega_1=\tilde{\mathrm{E}}\, y_2 y_1/\tilde{\mathrm{E}}\, y_t^2$, $t\geqq 2$. And let $\hat{\omega}_i$, $i=0, 1$, be a measurable function of $(x_1, y_2, \cdots, y_T)$. We shall assume that $\hat{\omega}_i$ converges to a random variable ω_i' in probability as $T\to\infty$, where ω_i' satisfies $\omega_0'-\omega_1'^2\cdot(1+\rho\theta)^{-1}>0$ almost everywhere. In fact, we see that $\omega_0-\omega_1^2(1+\rho\theta)^{-1}>0$. Since $\alpha_{i,T}$ is the function of ρ, ω_0, and ω_1, as is seen from the next proposition, it will be natural to set $\hat{\alpha}_{i,T}=\alpha_{i,T}(\hat{\rho}_T, \hat{\omega}_0, \hat{\omega}_1)$. Then we have the following result.

PROPOSITION 1. Each of the following expressions converges to zero in probability as $T\to\infty$.

(5) $$T^{1/2}\left\{\sum_{i=1}^{T}(\hat{\alpha}_{i,T}-\beta_{i,T})z_{T+1-i}+\beta_{T,T}m-\sum_{i=1}^{T-1}(\hat{\beta}_{i,T-1}-\beta_{i,T-1})y_{T+1-i}\right\},$$

(6) $$T^{1/2}\sum_{i=1}^{T}(\beta_{i,T}-\beta_{i,T-1})y_{T+1-i}.$$

PROOF. From the definition of $\alpha_{i,T}$, we have

$$\begin{pmatrix}\rho\\0\\\vdots\\0\\0\end{pmatrix}=\begin{pmatrix}1&\rho&0\cdots0&0\\\rho&1&\rho\cdots0&0\\\vdots&\vdots&\ddots\quad\vdots&\vdots\\0&0&0\cdots1&\omega_1\\0&0&0\cdots\omega_1&\omega_0\end{pmatrix}\begin{pmatrix}\alpha_{1,T}\\\alpha_{2,T}\\\vdots\\\alpha_{T-1,T}\\\alpha_{T,T}\end{pmatrix}.$$

Using the modified method of Anderson [1], p. 187, we obtain

(7) $$\begin{aligned}&\alpha_{i,T}=\rho\mu_{T-1,T-1}\mu_{T-i,T-1}/(\omega_0-\omega_1\mu_{1,T-1})+\beta_{i,T-1}, \qquad 1\leqq i\leqq T-1,\\&\alpha_{T,T}=-\rho\mu_{T-1,T-1}/(\omega_0-\omega_1\mu_{1,T-1}),\end{aligned}$$

where $\mu_{i,T-1}=A_{T-1}\theta^i+B_{T-1}\theta^{-i}$ and A_{T-1}, B_{T-1} are the solutions of the following equation,

$$\begin{aligned}\omega_1&=(\theta+\rho\theta^2)A_{T-1}+(\theta^{-1}+\rho\theta^{-2})B_{T-1},\\0&=(\rho\theta^{T-2}+\theta^{T-1})A_{T-1}+(\rho\theta^{-(T-2)}+\theta^{-(T-1)})B_{T-1}.\end{aligned}$$

$\beta_{i,T}$ is immediately obtained if ω_0 is replaced by $\omega_2=\tilde{\mathrm{E}}\,y_1^2/\tilde{\mathrm{E}}\,y_t^2$, $t\geqq 2$, in (7). Putting $A_{T-1}=\omega_1A'_{T-1}(\rho)$, $B_{T-1}=\omega_1B'_{T-1}(\rho)$, we see that (5) is equal to

$$\begin{aligned}&\hat{\rho}_TT^{1/2}\hat{\mu}_{T-1,T-1}(\hat{\omega}_0-\hat{\omega}_1\hat{\mu}_{1,T-1})^{-1}\\&\qquad\times\left[\hat{\omega}_1\sum_{i=1}^{T-1}\{A'_{T-1}(\hat{\rho}_T)\hat{\theta}_T^i+B'_{T-1}(\hat{\rho}_T)\hat{\theta}_T^{-i}\}y_{i+1}-x_1\right]\\&-\rho T^{1/2}\mu_{T-1,T-1}(\omega_2-\omega_1\mu_{1,T-1})^{-1}\\&\qquad\times\left[\omega_1\sum_{i=1}^{T-1}\{A'_{T-1}(\rho)\theta^i+B'_{T-1}(\rho)\theta^{-i}\}y_{i+1}-x_1\right]+T^{1/2}\beta_{T,T}m.\end{aligned}$$

