

Estimation for random coefficient autoregressive model

Ju Sung Kim^a · Sung Duck Lee^{a,1} · Na Rae Jo^a · In Suk Ham^b

^aDepartment of Information and Statistics, Chungbuk National University;

^bDepartment of Nursing Science, Chungbuk National University

(Received December 31, 2015; Revised January 3, 2016; Accepted January 3, 2016)

Abstract

Random Coefficient Autoregressive models (RCA) have attracted increased interest due to the wide range of applications in biology, economics, meteorology and finance. We consider an RCA as an appropriate model for non-linear properties and better than an AR model for linear properties. We study the methods of RCA parameter estimation. Especially we proposed the special case that a random coefficient $\phi(t)$ has the initial value $\phi(0)$ in the RCA model. In practical study, we estimated the parameters and compared Prediction Error Sum of Squares (PRESS) criterion between AR and RCA using Korean Mumps data.

Keywords: Random Coefficient Autoregressive Model, Subsample, Norli's estimation, least squares estimate, Korean Mumps data.

1. 서론

선형 시계열 모형들은 실제 시계열 자료들에 아주 잘 적합 될 뿐만 아니라 분석을 하는 데도 용이하다. 그러나 선형 시계열 모형들은 강한 비대칭성을 나타내는 자료들이나 또는 시간을 역행할 수 없는 자료들에 대해서는 관념적으로 적당하지 않다고 알려져 있다 (Tong, 1990). 여러 분야에서 얻어지는 다양한 시계열 자료의 일반적인 도약 현상(jump phenomena)과 한계 순환(limit cycles) 같은 것들은 선형 시계열 모형으로는 잘 설명되어질 수가 없어서 이러한 선형 시계열 모형들의 제한들과 비선형 시계열 모형들의 이점들을 이용한 비선형 시계열 모형들에 관한 연구가 이루어져 왔다. 시계열 자료가 자기회귀 모형(Autoregressive model; AR)을 따르지만 통제되지 않는 다양한 조건들에 의해 비선형성을 갖는 경우, AR 모형에 적합시키는 것은 좋은 결과를 가져오지 못할 수도 있다. 이런 경우, 먼저 평활법(smoothing method)을 사용하면 데이터 양식을 쉽게 개선할 수도 있지만, 또 다른 접근으로 확률계수 자기회귀 모형(Random Coefficient Autoregressive model; RCA)을 고려해 볼 수 있다. 확률계수 자기회귀 모형은 생물학, 기상학, 경제학 그리고 최근에는 금융분야와 같은 다양한 범위에서 분석의 용이성 때문에 많이 연구되고 있다 (Norli 등, 2011). Nicholls와 Quinn (1981)은 확률계수 자기회귀 모형을 처음 소개하였으며, Son (1996), Hwang과 Basawa (1998), Rahiala (1999), Wang과 Ghosh (2002), Aue 등 (2006), Norli 등 (2011)에 의해 다양하게 연구되었다.

This work was supported by the research grant of Chungbuk National University in 2013.

¹Corresponding author: Department of Information and Statistics, Chungbuk National University, 1 Chungdae-ro, Seowon-gu, Cheongju, Chungbuk 28644, Korea. E-mail: sdlee@chungbuk.ac.kr

본 논문은 선형 시계열 분석기법인 AR 모형과 비선형 시계열 분석기법인 RCA 모형에 자료를 적합시켜 모수를 추정하고 예측의 우수성을 비교하고자 한다. 2장에서는 확률계수 자기회귀 모형의 모수 추정에 대한 통계적 속성을 다룬 정리를 반복(귀납)적 회귀 추정 절차를 통해 증명하고, 이 방법을 이용하여 초기값과 미지의 분산을 추정하였다. 3장에서는 모의실험을 통하여 2장에서 제시한 방법으로 모수를 추정하였고 RCA(1) 모형의 시뮬레이션을 보였다. 4장 실증분석에서는 한국의 Mumps 자료를 이용하여 AR 모형과 RCA 모형에 적합시켜 모수를 추정하고, 각 예측값들을 PRESS로 비교하였다.

