

# Bootstrap Confidence Intervals for the INAR( $p$ ) Process

Hee-Young Kim<sup>1)</sup> and Yousung Park<sup>2)</sup>

## Abstract

The distributional properties of forecasts in an integer-valued time series model have not been discovered yet mainly because of the complexity arising from the binomial thinning operator. We propose two bootstrap methods to obtain nonparametric prediction intervals for an integer-valued autoregressive model : one accommodates the variation of estimating parameters and the other does not. Contrary to the results of the continuous ARMA model, we show that the latter is better than the former in forecasting the future values of the integer-valued autoregressive model.

*Keywords* : Stationary process; Integer valued time series; Prediction interval; Sieve Bootstrap.

## 1. 서론

자료가 이산(discrete) 시계열자료인 경우, 오차항의 정규성을 가정하는 ARMA모형으로 모형화 하는 것은 자료의 분포를 충분히 반영하지 않는 문제점을 안고 있다. 예를 들어, 공장에서 근로자의 월별사고 발생건수, 특이한 질병에 대한 환자의 월별 발생건수, 관공서등에서 기다리는 민원인의 숫자 등과 같이 실제로 많은 경우에 연속형 자료가 아닌 이산형 자료를 모형화 할 필요가 있다. 실제 현상에서 관찰값들이 정규성을 만족하지 않을 때 정규이론에 근거한 추정법들은 이론적인 오차보다 훨씬 많은 오류를 내포할 수 있다.

시계열 모형을 구분하는 기준은 여러 가지 있으나 Cox (1981)는 "관측치에 의한 모형 (observation-driven model)"과 "모수에 의한 모형 (parameter-driven model)"으로 분류하였다. 전자는 과거 관측값을 예측변수 (predictor variable)로 취함으로써 시계열자료의 자기상관을 설명하는 것이다. 즉, 과거의 관측값  $y_1, \dots, y_{t-1}$ 과 공변량들 (covariate)이 주어졌을 때, 이들에 대한 조건부 적률 (moment) 혹은 조건부분포를 모형화 하는 것이다. 후자는 자기상관을 갖는 잠재변수 (latent variable)  $\epsilon_t$ 를 모형에 도입하여 시계열자료의 상관을 모형화 하는 것이다. 본 연구에서는 관측치에 의한 모

---

1) Research Assistant Professor, Institute of Statistics, Korea University, Seoul 136-701, Korea. Correspondence : starkim@korea.ac.kr

2) Professor, Department of Statistics, Korea University, Seoul 136-701, Korea.

형 (observation-driven model)의 하나인 McKenzie (1985)와 Al-Osh와 Alzaid (1987)에 의해 처음 정의된 차수가  $p$ 인 정수값을 갖는 자기회귀 (integer-valued autoregressive, INAR( $p$ )) 모형에 초점을 두고자 한다. INAR모형은 이항디닝 (binomial thinning) 연산자를 이용함으로써 비음의 정수값을 가지는 시계열자료의 모형화가 가능하게 했을 뿐만 아니라, 연속형 AR( $p$ ) 모형과 동일한 자기상관함수를 가진다.

Steutel과 Van Harn (1979)이 처음 정의한 이항디닝 연산자를 모형에 이용하면 포아송분포 (Poisson), 복합포아송분포 (compound Poisson), 음이항분포 (negative binomial)등을 주변분포로 갖는 이산형 시계열모형이 가능하다. 특히, INAR(1) 모형은 고전적인 확률과정인 유입 (immigration)이 있는 분기과정 (branching process)과 대응되며, 또한, 오차항이 포아송분포인 경우에 INAR(1)모형은 대기행렬과정 (queueing process)의 하나인  $M/M/\infty$ 과 일치한다 (McKenzie, 1988). 그러나, INAR 모형에 관한 연구는 에르고딕성 (ergodicity), 고차적률 (higher-order moment), 누울 (cumulant), 존재성 (existence) 등과 같은 확률론적 관점이 주로 논의 대상이었다.

시계열 분석에 있어서 중요한 것 중 하나는 과거의 관찰값으로부터 미래를 예측하는 것이다. 그러나, 비음의 정수값을 모형화 하는 INAR( $p$ ) 모형의 미래시점의 예측에 관한 연구결과들은 많지 않다. Freeland와 McCabe (2004)는 주변분포가 포아송분포인 INAR(1)모형에서  $h$ 시점앞 ( $h$ -step ahead)의 확률질량함수 (probability mass function)를 계산하고, 모수의 최대가능도추정량 (maximum likelihood estimator)과 델타방법 (delta method)을 사용하여  $h$ 시점앞의 확률질량함수의 신뢰구간을 구하였다. 또한, McCabe와 Martin (2005)은 베이지안방법, Jung과 Tremayne (2006)은 붓스트랩 방법으로 INAR(1)과 INAR(2) 모형의  $h$ 시점앞의 확률질량함수의 신뢰구간을 연구하였다. 그러나, 이 방법들은 예측치의 분산을 구하지 않고 예측치의 확률질량함수에 대한 신뢰구간을 구하게 되어 예측치의 값이 0, 1 또는 2와 같이 매우 작은 정수값일 경우이외에는 사용하기 어렵다. 그리하여, Kim과 Park (2006)은 주변분포가 포아송인 INAR(1) 모형에서  $h$ -시점앞의 예측평균제곱오차 (mean square error)를 모수의 변동성을 반영하여 구하였지만, 이 결과를 INAR( $p$ ) 모형에서 수리적으로 확장하기는 이항디닝 연산자의 성질에 의해 쉽지 않다.

한편, 자료가 연속형 시계열인 경우에는 예측치의 신뢰구간에 대한 연구가 상당히 많다. 특히, 분포무관 (distribution free)한 붓스트랩 신뢰구간의 연구는 자료의 정규성을 가정하지 않는다는 장점이 있다. 기존의 연구결과들 중 AR( $p$ ) 모형에서 대표적인 것만 살펴보면, Stine (1987), Grigoletto (1998)은 모수추정치 변동성을 반영하는 예측평균제곱오차 (prediction mean error)의 추정을 붓스트랩 방법으로 구하였으며, Thombs와 Schucany (1990)은 자료가 시간역류성 (time reversible)을 만족할 때 예측치의 분포를 붓스트랩 방법으로 추정하였다. 또한, Garcia-Jurado 등 (1995)은 Thombs와 Schucany (1990)의 방법을 ARI 모형에서 확장하였다. Alonso 등 (2002)은 예측치의 분포를 모수추정의 변동성을 고려하는 붓스트랩 방법을 제시하였고, 특히 이 방법은 주어진 시계열에 시간역류성 (time reversible)을 요구하지 않는 장점이 있다. 자료가 시간역류성을 만족할 필요가 없으므로 보다 광범위한 시계열 자료에 적용할 수 있으나, 후방예측(back forecasting)을 할 수 없다는 단점이 있다. 일변량 시계

열 자료의 시간역류성에 관한 연구는 Findley (1986), Hallin 등 (1988), Cheng (1999) 등을 참고할 수 있다. 최근 Pascual 등 (2004)는 Alonso 등 (2002)의 방법을 ARIMA 모형에서 확장하여 제시하였다.