Now we shall show that all terms converge to zero in probability as $T\to\infty$. As for the first one, by noting that $\mu_{T-1,T-1}$ is equal to ω_1 multiplied by the term of order θ^T, the assumption asserts that $p-\lim_{T\to\infty}\hat{\rho}_TT^{1/2}\hat{\mu}_{T-1,T-1}=0$ and $p-\lim_{T\to\infty}\hat{\omega}_0-\hat{\omega}_1\hat{\mu}_{1,T-1}=\omega'_0-\omega'^2_1(1+\rho\theta)^{-1}$. Moreover, using Lemma 2.(i), we have

$$p-\lim_{T\to\infty}\sum_{i=1}^{T-1}\{A'_{T-1}(\hat{\rho}_T)\hat{\theta}_T^i-A'_{T-1}(\rho)\theta^i+B'_{T-1}(\hat{\rho})_T\hat{\theta}_T^{-i}-B'_{T-1}(\rho)\theta^{-i}\}y_{T+1-i}=0.$$

Then the first term converges to zero in probability as $T\to\infty$. The other terms also converge to zero in probability because $\beta_{T,T}$ and $\mu_{T-1,T-1}$ go to zero exponentially as $T\to\infty$. The first half of the proof is now finished.

Next, (6) is equal to

$$\rho T^{1/2}\mu_{T-1,T-1}(\omega_2-\omega_1\mu_{1,T-1})^{-1}\left(\sum_{i=1}^{T-1}\mu_{i,T-1}y_{i+1}-y_1\right),$$

which clearly converges to zero in probability when $T\to\infty$, as was proved.

Remark 3. Using Proposition 1, we find that

$$p-\lim_{T\to\infty}\left(y_{T+1}-\sum_{i=1}^{T}\beta_{i,T}y_{T+1-i}\right)=p-\lim_{T\to\infty}\left(y_{T+1}-\sum_{i=1}^{T-1}\beta_{i,T-1}y_{T+1-i}\right).$$

Accordingly the asymptotic variance of the first term in (1) and (2) is one-step ahead prediction error of an ordinary MA process given all the present and past observations.

The last two results give the answers for the two problems mentioned at the beginning of this section. First it follows from Proposition 1 that x_1 does not influence on the predictor up to the order of $T^{-1/2}$. Thus it seems useless to make much efforts to estimate ω_i if our main purpose is to predict future values. Next, if we neglect x_1, noting Theorem 1, we should increase k as $T\to\infty$ so that $\lim_{T\to\infty}T\theta^{2k}=0$ in order to minimize the error of prediction.

To end this section, we shall comment on the choice of estimators $\hat{\beta}_{i,k}$. Though Whittle [14] indicated that $\hat{\rho}_T$ employed here is much less efficient than the maximum likelihood estimator, we can also obtain the result corresponding to Theorem 1 even if the improved estimator proposed by Walker [13] is used. Further with the more improved estimator of Mentz [10] having the same asymptotic variance as that of the maximum likelihood estimator, we can see that the same result still holds. So we shall omit the detail.

