信頼性應用硏究
제3권, 제2호, pp. 127-136, 2003

오차항이 MA(1) 과정을 따르는 회귀모형에서의 Leverage*

이종협

성신여자대학교 통계학과

Leverage in Regression Models with MA(1) Errors

Jong Hyup Lee

Department of Statistics, Sungshin Women's University

Abstract

This paper investigates the effect of individual observations in regression models with MA(1) errors through the 'hat matrix'. It shows that the first observation has the largest hat matrix diagonal component for $\theta < 0$ in the regression model with an intercept. This provides additional evidence for retaining the first observation in performing estimation in this setting. When the regression model goes to the origin and the independent variable has a deterministic trend, the last observation has the greatest leverage for $|\theta| < 1$ and may have potentially large impact on parameter estimation.

* 이 논문은 2002년도 성신여자대학교 학술연구 조성비 지원에 의하여 연구되었음.

1. 서 론

Puterman(1988)은 오차항이 AR(1) 과정을 따르는 회귀모형에서 'hat matrix'와 다른 회귀 진단 통계량을 사용하여 추정에 있어서의 첫 번째 관측 값의 효과를 연구하였다. 절편만 있는 모형과 원점을 지나는 선형회귀 모형 하에서 양의 상관이 존재하는 경우 첫 번째 관측 값은 큰 값의 leverage를 가지며 추정에 상당한 영향을 미치므로 실제 자료 분석시 첫 번째 변환된 관측 값을 포함시키는 추정 방법의 사용을 주장하였다. 그러나 재정 분야의 자료에 적합한 모형으로 오차항이 MA(1) 과정을 따르는 회귀모형이 있다.

오차항이 AR(1) 과정을 따르는 선형 회귀모형에서 첫 번째 관측 값을 제거하는 Cochrane-Orcutt-Transformation(COT)는 응용분야에서 계산의 편리함과 모형의 단순화라는 측면에서 많이 사용되고 있다(Judge et. al., 1985). 그러나 이 경우 COT 추정량이 보통 최소제곱(OLS) 추정량에 비해 비효율적인 것이 밝혀졌다(Puterman, 1988). 오차항이 MA(1) 과정을 따르는 경우에도 Balestra(1980)는 계산의 편의를 위해 첫 번째 관측값을 제거하는 COT와 유사한 변환 방법을 제시하고 이를 이용하여 회귀계수에 대한 추정방법을 제시하였다. 그 이후 Park and Heikes(1983), Choudhury and Louis(1990), Choudhury(1994) 등은 Balestra 변환 방법을 수정 보완하여 보다 효율적인 추정방법을 제시하였다. 특히 Song, et al (1999)은 오차항에 대한 이동평균 모수에 대한 정보가 알려져 있지 않은 경우 일반화 최소제곱(GLS) 추정량, Balestra 변환 추정량, Park and Heikes 변환 추정량에 대한 OLS 추정량의 상대효율을 비교 분석하고 OLS 추정량의 사용이 효율적인 회귀모형을 제안하였다.

본 논문은 오차항이 AR(1) 과정을 따르는 회귀모형에서 첫 번째 관측 값이 상대적으로 큰 leverage를 가지고 있는 것으로 나타났는데 과연 오차항이 MA(1) 과정을 따르는 회귀모형에서도 동일한 현상이 일어나는 지를 다양한 회귀모형에서 조사한다. 이를 통해 기존에 알려진 첫 번째 관측 값을 제거한 회귀계수 추정방법이 근사적 추정방법으로 여전히 유용한 지를 진단 통계량인 leverage를 사용하여 다양한 θ 와 n 에 대해 그 효과를 연구하여 바람직한 추정방법을 제시한다.