본 연구에서는 INAR( $p$ ) 모형에서 미래를 예측하기 위한 방법으로 수리적인 접근 대신 모수추정량의 변동성을 고려한 붓스트랩 예측구간을 제안하고자 한다. 논문의 구성은 다음과 같다. 2절에서는 간략히 INAR( $p$ ) 모형을 소개하고, 3절에서는 붓스트랩 예측구간을 두 가지 방법으로 제시한다. 4절에서는 오차항의 분포가 포아송분포, 음이항분포를 따를 때 3절에서 제시한 두 가지 붓스트랩 방법을 INAR(1), INAR(2) 모형에서 모의실험을 통하여 비교한다. 4절에서는 실제 자료분석을 하며, 마지막으로 5절은 향후 연구과제와 결론을 포함한다.

## 2. INAR( $p$ ) 모형의 소개

INAR모형을 정의하기 전에 Steutel과 Van Harn (1979)에 의해 처음 정의된 이항디닝 연산자 ‘ $\circ$ ’를 소개한다. 비음의 정수값을 갖는 확률변수  $X$ 와 임의의  $\alpha \in [0,1]$ 에 대하여 ‘ $\circ$ ’는 다음과 같이 정의된다.

$$\alpha \circ X = \sum_{i=1}^X w_i, \tag{2.1}$$

단,  $w_i$ 는 성공확률이  $\alpha$ 인 *i.i.d* 베르누이 확률변수이고,  $X$ 와는 서로 독립이다.

식 (2.1)에 정의된 이항디닝 연산자의 기본성질은  $E(\alpha \circ X|X) = \alpha X$ ,  $E(\alpha \circ X) = \alpha E(X)$ ,  $Var(\alpha \circ X|X) = \alpha(1-\alpha)X$ ,  $Var(\alpha \circ X) = \alpha^2 Var(X) + \alpha(1-\alpha)E(X)$ 임을 쉽게 알 수 있다. Al-Osh와 Alzaid (1987)는 이항디닝 연산자를 이용하여 다음의 INAR(1) 모형을 정의하였다.

$$X_t = \alpha \circ X_{t-1} + \epsilon_t, \tag{2.2}$$

여기서 오차항  $\epsilon_t$ 은 *i.i.d*한 비음의 정수값을 가지며 평균이  $\mu_\epsilon$ , 분산은  $\sigma_\epsilon^2$ 이고,  $X_{t-1}$ 과  $\epsilon_t$ 는 서로 독립이다. 식 (2.2)의 INAR(1) 모형은 고전적 확률모형인 아래 식 (2.3)과 같은 유입이 있는 분기과정 (branching process with immigration)의 특별한 경우이다.

$$X_t = \sum_{i=1}^{X_{t-1}} Y_{t,i} + I_t, \tag{2.3}$$

이때  $\{Y_{t,i}\}$ 와  $\{I_t\}$ 는 서로 독립이며 비음의 정수값을 갖는 확률변수이다.

식 (2.2)의 INAR(1) 모형은  $\alpha \in (0,1)$  경우에 약정상과정 (weakly stationary process)이며  $E(X_t|X_{t-1}) = \alpha X_{t-1} + \mu_\epsilon$ ,  $E(X_t) = \mu_\epsilon / (1-\alpha)$ ,  $Var(X_t|X_{t-1}) = \alpha(1-\alpha) \times X_{t-1}$ ,  $Var(X_t) = (\alpha\mu_\epsilon + \sigma_\epsilon^2) / (1-\alpha^2)$ 이다. 또한, 자기상관함수 (autocorrelation function)는  $\rho_X(k) = \alpha^k$ 으로 연속형 AR(1)과 같다. 그리고, Steutel과 Van Harn (1979)이 밝힌 바와 같이 이항자기분기 (discrete self-decomposable) 분포는 식 (2.2)의 주변분

포가 될 수 있다. 이러한 이항자기분기 (discrete self-decomposable) 분포로는 포아송, 기하분포, 음이항분포가 포함되며, 이항분포는 제외된다 (McKenzie, 1985; Al-Osh와 Alzaid, 1987; Park과 Kim, 2000).

INAR(1) 모형의 자연스러운 확장으로 Du와 Li (1991)는 아래와 같이 INAR( $p$ ) 모형을 정의하였다.

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

여기서 오차항  $\{\epsilon_t\}$ 의 가정은 INAR(1) 모형과 같고,  $\alpha_i \circ X_{t-i} = \sum_{j=1}^{X_{t-i}} w_{ij}$ 에 포함된 계수 과정 (counting process)인  $\{w_{ij}\}$ 는 상호독립(mutually independent)이고,  $\alpha_i \in [0,1] (i=1, \dots, p)$ 이다. 또한,  $\{\epsilon_t\}$ 는 계수과정  $\{w_{ij}\}$ 과 서로 독립이다. Du와 Li (1991)와는 다른 방식으로 Alzaid와 Al-Osh (1990)는 INAR( $p$ ) 모형을 정의하였다. 즉, Alzaid와 Al-Osh (1990)는  $X_t = x$ 로 주어졌을 때 벡터  $(\alpha_1 \circ X_t, \dots, \alpha_p \circ X_t)$ 는 모수가  $(\alpha_1, \dots, \alpha_p, x)$ 인 다항분포를 따르며, 과거 시점의 확률과정  $\{X_t\}$ 과는 서로 독립이다. 즉,  $X_t = x$  일 때  $\alpha_i \circ X_t$ 는  $X_{t-k}$ 와  $\alpha_j \circ X_{t-k} (i, j=1, \dots, p; k > 0)$ 와는 서로 독립이다. Alzaid와 Al-Osh (1990)과 Du와 Li (1991)의 INAR( $p$ )모형이 약정상 확률과정이 되기 위한 조건은 특성함수  $\phi(z) = 1 - \phi_1 z - \cdots - \phi_p z^p$ 의 모든 근이 단위 원 밖에 있어야 함으로 동일하다.

