*Journal of the Korean Statistical Society (2006),* **35**: *1, pp 37–47*

# PREDICTION MEAN SQUARED ERROR OF THE POISSON INAR(1) PROCESS WITH ESTIMATED PARAMETERS

HEE-YOUNG KIM[1] AND YOUSUNG PARK[2]

ABSTRACT

Recently, as a result of the growing interest in modeling stationary processes with discrete marginal distributions, several models for integer valued time series have been proposed in the literature. One of these models is the integer-valued autoregressive (INAR) models. However, when modeling with integer-valued autoregressive processes, the distributional properties of forecasts have been not yet discovered due to the difficulty in handling the Steutal Van Harn thinning operator " $\circ$ "(Steutal and van Harn, 1979). In this study, we derive the mean squared error of $h$-step-ahead prediction from a Poisson INAR(1) process, reflecting the effect of the variability of parameter estimates in the prediction mean squared error.

*AMS 2000 subject classifications.* Primary 60G10; Secondary 37M10.
*Keywords.* Stationary process, integer valued time series, mean-squared prediction errors.

## 1. INTRODUCTION

There has been a growing research in modeling discrete time stationary processes with discrete marginal distributions. The usual linear models for time series, ARMA models, are suitable for modeling stationary dependent sequences under the Gaussian assumption. However, the Gaussian assumption is often inappropriate for modeling counting data. Thus, motivated by the need for modeling correlated series of counts, several models for integer-valued time series have been proposed in the literature.

Received January 2006; accepted February 2006.
[1]Corresponding author. Institute of Statistics, Korea University, Seoul 136-701, Korea (e-mail : starkim@korea.ac.kr)
[2]Department of Statistics, Korea University, Seoul 136-701, Korea (e-mail : yspark@korea.ac.kr)

Following Cox (1981), the models are divided into two broad categories: observation-driven and parameter-driven models. Our interest is in a special class of observation-driven models, the so-called integer valued autoregressive (INAR) process introduced by McKenzie (1985) and Al-Osh and Alzaid (1987). Especially, the first-order INAR (INAR(1)) model is attractive because it is a typical example of the branching process of population: the present size of population is the sum of those that remain (or survive) from the previous period and immigrants that enter in the intervening period. Moreover, the Poisson distribution in the INAR(1) process is the counterpart of the Gaussian distribution in the continuous AR(1) process. Mckenzie (1988) showed that the Poisson INAR(1) represents a $M/M/\infty$ queueing process observed at regularly spaced intervals of time. Therefore, in the sequel, our main focus is on the Poisson INAR(1) process.

The INAR has been extensively studied for its properties such as the ergodicity, the higher-order moments, and cumulants of the process. However, the forecasting aspect in the INAR process has rarely discussed mainly because of distributional complexity incurred from the binomial thinning operator, called as the Steutal Van Harn operator "$\circ$". To overcome this difficulty, Freeland and McCabe (2004) studied the estimation of the probabilities for integer-valued $h$-step-ahead forecasts in a Poisson INAR(1) process and proposed a method calculating the confidence intervals for the h-step ahead probability mass. As alternative methods, two approaches, *i.e.* Bayesian approach (McCabe and Martin, 2005) and bootstrap approach (Jung and Tremayne, 2006), have been applied to estimate the predictive mass function of $h$-step-ahead forecasts in INAR(1) or INAR(2) process. However, these approaches are worthwhile only when the data comprise very low count values, for example, 0, 1 or 2 as in Freeland and McCabe (2004) and McCabe and Matin (2005) since their methods estimate the probability that the $h$-step-ahead forecast has such a small count, 0, 1 or 2.

We usually use the conditional expectation as a forecast for general integer-valued process. In this case, we need the mean square error of the forecast to construct its confidence interval. Therefore, after briefly discussing the basic properties of the INAR process in Section 2, we derive the mean square error for the $h$-step-ahead forecast in the Poisson INAR(1) and also measure the effect of the variability arising from estimating parameters in the Poisson INAR(1) in Section 3. Some concluding remarks are included in Section 4.