4. Prediction of IMA processes (General case)

In this section, we shall show that the results obtained in the previous section can be extended to general IMA processes of order (d, q). As $x_{T+1}=\sum_{l=1}^{d}(-1)^{l-1}\binom{d}{l}x_{T+1-l}+y_{T+1}$ and $L^2\{x;1,T\}=L^2\{x_1,\cdots,x_d,y_{d+1},\cdots,y_T\}$, the prediction of x_{T+1} given $\{x_1,x_2,\cdots,x_T\}$ reduces to that of y_{T+1} based on $\{x_1,\cdots,x_d,y_{d+1},\cdots,y_T\}$. Hereafter we assume $q>d$. For $q\leqq d$, the results are given in a similar way. Let $\sum_{i=1}^{T}\alpha_{i,T}z_{T+1-i}$ and

$\sum_{i=1}^{T} \delta_{i,t}\tilde{z}_{T+1-i}$ be the projection of y_{T+1} on $L^2\{z; 1, T\}$ and $L^2\{\tilde{z}; 1, T\}$ respectively where $z_i = x_i$, $\tilde{z}_i = x_i - \tilde{\mathrm{E}}\, x_i$, $1 \leqq i \leqq d$ and $z_i = \tilde{z}_i = y_i$, $d+1 \leqq i \leqq T$. On the other hand, $\sum_{i=1}^{k} \beta_{i,k} y_{T+1-i}$, $1 \leqq i \leqq T-q$, have the same definition as in Section 3.

If $\hat{\alpha}_{i,T}$ and $\hat{\beta}_{i,k}$ are the functions of z_i, $1 \leqq i \leqq T$ and y_i, $q+1 \leqq i \leqq T$ respectively, we have the following equations corresponding to (1), (2)

$$\begin{aligned}
&\tilde{\mathrm{E}} \left\{ y_{T+1} - \sum_{i=1}^{T} \hat{\alpha}_{i,T} z_{T+1-i} \right\}^2 \\
&\quad = \tilde{\mathrm{E}} \left\{ y_{T+1} - \sum_{i=1}^{T} \delta_{i,T} \tilde{z}_{T+1-i} \right\}^2 + \tilde{\mathrm{E}} \left\{ \sum_{i=1}^{T} (\hat{\alpha}_{i,T} z_{T+1-i} - \delta_{i,T} \tilde{z}_{T+1-i}) \right\}^2 , \\
&\tilde{\mathrm{E}} \left\{ y_{T+1} - \sum_{i=1}^{k} \hat{\beta}_{i,k} z_{T+1-i} \right\}^2 \\
&\quad = \tilde{\mathrm{E}} \left\{ y_{T+1} - \sum_{i=1}^{T} \delta_{i,T} \tilde{z}_{T+1-i} \right\}^2 + \tilde{\mathrm{E}} \left\{ \sum_{i=1}^{T} (\delta_{i,T} \tilde{z}_{T+1-i} - \beta_{i,T-q} y_{T+1-i}) \right\}^2 \\
&\qquad + \tilde{\mathrm{E}} \left\{ \sum_{i=1}^{k} (\hat{\beta}_{i,k} - \beta_{i,k}) y_{T+1-i} + \sum_{i=1}^{T-q} (\beta_{i,k} - \beta_{i,T-q}) y_{T+1-i} \right\}^2 .
\end{aligned}$$

In order to generalize the results in Section 3, we shall show that M_j and h_j, which are the functions of ρ_i, satisfy the assumption of the theorem (ii) in Section 6a.2 of Rao [11], p. 387 and that the partial derivatives of β_i with respect to ρ_j, $1 \leqq j \leqq q$, exist.

LEMMA 4. *Let us use the same notations and assumptions as in Lemma 1. Then M_j and h_j, accordingly also β_i, are totally differentiable functions of $(\rho_1, \rho_2, \cdots, \rho_q)$ where $\rho_i = \tilde{\mathrm{E}}\, y_t y_{t+i} / \tilde{\mathrm{E}}\, y_t^2$, $t \geqq q+1$.*

PROOF. First we show that h_j is totally differentiable. For this purpose it suffices to show that the Jacobian $\partial(\rho)/\partial(h)$ of $(\rho_1, \cdots, \rho_q)$ with respect to $(h_1, \cdots, h_q)$ never vanishes. If this assertion is true, h_j is a C^1-class function of $(\rho_1, \cdots, \rho_q)$, and hence is totally differentiable. Let us introduce the following polynomial

$$U(w) = 1 + \sum_{i=1}^{q} \rho_i (w^i + w^{-i}) = \sum_{i=0}^{q} v_i z^i = V(z) , \qquad z = w + w^{-1} .$$