2. 확률계수 자기회귀 모형

확률계수 자기회귀 모형(Random Coefficient Autoregressive model; RCA)은 Nicholls와 Quinn (1981)에 의해 처음 연구되었으며, 일반화된 p 차 RCA 모형, RCA(p)는 다음과 같다.

$$H(X_{t-1}, Z_{tj}; \theta) = (\theta_1 + Z_{t1})X_{t-1} + \cdots + (\theta_p + Z_{tp})X_{t-p} + \epsilon_t,$$

여기서 $\{Z_{tj}, j = 1, \dots, p\}$ 는 평균이 0이고 분산이 유한한, 독립이고 동일한 분포를 갖는 확률변수이며 $\{\epsilon_t\}$ 와 독립이다. 또한 $\{\epsilon_t\}$ 는 평균이 0이고, 분산이 σ^2 인 독립이며 동일한 분포를 갖는 확률변수의 수열이고, $\theta_1, \dots, \theta_p$ 는 모수들이다. 위 식의 특수한 경우인 1차 RCA 모형, RCA(1)은 다음과 같다.

$$X_t = (\theta + Z_t)X_{t-1} + \epsilon_t, \quad t = 1, 2, \dots,$$

여기서 $\{X_t\}$ 는 $\theta^2 + EZ_t^2 < 1$ 조건하에서 정상(stationary)이며 에르고딕(ergodic)이다.

2.1. Norli의 RCA 모형과 추정

Norli 등 (2011)은 Nicholls와 Quinn (1981)의 최소제곱법을 이용한 RCA(1) 모형의 모수 추정 절차를 다음과 같이 정리하였다. Norli 등 (2011)이 정리한 RCA(p) 모형의 식은 다음과 같다.

$$y_t = \sum_{i=1}^p [\theta_i + b_i(t)]y_{t-i} + \epsilon_t,$$

여기서 $\theta = (\theta_1, \theta_2, \dots, \theta_p)'$ 인 $p \times 1$ 상수벡터이고, $\{\epsilon_t\} \sim IN(0, \sigma^2)$ 이다. 그리고 $b(t) = (b_1(t), \dots, b_p(t))'$ 이며 평균이 0이고, 각 t 에 대해 $E[b(t)b'(t)] = \sum \text{diag}(\sigma_{b_1^2}, \sigma_{b_2^2}, \dots, \sigma_{b_p^2})$ 이며, $\{\epsilon_t\}$ 와 독립이다. 그리고 $\theta'\theta + \text{tr}(\sum) < 1$ 일 때, 정상적이며 에르고딕이다.

또한 Norli 등 (2011)이 제시한 $p = 1$ 인 경우인 RCA(1) 모형의 모수 추정식은 다음과 같다.

$$\begin{aligned} \hat{\theta} &= \frac{\sum_{t=2}^n y_t y_{t-1}}{\sum_{t=2}^n y_{t-1}^2}, \\ \hat{\sigma}_b^2 &= \frac{\sum_{t=2}^n \hat{u}_t^2 (y_{t-1}^2 - \bar{z})}{\sum_{t=2}^n y_{t-1}^4 - \bar{z}^2}, \\ \hat{\sigma}_\epsilon^2 &= \frac{\sum_{t=2}^n \hat{u}_t^2 - \hat{\sigma}_b^2 \bar{z}}{n-1}, \end{aligned} \quad (2.1)$$

여기서 $\bar{z} = (\sum_{t=2}^n y_{t-1}^2)/(n-1)$, 그리고 $\hat{u}_t = y_t - \hat{\theta}y_{t-1}$ 이다.

2.2. Nicholls와 Quinn의 RCA 모형과 추정

다음 모형을 고려해보자.

$$Y_t = \underline{\phi}'(t)X_t + \epsilon_t \tag{2.2}$$

$$\underline{\phi}(t) = \underline{\phi}(t-1) + \eta_t, \quad t = 0, \pm 1, \pm 2, \dots,$$

여기서 $X_t = [Y_{t-1}, Y_{t-2}, \dots, Y_{t-p}]'$, $\underline{\phi}(t) = [\phi_1(t), \phi_2(t), \dots, \phi_p(t)]'$, 그리고 $\eta_t = [\eta_{1t}, \eta_{2t}, \dots, \eta_{pt}]'$ 이다. $\{\epsilon_t\}$ 는 평균이 0이고 분산이 σ^2 인 독립적인 정규분포 $IN(0, \sigma^2)$ 이다. $\{\eta_t\}$ 는 $IN(0, \alpha\sigma^2 I_p)$ 이고 $\{\epsilon_t\}$ 와 독립이며, I_p 는 순서 p 의 단위 행렬이다. α 와 σ^2 은 모두 알려져 있지 않고, $\alpha < 1$ 이다.

여기서 주안점은 초기값(initial value) $\underline{\phi}(0) = [\phi_1(0), \phi_2(0), \dots, \phi_p(0)]'$ 을 갖는 특별한 경우이다.

$$\underline{\phi}(t) = \underline{\phi}(0) + \eta_t. \tag{2.3}$$

2.2.1. $\underline{\phi}(t)$ 의 추정 회귀 공식에 기초한 반복적인 추정 절차(recursive estimation procedure)는 다음과 같다.