## 2. TIME SERIES MODELS FOR COUNTS

Many studies have been done for a family of stochastic processes which have discrete distributions of a certain type known as self-decompossible. This family of processes shares many features in common with the standard ARMA models such as the form of the process expressed as a difference equation. Indeed, the essential difference of the discrete variate process from the standard ARMA model is that the discrete variate process uses the thinning operator defined below in the place of the multiplication in the ARMA model.

Suppose $X$ is a random variable taking values in $\mathcal{N}_0 = \{0, 1, \ldots\}$ with probability generating function $p(s) = Es^X$. Then the distribution of $X$ is discrete self-decompossible if

$$p(s) = p(1 - \alpha + \alpha s)p_\alpha(s), \ |s| \leq 1, \ \alpha \in (0, 1), \tag{2.1}$$

where $p_\alpha(s)$ is also a probability generating function. Let $w_i$ be an *iid* sequence of Bernoulli random variables with $P(w_i = 1) = \alpha$. By noting that $\sum_{i=1}^{X} w_i$ has the probability generating function $p(1 - \alpha + \alpha s)$, we see that the decomposition (2.1) can be written as

$$X = \alpha \circ X + \varepsilon_\alpha, \tag{2.2}$$

where $\alpha \circ X$ denotes $\sum_{i=1}^{X} w_i$ and $\varepsilon_\alpha$ is a random variable independent of $\alpha \circ X$, having the probability generating function $p_\alpha(s)$. Further details of the self-decompossible class of distributions can be found in Steutel and Van Harn (1979). By virtue of (2.2), a stationary integer-valued time series process $X_t$ is defined by

$$X_t = \alpha \circ X_{t-1} + \varepsilon_t, \tag{2.3}$$

where $X_{t-1}$ is independent of $\varepsilon_t$ which is an *iid* non-negative integer-valued random variables with mean $E(\varepsilon_t) = \mu_\varepsilon, \mathrm{Var}(\varepsilon_t) = \sigma_\varepsilon^2$. The $\circ$-operation is referred as binomial thinning and model (2.3) is referred as integer-valued autoregressive process with order 1 (INAR(1)). McKenzie (1985), Al-Osh and Alzaid (1987) and Park and Kim (2000) discussed some basic and asymptotic properties of the INAR(1) process under various assumptions of the marginal distribution.

Under the assumptions of thinning operator, the basic properties are $E(\alpha \circ X|X) = \alpha X$, $E(\alpha \circ X) = \alpha E(X)$, $\mathrm{Var}(\alpha \circ X|X) = \alpha(1-\alpha)X$ and $\mathrm{Var}(\alpha \circ X) = \alpha^2 \mathrm{Var}(X) + \alpha(1-\alpha)E(X)$.

Since the process $\{X_t\}$ satisfying (2.3) is a second-order stationary if $0 \leq \alpha < 1$, it is easy to show that $E(X_t|X_{t-1}) = \alpha X_{t-1} + \mu_\varepsilon$, $E(X_t) = \mu_\varepsilon/(1-\alpha)$, $\mathrm{Var}(X_t|X_{t-1}) = \alpha(1-\alpha)X_{t-1} + \sigma_\varepsilon^2$ and $\mathrm{Var}(X_t) = (\alpha\mu_\varepsilon + \sigma_\varepsilon^2)/(1-\alpha^2)$. Using these two moments, we have that the autocorrelation function of INAR(1) is $\rho_X(k) = \alpha^k$, for $k = 0, 1, \ldots$