2. 선형 회귀 모형과 추정량

다음과 같은 선형회귀 모형을 고려하자.

$$y = X\beta + e \tag{2.1}$$

여기서 y는 $n\times 1$인 반응 변수 벡터, X는 $n\times p$인 독립 변수 행렬이고 X의 계수 $= p \le n$, β는 $p\times 1$인 회귀 계수 벡터, 그리고 e는 $n\times 1$인 오차 벡터로 $E(e)=0$, $Cov(e)=\sigma^2 V$, $V\neq I$, V는 양정치 행렬이라 가정한다.

만일 오차항이 MA(1) 과정을 따른다면 모형 (2.1)은 다음과 같이 표현될 수 있다.

$$y_t = x_t'\beta + e_t, \quad t = 1, \cdots, n$$
$$e_t = a_t - \theta a_{t-1}, \quad |\theta| < 1, \quad a_t \sim \mathrm{IN}(0, \sigma_a^2) \tag{2.2}$$

모형 (2.1)에서 회귀 계수에 대한 GLS 추정량은 다음과 같다.

$$\hat{\beta}_{GLS} = (X'V^{-1}X)^{-1}X'V^{-1}y \tag{2.3}$$

여기서

$$V = \begin{pmatrix} 1+\theta^2 & -\theta & 0 & \cdots & 0 \\ -\theta & 1+\theta^2 & -\theta & \cdots & 0 \\ 0 & -\theta & 1+\theta^2 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & -\theta & 1+\theta^2 \end{pmatrix} \tag{2.4}$$

또한 역행렬 V^{-1}은 다음과 같다(Balestra, 1980).

$$\begin{aligned} V^{-1} &= A - \frac{\theta^2}{(1-\theta^{2n+2})(1-\theta^2)}(z_1z_1' + z_2z_2') \\ &+ \frac{\theta^{n+3}}{(1-\theta^{2n+2})(1-\theta^2)}(z_1z_1' + z_2z_2') \end{aligned} \tag{2.5}$$

여기서

$$A = \begin{pmatrix} 1 & \theta & \theta^2 & \cdots & \theta^{n-1} \\ \theta & 1 & \theta & \cdots & \theta^{n-2} \\ \theta^2 & \theta & 1 & \cdots & \theta^{n-3} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \theta^{n-1} & \theta^{n-2} & \theta^{n-3} & \cdots & 1 \end{pmatrix} \quad z_1 = (1\ \theta\ \cdots\ \theta^{n-1})' \quad z_2 = (\theta^{n-1}\ \theta^{n-2}\ \cdots\ 1)$$

식 (2.5)에서 V와 V^{-1}은 양정치 행렬이므로 V^{-1}은 다음과 같이 분해할 수 있다.

$$V^{-1} = Q_1'Q_1 \tag{2.6}$$

여기서

$$Q_1 = D^{-1/2}P, \qquad P = \begin{pmatrix} 1 & 0 & 0 & \cdots & 0 \\ \theta & c_1 & 0 & \cdots & 0 \\ \theta^2 & c_1\theta & c_2 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \theta^{n-1} & c_1\theta^{n-2} & c_2\theta^{n-3} & \cdots & c_{n-1} \end{pmatrix} \tag{2.7}$$

$$D = diag(c_0c_1, c_1c_2, c_2c_3, \cdots, c_{n-1}c_n), \quad c_s = 1 + \theta^2 + \cdots + \theta^{2s}, \quad c_0 = 1.$$

그러나 현실적으로 θ 값이 알려져 있지 않은 경우가 대부분이므로 GLS 추정량을 구할 수가 없기 때문에 이론적인 추정량에 불과한 경우가 대부분이다. 따라서 이동평균 모수 θ에 대한 좋은 일치 추정량을 계산하여 사용하는 것이 일반적이다. 다행히도 Pierce(1971)는 모형 (2.2)하에서 비선형 최소제곱 추정량이 θ의 일치추정량임을 증명하였으며, Judge et. al. (1985, 식(8.2.78))은 OLS 잔차에 근거한 θ의 추정량을 제시하였다.