시간 $t = 1$ 에서 식 (2.2)는 $Y_1 = \epsilon_1$, $\underline{\phi}(1) = \underline{\phi}(0) + \eta_1$ 이므로 $\underline{\phi}(1)$ 의 가중 최소 제곱 추정치(weighted least squares estimate)는 $\underline{g}(1) = \underline{\phi}(0)$ 으로 주어진다. 시간 $t = 2$ 에서 식 (2.2)는 $Y_2 = \underline{\phi}'(2)X_2 + \epsilon_2$, $\underline{\phi}(2) = \underline{\phi}(1) + \eta_2$ 이고, 여기서 $X_2 = [Y_1, 0, \dots, 0]'$ 이다. 이를 바탕으로 다음과 같이 나타낼 수 있다.

$$\begin{bmatrix} Y_2 \\ \underline{g}(2|1) \end{bmatrix} = \begin{bmatrix} X_2' \\ I_p \end{bmatrix} \underline{\phi}(2) + \begin{bmatrix} \epsilon_2 \\ \underline{g}(2|1) - \underline{\phi}(2) \end{bmatrix},$$

여기서 $\underline{g}(2|1)$ 은 $\underline{g}(1)$ 을 알고 난 후의 $\underline{\phi}(2)$ 의 조건부 추정치로, 식 (2.2)에서 $\underline{g}(2|1) = \underline{g}(1) + \eta_1$ 이다. $\underline{g}(2|1)$ 의 분산은

$$V[\underline{g}(2|1)] = V[\underline{g}(1)] + V[\eta_1] \tag{2.4}$$

이다. 실제로 $\underline{g}(2|1)$ 은 η_1 을 알지 못하기 때문에 $\underline{g}(2|1) = \underline{g}(1)$ 으로 계산하여 $\underline{g}(1|0) = \underline{\phi}(0)$ 이다. 만약 $U_{2|1} = V[\underline{g}(2|1)]$ 이라 하면 식 (2.4)는

$$U_{2|1} = U_1 + \alpha\sigma^2 I_p$$

$$= \alpha\sigma^2 I_p$$

이고, 여기서 $U_1 = V[\underline{g}(1)] = 0$ 이다.

그러므로 $\underline{\phi}(2)$ 의 가중 최소 제곱 추정치는 다음과 같이 주어진다.

$$\underline{g}(2) = \left[\begin{bmatrix} X_2 & I_p \end{bmatrix} \begin{bmatrix} \sigma^2 & 0 \\ 0 & U_{2|1} \end{bmatrix}^{-1} \begin{bmatrix} X_2' \\ I_p \end{bmatrix} \right]^{-1} \begin{bmatrix} X_2 & I_p \end{bmatrix} \begin{bmatrix} \sigma^2 & 0 \\ 0 & U_{2|1} \end{bmatrix}^{-1} \begin{bmatrix} Y_2 \\ \underline{g}(2|1) \end{bmatrix}$$

$$= \left[X_2 X_2' + V_{2|1}^{-1} \right]^{-1} \left[X_2 Y_2 + V_{2|1}^{-1} \underline{g}(2|1) \right],$$

여기서 $V_{2|1} = (\sigma^2)^{-1} U_{2|1}$ 이다.

반복추정의 절차는 $\underline{\phi}(t)$ 의 최소 제곱 추정치를 얻기 위한 시간 t 까지 반복되며 다음과 같다.

$$\underline{g}(t) = \left[X_t X_t' + V_{t|t-1}^{-1} \right]^{-1} \left[X_t Y_t + V_{t|t-1}^{-1} \underline{g}(t-1) \right], \tag{2.5}$$

여기서

$$\underline{X}_t = \begin{cases} (Y_{t-1}, Y_{t-2}, \dots, Y_{t-p-1}, 0, \dots, 0), & t = 1, 2, \dots, p+1, \\ (Y_t, \dots, Y_{t-p}), & t = p+2, \dots, n, \end{cases}$$

$V_{t|t-1} = V_{t-1} + \alpha I_p$, $V_t = [\underline{X}_t \underline{X}_t' + V_{t|t-1}^{-1}]^{-1}$ 이고, $\underline{g}(t|t-1) = \underline{g}(t-1) + \underline{\eta}_{t-1}$ 이다.