The INAR($p$) process $\{X_t\}$, a seemingly natural extension of the INAR(1) process, is defined by

$$X_t = \alpha_1 \circ X_{t-1} + \alpha_2 \circ X_{t-2} + \cdots + \alpha_p \circ X_{t-p} + \varepsilon_t, \tag{2.4}$$

where the assumptions for $\{\varepsilon_t\}$ are the same as those of INAR(1), all counting series $\{w_{ij}\}$ involved in $\alpha_i \circ X_{t-i} = \sum_{j=1}^{X_{t-i}} w_{ij}$ are mutually independent, and $\alpha_i \in [0,1]$ for $i = 1, \ldots, p$. Furthermore, $\{\varepsilon_t\}$ are independent of all counting series. This definition has been proposed by Du and Li (1991), which is different from that of Alzaid and Al-Osh (1990). That is, they assumed that the conditional distribution of the vector $(\alpha_1 \circ X_t, \ldots, \alpha_p \circ X_t)$ given $X_t = x$ is multinomial with parameter $(\alpha_1, \ldots, \alpha_p, x)$ and is independent of the past history of the process. The two different formulations lead to different second-order structures. In particular, the INAR($p$) under Du and Li (1991) has the same autocorrelation structure as that of the continuous countpart AR($p$) and the conditional expectation of $X_t$ given past information is linear, whereas the INAR($p$) under Alzaid and Al-Osh (1990) not only has the same autocorrelation structure as that of the ARMA($p, p-1$) but also the conditional expectation of $X_t$ is non-linear, thus necessitating non-linear least squares estimator to be used, and should be extremely complex. For theses reasons, several studies (Silva and Oliveira, 2005; Latour, 1998; Park *et al*, 2006) focused only on INAR($p$) under Du and Li (1991)'s assumptions.

## 3. Properties of Predictor for Poisson INAR(1) Process

The most common procedure for predicting in time series models is to use conditional expectation, which is the predictor with minimum mean squared error (MSE). Let $\tilde{X}_{T+h} \equiv E(X_{T+h}|X_1, \ldots, X_T)$. Then we have the prediction mean squared error (PMSE) of $\tilde{X}_{T+h}$ as given by

LEMMA 3.1. *Let $\{X_t\}$ be a stationary INAR(1) process of (2.3), where $\{\varepsilon_t\}$ is iid Poisson random variable with parameter $\lambda$ . Then,*

$$\tilde{X}_{T+h} = E(X_{T+h}|\mathcal{F}_T) = \alpha^h X_T + \frac{\lambda(1-\alpha^h)}{1-\alpha},$$

*where* $\mathcal{F}_T = \sigma(X_1, \ldots, X_T)$ *and the PMSE of* $\tilde{X}_{T+h}$ *is given by*

$$E(X_{T+h} - \tilde{X}_{T+h})^2 = \frac{\lambda(1-\alpha^{2h})}{1-\alpha}. \tag{3.1}$$

The proof and all subsequent theoretical proofs are provided in the Appendix.

Lemma 3.1 shows the PMSE of $\tilde{X}_{T+h}$ under the assumption that the parameters $\boldsymbol{\theta} = (\alpha, \lambda)'$ are known. However, in the practical sense, $\boldsymbol{\theta}$ is unknown in advance. Thus, we estimate the PMSE of $\tilde{X}_{T+h}$ by plugging an estimator of $\boldsymbol{\theta}$ into (3.1). To do this, we need the asymptotic property for an estimator of $\boldsymbol{\theta}$ as shown below with the least squares estimator of $\boldsymbol{\theta}$.

LEMMA 3.2. *Let* $\{X_t\}$ *be the same INAR*(1) *process as in Lemma* 3.1, *and* $\hat{\alpha}$ *and* $\hat{\lambda}$ *be the least squares estimators minimizing* $\sum_{t=2}^{T}(X_t - \alpha X_{t-1} - \lambda)^2$. *Then*

$$\sqrt{T}\begin{pmatrix} \hat{\alpha} - \alpha \\ \hat{\lambda} - \lambda \end{pmatrix} \xrightarrow{d} N(\mathbf{0}, \mathbf{J}^{-1}),$$

*where*

$$\mathbf{J}^{-1} = \begin{pmatrix} \frac{\alpha(1-\alpha^2)}{\lambda} + (1-\alpha)(1+\alpha) & -(1+\alpha)\lambda \\ -(1+\alpha)\lambda & \lambda + \frac{1+\alpha}{1-\alpha}\lambda^2 \end{pmatrix}.$$