3. Leverage

회귀 분석시 보통 최소제곱 가정 하에서 영향력이 큰 관측값을 발견하기 위해 사용되는 통계량으로 "hat matrix" $H = X(X'X)^{-1}X'$가 있다. 이에 대응되는 것으로 오차항이 MA(1) 과정을 따르는 회귀모형에서 hat matrix는 다음과 같이 주어진다.

$$H^* = X^*(X^{*'}X^*)^{-1}X^{*'} = Q_1X(X'Q_1'Q_1X)^{-1}X'Q_1'$$

H^*의 대각원소를 h^*_i라 놓자. 그러면 이 대각원소들은 어떤 특정 관측 값이 회귀 계수 추정치에 미치는 영향을 평가하는 진단 통계량으로 널리 사용되고 있다(Belsley et. al., 1980).

큰 값 h^*_i을 가지는 관측 값은 독립 변수에서 이상치이며 잠재적으로 회귀 계수 추정치에 큰 영향을 주는 것으로 알려졌다. 어떤 상황하에서 어느 관측 값이 큰 leverage를 가지며 결과적으로 이를 제거하는 것이 모수 추정에 상당한 영향을 주는 지를 살펴보고자 한다. 우선 오차항의 일차 이동평균 모수가 알려져 있다고 가정하자. 그리고 모수 추정시 모든 관측값을 사용하는 GLS 방법을 사용하기로 하자.

3.1 절편항만 있는 회귀모형

모형 (2.1)에서 X가 단지 절편항만을 포함하는 경우, 즉 $X = (1, 1, \cdots, 1)'$인 모형을 고

려해 보자. 왜냐하면 실제 이런 모형을 따르는 자료가 종종 발견되며, H^* 행렬의 대각원소를 정확히 표현할 수 있으며 민감도를 조사할 모수의 수가 적기 때문이다.

$$\begin{aligned} y_t &= \beta_0 + e_t \\ e_t &= a_t - \theta a_{t-1}, \quad t=1,2,\cdots,n. \end{aligned} \tag{3.1}$$

모형(3.1)의 경우 GLS 추정 방법에 의한 H^*의 대각 원소는 다음과 같다.

$$h_i^* = \frac{[(1-\theta^{2i}) + \sum_{k=1}^{i-1}(1-\theta^{2k})\theta^{i-k}]^2/[(1-\theta^{2i})(1-\theta^{2i+2})]}{\sum_{j=1}^{n}[\{(1-\theta^{2j}) + \sum_{k=1}^{j-1}(1-\theta^{2k})\theta^{j-k}\}^2/\{(1-\theta^{2j})(1-\theta^{2j+2})\}]} \tag{3.2}$$

< 표 1 > 모형 (3.1)하에서의 Leverage

	$n=20$			$n=30$			$n=50$			$n=100$		
θ	h_1^*	r_m	r_a	h_1^*	r_m	r_a	h_1^*	r_m	r_a	h_1^*	r_m	r_a
-.9	.10	1.51	2.00	.06	1.51	2.00	.04	1.51	2.00	.02	1.51	2.00
-.7	.09	1.55	1.95	.06	1.55	1.95	.04	1.55	1.94	.02	1.55	1.94
-.5	.09	1.59	1.81	.06	1.59	1.81	.04	1.59	1.80	.02	1.57	1.80
-.3	.08	1.49	1.56	.05	1.49	1.56	.03	1.49	1.55	.02	1.49	1.55
-.1	.06	1.20	1.20	.04	1.20	1.20	.02	1.20	1.20	.01	1.20	1.20
	h_n^*	r_m	r_a	h_n^*	r_m	r_a	h_n^*	r_m	r_a	h_n^*	r_m	r_a
.0	.05	1.00	1.00	.03	1.00	1.00	.02	1.00	1.00	.01	1.00	1.00
.1	.05	1.00	1.01	.03	1.00	1.01	.02	1.00	1.01	.01	1.00	1.00
.3	.05	1.00	1.05	.03	1.00	1.03	.02	1.00	1.02	.01	1.00	1.01
.5	.06	1.00	1.12	.04	1.00	1.07	.02	1.00	1.04	.01	1.00	1.02
.7	.07	1.00	1.32	.04	1.00	1.19	.02	1.00	1.11	.01	1.00	1.05
.9	.11	1.05	2.33	.06	1.02	1.97	.03	1.00	1.54	.01	1.00	1.22