행렬 이론으로부터 다음을 보일 수 있다.

$$\begin{aligned} V_t &= [\underline{X}_t \underline{X}_t' + V_{t|t-1}^{-1}]^{-1} \\ &= V_{t|t-1} + V_{t|t-1} \underline{X}_t M_t^{-1} \underline{X}_t' V_{t|t-1}, \end{aligned}$$

여기서 $M_t = [1 + \underline{X}_t' V_{t|t-1} \underline{X}_t]$ 이다. 식 (2.5)를 V_t 와 $V_{t|t-1} \underline{X}_t M_t^{-1} \underline{X}_t' Y_t$ 를 이용하여 정리하면 다음과 같다.

$$\begin{aligned} \underline{g}(t) &= \underline{g}(t|t-1) + V_{t|t-1} \underline{X}_t Y_t - V_{t|t-1} \underline{X}_t M_t^{-1} \underline{X}_t' V_{t|t-1} \underline{X}_t Y_t - V_{t|t-1} \underline{X}_t M_t^{-1} \underline{X}_t' \underline{g}(t|t-1) \\ &= \underline{g}(t|t-1) + V_{t|t-1} \underline{X}_t M_t^{-1} (Y_t - \underline{X}_t' \underline{g}(t|t-1)) + V_{t|t-1} \underline{X}_t (1 - M_t^{-1} \underline{X}_t' V_{t|t-1} \underline{X}_t - M_t^{-1}) Y_t \\ &= \underline{g}(t|t-1) + V_{t|t-1} \underline{X}_t M_t^{-1} (Y_t - \underline{X}_t' \underline{g}(t|t-1)), \end{aligned}$$

여기서 $1 - M_t^{-1} (\underline{X}_t' V_{t|t-1} \underline{X}_t + 1) = 0$ 이다.

따라서 크기가 n 인 식 (2.2)와 같은 모형이 있을 때, $Y_{-h} = 0$, $h = 0, 1, \dots$ 이라 하자. 그러면 시간 t 에서 $\underline{\phi}(0)$ 와 α 가 주어졌을 때, $\underline{\phi}(t)$ 의 MVUE(Minimum Variance Unbiased Estimator)는 다음과 같이 주어진다.

$$\underline{g}(t) = \underline{g}(t|t-1) + V_{t|t-1} \underline{X}_t M_t^{-1} (Y_t - \underline{X}_t' \underline{g}(t|t-1)), \quad (2.6)$$

여기서 $\underline{g}(1|0) = \underline{\phi}(0)$, $M_t = 1 + \alpha \underline{X}_t' \underline{X}_t$, $V_{t|t-1} = V_{t-1} + \alpha I_p$, $V_t = V_{t|t-1} - V_{t|t-1} \underline{X}_t M_t^{-1} \underline{X}_t' V_{t|t-1}$, $V_0 = 0$ 이다.

시간 t 에서 $\underline{g}(t)$ 는 $\underline{\phi}(t)$ 의 가중 최소 제곱 추정치이기 때문에 $E[\underline{g}(t)] = \underline{\phi}(t)$ 이고, $\underline{g}(t)$ 는 $\underline{\phi}(t)$ 의 MVUE이다. 그러므로 $\underline{g}(t)$ 는 $N[\underline{\phi}(t), V_t]$ 의 분포를 따른다.

2.2.2. $\underline{\phi}(0)$ 의 추정 n 개의 관측치들을 크기가 m 인 똑같은 표본 조각들(부표본)로 나눈다고 한다. 여기서 부표본의 크기 m 은 $m > p$ 이고 $m < 2n$ 이다. 그러면 이 부표본들의 수는 $h = n/m$ 이고, 이 추정치들은 다음과 같이 만들어진다.

$$Y_1, Y_2, \dots, Y_m; Y_{m+1}, \dots, Y_{2m}; \dots; Y_{n-m+1}, \dots, Y_n.$$

시간이 주어졌을 때, $\underline{\phi}(t)$ 는 상수로 가정할 수 있고 간격 $1 + m(j-1) \leq t \leq mj$, $j = 1, 2, \dots, h$ 에서 $\underline{\phi}(j)$ 와 같다.

그러므로 j 번째 부표본에서 다음과 같은 모형으로 표현할 수 있다.

$$Y_{k+m(j-1)} = \underline{\phi}'(j) \underline{X}_{k+m(j-1)} + \epsilon_{k+m(j-1)}, \quad k = 1, 2, \dots, m, \quad j = 1, 2, \dots, h,$$

여기서 $\underline{X}_{k+m(j-1)} = [Y_{k+m(j-1)-1}, Y_{k+m(j-1)-2}, \dots, Y_{k+m(j-1)-p}]'$, $\underline{\phi}(j) = [\underline{\phi}_1(j), \underline{\phi}_2(j), \dots, \underline{\phi}_p(j)]'$ 그리고 $\epsilon_{k+m(j-1)} \sim IN(0, \sigma_j^2)$ 이다.

j 번째 부표본에서 $\underline{\phi}(j)$ 의 최소 제곱 추정치는 다음과 같이 구할 수 있다.

$$\hat{\underline{\phi}}(j) = A_m^{-1} \sum_{k=p+1}^m \underline{X}_{k+m(j-1)} \underline{Y}_{k+m(j-1)},$$

여기서 $A_m = \sum_{k=p+1}^m \underline{X}_{k+m(j-1)} \underline{X}'_{k+m(j-1)}$ 이다. 그러므로 $\underline{\phi}(0)$ 의 추정치는 다음과 같다.