From Lemma 3.1, the predictor of $X_{T+h}$ replaced $\alpha$ and $\lambda$ with the least squares estimators $\hat{\alpha}$ and $\hat{\lambda}$ is now

$$\hat{X}_{T+h} = \hat{\alpha}^h X_T + \hat{\lambda}\frac{1-\hat{\alpha}^h}{1-\hat{\alpha}}. \tag{3.2}$$

The Taylor expansion of (3.2) at $\hat{\boldsymbol{\theta}} = \boldsymbol{\theta}$ gives

$$\begin{aligned} \hat{X}_{T+h} &= \alpha^h X_T + \lambda\frac{1-\alpha^h}{1-\alpha} \\ &\quad + \Big\{h\alpha^{h-1}X_T + \lambda\frac{h\alpha^{h-1}(\alpha-1) + 1 - \alpha^h}{(1-\alpha)^2}\Big\}(\hat{\alpha}-\alpha) \\ &\quad + \frac{1-\alpha^h}{1-\alpha}(\hat{\lambda}-\lambda) + o_p(1). \end{aligned}$$

Therefore, the prediction error, $\hat{X}_{T+h} - X_{T+h}$, is

$$\begin{aligned} \hat{X}_{T+h} - X_{T+h} &= \tilde{X}_{T+h} - X_{T+h} \\ &\quad + (\hat{\alpha}-\alpha \quad \hat{\lambda}-\lambda)\begin{pmatrix} h\alpha^{h-1} & \lambda\frac{h\alpha^{h-1}(\alpha-1)+(1-\alpha^h)}{(1-\alpha)^2} \\ 0 & \frac{1-\alpha^h}{1-\alpha} \end{pmatrix}\begin{pmatrix} X_T \\ 1 \end{pmatrix} \\ &= \tilde{X}_{T+h} - X_{T+h} + (\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})'\mathbf{M}\mathbf{X}_T + o_p(1), \end{aligned} \tag{3.3}$$

where

$$\mathbf{M} = \begin{pmatrix} h\alpha^{h-1} & \lambda\frac{h\alpha^{h-1}(\alpha-1)+(1-\alpha^h)}{(1-\alpha)^2} \\ 0 & \frac{1-\alpha^h}{1-\alpha} \end{pmatrix} \quad \text{and} \quad \mathbf{X}_T = \begin{pmatrix} X_T \\ 1 \end{pmatrix}.$$

We employ the assumption that there are two independent processes: one is the process for estimating parameters involved in the Posson INAR(1) model and the other one is the process for predicting the future values of $X_t$. This assumption has used to derive the prediction mean squared error reflecting estimation error of the Gaussian AR($p$) process (Bloomfield, 1972; Bhansali, 1974; Schmidt, 1974; Yamamoto, 1976). Under this assumption, using Lemma 3.1 and Lemma 3.2, we have the following main result.

THEOREM 3.1. *Let $\{X_t\}$ be the same INAR(1) process as in Lemma 3.1. Then the asymptotic mean squared error of $\hat{X}_{T+h}$ is*

$$E(\hat{X}_{T+h} - X_{T+h})^2 = \frac{\lambda(1-\alpha^{2h})}{1-\alpha} + tr\left\{ (\mathbf{M}'\frac{1}{T}\mathbf{J}^{-1}\mathbf{M}) \begin{pmatrix} E(X_T^2) & E(X_T) \\ E(X_T) & 1 \end{pmatrix} \right\} + o(1) .$$

## 4. CONCLUSION

An important concern in the time series analysis is constructing prediction intervals to capture the future values with the nominal coverage given a realization of the past variables. We derived a h-step-ahead predictor with the minimum mean squared error in a Poisson INAR(1) process and its prediction mean squared error for the confidence limits. In practice, since we never know the parameters involved in the prediction mean squared error, we further derived the prediction mean squared error reflecting the variation of estimating the parameters in the Poisson INAR(1) process.