그러나, 두 INAR( $p$ ) 모형의 차이점은 Alzaid와 Al-Osh (1990)의 모형은 연속형 ARMA( $p, p-1$ )과 같은 자기상관함수 (autocorrelation function)을 가지며, Du와 Li (1991) 모형은 연속형 AR( $p$ )와 동일한 자기상관함수를 가진다. 또한 Alzaid와 Al-Osh (1990)의 모형의 조건부 기대치는 과거자료의 비선형함수이나 Du와 Li의 모형은 과거자료의 선형함수이다. 그러므로, Alzaid와 Al-Osh의 INAR( $p$ ) 모형은 선형 회귀에 의해 접근이 안되므로 모수의 추정과 검정 등 통계적인 측면에 관한 연구는 쉽지 않다. 따라서, Silva와 Oliveira (2005), Latour (1998), Park 등 (2006) 등의 많은 연구들이 Du와 Li (1991)의 INAR( $p$ ) 모형으로 후속연구를 진행하였다.

### 3. INAR( $p$ ) 모형의 붓스트랩 예측구간

시계열자료  $\{Y_t\}$ 가 시점  $T$ 까지  $\mathbf{Y} = (Y_1, \dots, Y_T)'$ 로 주어졌을 때  $T+h$ 시점에서  $Y_{T+h}$ 의  $100 \times (1-\alpha)\%$  신뢰구간  $(L(\mathbf{Y}), U(\mathbf{Y}))$ 의 구축은 다음을 의미한다.

$$E_{\mathbf{Y}} \int_{L(\mathbf{Y})}^{U(\mathbf{Y})} dF(Y_{T+h} | \mathbf{Y}) = 1 - \alpha. \quad (3.1)$$

즉, 자료가  $T$ 시점까지 주어졌을 때  $T+h$ 시점의 조건부 분포를 유도해야 한다.

식 (2.2)의 INAR(1)모형에서는 이항디닝 연산자의 정의에 의하면  $Y_T = y_T$  로 주어졌을 때  $Y_{T+1}$ 의 조건부 분포는 아래와 같다.

$$P(Y_{T+1} = y_{T+1} | Y_T = y_T) = \sum_{i=0}^{\min(y_T, y_{T+1})} P[B_{y_T}^\alpha = i] P[\epsilon_{T+1} = y_{T+1} - i], \quad (3.2)$$

여기서  $B_n^\alpha$ 는 이항분포  $B(n, \alpha)$ 을 나타낸다. 즉, 식 (3.2)는 이항분포  $B(y_T, \alpha)$ 와  $\epsilon_{T+1}$ 의 합성곱(convolution)이 된다. 유사하게  $Y_T = y_T$ 가 주어졌을 때,  $Y_{T+h}$ 의 조건부 분포는 이항분포  $B(y_T, \alpha^p)$ 와  $\sum_{j=0}^{h-1} \alpha^j \circ \epsilon_{T+h-j}$ 의 합성곱 (convolution)으로

$$P(Y_{T+h} = y_{T+h} | Y_T = y_T) = \sum_{i=0}^{\min(y_T, y_{T+h})} B[y_T^\alpha = i] P[\sum_{j=0}^{h-1} \alpha^j \circ \epsilon_{T+h-j} = y_{T+h} - i] \quad (3.3)$$

이 된다. 식 (3.3)에서 우변의  $P[\sum_{j=0}^{h-1} \alpha^j \circ \epsilon_{T+h-j} = y_{T+h} - i]$ 항은 오차항  $\{\epsilon_t\}$ 이 독립이므로 다음과 같은 합성곱으로 표현된다.

$$\{P[\alpha^{h-1} \circ \epsilon_{T+1} = y_{T+h} - i]\} * \dots * \{P[\alpha^0 \circ \epsilon_{T+h} = y_{T+h} - i]\}, \quad (3.4)$$

여기서  $\{c_n\} = \{a_n\} * \{b_n\} = \sum_{i=0}^n a_i b_{n-i}$  임을 의미한다. 이러한 복잡한 합성곱에 대한 분포를 유도하는 것은 대부분의 경우 불가능할 뿐만 아니라 설령 가능하다 할지라도 (즉  $\{\epsilon_t\}$ 의 분포가 포아송분포인 경우), 그 분포가 매우 복잡하여  $Y_{T+h}$ 의 조건부 분포를 수리적으로 유도하는 것은 대부분 가능하지 않다.

최근에, Kim과 Park (2006)은 주변분포가 포아송분포를 따르는 INAR(1) 모형에서 예측오차  $\tilde{e}_{T+h} \equiv Y_{T+h} - \tilde{Y}_{T+h}$  (여기서  $\tilde{Y}_{T+h} = E(Y_{T+h} | Y_1, \dots, Y_T)$ )의 예측평균제곱오차 (prediction mean square error)와  $\hat{e}_{T+h} \equiv Y_{T+h} - \hat{Y}_{T+h}$  (여기서  $\hat{Y}_{T+h}$ 는  $\tilde{Y}_{T+h}$ 에서 모수대신 최소제곱추정량 대입)의 평균제곱오차를 유도하였다. 이 결과를 INAR( $p$ ) 모형으로 일반화 하는 것 역시 식 (3.4)의 합성곱의 분포에 의하여 매우 어려운 작업으로 보인다.

따라서, 본 절에서는 INAR( $p$ ) 모형의 붓스트랩 예측구간을 구축하는 방법을 제안하고자 한다. 2절에서 언급한바와 같이 Du와 Li (1991)의 INAR( $p$ ) 모형은 연속형 AR( $p$ )와 같은 자기상관함수를 가진다. 이를 이용하여 본 논문에서는 연속형 AR( $p$ ) 모형에서 붓스트랩 예측구간의 기존의 연구결과들 중에서 적절한 방법을 선택하여 INAR( $p$ ) 모형으로 응용하고자 한다. 1절 서론에서 언급한 붓스트랩 방법들 중 Alonso 등 (2002)이 연속형 AR( $p$ )에서 제안한 붓스트랩 예측방법을 INAR( $p$ )모형에

확장하여 응용하고자 한다. 이 방법은 연속형 AR( $p$ ) 모형에서는 시간역류성을 만족할 필요가 없으며, 모수추정의 변동성을 예측구간에 반영할 수 있는 장점이 있다.

[단계1] 다음과 같은 오차제곱합  $Q$ 을 최소화하는 추정량을  $\hat{\mu}_\epsilon, \hat{\alpha}_1, \dots, \hat{\alpha}_p$ 라 하자.

$$Q(\mu_\epsilon, \alpha_1, \dots, \alpha_p) = \sum_{t=p+1}^T (X_t - E(X_t | X_{t-1}, \dots))^2 = \sum_{t=p+1}^T (X_t - \alpha_1 X_{t-1} - \dots - \alpha_p X_{t-p} - \mu_\epsilon)^2.$$

[단계2] 잔차  $\hat{\epsilon}_t = X_t - (\hat{\alpha}_1 X_{t-1} + \dots + \hat{\alpha}_p X_{t-p})$ ,  $t = p+1, \dots, T$  를 생성한다.