$$\hat{\underline{\phi}}(0) = (h)^{-1} \sum_{j=1}^h \hat{\underline{\phi}}(j). \tag{2.7}$$

2.2.3. σ^2 과 α 의 추정 위의 결과로부터 j 번째 부표본의 σ_j^2 의 최소 제곱 추정치는 다음과 같이 구할 수 있다.

$$\hat{\sigma}_j^2 = (m - p - 1)^{-1} \sum_{k=p+1}^m \left[Y_{k+m(j-1)} - \hat{\phi}_1(j) Y_{k+m(j-1)-1} - \cdots - \hat{\phi}_p(j) Y_{k+m(j-1)-p} \right]^2.$$

그리고 이 분산의 평균은 미지의 분산 σ^2 을 추정하는 기대치가 될 것이다. 그러므로 σ^2 의 추정치는

$$\hat{\sigma}^2 = (h)^{-1} \sum_{j=1}^h \hat{\sigma}_j^2 \tag{2.8}$$

이다. 위의 유사한 방법으로 α 를 추정할 수 있다.

$$\hat{\alpha} = \frac{\hat{V}[\hat{\underline{\phi}}(j)]}{\hat{\sigma}^2}, \tag{2.9}$$

여기서 $V[\hat{\underline{\phi}}(j)] = (h)^{-1} [\sum_{j=1}^h [\hat{\underline{\phi}}(j)]^2 - [\sum_{j=1}^h \hat{\underline{\phi}}(j)]^2/h]$ 이다.

3. 시뮬레이션 분석

R프로그램을 이용하여 식 (3.1)과 같이 초기값 $\phi(0)$ 을 갖는 RCA(1) 모형을 따르는 난수를 발생시켜 시뮬레이션 분석을 실시였다.

$$\begin{aligned} Y_t &= \phi(t)Y_{t-1} + \epsilon_t, \\ \phi(t) &= \phi(0) + \eta_t, \end{aligned} \tag{3.1}$$

여기서 $\{\epsilon_t\} \sim IN(0, 1)$ 이고 $\{\eta_t\}$ 는 $IN(0, \alpha)$ 이며 $\{\epsilon_t\}$ 와 독립이다.

R 프로그램을 이용하여 $\text{Var}(\epsilon_t) = 1$ 이고, $\phi(0)$ 와 α 의 다양한 값을 갖는 표본사이즈(n)가 500인 시계열을 생성하여 6가지 RCA(1) 모형의 시계열 그림을 Figure 3.1에 제시하였다.

난수발생한 데이터를 활용하여 부표본을 이용한 2.2절의 Nicholls와 Quinn 모형의 추정으로 모수를 추정하기 위하여 표본사이즈가 500인 데이터를 부표본 크기(subsample size)가 50인 10개의 부표본으로 나누었으며, 부표본 모형은 다음과 같다.

$$\begin{aligned} Y_{k+50(j-1)} &= \phi(j)Y_{k+50(j-1)-1} + \epsilon_{k+50(j-1)}, \\ k &= 1, 2, \dots, 50, \quad j = 1, 2, \dots, 10, \\ \epsilon_{k+50(j-1)} &\sim IN(0, \sigma_j^2). \end{aligned} \tag{3.2}$$