It is not immediate to extend the results of the Poisson INAR(1) process discussed in this paper to a general Poisson INAR($p$) process since more complex the distributional property, larger the $p$ in INAR($p$) mainly due to the binomial thinning operator again. To avoid this complication in forecasting future values while conserving the characteristics of a count data, one possible way is a bootstrap approach as a distribution free alternative. For example, one can use the bootstrap method of Alonso *et al.* (2002) by replacing the multiplication used in AR($p$) model with the thinning operator used in INAR($p$) model. Results for the bootstrap method will be reported elsewhere in due course.

## APPENDIX : MATHEMATICAL PROOFS

*Proof of Lemma 3.1.*

As shown in Al-Osh and Alzaid (1987), since

$$(X_T, X_{T+h}) \overset{d}{\equiv} (X_T,\ \alpha^h \circ X_T + \sum_{i=0}^{h-1} \alpha^i \circ \varepsilon_{T+h-i}) \tag{A.1}$$

and $\alpha \circ \varepsilon_t$ is Poisson($\alpha\lambda$) when $\varepsilon_t$ is Poisson($\lambda$) by the probability generating functions, we have

$$\sum_{i=0}^{h-1} \alpha^i \circ \varepsilon_{T+h-i} \text{ is } \mathrm{Poisson}(\sum_{i=0}^{h-1} \alpha^i \lambda). \tag{A.2}$$

Thus, (A.1) and (A.2) implies that the distribution of $X_{T+h}$ given $\mathcal{F}_T$ is a convolution of binomial distribution with parameters $(X_T, \alpha^h)$ and Poisson($\sum_{i=0}^{h-1} \alpha^i \lambda$).

Hence, we have

$$E(X_{T+h}|\mathcal{F}_T) = \alpha^h X_T + \frac{\lambda(1-\alpha^h)}{1-\alpha} = \alpha^h\Big(X_T - \frac{\lambda}{1-\alpha}\Big) + \frac{\lambda}{1-\alpha}. \tag{A.3}$$

$\{X_T - \lambda/(1-\alpha)\}$ of the first term in the right-hand side of (A.3) measures how much $X_T$ deviates from the marginal mean of the process. Furthermore, since as $h \to \infty$, the first part of (A.3) becomes zero, the $h$-step-ahead forecast approaches the marginal mean of the process.

Using (A.1) and (A.3), we have the PMSE of $\tilde{X}_{T+h} = E(X_{T+h}|\mathcal{F}_T)$ as

$$E\{X_{T+h} - E(X_{T+h}|\mathcal{F}_T)\}^2 = E\Big\{\alpha^h \circ X_T + \sum_{i=0}^{h-1}(\alpha^i \circ \varepsilon_{T+h-i}) - (\alpha^h X_T + \sum_{i=0}^{h-1} \alpha^i \lambda)\Big\}^2.$$

Thus,

$$\begin{aligned}
E\{X_{T+h} - E(X_{T+h}|\mathcal{F}_T)\}^2 &= E\Big\{(\alpha^h \circ X_T - \alpha^h X_T) + \sum_{i=0}^{h-1}(\alpha^i \circ \varepsilon_{T+h-i} - \alpha^i \lambda)\Big\}^2 \\
&= E(\alpha^h \circ X_T - \alpha^h X_T)^2 + \sum_{i=0}^{h-1} E(\alpha^i \circ \varepsilon_{T+h-i} - \alpha^i\lambda)^2 \\
&\quad +2\sum_{i=0}^{h-1} E(\alpha^h \circ X_T - \alpha^h X_T)(\alpha^i \circ \varepsilon_{T+h-i} - \alpha^i\lambda) \\
&\quad + \sum_{\substack{0 \le i,j \le h-1 \\ i \ne j}} E(\alpha^i \circ \varepsilon_{T+h-i} - \alpha^i\lambda)(\alpha^j \circ \varepsilon_{T+h-j} - \alpha^j\lambda).
\end{aligned} \tag{A.4}$$