[단계3] [단계2]의 잔차  $\hat{\epsilon}_t$  는 음수 혹은 분수가 될 수 있으므로 아래와 같은 수정된 잔차  $\tilde{\epsilon}_t$ 를 정의한다.

$$\tilde{\epsilon}_t = \begin{cases} [\hat{\epsilon}_t] : \hat{\epsilon}_t > 0 \\ 0 : \hat{\epsilon}_t \leq 0, \end{cases}$$

여기서  $[\cdot]$  함수는 가우스함수이다. 수정된 잔차의 경험적 분포를  $\hat{F}_\epsilon(x) =$

$$(T-p)^{-1} \sum_{t=p+1}^T 1_{\{\tilde{\epsilon}_t \leq x\}}$$
 정의하자.

[단계4] [단계3]의  $\hat{F}_\epsilon(\cdot)$ 에서 *i.i.d*한  $\epsilon_t^*$ 를 생성한다.

[단계5] 다음과 같이 붓스트랩 표본을 생성한다.

$$X_t^* = \sum_{i=1}^p \hat{\alpha}_i \circ X_{t-i}^* + \epsilon_t^*, \quad t = 1, 2, \dots, T$$

[단계6] 생성된 붓스트랩 표본  $\{X_1^*, \dots, X_T^*\}$ 을 이용하여 INAR( $p$ )모형의 모수  $\alpha_1, \dots, \alpha_p$ 를 [단계2]와 같이 재추정하고 추정량을  $(\hat{\alpha}_1^*, \dots, \hat{\alpha}_p^*)$ 로 표시하자.

[단계7]  $T+h$ 시점의 붓스트랩 표본은 다음과 같이 생성한다.

$$X_{T+h}^* = \sum_{i=1}^p \hat{\alpha}_i^* \circ X_{T+h-i}^* + \epsilon_t^*,$$

여기서  $h > 0$  이고  $t \leq T$ 인 경우  $X_t^* = X_t$  이다.

위 모의 실험 과정 중 몇가지 주의할 사항은 다음과 같다. [단계2]에서 생성한 잔차는  $X_t$ 와  $E(X_t | X_{t-1}, \dots)$ 의 차가 아니다. 또한, [단계7]에서는 예를 들어 미래시점  $T+1$ 의 붓스트랩 표본  $X_{T+1}^*$ 을 얻기 위하여 주어진 관측자료  $x_T, x_{T-1}, \dots, x_{T+1-p}$ 를 고정하였으나, [단계5]에서는 붓스트랩 표본  $X_t^*$ 는 붓스트랩 표본  $X_{t-1}^*, \dots, X_{t-p}^*$

을 이용하여 축차적으로 생성한다.

$X_{T+h}^*$ 의 붓스트랩분포를  $F_{X_{T+h}^*}^*(x)$ 라 할 때 [단계4]-[단계7]을  $B$ 번 반복하여  $\hat{F}_{X_{T+h}^*}^*(x) = \#\{X_{T+h}^{*,b} \leq x\} / B, b \in \{1, \dots, B\}$ 를 얻고, 이로부터  $X_{T+h}$ 의  $100 \times (1 - \alpha)\%$  붓스트랩 예측구간은

$$[Q^*(\alpha/2), Q^*(1 - \alpha/2)] \tag{3.5}$$

을 구한다. 단,  $Q^*(\cdot)$ 은  $\hat{F}_{X_{T+h}^*}^*$ 의 분위수이다.

본 절에서 제안한 INAR(p) 모형의 예측구간의 붓스트랩방법과 Alonso 등 (2002)의 AR(p) 모형에서 붓스트랩 예측구간의 차이는 [단계2]와 [단계3]으로, 잔차  $\hat{\epsilon}_t$ 를 중심화 (centering)하거나 재척도화 (rescaling)할 필요가 없다. 그 이유는, INAR(p) 모형은 평균이  $\mu_c (\neq 0)$ 인 오차항  $\epsilon_t$ 을 가지기 때문이다. 또한, 잔차  $\hat{\epsilon}_t$ 는 분수, 음수일 수 있으므로 [단계3]에서 수정된 잔차  $\tilde{\epsilon}_t$ 를 고려하였다. 위의 절차에서 만약 [단계6]이 생략되면 Alonso 등 (2002)의 논문에서와 같이 CS (conditional sieve bootstrap) 방법으로, [단계6]이 포함되면 VS (sieve bootstrap that includes the variability due to parameter estimation) 방법으로 지칭하자. CS방법과 VS방법의 차이는 [단계6]으로 모수추정의 변동성을 예측구간에 반영하는지 여부이다.

#### 4. 모의실험

본 절에서는 앞 절에서 언급한 2가지 방법의 특징을 모의실험을 통하여 비교하고자 한다. 고려한 모형은 다음과 같다.

$$\text{<모형1> } X_t = \alpha_1 \circ X_{t-1} + \epsilon_t.$$

$$\text{<모형2> } X_t = \alpha_1 \circ X_{t-1} + \alpha_2 \circ X_{t-2} + \epsilon_t.$$

오차항의 분포로는 평균이 3, 5, 10인 포아송, 음이항분포를 고려하였다. 일반적으로 이산형 확률과정에서 포아송분포는 평균과 분산이 같은 경우를 음이항분포는 분산이 평균보다 큰 과대산포분포 (overdispersed distribution)를 나타낸다. 본 논문에서 고려한 음이항 분포는 평균이 3, 5, 10이며, 이때 분산은 각각 12, 8, 43이다.

2절에서 언급한 바와 같이 INAR(p) 모형이 약정상성이 되기 위한 조건은 연속형 AR(p)와 같으므로, <모형1>에서  $\alpha_1 = 0.3, 0.5, 0.7$ 으로, <모형2>에서는  $\alpha_1, \alpha_2$ 를 다항

식  $1 - \sum_{i=1}^2 \alpha_i z^i = 0$ 의 근의 절대값이 1보다 크도록 선택하였다. 사용한 표본수는  $T = 25, 50$ 이고, 선행시차 (lead time)  $h$ 는 1부터 5까지이고,  $1 - \alpha$ 는 0.95로 정하였다. CS방법과 VS방법을 비교하기 위하여 다음의 순서로 모의실험을 진행하였다. 계산 프로그램은 SAS/IML을 이용하였다.

[단계1] 고정된 모형과 모수, 표본의 수, 오차항의 분포에서  $x_{-100}, \dots, x_0, x_1, \dots, x_T$ 을 생성한다. 생성된 자료중  $x_1, \dots, x_T$ 을 이용하여  $T+h$ 시점의  $X_{T+h}$  값  $R = 1000$  개를 생성한다.

[단계2] [단계1]에서 만들어진 자료  $x_1, \dots, x_T$ 을 이용하여 3절의 붓스트랩 방법을  $B = 1000$ 번 반복하여  $T+h$ 시점의  $X_{T+h}$ 의  $100 \times (1 - \alpha)\%$  붓스트랩 예측구간을 식 (3.5)의  $[Q_M^*(\alpha/2), Q_M^*(1 - \alpha/2)]$ 로 구한다.