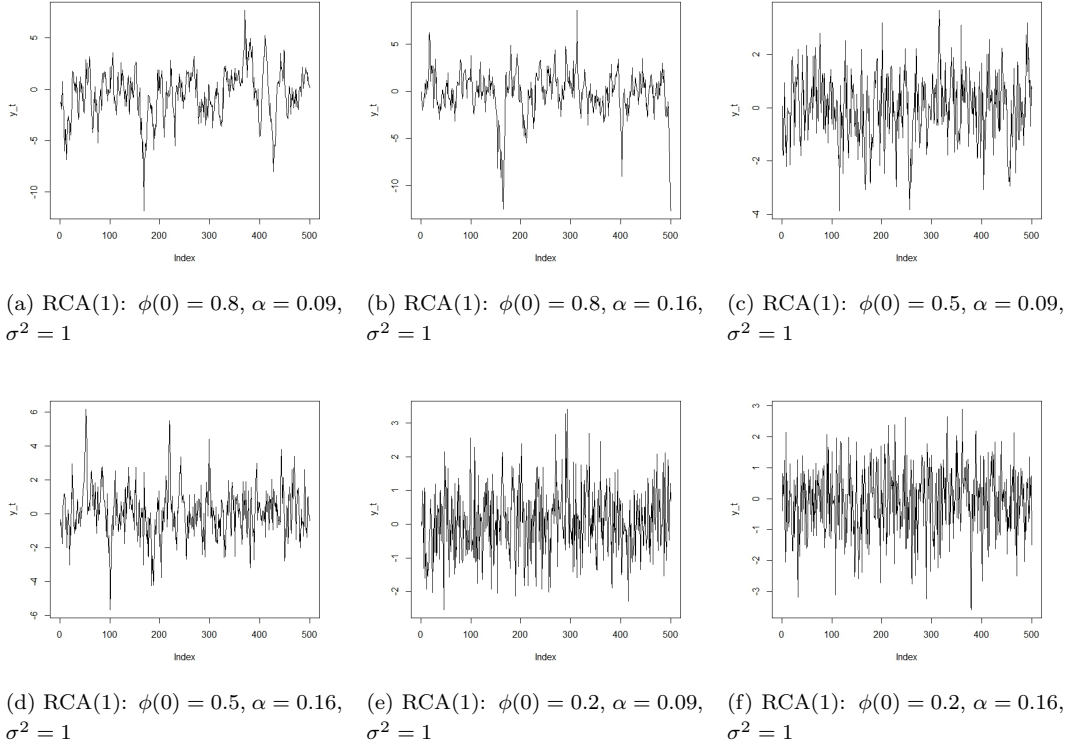


Figure 3.1. Time series plots of RCA(1) simulations.

Table 3.1. Parameter estimation result for simulation

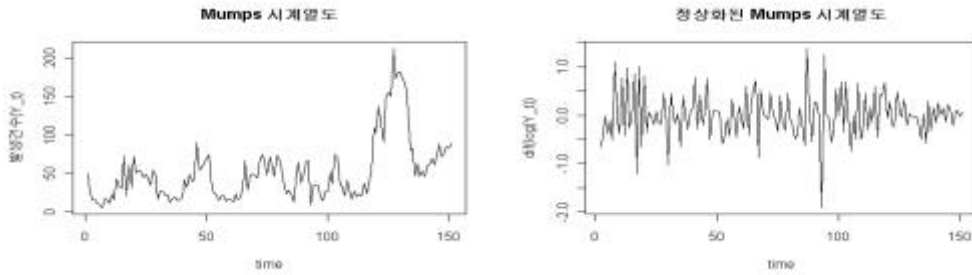
데이터	$\phi(0)$ $\hat{\phi}(0)$	σ^2 $\hat{\sigma}^2$	α $\hat{\alpha}$	데이터	$\phi(0)$ $\hat{\phi}(0)$	σ^2 $\hat{\sigma}^2$	α $\hat{\alpha}$
(a)	0.8	1	0.09	(b)	0.8	1	0.16
	0.7996	1.5335	0.0673		0.7869	1.8580	0.1010
(c)	0.5	1	0.09	(d)	0.5	1	0.16
	0.5311	1.0274	0.1070		0.5621	1.2150	0.1469
(e)	0.2	1	0.09	(f)	0.2	1	0.16
	0.2192	0.9042	0.0987		0.1896	1.1833	0.1447

Figure 3.1의 난수 데이터를 2.2절의 부표본을 활용한 Nicholls와 Quinn 모형의 추정모수 추정한 결과이다. 난수 발생 시 지정한 모수값과 부표본을 활용하여 추정한 추정값의 차이가 크지 않음을 확인할 수 있다.

난수 발생한 시계열을 2.1절의 Norli 모형 추정과 2.2절의 부표본을 활용한 Nicholls와 Quinn 모형의 추정모수 추정한 결과를 비교한 결과는 Table 3.2와 같다. 시계열을 부표본으로 나누어 모수를 추정한 결과, Norli 모형의 추정값보다 주어진 모수값에 더 가까운 것을 알 수 있다. 따라서 부표본을 이용한 Nicholls와 Quinn 모형의 추정모수 식 (3.1)을 추정하는 것이 더 바람직함을 확인하였다.