Since $\alpha^i \circ \varepsilon_{T+h-i} \sim$ Poisson$(\alpha^i\lambda)$ and $\alpha^j \circ \varepsilon_{T+h-j} \sim$ Poisson$(\alpha^j\lambda)$ which are independent from independence of $\varepsilon_t$, we have

$$\sum_{\substack{0\le i,j\le h-1\\ i\ne j}} E(\alpha^i \circ \varepsilon_{T+h-i} - \alpha^i\lambda)(\alpha^j \circ \varepsilon_{T+h-j} - \alpha^j\lambda) = 0. \tag{A.5}$$

The independence of $X_T$ and $\varepsilon_{T+h}, h > 0$ gives

$$\sum_{i=0}^{h-1} E(\alpha^h \circ X_T - \alpha^h X_T)(\alpha^i \circ \varepsilon_{T+h-i} - \alpha^i\lambda) = 0. \tag{A.6}$$

These (A.5) and (A.6) reduce (A.4) to

$$E(\alpha^h \circ X_T - \alpha^h X_T)^2 + \sum_{i=0}^{h-1} E(\alpha^i \circ \varepsilon_{T+h-i} - \alpha^i\lambda)^2. \tag{A.7}$$

Since the first term of (A.7) is $E[\{E(\alpha^h \circ X_T - \alpha^h X_T)^2|X_T\}] = E\{\alpha^h(1-\alpha^h)X_T\} = \alpha^h(1-\alpha^h)\lambda/(1-\alpha)$, and the second term is the variance of $\alpha^i \circ \varepsilon_{T+h-i}$, which follows Poissson$(\alpha^i\lambda)$, the claim holds. □

*Proof of Lemma 3.2.*

Observe that the least squares estimator $(\hat{\alpha}, \hat{\lambda})'$ of $(\alpha, \lambda)'$ is the solution of the equation

$$\Psi(\alpha,\lambda) = \sum_{t=2}^{T} \psi_t = \sum_{t=2}^{T} \begin{pmatrix} \psi_{\alpha t} \\ \psi_{\lambda t} \end{pmatrix} = \sum_{t=2}^{T} \begin{pmatrix} (X_t - \alpha X_{t-1} - \lambda)X_{t-1} \\ (X_t - \alpha X_{t-1} - \lambda) \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}.$$

It is easy to show that $\Psi(\alpha, \lambda)$ is a martingale, and $\psi_t$ is a martingale differences. Thus, by Klimko and Nelson (1978), the least squares estimator $(\hat{\alpha}, \hat{\lambda})'$ follows asymptotic normality with mean $(\alpha, \lambda)'$ and variance $\mathbf{V}^{-1}\mathbf{W}\mathbf{V}^{-1}$, where

$$\mathbf{V} = E\begin{pmatrix} \frac{\partial}{\partial\alpha}\psi_{\alpha t} & \frac{\partial}{\partial\lambda}\psi_{\alpha t} \\ \frac{\partial}{\partial\alpha}\psi_{\lambda t} & \frac{\partial}{\partial\lambda}\psi_{\lambda t} \end{pmatrix} \quad \text{and} \quad \mathbf{W} = E(\psi_t\psi_t') = E\begin{pmatrix} \psi_{\alpha t}^2 & \psi_{\alpha t}\psi_{\lambda t} \\ \psi_{\alpha t}\psi_{\lambda t} & \psi_{\lambda t}^2 \end{pmatrix}.$$

Hence, it suffices to show that

$$\mathbf{V}^{-1}\mathbf{W}\mathbf{V}^{-1} = \begin{pmatrix} \frac{\alpha(1-\alpha^2)}{\lambda} + (1-\alpha)(1+\alpha) & -(1+\alpha)\lambda \\ -(1+\alpha)\lambda & \lambda + \frac{1+\alpha}{1-\alpha}\lambda^2 \end{pmatrix}.$$