여기서 $r_m = h_n^*/h^*_{\max}$ 또는 $h_1^*/h^*_{\max}$; $h^*_{\max}$ 는 행렬 H^*에서 h_n^* 또는 h_1^* 을 제외한 대각원소 들 중 최대값을 의미하며, $r_a = h_1^*/h_{ave}$ 또는 h_n^*/h_{ave} ; h_{ave} 는 행렬 H^*에서 h_n^* 또는 h_1^* 을 제외한 대각원소들의 평균값을 의미함.

<표 1>에서 양의 상관 관계가 존재하는 경우 즉, $\theta<0$ 경우 첫 번째 관측 값이 가장 큰 leverage 값을 가지고 있으며, 이는 오차항의 일차 자기상관 계수가 1/3보다 큰 경우 첫 번째 관측 값이 큰 leverage 값이 걸리고 이는 모수 추정치에 큰 영향을 미친다는 Putterman(1988)의 결과와 일치한다. 또한 이는 다른 어떤 관측 값보다 1.5배 정도 큰 값을 가지고 있으며 θ 가 -1에 가까워짐에 따라 다른 모든 h_j^*들의 평균값보다 2배 만큼 큼을 알 수 있다. 따라서 $\theta<0$일 때 Balestra(1980) 가 제안한 첫 번째 관측 값을 제거한 추정 방법을 적용하는 데 주의를 기울여야 한다 반면에 $\theta\geq 0$에 대해서는 마지막 관측 값의 leverage h^*_n가 가장 큰 값을 가지고 있지만 r_m 이나 r_a 가 거의 1에 가까워 leverage 값이 거의 유사하다고 볼 수 있다.

3.2 원점을 지나는 선형추세 회귀모형

모형 (2.1)에서 X가 원점을 지나는 선형추세, 즉 $X=(1,2,\cdots,n)'$인 모형을 고려해 보자.

$$\begin{aligned} y_t &= \beta_1 x_t + e_t, \quad x_t = t, \quad t=1,2,\cdots,n \\ e_t &= a_t - \theta a_{t-1} \end{aligned} \tag{3.3}$$

모형(3.3)의 경우 GLS 추정 방법에 의한 H^*의 대각 원소는 다음과 같다.

$$h_i^* = \frac{[\,i(1-\theta^{2i}) + \sum_{k=1}^{i-1} k(1-\theta^{2k})\theta^{i-k}]^2/[\,(1-\theta^{2i})(1-\theta^{2i+2})]}{\sum_{j=1}^{n}[\,\{j(1-\theta^{2j}) + \sum_{k=1}^{j-1} k(1-\theta^{2k})\theta^{j-k}\}^2/\{(1-\theta^{2j})(1-\theta^{2j+2})\}]} \tag{3.4}$$

식(3.4)를 이용하여 leverage 를 계산한 결과가 <표 2>에 주어져 있는데, 모든 θ와 n에 대해서 마지막 관측값이 가장 큰 leverage 값을 가지고 있으며, h^*_i는 i가 커짐에 따라 증가하는 단조 증가함수임을 보이고 있다.