[단계3] 붓스트랩 방법의 포함비율 (coverage)은 아래와 같다.

$$C_M = \frac{\#\{Q_M^*(\alpha/2) \leq X_{T+h}^r \leq Q_M^*(1 - \alpha/2)\}}{R},$$

여기서  $X_{T+h}^r (r = 1, \dots, R)$ 은 [단계1]에서 생성한  $T+h$ 시점의 생성된 자료이다.

본 논문의 모의실험은 선행시차  $h$ 를 최대 5를 고려하였고, [단계1]에서 생성된  $x_1, \dots, x_T$ 를 이용하여  $(X_{T+1}, X_{T+2}, \dots, X_{T+5})$ 를  $R = 1000$ 개 생성하였다. 단계1로부터  $X_{T+h}$ 의 경험적인 분포 (empirical distribution)의 구간길이  $L_e = X_{T+h}^{[R(1 - \frac{\alpha}{2})]} - X_{T+h}^{[R(\frac{\alpha}{2})]}$ 를 얻고, [단계2]에서는 붓스트랩 구간길이  $L_M = Q_M^*(1 - \alpha/2) - Q_M^*(\alpha/2)$ 를 구한다.

앞의 [단계1]-[단계3]을  $S = 200$  번 반복하여  $C_{M,i}, L_{M,i}, i = 1, \dots, S$ 를 측정 한 후에 CS방법과 VS방법을 비교하는 요약통계량으로 아래의 통계량들을 이용한다 (Alonso 등, 2002).

$$\begin{aligned} \bar{C}_M &= \sum_{i=1}^S C_{M,i} / S \\ SE(\bar{C}_M) &= \left[ \sum_{i=1}^S (C_{M,i} - \bar{C}_M)^2 / S(S-1) \right]^{1/2} \\ \bar{L}_M &= \sum_{i=1}^S L_{M,i} / S \end{aligned} \tag{4.1}$$

$$SE(\bar{L}_M) = \left[ \sum_{i=1}^S (L_{M,i} - \bar{L}_M)^2 / S(S-1) \right]^{1/2}$$

$$CQ_M = |1 - \bar{C}_M / \bar{C}_T| + |1 - \bar{L}_M / \bar{L}_T| ,$$

즉,  $\bar{C}_M$ 과  $SE(\bar{C}_M)$ 는  $S$ 개의 붓스트랩 포함비율의 평균과 표준편차이고,  $\bar{L}_M$  과  $SE(\bar{L}_M)$ 는  $S$ 개의 붓스트랩 구간길이의 평균과 표준편차를 나타낸다. 식 (4.1)에서



$\bar{L}_T = \sum_{i=1}^S L_{e,i} / S$  으로 추정된 참평균길이 (true mean interval length),  $\bar{C}_T = 1 - \alpha$  으로 명목포함비율 (nominal coverage) 이다. Alonso 등 (2002)이 언급한바와 같이 통계량  $CQ_M$ 는 붓스트랩 표본의 평균포함비율 (mean coverage)와 평균길이 (mean length)을 모두 고려하는 지표로  $CQ_M$ 이 작은 값을 갖는 붓스트랩 방법이 보다 우수하다.

<표 1>-<표 4>는 다양한 모의실험 결과 중 일부만을 정리한 표이며, se는 식 (4.1)에서의 표준편차를 나타낸다.

<표 1>을 보면 CS, VS방법 모두 평균포함비율 ( $\bar{C}_M$ )이 명목포함비율인 0.95보다 크므로 상향추정을 하고 있으나, CS방법이 명목포함비율에 보다 근접함을 알 수 있다. 이와는 다르게 Alonso 등 (2002)의 연속형 AR(p)의 붓스트랩 방법의 예측구간의 모의 실험에서는 CS, VS 두 방법 모두 하향추정하는 결과를 보여 주었다. 또한, <표 1>에서 CS, VS방법 모두 붓스트랩 구간길이 평균 ( $\bar{L}_M$ )은 추정된 참평균길이 ( $\bar{L}_T$ )보다 크지만, CS방법의  $\bar{L}_M$ 이  $\bar{L}_T$ 에 보다 가까움을 나타낸다. 따라서, CS방법의  $CQ_M$ 이 표본수 (T), 선행시차 (lead)에 관계없이 VS방법의  $CQ_M$ 보다 작게 된다.

<표 1> 모형1의 모의실험결과- $\alpha_1 = 0.3$ , 오차항은 평균이 10인 포아송분포

lead	sample size	method	$\bar{C}_M (se)$	$L_T$	$\bar{L}_M (se)$	$CQ_M$
<i>h</i>	<i>T</i>					0
1	25	CS	0.965 (0.00236)	13.97	17.59 (0.1734)	0.2751
		VS	0.980 (0.00140)		19.48 (0.1670)	0.4257
	50	CS	0.974 (0.00148)	14.00	16.49 (0.1226)	0.2027
		VS	0.981 (0.00098)		17.57 (0.1374)	0.2868
<i>h</i>	<i>T</i>					0
3	25	CS	0.962 (0.00272)	14.70	18.38 (0.1521)	0.2626
		VS	0.989 (0.00091)		21.70 (0.1453)	0.5175
	50	CS	0.970 (0.00157)	14.73	17.06 (0.1226)	0.1789
		VS	0.986 (0.00079)		19.24 (0.1274)	0.3450
<i>h</i>	<i>T</i>					0
5	25	CS	0.961 (0.00294)	14.77	18.46 (0.1447)	0.2619
		VS	0.993 (0.00063)		22.43 (0.1518)	0.5640
	50	CS	0.970 (0.00156)	14.81	17.07 (0.1272)	0.1732
		VS	0.987 (0.00080)		19.33 (0.1349)	0.3440

<표 2>는 오차항이 과대산포분포인 음이항분포를 따르는 경우로, <표 1>과 몇가지 다른 점을 파악할 수 있다. 첫째, VS방법은 여전히  $\bar{C}_M$ 이 명목포함비율보다 크지만, CS방법의  $\bar{C}_M$ 은 그런 경향은 보이지 않고 <표 1>과 비교하여 명목포함비율인 0.95에 더 가까움을 보여준다. 둘째로, <표 1>과 비교하여 CS, VS방법의  $CQ_M$ 이 모두

상당히 작아졌음을 나타낸다. 오차항이 음이항분포를 따르는 경우에도  $CQ_M$ 지표의 측면으로 보면 CS방법이 VS방법에 비하여 우수함을 알 수 있다.