Using the thinning property of Poisson distribution, it is immediate that $E(X_t) = \lambda/(1-\alpha)$, $E(X_t^2) = \lambda(\lambda+1-\alpha)/(1-\alpha)^2$ and $E(X_t^3) = \{\lambda^3 + \lambda(1-\alpha)(3\lambda+1-\alpha)\}/(1-\alpha)^3$. Thus, we have

$$\begin{aligned}\mathbf{V} &= E\begin{pmatrix} \frac{\partial}{\partial\alpha}\psi_{\alpha t} & \frac{\partial}{\partial\lambda}\psi_{\alpha t} \\ \frac{\partial}{\partial\alpha}\psi_{\lambda t} & \frac{\partial}{\partial\lambda}\psi_{\lambda t}\end{pmatrix} \\ &= -E\begin{pmatrix} X_{t-1}^2 & X_{t-1} \\ X_{t-1} & 1\end{pmatrix} = -\begin{pmatrix} \frac{\lambda}{1-\alpha} + (\frac{\lambda}{1-\alpha})^2 & \frac{\lambda}{1-\alpha} \\ \frac{\lambda}{1-\alpha} & 1\end{pmatrix}.\end{aligned}$$

For $\mathbf{W}$, since $E\{(X_t - \alpha X_{t-1} - \lambda)^2 | X_{t-1}\} = \mathrm{Var}(X_t|X_{t-1}) = \alpha(1-\alpha)X_{t-1} + \lambda$, and $X_t$ is stationary,

$$\begin{aligned}E(\psi_{\alpha t}^2) &= E\{(X_t - \alpha X_{t-1} - \lambda)^2 X_{t-1}^2\} \\ &= E[X_{t-1}^2 E\{(X_t - \alpha X_{t-1} - \lambda)^2 | X_{t-1}\}] \\ &= E[X_{t-1}^2\{\alpha(1-\alpha)X_{t-1} + \lambda\}] \\ &= E\{\alpha(1-\alpha)X_t^3 + \lambda X_t^2\}.\end{aligned}$$

Similarly, using the conditional expectation technique, we have

$$\begin{aligned}E(\psi_{\alpha t}\psi_{\lambda t}) &= E\{(X_t - \alpha X_{t-1} - \lambda)X_{t-1}(X_{t-1} - \alpha X_{t-1} - \lambda)\} \\ &= E\{(X_t - \alpha X_{t-1} - \lambda)^2 X_{t-1}\} \\ &= E[X_{t-1}E\{(X_t - \alpha X_{t-1} - \lambda)^2 | X_{t-1}\}] \\ &= \alpha\lambda + \frac{\lambda^2(1+\alpha)}{1-\alpha}\end{aligned}$$

and

$$\begin{aligned}E(\psi_{\lambda t}^2) &= E(X_t - \alpha X_{t-1} - \lambda)^2 = E[E\{(X_t - \alpha X_{t-1} - \lambda)^2 | X_{t-1}\}] \\ &= E\{\alpha(1-\alpha)X_{t-1} + \lambda\} = (1+\alpha)\lambda.\end{aligned}$$

This completes the proof. □

*Proof of Theorem 3.1.*

Since $\hat{X}_{T+h} - X_{T+h} = \tilde{X}_{T+h} - X_{T+h} + (\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})'\mathbf{M}\mathbf{X}_T + o_p(1)$ from (3.3), the asymptotic mean squared error of $\hat{X}_{T+h}$ is

$$\begin{aligned}&E(\hat{X}_{T+h} - X_{T+h})^2 \\ &= E(\tilde{X}_{T+h} - X_{T+h})^2 + E\{\mathbf{X}_T'\mathbf{M}'(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})'\mathbf{M}\mathbf{X}_T\} + o(1). \qquad (A.8)\end{aligned}$$

The first term of (A.8) is $E(\tilde{X}_{T+h}-X_{T+h})^2 = \lambda(1-\alpha^{2h})/(1-\alpha)$ from Lemma 3.1, and the second term is

$$\begin{aligned} E[tr\{\mathbf{X}_T'\mathbf{M}'(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta})(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta})'\mathbf{M}\mathbf{X}_T\}] &= E[tr\{\mathbf{M}'(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta})(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta})'\mathbf{M}\mathbf{X}_T\mathbf{X}_T'\}] \\ &= tr[E\{\mathbf{M}'(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta})(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta})'\mathbf{M}\}E(\mathbf{X}_T\mathbf{X}_T')] \\ &= tr\left\{(\mathbf{M}'\frac{1}{T}\mathbf{J}^{-1}\mathbf{M})\begin{pmatrix} E(X_T^2) & E(X_T) \\ E(X_T) & 1 \end{pmatrix}\right\}, \end{aligned} \tag{A.9}$$

where we used the independence assumption between the processes in estimation and prediction for the second equality, and the third equality follows from Lemma 3.2. This completes the proof. □