< 표 2 > 모형 (3.3) 하에서의 Leverage

	$n=20$			$n=30$			$n=50$			$n=100$		
θ	h^*_n	r_m	r_a	h^*_n	r_m	r_a	h^*_n	r_m	r_a	h^*_n	r_m	r_a
-.9	.14	1.11	3.00	.09	1.07	3.00	.06	1.04	3.00	.03	1.02	3.00
-.7	.14	1.11	3.01	.09	1.07	3.01	.06	1.04	3.00	.03	1.02	3.00
-.5	.14	1.11	3.02	.09	1.07	3.02	.06	1.04	3.01	.03	1.02	3.00
-.3	.14	1.11	3.03	.09	1.07	3.03	.06	1.04	3.02	.03	1.02	3.01
-.1	.14	1.11	3.06	.09	1.07	3.04	.06	1.04	3.02	.03	1.02	3.01
.0	.14	1.11	3.08	.10	1.07	3.05	.06	1.04	3.03	.03	1.02	3.02
.1	.14	1.11	3.10	.10	1.07	3.06	.06	1.04	3.04	.03	1.02	3.02
.3	.14	1.11	3.15	.10	1.07	3.10	.06	1.04	3.06	.03	1.02	3.03
.5	.15	1.11	3.25	.10	1.07	3.16	.06	1.04	3.09	.03	1.02	3.05
.7	.16	1.12	3.50	.10	1.08	3.32	.06	1.04	3.18	.03	1.02	3.09
.9	.19	1.18	4.56	.12	1.10	4.13	.07	1.05	3.67	.03	1.02	3.31

<표 2>에 의하면 모든 $|\theta|<1$에 대해 h^*_n이 가장 큰 값을 보이고 있으므로 계산상의 편의를 위해 회귀계수 추정방법으로 아래에 제시된 Balestra(1980)의 변환 추정량 (3.5) 또는 Park and Heikes(1983) 등의 추정량을 사용하여도 무방하다. 또한 모든 $|\theta|<1$에 대해 h^*_1 은 거의 0에 가까워 회귀계수의 추정에 거의 영향을 미치지 않는다고 할 수 있다.

AR(1) 오차를 가지고 있는 회귀모형에서 흔히 사용되는 COT 추정량처럼 식(2.7)의 Q_1 행렬의 첫 번째 행을 제거한 후 Balestra(1980)은 다음과 같은 근사적 변환 모형을 제안하였다.

$$y^o = X^o\beta + e^o$$

여기서 $y^o = Q_2 y$, $X^o = Q_2 X$, $e^o = Q_2 e$ 이며 Q_2는 $(n-1)\times n$인 행렬로 다음과 같다.

$$Q_2 = \begin{pmatrix} \theta & 1 & 0 & 0 & \cdots & 0 \\ \theta^2 & \theta & 1 & 0 & \cdots & 0 \\ \theta^3 & \theta^2 & \theta & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ \theta^{n-1} & \theta^{n-2} & \theta^{n-3} & \theta^{n-4} & \cdots & 1 \end{pmatrix}$$

위 근사적 변환 모형에 보통최소제곱법을 적용하여 구한 추정량은 다음과 같다.

$$\tilde{\beta}_{BT} = (X^{o\prime}X^o)^{-1}X^{o\prime}y^o = (X'Q_2'Q_2X)^{-1}X'Q_2'Q_2y \tag{3.5}$$

3.3 절편과 선형 추세가 있는 회귀모형

모형 (2.1)에서 X가 단지 절편항만을 포함하는 경우, 즉 $X=\begin{pmatrix} 1 & 1 & \cdots & 1 \\ 1 & 2 & \cdots & n \end{pmatrix}'$ 인 모형을 고려해 보자. 즉

$$\begin{aligned} y_t &= \beta_0+\beta_1 t+e_t \\ e_t &= a_t-\theta a_{t-1}, \quad t=1,\cdots,n \end{aligned} \tag{3.6}$$