Then the roots of $V(z)=0$ are $z_i = h_i + h_i^{-1}$, $1 \leqq i \leqq q$. As $\partial(\rho)/\partial(h) = \partial(\rho)/\partial(v) \times \partial(v)/\partial(z) \times \partial(z)/\partial(h)$, let us show that $\partial(\rho)/\partial(v)$, $\partial(v)/\partial(z)$ and $\partial(z)/\partial(h)$ never vanish. First there is the relation between ρ_i and v_i that $v_i = \sum_{j=i}^{q} c_{ij} \rho_j$ where c_{ij} are the constants with $c_{ii} = 1$ (see Rose [12]). Thus we have $\partial(\rho)/\partial(v) = 1$. Next since $h_i \neq h_j$ for any i, j with $i \neq j$, it follows that $\{z_i\}$ are also distinct each other. Then it can be easily seen that $\partial(v)/\partial(z) = \text{const.} \prod_{i<j} (z_i - z_j)$ never vanishes. Finally, as $|h_j| < 1$ for any j, $\partial(z)/\partial(h)=\prod_{j=1}^{q}(1-h_i^{-2})$ is non-zero.

On the other hand, about M_j, we have the identity $\sum_{j=1}^{q} M_j \prod_{\substack{i=1\\ i\neq j}}^{q}(1-h_i z)=1$. Hence M_j is totally differentiable with respect to $(h_1,\cdots,h_q)$ and also with respect to $(\rho_1,\cdots,\rho_q)$, by noting $\partial(\rho)/\partial(h)\neq 0$. We have finished the proof.

Now if we set $\hat{\rho}_i=\left\{\sum_{t=q+1}^{T-i} y_t y_{t+i}/(T-i-q)\right\}\Big/\left\{\sum_{t=q+1}^{T} y_t^2/(T-q)\right\}$ and $\hat{\beta}_{i,k}=\beta_{i,k}(\hat{\rho}_1,\cdots,\hat{\rho}_q)$, the general results corresponding to Theorem 1 and Proposition 1 are given. We state them without proofs because they can be obtained essentially in the same way by using the lemmata in Section 3.

THEOREM 2. (i) *The limit distribution as* $T\to\infty$ *of* $T^{1/2}\left\{\sum_{i=1}^{k}(\hat{\beta}_{i,k}-\beta_{i,k})+\sum_{i=1}^{T-q}(\beta_{i,k}-\beta_{i,T-q})\right\}y_{T+1-i}$ *is the same as the distribution of* $\sum_{i=1}^{q} W_i X_i+Y$ *if* $\lim_{T\to\infty} T(\max_j |h_j|)^{2k}\ (\neq 0)<\infty$, $\sum_{i=1}^{q} W_i X_i$ *if* $\lim_{T\to\infty} T(\max_j |h_j|)^{2k}=0$, *where* W_i, X_i, *and* Y *are random variables having the normal distributions which are the limit distributions of* $T^{1/2}(\hat{\rho}_i-\rho_i)$, $\sum_{j=1}^{k}\partial\beta_j/\partial\rho_i\times y_{T+1-i}$ *and* $T^{1/2}\sum_{i=1}^{T-q}(\beta_{i,k}-\beta_{i,T-q})y_{T+1-i}$ *respectively.*

(ii) $\tilde{\mathrm{E}}\left\{\sum_{i=1}^{q} W_i X_i+Y\right\}^2=\sum_{i=1}^{q}\sum_{j=1}^{q}\tilde{\mathrm{E}}\,W_iW_j\,\tilde{\mathrm{E}}\,X_iX_j+\tilde{\mathrm{E}}\,Y^2$.

PROPOSITION 2. Define the following quantities and matrices; $\omega_{ij}=\tilde{\mathrm{E}}\,z_iz_j/\tilde{\mathrm{E}}\,y_t^2$ where at least one of $1\leqq i\leqq q$ or $1\leqq j\leqq q$ holds, $A_{11}(\rho)=(\rho_{|i-j|})$, $1\leqq i,\ j\leqq T-q$, $A_{12}(\omega)=(w_{T+1-i,q+1-j})$, $1\leqq i\leqq T-q$, $1\leqq j\leqq q$, $A_{21}(w)={}^TA_{12}(w)$, $A_{22}(w)=(w_{i,j})$, $1\leqq i,\ j\leqq q$.