<표 2> 모형1의 모의실험결과- $\alpha_1 = 0.3$ , 오차항은 평균이 10인 음이항분포

lead	sample size	method	$\bar{C}_M (se)$	$L_T$	$\bar{L}_M (se)$	$CQ_M$
$h$	$T$					
1	25	CS	0.946 (0.00367)	26.33	27.38 (0.4656)	0.0441
		VS	0.956 (0.00281)		28.62 (0.4451)	0.0936
	50	CS	0.957 (0.00190)	26.27	26.90 (0.2930)	0.0313
		VS	0.965 (0.00154)		28.14 (0.3054)	0.0875
3	25	CS	0.948 (0.00348)	27.39	28.57 (0.4534)	0.0451
		VS	0.960 (0.00270)		30.12 (0.4112)	0.1104
	50	CS	0.958 (0.00195)	27.39	28.37 (0.3666)	0.0442
		VS	0.967 (0.00168)		29.77 (0.3208)	0.1051
5	25	CS	0.956 (0.00350)	27.48	28.50 (0.4570)	0.0403
		VS	0.962 (0.00262)		30.53 (0.4179)	0.1233
	50	CS	0.959 (0.00194)	27.43	28.28 (0.3185)	0.0405
		VS	0.968 (0.00162)		30.15 (0.3375)	0.1186

<표 3> 모형2의 모의실험결과- $\alpha_1 = 0.3, \alpha_2 = 0.3$ , 오차항은 평균이 10인 포아송분포

lead	sample size	method	$\bar{C}_M (se)$	$L_T$	$\bar{L}_M (se)$	$CQ_M$
$h$	$T$					
1	25	CS	0.964 (0.00322)	17.72	24.05 (0.2835)	0.3716
		VS	0.992 (0.00428)		37.70 (0.4825)	1.1719
	50	CS	0.979 (0.00125)	17.49	23.04 (0.2101)	0.3473
		VS	0.997 (0.00051)		33.93 (0.3374)	0.9881
3	25	CS	0.954 (0.00423)	19.75	26.47 (0.2230)	0.3447
		VS	0.995 (0.00407)		54.68 (0.7658)	1.8161
	50	CS	0.972 (0.00183)	19.61	25.09 (0.1595)	0.3027
		VS	0.999 (0.00028)		50.26 (0.5287)	1.6147
5	25	CS	0.947 (0.00506)	20.26	26.94 (0.2160)	0.3332
		VS	0.996 (0.00388)		70.29 (1.0731)	2.5175
	50	CS	0.967 (0.00214)	20.28	25.60 (0.1738)	0.2804
		VS	0.999 (0.00020)		65.71 (0.7659)	2.2931

<표 3>은 오차항이 포아송분포인 INAR(2) 모형에서  $\alpha_1 = 0.3, \alpha_2 = 0.3$  인 경우로,

<표 1>과 비교하여 VS방법의 붓스트랩 평균길이  $\bar{C}_M$ 은 0.95보다 매우 크므로 상향 추정의 정도가 <표 1>보다 큼을 보여준다.

<표 4> 모형2의 모의실험결과- $\alpha_1 = 0.3, \alpha_2 = 0.3$ ,오차항은 평균이 10인 음이항분포

lead	sample size	method	$\bar{C}_M (se)$	$L_T$	$\bar{L}_M (se)$	$CQ_M$
$h$	$T$					
1	25	CS	0.943 (0.00408)	28.62	31.71 (0.5141)	0.1157
		VS	0.984 (0.00193)		44.73 (0.7251)	0.5987
	50	CS	0.957 (0.00219)	28.44	30.90 (0.3649)	0.0993
		VS	0.983 (0.00145)		40.16 (0.5167)	0.4473
$h$	$T$					
3	25	CS	0.932 (0.00506)	32.02	34.38 (0.4821)	0.0925
		VS	0.994 (0.00153)		62.00 (0.9695)	0.9827
	50	CS	0.953 (0.00267)	31.81	33.87 (0.3436)	0.0676
		VS	0.996 (0.00081)		56.83 (0.6990)	0.8343
$h$	$T$					
5	25	CS	0.928 (0.00556)	32.86	35.14 (0.4766)	0.0926
		VS	0.998 (0.00064)		78.24 (1.2401)	1.4314
	50	CS	0.950 (0.00291)	32.78	34.64 (0.3486)	0.0571
		VS	0.998 (0.00047)		72.62 (0.9187)	1.2656

또한, CS, VS방법의  $\bar{L}_M$ 이 모두  $\bar{L}_T$ 보다 큰 점에서는 <표 1>과 동일하지만, VS방법의  $\bar{L}_M$ 은  $\bar{L}_T$ 보다 매우 크고, CS방법의  $CQ_M$ 은 <표 1>에 비하여 대부분 근소하게 커진 반면, VS방법의  $CQ_M$ 은 예를 들어 선행시차가 5, 표본수 50일 때 <표 1>에서는 0.3440, <표 3>에서는 2.2931로 차이가 작지 않음을 보인다.

<표 4>는 오차항이 음이항 분포인 INAR(2) 모형에서  $\alpha_1 = 0.3, \alpha_2 = 0.3$  인 경우이다. 오차항이 포아송분포를 따르는 <표 3>과 비교하여 <표 4>에서 살펴볼 수 있는 특징은 평균포함비율과 평균길이를 함께 고려한  $CQ_M$ 이 CS, VS방법 모두 작아졌음을 보인다. 이러한 경향은 INAR(1)모형을 비교한 <표 1>과 <표 2>에서도 마찬가지로 오차항이 음이항분포를 따르는 경우가 포아송 분포일 때 보다 각각 CS, VS방법의  $CQ_M$ 이 작음을 나타내었다.