모형 (3.6)하에서의 관측값에 대한 영향력 평가를 위한 hat matrix의 대각원소 h_i^*를 GLS 방법을 적용한 경우

$$\begin{aligned} h_i^* = \{ & \delta_2 \,[(1-\theta^{2i})+\sum_{k=1}^{i-1}(1-\theta^{2k})\theta^{i-k}]^2 + \delta_1\,[i(1-\theta^{2i})+\sum_{k=1}^{i-1}k(1-\theta^{2k})\theta^{i-k}]^2 \\ & -2\,\delta_3\,[(1-\theta^{2i})+\sum_{k=1}^{i-1}(1-\theta^{2k})\theta^{i-k}][i(1-\theta^{2i})+\sum_{k=1}^{i-1}k(1-\theta^{2k})\theta^{i-k}]\}/ \\ & \{(\delta_1\delta_2-\delta_3^2)(1-\theta^{2i})(1-\theta^{2i+2})\} \end{aligned} \tag{3.7}$$

이 된다. 여기서

$$\delta_1=\sum_{j=1}^{n}\frac{[(1-\theta^{2j})+\sum_{k=1}^{j-1}(1-\theta^{2k})\theta^{j-k}]^2}{(1-\theta^{2j})(1-\theta^{2j+2})},$$

$$\delta_2=\sum_{j=1}^{n}\frac{[j(1-\theta^{2j})+\sum_{k=1}^{j-1}k(1-\theta^{2k})\theta^{j-k}]^2}{(1-\theta^{2j})(1-\theta^{2j+2})}$$

$$\delta_3=\sum_{j=1}^{n}\frac{[(1-\theta^{2j})+\sum_{k=1}^{j-1}(1-\theta^{2k})\theta^{j-k}][j(1-\theta^{2j})+\sum_{k=1}^{j-1}k(1-\theta^{2k})\theta^{j-k}]}{(1-\theta^{2j})(1-\theta^{2j+2})}.$$

식(3.7)을 이용하여 다양한 n 과 θ 값에 대해 가장 큰 leverage값을 가지는 관측 값과 두 번째로 큰 leverage 값을 가지는 관측 값에 대한 상대적 크기가 <표3>에 나타나 있다.

< 표 3 > 모형 (3.6)하에서의 Leverage

	$n=20$			$n=30$			$n=50$			$n=100$		
θ	h_1^*	r_m	r_a	h_1^*	r_m	r_a	h_1^*	r_m	r_a	h_1^*	r_m	r_a
-.9	.33	1.84	3.76	.23	1.85	3.85	.15	1.70	3.91	.08	1.60	3.95
-.7	.33	1.82	3.70	.23	1.86	3.77	.14	1.76	3.82	.07	1.65	3.85
-.5	.31	1.71	3.46	.22	1.74	3.52	.14	1.76	3.55	.07	1.69	3.58
-.3	.27	1.50	2.99	.19	1.51	3.03	.12	1.53	3.06	.06	1.54	3.08
-.1	.22	1.18	2.32	.15	1.19	2.35	.09	1.19	2.37	.05	1.19	2.38
.0	.19	1.17	1.94	.13	1.11	1.96	.08	1.06	1.98	.04	1.03	1.99
	h_n^*	r_m	r_a	h_n^*	r_m	r_a	h_n^*	r_m	r_a	h_n^*	r_m	r_a
.1	.19	1.16	1.97	.13	1.11	1.98	.08	1.06	1.98	.04	1.03	1.99
.3	.19	1.18	2.03	.13	1.11	2.02	.08	1.06	2.01	.04	1.03	2.01
.5	.20	1.19	2.16	.14	1.12	2.11	.08	1.07	2.06	.04	1.03	2.03
.7	.23	1.23	2.47	.15	1.13	2.31	.09	1.07	2.18	.04	1.03	2.09
.9	.30	1.35	3.30	.20	1.20	3.14	.11	1.09	2.81	.05	1.04	2.41