## REFERENCES


ALONSO, A. M., PEÑA, D. AND ROMO, J. (2002). "Forecasting time series with sieve bootstrap", *Journal of Statistical Planning and Inference*, **100**, 1–11.

AL-OSH, M. A. AND ALZAID, A. (1987). "First-order integer-valued autoregressive (INAR(1)) process", *Journal of Time Series Analysis*, **8**, 261–275.

ALZAID, A. A. AND AL-OSH, M. (1990). "An integer-valued $p$th-order autoregressive structure (INAR($p$)) process", *Journal of Applied Probability*, **27**, 314–324.

BHANSALI, R. J. (1974). "Asymptotic mean-square error of predicting more than one-step ahead using the regression method", *Journal of Royal Statistical society*, Ser. C, **23**, 35–42.

BLOOMFIELD, P. (1972). "On the error of prediction of a time series", *Biometrika*, **59**, 501–507.

COX, D. R. (1981). "Statistical analysis of time series : some recent developments (with discussion)", *Scandinavian Journal of Statistics. Theory and Applications*, **8**, 93–115.

DU, J.GUAN AND LI, Y. (1991). "The integer-valued autoregressive (INAR($p$)) model", *Journal of Time Series Analysis*, **12**, 129–142.

FREELAND, R. K. AND MCCABE, B. P. M. (2004). "Forecasting discrete valued low count time series", *International Journal of Forecasting*, **20**, 427–434.

JUNG, R. C. AND TREMAYNE, A. R. (2006). "Coherent forecasting in integer time series models", *International Journal of Forecasting*, to appear.

KLIMKO, L. A. AND NELSON, P. I. (1978). "On conditional least squares estimation for stochastic processes", *The Annals of Statistics*, **6**, 629–642.

LATOUR, A. (1998). "Existence and stochastic structure of non-negative inter-valued autoregressive process", *Journal of Time Series Analysis*, **19**, 439–455.

MCCABE, B. P. M. AND MARTIN, G. M. (2005). "Bayesian predictions of low count time series", *International Journal of Forecasting*, **21**, 315–330.

MCKENZIE, E. (1985). "Some simple models for discrete variate time series", *Water Resources Bulletin*, **21**, 645–650.

MCKENZIE, E. (1988). "Some ARMA models for dependent sequences of Poisson counts", *Advances in Applied Probability*, **20**, 822–835.

PARK, Y. S. AND KIM, H. Y. (2000). "On the autocovariance function of INAR(1) process with a negative binomial or a Poisson marginal", *Journal of the Korean Statistical Society*, **29**, 269–284.

PARK, Y. S., CHOI, J. W. AND KIM, H. Y. (2006). "Forecasting Cause-Age specific mortality using two random processes", *Journal of American Statistical Association*, to appear.

SCHMIDT, P. (1974). "The asymptotic distribution of forecasts in the dynamic simulation of an econometric model", *Econometrika, Journal of Econometric Society*, **42**, 303–309.

SILVA, M. E. AND OLIVEIRA, V. L. (2005). "Difference equations for the higher order moments and cumulants of the INAR($p$) model", *Journal of Time Series Analysis*, **26**, 17–36.

STEUTEL, F. W. AND VAN HARN, K. (1979). "Discrete analogues of self-decomposability and stability", *The Annals of Probability*, **7**, 893–899.

YAMAMOTO, T. (1976). "Asymptotic mean square prediction error for an autoregressive model with estimated coefficients", *Journal of the Royal Statistical Society*, Ser. C, **25**, 123–127.