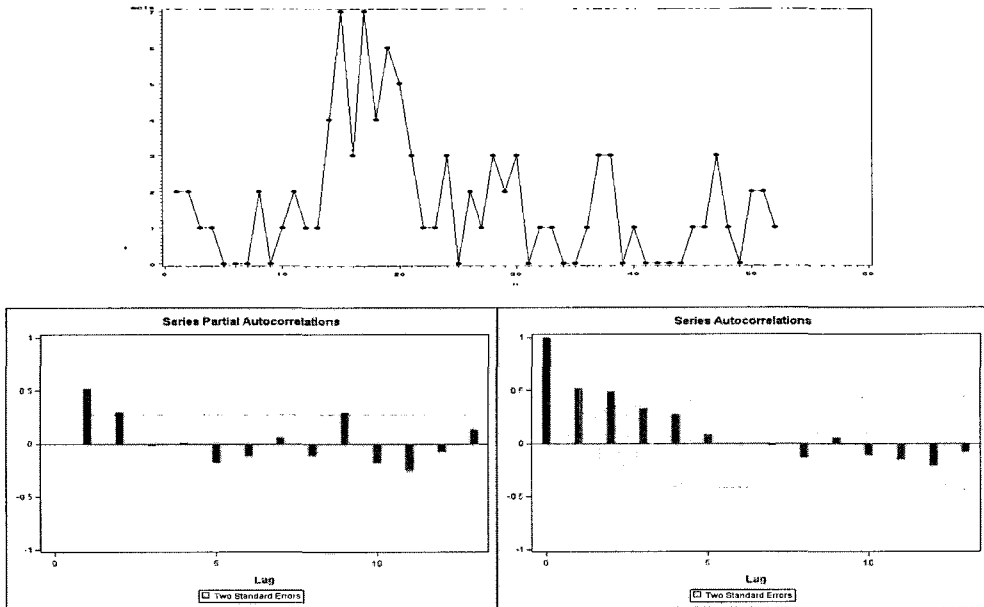
본 논문에는 수록하지 않았지만 대부분의 경우 CS방법이 CQ지표를 기준으로 할 때 VS방법보다 우수했다. 연속형 AR( $p$ )에서 CS, VS방법을 비교한 Alonso 등 (2002)에서는 VS방법이 우수했는데, 본 논문의 INAR모형의 모의실험결과는 다른 결과를 보여주고 있다. 그 이유는 INAR모형은 이항디닝 연산자를 포함함으로써 계수(coefficient)의 변동성(variability)을 이미 모형에 반영하였는데, 이에 추가적으로 모수추정의 변동성을 고려한 VS방법을 적용할 경우 평균포함비율은 명목포함비율보다 커지고, 붓스트랩 예측구간의 길이는 경험적인 분포(empirical distribution)의 구간길

이 보다 커지기 때문이다.

### 5. 실증분석

5절에서는 이산형 시계열 자료를 Box와 Jenkins의 ARMA모형으로 적합한 후의 예측과 본 논문에서 제안한 붓스트랩 방법의 예측을 비교하고자 한다. 실증분석 자료로 1982년 일본의 Totori 현(縣)에서 52주 동안 발생한 특정질병 (muco-cutaneous lymph node syndrome: MCLS)의 발생건수를 이용하였다. 이 자료는 Kashiwagi와 Yanagimoto (1992)가 상태공간모형으로 이용하였고, 또한 Grunwald 등 (2000)이 모형화 하였다. 즉, Grunwald 등 (2000)은 오차항 분포가 평균  $\lambda$ 인 포아송분포를 따르는 INAR(1) 모형으로, 그리고  $X_t|X_{t-1}=x$ 의 분포로 평균이  $\alpha x + \lambda$ 인 포아송분포로 모형화하였다. 이론적으로 Grunwald 등 (2000)이 고려한 두 모형은 동일한 최소제곱추정량  $\hat{\alpha}$ ,  $\hat{\lambda}$ 과 주변평균 (marginal mean)  $\mu \equiv \lambda/(1-\alpha)$ 을 제공한다. 두 모형의 차이점은 주변분산 (marginal variance)으로 전자는  $Var(X_t) = \mu$ 이고, 후자는  $Var(X_t) = \mu/(1-\alpha^2)$ 이다.

<그림 1>에서 보는 바와 같이 MCLS자료는 최대값이 7을 취하는 이산형자료이며, 표본자기상관함수 (sample autocorrelation function)와 표본편자기상관함수 (sample partial autocorrelation function)로 볼 때 INAR(1)모형을 고려할 수 있다. 2절에서 서술한바와 같이 INAR(p)의 자기상관함수는 연속형 AR(p)와 같으므로 INAR모형의



<그림 1> 1982년 일본의 Totori 현(縣)에서 발생한 주별 MCLS의 발생건수, 표본자기상관함수, 표본편자기 상관함수

적절한 차수  $p$ 를 선택할 수 있다. 모형적합을 평가하기 위하여 자료에서 마지막 12시점의 관찰값들은 제외하고 모형 적합을 실시하였다. <표 5>는 40개 자료를 분석에 활용하여 INAR(1) 모형의 CS붓스트랩방법, 자료를 연속형 AR(1)모형으로 간주한 CS붓스트랩방법으로 12시점앞의 예측값과 95% 예측구간으로 붓스트랩 반복 횟수  $B=1000$ 이다. AR(1) 모형의 붓스트랩표본 생성은 Pascual 등 (2004)에서 밝힌 바와 같이 잔차를 중심화 (centering)와 재척도화 (rescaling) 한 후 재표집 하였다. <표 5>에서 AR(1) 모형의 예측치 (forecast)는  $B$ 회의 붓스트랩 표본평균, INAR(1) 모형의 예측치는  $B$ 회의 붓스트랩 표본 평균과 중위수를 같이 기술하였다. <표 5>에서 보는 바와 같이 INAR(1) 모형으로 적합시 CS 붓스트랩 예측구간은 자료의 속성을 그대로 유지한 비음의 정수값이며, 또한 모든 선행시차에서 INAR(1)모형의 붓스트랩 예측구간 길이는 AR(1)모형의 붓스트랩 예측구간 길이보다 짧음을 알 수 있다. 자료를 INAR(1) 모형과 연속형 AR(1) 모형으로 적합할 때 아래의 평균제곱오차 (mean square error: MSE)로 예측능력을 비교하고자 한다.

$$MSE = \sum_{t=1}^{12} \frac{(\hat{Y}_t - Y_t)^2}{12}, \tag{5.1}$$

여기서  $\hat{Y}_t$ 는  $Y_t$ 의 예측치이다. 결과는 AR(1)의 MSE는 1.633, INAR(1)에서는 붓스트랩표본의 평균을 예측치로 사용할 때 MSE는 1.294이고, 붓스트랩표본의 중위수를 예측치로 사용할 때 MSE는 0.917이다.

<표 5> MCLS자료의 95% 신뢰구간과 예측치

lead	observed value	AR(1)			INAR(1)			
		forecast	lower limit	upper limit	forecast (mean)	forecast (median)	lower limit	upper limit
1	0	1.324	-1.054	5.521	1.322	1	0	5
2	0	1.698	-1.321	6.035	1.455	1	0	5
3	0	1.773	-1.340	6.163	1.573	1	0	5
4	0	1.850	-1.273	6.220	1.614	1	0	5
5	1	1.995	-1.109	6.290	1.721	1	0	6
6	1	1.935	-1.117	6.409	1.648	1	0	6
7	3	1.935	-1.124	6.266	1.611	1	0	5
8	1	1.888	-1.318	6.086	1.613	1	0	6
9	0	1.903	-1.204	6.627	1.591	1	0	6
10	2	1.871	-1.140	6.353	1.541	1	0	5
11	2	1.882	-1.164	6.142	1.645	1	0	6
12	1	1.980	-1.195	6.414	1.686	1	0	5

## 6. 결론

시계열자료 분석 시 중요한 점은 미래시점의 예측이며, 구간추정이 점추정보다 선호된다. 미래시점의 예측을 위해서는 자료의 속성을 잘 반영하는 모형을 세우는 일이 가장 우선적으로 선행되어야 한다. 그러나, 대부분의 이산형 시계열자료 분석 시 정규 분포를 가정한 Box와 Jenkins의 ARMA모형으로 접근하고 있는 실정이다.