$\theta \leq 0$ 일 때 h_1^*가 가장 큰 leverage 값을 가지고 있으며 이는 모수 추정치에 큰 영향을 미친다는 Puterman(1988)의 결과와 동일하다고 할 수 있다. 따라서 식(3.5)에 의한 Balestra 추정 방법을 적용하는 데 주의를 기울여야 한다. θ 가 -1에 가까워질수록 h_1^*가 다른 leverage의 평균값보다 약 4배정도 큰 값을 가지고 있으며 n이 증가함에 따라 이 값이 약간 커지고 있다. 또한 다른 어떤 관측값보다 약 2배 정도 큼을 알 수 있다. 반면에 $\theta > 0$인 경우 최대값은 마지막 관측값에 나타나는 데 θ가 +1에 갈수록 r_a는 증가하는 패턴을 보이고 있다. 표 3에서는 나타나지 않지만 $0 < \theta \leq 0.5$의 경우 leverage 값은 중간시점을 저점으로 증가하는 패턴을 보이고 있었다. 또한 Belsley et al(1980)의 기준으로 판단하면 h_1^* 또는 h_n^* 이 영향력이 큰 관측 값이라 할 수 있기에 회귀계수의 추정치에 영향을 미치므로 미래값 예측시 새로운 관찰 값이 출현하면 이를 포함시켜 모형을 재 추정하여 예측함이 좋을 듯 하다.

4. 결 론

MA(1) 오차를 가지는 절편만 있는 회귀모형에 대해 $-1<\theta<0$에 대해 첫 번째 관측 값을 제거하는 Balestra(1980) 변환 추정량을 사용하는데 주의를 기울여야 하며, 이 경우에는 Park and Heikes(1983), Choudhury and Louis(1990), Choudhury(1994) 등이 제안한 추정량이 바람직한 것으로 보인다. 특히 $|\theta|$ 가 그리 크지 않은 경우 Song et al(1999)의 결과가 이를 뒷받침하고 있다. $0<\theta<1$에 대해서는 본 논문에서 고려한 모든 회귀모형에서 h^*_n가 가장 큰 leverage 값을 가지고 있으므로 AR(1) 경우의 COTE와 마찬가지로 Balestra 추정량을 사용해도 무방할 것이다. 특히 추세가 있는 회귀모형에서 마지막 관측 값이 가장 큰 leverage 값을 가지고 있으며, 이는 잠재적으로 영향력이 큰 관측 값으로 간주될 수 있기 때문에 새로운 관찰값이 얻어지면 회귀계수를 재 추정하여 예측함이 바람직하다.

참고문헌

[1] Balestra, P. (1980). A note on the exact transformation associated with the first order moving average process. *Journal of Econometrics,* 14, 381-394.

[2] Belsley, D.A., Kuh, E. and Welsch, R.E. (1980). *Regression Diagnostics: Identifying Influential Data and Sources of Collinearity,* John Wiley & Sons, New York.

[3] Choudhury, A.H. and St. Louis, R.D. (1990). A note on Park and Heikes'(1983) modified approximate estimator for the first order moving average process. *Journal of Econometrics*, 46, 399-406.

[4] Choudhury, A.H. (1994). Untransformed first observation problem in regression model with moving average process. *Communications in Statistics, Theory and Method,* 23, 2297-2937.

[5] Judge, G.G., Griffiths, W.E., Hill, R.C., Lutkepohl, H. and Lee, T.C. (1985). *The Theory and Practice of Econometrics*. John Wiley, New York.

[6] Park, C.Y. and Heikes, R.G. (1983). A note on Balestra's(1980) approximate estimator for the first order moving average process. *Journal of Econometrics*, 21, 387-388.

[7] Pierce, D.A. (1971), "Distribution of residual autocorrelations in the regression model with autoregressive-moving average errors," *Journal of the Royal Statistical Society*, Series B, 33, 140-146.

[8] Puterman, M.L. (1988), "Leverage and influence in autocorrelated regression models," *Applied Statistics,* 37(1), 76-86.

[9] Song, S.H., Lee, J.H. and Kim, K.W. (1999). Efficient estimation of regression coefficients in regression model with moving average process. *The Korean Journal of Applied Statistics,* 12(1), 109-124.