Assume that $\hat{\omega}_{ij}$ is a function of $(x_1,x_2,\cdots,x_d,y_{d+1},\cdots,y_T)$ such that $p-\lim_{T\to\infty}\hat{\omega}_{ij}=\omega'_{ij}$ where ω'_{ij} is a random variable and $A_{22}(\omega')-A_{21}(\omega')\cdot A_{11}^{-1}(\rho)A_{12}(\omega')$ is positive definite almost everywhere. Further let define $\hat{\alpha}_{i,T}=\alpha_{i,T}(\hat{\rho}_1,\cdots,\hat{\rho}_q,\hat{\omega}_{j,k})$. Then each of the following expressions converges to zero in probability as $T\to\infty$.

$$T^{1/2}\left\{\sum_{i=1}^{T}(\hat{\alpha}_{i,T}z_{T+1-i}-\delta_{i,T}\tilde{z}_{T+1-i})-\sum_{i=1}^{T-q}(\hat{\beta}_{i,T-q}-\beta_{i,T-q})y_{T+1-i}\right\},$$

$$T^{1/2}\sum_{i=1}^{T}(\delta_{i,T}\tilde{z}_{T+1-i}-\beta_{i,T-q}y_{T+1-i}).$$

Thus we can extend the assertions in the previous section to general IMA processes of order (d,q). First the initial values do not influence on the predictor up to the order of $T^{-1/2}$. Next, if we neglect the initial values, we should increase k as $T \to \infty$ so that $\lim_{T \to \infty} T(\max_j |h_j|)^{2k} = 0$ to minimize the error of prediction.

Acknowledgement

The author is very glad to express his gratitude to Professor M. Huzii for his valuable advices and encouragements.

TOKYO INSTITUTE OF TECHNOLOGY

REFERENCES

[1] Anderson, T. W. (1970). *The Statistical Analysis of Time Series*, Wiley, New York.

[2] Anderson, T. W. and Walker, A. M. (1964). On the asymptotic distribution of the autocorrelations of a sample from a linear stochastic process, *Ann. Math. Statist.*, **35**, 1296-1303.

[3] Billingsley, P. (1968). *Convergence of Probability Measures*, Wiley, New York.

[4] Bhansali, R. J. (1978). Linear prediction by autoregressive model fitting in the time domain, *Ann. Statist.*, **6**, 224-231.

[5] Bloomfield, P. (1972). On the error of a prediction of a time series, *Biometrika*, **59**, 501-508.

[6] Box, G. E. P. and Jenkins, J. M. (1970). *Time Series Analysis, Forecasting and Control* Holden Day, San Francisco.

[7] Grenander, U. and Rosenblatt, M. (1957). *Statistical Analysis of Stationary Time Series*, Wiley, New York.

[8] Grenander, U. and Szegö, G. (1958). *Toeplitz Forms and Their Applications*, Univ. of California Press, Berkeley.

[9] Mentz, R. P. (1976). On the inverse of some covariance matrices of Toeplitz type, *SIAM J. Appl. Math.*, **31**, 426-437.

[10] Mentz, R. P. (1977). Estimation in the first order moving average model based on sample autocorrelations, *Ann. Statist.*, **5**, 1250-1257.

[11] Rao, C. R. (1973). *Linear Statistical Inference and Its Applications*, 2nd ed., Wiley, New York.

[12] Rose, D. E. (1977). Forecasting aggregates of independent ARIMA processes, *J. Econometrics*, **5**, 323-345.

[13] Walker, A. M. (1961). Large sample estimation of parameters for moving average models, *Biometrika*, **48**, 343-357.

[14] Whittle, P. (1953). Estimation and information in stationary time series, *Ark. Mat.*, **2**, 423-434.

[15] Yamamoto, T. (1976). Asymptotic mean square prediction error for autoregressive model with estimated coefficients, *Appl. Statist.*, **25**, 123-127.