Table 3.2. Comparison of estimates

데이터	모수	Norli 모형의 추정	Nicholls와 Quinn 모형의 추정
(a)	$\phi(0) = 0.8$	$\hat{\theta} = 0.8555$	$\hat{\phi}(0) = 0.7996$
	$\sigma^2 = 1$	$\hat{\sigma}_\epsilon^2 = 1.5436$	$\hat{\sigma}^2 = 1.5335$
	$\alpha\sigma^2 = 0.09$	$\hat{\sigma}_b^2 = 0.0307$	$\alpha\hat{\sigma}^2 = 0.1032$
(b)	$\phi(0) = 0.8$	$\hat{\theta} = 0.8500$	$\hat{\phi}(0) = 0.7869$
	$\sigma^2 = 1$	$\hat{\sigma}_\epsilon^2 = 1.8983$	$\hat{\sigma}^2 = 1.8580$
	$\alpha\sigma^2 = 0.16$	$\hat{\sigma}_b^2 = 0.0823$	$\alpha\hat{\sigma}^2 = 0.1877$
(c)	$\phi(0) = 0.5$	$\hat{\theta} = 0.5396$	$\hat{\phi}(0) = 0.5311$
	$\sigma^2 = 1$	$\hat{\sigma}_\epsilon^2 = 1.0476$	$\hat{\sigma}^2 = 1.0274$
	$\alpha\sigma^2 = 0.09$	$\hat{\sigma}_b^2 = 0.0178$	$\alpha\hat{\sigma}^2 = 0.1100$
(d)	$\phi(0) = 0.5$	$\hat{\theta} = 0.6302$	$\hat{\phi}(0) = 0.5621$
	$\sigma^2 = 1$	$\hat{\sigma}_\epsilon^2 = 1.2727$	$\hat{\sigma}^2 = 1.2150$
	$\alpha\sigma^2 = 0.16$	$\hat{\sigma}_b^2 = 0.0684$	$\alpha\hat{\sigma}^2 = 0.1785$
(e)	$\phi(0) = 0.2$	$\hat{\theta} = 0.2277$	$\hat{\phi}(0) = 0.2192$
	$\sigma^2 = 1$	$\hat{\sigma}_\epsilon^2 = 0.8922$	$\hat{\sigma}^2 = 0.9042$
	$\alpha\sigma^2 = 0.09$	$\hat{\sigma}_b^2 = 0.0006$	$\alpha\hat{\sigma}^2 = 0.0893$
(f)	$\phi(0) = 0.2$	$\hat{\theta} = 0.2131$	$\hat{\phi}(0) = 0.1896$
	$\sigma^2 = 1$	$\hat{\sigma}_\epsilon^2 = 1.1965$	$\hat{\sigma}^2 = 1.1833$
	$\alpha\sigma^2 = 0.16$	$\hat{\sigma}_b^2 = 0.1076$	$\alpha\hat{\sigma}^2 = 0.1712$



(a) Original Mumps data

(b) Stationary Mumps data

Figure 4.1. Time series plot.

4. 실증분석

실증분석에서는 한국 Mumps(유행성 이하선염) 자료를 이용하여 AR 모형과 RCA 모형에 각각 적합시켜 모수를 추정하고 예측값을 구하였으며 각 모형의 예측값을 이용하여 PRESS(Predicted Error of Sum of Square)를 비교하였다.

자료는 한국 질병관리본부에서 공개한 2005년 1주부터 2007년 46주까지 151개의 주별 한국 Mumps 발생건수를 사용하였다. 그 시계열도는 Figure 4.1의 (a)와 같고 원자료가 시간의 흐름에 따라 증가와 감소를 반복하다가 시계열의 값이 갑자기 커지는 경향이 있으므로 분산 안정화와 차분이 필요함을 알 수 있다. 자료는 로그변환을 통해 분산을 안정화 시키고, 차분을 통해 정상성을 만족하였으며 정상화된 Mumps 자료의 시계열도는 Figure 4.1의 (b)와 같다.

Table 4.1. Comparison of prediction for AR(1) and RCA(1)

	Mumps	AR(1)		RCA(1)	
		예측값	예측오차	예측값	예측오차
1	107	87.1803	19.8197	83.6416	23.3583
2	106	88.5067	17.4933	88.3922	17.6077
3	152	88.5921	63.4079	84.1530	67.8469
4	131	89.1745	41.8255	87.9142	43.0858
5	112	89.5627	22.4373	84.5600	27.4399
6	88	99.0312	-2.0312	87.5376	0.4623
7	60	90.4708	-30.4708	84.8835	-24.8835
8	64	90.9251	-26.9251	87.2407	-23.2408
9	50	91.3767	-41.3767	85.1404	-35.1405
10	52	91.8324	-39.8324	87.0065	-35.0065
		PRESS	11928.41	PRESS	11688.03

정상 자료로 변환한 Mumps 데이터를 이용하여 AR(1) 모형과 RCA(1) 모형에 각각 적합시켰다. 먼저 정상화된 Mumps 자료를 ACF와 PACF로 모형식별하여 AR(1) 모형에 적합시켰으며 최종모형은 다음과 같다.

$$Y_t = -0.3927Y_{t-1} + \epsilon_t, \quad \{\epsilon_t\} \sim IN(0, 0.1828). \quad (4.1)$$

자료의 비선형성을 고려하여 정상화된 Mumps 자료를 RCA(1) 모형에 적합시켜 보았으며, 자료 150개를 부표본 크기가 15인 10개의 부표본으로 나누어 2.2절의 Nicholls와 Quinn 모형의 추정을 이용하여 추정하였다. 추정된 모형은

$$\begin{aligned} Y_t &= \phi(t)Y_{t-1} + \epsilon_t, \quad \{\epsilon_t\} \sim IN(0, 0.1851), \\ \phi(t) &= -0.3131 + \eta_t, \quad \{\eta_t\} \sim IN(0, 0.0406) \end{aligned} \quad (4.2)$$

이며, $\hat{\alpha} = 0.2196$ 이다.