그러나, 최근 1990년 대 이후 비음의 정수값을 가지는 시계열자료를 모형화하기 위해 이항디닝 연산자를 이용한 모형의 연구가 활발히 진행되고 있다. 이중 특히 INAR( $p$ ) 모형은 연속형 AR( $p$ )과 같은 형태의 자기상관함수를 가지는 특징이 있으나, INAR( $p$ )에서  $p$ 가 2이상일 때 예측구간을 수리적으로 구하는 연구는 이제까지 없는 실정이다. 따라서, 본 연구에서는 예측구간을 구하기 위해 두 가지 붓스트랩방법 즉, CS (conditional sieve bootstrap)방법과 모수추정량의 변동성을 반영하는 VS (sieve bootstrap that includes the variability due to parameter estimation)방법을 제안하고, 모의실험을 통하여 제시한 방법들을 비교하였다. 모의실험 결과 INAR모형에서는 연속형 AR모형과 다르게 CS방법이 오차항의 분포가 포아송, 음이항 분포, 선행시차, 표본수에 관계없이 VS방법보다 평균포함비율과 평균길이를 함께 고려하는  $CQ_M$ 지표측면에서 우수했다. 또한, 오차항의 분포가 과대산포분포인 음이항분포인 경우가 포아송 분포에 비하여  $CQ_M$ 이 보다 작은 수치를 제공하였다.

실제 자료분석에서는 이산시계열자료를 Box와 Jenkins의 ARMA모형으로 적합 시 붓스트랩 예측구간과 INAR모형으로 적합 시 본 논문에서 제시한 붓스트랩 예측구간을 비교하였다. 그 결과 INARA모형이 평균제곱오차, 예측구간의 길이 측면에서 보다 좋음을 확인할 수 있었다.

향후 연구과제로는 본 연구에서 수행된 모의 실험결과를 뒷받침할 이론적인 연구가 진행되어야 할 것이다.

## 참고문헌

- [1] Alzaid, A.A. and Al-Osh, M. (1990). An integer-valued  $p$ th-order autoregressive structure (INAR( $p$ )) process. *Journal of Applied Probability*, Vol. 27, 314-324.
- [2] Alonso, A.M., Peña, D. and Romo, J. (2002). Forecasting time series with sieve bootstrap. *Journal of Statistical Planning and Inference*, Vol. 100, 1-11.
- [3] Al-Osh, M.A. and Alzaid, A.A. (1987). First-order integer-valued autoregressive (INAR(1)) process. *Journal of Time Series Analysis*, Vol. 8, 261-275.
- [4] Cheng, Q.S. (1999). On time reversibility of linear processes. *Biometrika*, Vol. 86, 483-486.
- [5] Cox, D.R. (1981). Statistical analysis of time series: Some recent developments. *Scandinavian Journal of Statistics*, Vol. 8, 93-115.

- [6] Du, J.-Guan and Li, Y. (1991). The integer-valued autoregressive (INAR( $p$ )) model. *Journal of Time Series Analysis*, Vol. 12, 129-142.
- [7] Findley, D.F. (1986). The uniqueness of moving average representations with independent and identically distributed random variables for non-Gaussian stationary time series. *Biometrika*, Vol. 73, 520-521.
- [8] Freeland, R.K. and McCabe, B.P.M. (2004). Forecasting discrete valued low count time series. *International Journal of Forecasting*, Vol. 20, 427-434.
- [9] García-Jurado, I., González-Manteiga W., Prada-Sánchez, J.M., Febrero-Bande, N., Febrero-Bande, M. and Cao, R. (1995). Predicting using Box-Jenkins, nonparametric bootstrap techniques. *Technometrics*, Vol. 37, 303-310.
- [10] Grigoletto, M. (1998). Bootstrap prediction intervals for autoregression: some alternative. *International Journal of Forecasting*, Vol. 14, 447-456.
- [11] Grunwald, G.K., Hyndman, R.J., Tedeso, L. and Tweedie, R. L. (2000). Non-Gaussian conditional linear AR(1) models. *Australian & New Zealand Journal of Statistics*, Vol. 42, 479-495.
- [12] Hallin, M., Lefevre, C. and Puri, M.L. (1988). On time-reversibility and the uniqueness of moving average representation for non-Gaussian stationary time series. *Biometrika*, Vol. 75, 170-171.
- [13] Jung, R.C. and Tremayne, A.R. (2006). Coherent forecasting in integer time series models. *International Journal of Forecasting*. To appear.
- [14] Kashiwagi, N., and Yanagimoto, T. (1992). Smoothing serial count data through a state-space model. *Biometrics*, Vol. 48, 1187-1194.
- [15] Kim, H.Y. and Park, Y.S. (2006). Prediction mean squared error of the Poisson INAR(1) process with estimated parameters. *Journal of the Korean Statistical Society*. Vol. 35, 37-47.
- [16] Latour, A. (1998). Existence and stochastic structure of non-negative inter-valued autoregressive processes. *Journal of Time Series Analysis*, Vol. 19, 439-455.
- [17] McCabe, B.P.M. and Martin, G.M. (2005). Bayesian predictions of low count time series. *International Journal of Forecasting*, Vol. 21, 315-330.
- [18] McKenzie, E. (1985). Some simple models for discrete variate time series. *Water Resources Bulletin*, Vol. 21, 645-650.
- [19] McKenzie, E. (1988). Some ARMA models for dependent sequence of Poisson counts. *Advances in Applied Probability*, Vol. 20, 822-835.
- [20] Pacual, L., Romo, J. and Ruiz, E. (2004). Bootstrap predictive inference for ARIMA processes. *Journal of Time Series Analysis*, Vol. 25, 449-465.
- [21] 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*, Vol. 101, 472-483.

- [22] 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*, Vol. 29, 269-284.
- [23] 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*, Vol. 26, 17-36.
- [24] Stine, R.A. (1987). Estimating properties of autoregressive forecasts. *Journal of American Statistical Association*. Vol. 82, 1072-1078.
- [25] Sueutel, F.W. and Van Harn, K. (1979). Discrete analogues of self-decomposability and stability. *The Annals of Probability*, Vol. 7, 893-899.
- [26] Thombs, L.A. and Schucany, W.R. (1990). Bootstrap prediction intervals for autoregression. *Journal of the American Statistical Association*, Vol. 85, 486-492.

[Received April 2006, Accepted June 2006]