Mumps 자료의 AR(1) 모형과 RCA(1) 모형에 의한 10개의 예측값과 예측오차, 그리고 계산된 PRESS는 Table 4.1과 같다. PRESS를 비교해 보면 RCA(1) 모형의 PRESS 값이 더 작으므로, 선형모형인 AR 모형으로 적합하는 것보다 비선형모형인 RCA 모형으로의 적합이 더 바람직하다고 할 수 있다.

5. 결론

시계열 자료가 자기회귀 모형(AR)을 따르지만, 통제되지 않는 여러가지 조건들에 의해 비선형성을 갖는 경우 비선형 모형인 확률계수 자기회귀 모형(RCA)에 적합시키는 것을 고려해 볼 수 있다. RCA 모형의 확률계수의 특별한 경우인 $\phi(t) = \phi(0) + \eta$ 에 주안점을 두고 반복적인 절차를 통해 $\phi(t)$ 의 최소제곱추정법에 의한 정리와 그에 따른 증명을 보였다. 또한 전체 데이터를 부표본으로 나누어 초기값 $\phi(0)$ 와 σ^2 , 그리고 α 를 추정하였다. 모의실험을 통해 데이터를 부표본으로 나누어 모수를 추정하는 것이 더 바람직함을 보였다. 실증분석에서는 갑자기 발생건수가 증가하여 비선형성을 갖는 한국의 주별 Mumps 발생건수 자료를 사용하여 선형모형인 AR 모형과 비선형모형인 RCA 모형에 각각 적합시켜 각 모형의 PRESS 값을 비교하여 선형모형인 AR 모형보다 비선형 모형인 RCA 모형의 예측이 더 우수함을 보였다.

References

- Aue, A., Horvath, L., and Steinebach, J. (2004). Estimation in random coefficient autoregressive models, *Journal of Time Series Analysis*, **27**, 61–76.
- Hwang, S. Y. and Basawa, I. V. (1998). Parameter estimation for generalized random coefficient autoregressive processes, *Journal of Statistical Planning and Inference*, **68**, 323–337.
- Nicholls, D. F. and Quinn, B. G. (1981). The estimation of random coefficient autoregressive models II, *Journal of Time Series Analysis*, **2**, 185–203.
- Norli, A. A., Ibrahim, M., Shelton, P., and Nor, A. A. (2011). A new iterative procedure for estimation of RCA parameters based on estimating functions, *Mathematical Sciences*, **5**, 193–202.
- Rahiala, M. (1999). Random coefficient auto regressive model for longitudinal data, *Biometrika*, **86**, 718–722.
- Son, Y. S. (1996). Estimation of random coefficient AR(1) model for panel data, *Journal of the Korean Statistical Society*, **25**, 529–544.
- Tong, H. (1990). *Non-Linear Time Series*, Oxford University Press, Oxford.
- Wang, D. and Ghosh, S. K. (2002). Bayesian analysis of random coefficient autoregressive models, *Model Assisted Statistics and Applications*, **3**, 281–295.

확률계수 자기회귀 모형의 추정

김주성^a · 이성덕^{a,1} · 조나래^a · 함인숙^b

^a충북대학교 정보통계학과, ^b충북대학교 간호학과

(2015년 12월 31일 접수, 2016년 1월 3일 수정, 2016년 1월 3일 채택)

요약

비선형 모형인 확률계수 자기회귀 모형의 모수를 추정하기 위해 전체 데이터를 부표본으로 나누어 확률계수 $\phi(t)$ 가 초기값, $\phi(0)$ 를 갖는 특별한 경우를 제안하고 추정하였다. 모의 실험으로 부표본으로 나누어 확률계수 자기회귀 모형을 추정하는 더 바람직함을 확인하였다. 실증분석에서는 한국 Mumps 자료를 선형 모형인 자기회귀 모형과 확률계수 자기회귀 모형에 각각 적합시켜 모수를 추정하고, PRESS 값을 비교하여 확률계수 자기회귀 모형의 예측이 더 우수함을 보였다.

주요용어: 확률계수 자기회귀 모형, 부표본, 최소제곱추정법, Norli의 추정법, 한국 Mumps 자료

이 논문은 2013년도 충북대학교 학술연구지원사업의 연구비 지원에 의하여 연구되었음.

¹교신저자: (28644) 충북 청주시 서원구 충대로 1(개신동), 충북대학교 정보통계학과.

E-mail: sdlee@chungbuk.ac.kr