

자기회귀모형에서의 로버스트한 모수 추정방법들에 관한 연구¹

강희정² · 김순영³

요약

본 논문에서는 가장 많이 사용되는 시계열 모형중의 하나인 자기회귀모형에서 모수를 추정하는 방법으로 최소 절대 편차 추정법(least absolute deviation estimation)을 포함한 로버스트한 추정방법(robust estimation)의 사용을 제안하고 모의 실험을 통하여 이러한 방법들을 기존의 최소 제곱 추정 방법과 예측의 관점에서 비교·검토하여 시계열 자료분석에서의 로버스트한 모수 추정 방법의 유효성을 확인해 보고자 한다.

주제어: 자기회귀모형, 최소절대편차 추정법, 로버스트 추정법, 최소제곱 추정법, 예측 오차.

1. 서론

관측된 시계열 자료 $\{X_t\}$ 가 다음과 같은 p 차의 자기 회귀모형 (p -th order autoregressive process)을 따른다고 가정하여 보자.

$$X_t = \alpha' Y_{t-1} + \epsilon_t, \quad t = 0, \pm 1, \pm 2, \dots \quad (1.1)$$

여기서, ϵ_t 는 서로 독립이고 분산이 유한하며 동일한 연속형 분포함수 $F(\cdot)$ 를 따르는 오차항을 나타내며 $\alpha = (\alpha_1, \dots, \alpha_p)'$, $Y_{t-1} = (X_{t-1}, \dots, X_{t-p})'$ 이고 p 차의 자기 회귀모형의 특성 방정식(characteristic equation) $\phi(B) = 1 - \alpha_1 B - \dots - \alpha_p B^p$ 의 모든 근은 단위원(unit circle)밖에 존재함을 가정하여 시계열 $\{X_t\}$ 가 정상성(stationary)조건을 만족시킨다고 가정한다.

¹ 이 논문은 1998년 한국학술진흥재단의 과학기술기초 중점연구지원과제 (기초과학연구소, 1998-015-D00048)의 일부 지원으로 수행되었음.

² 전북 전주시 덕진구 덕진동 1가 664-14, 전북대학교 자연과학대학 수학·통계정보과학부, 기초과학연구소, 조교수

³ 전북 전주시 덕진구 덕진동 전북대학교 자연과학대학 전산통계학과 박사과정

가정된 위의 시계열 모형 (1.1)을 완성시키기 위해서는 모수 벡터 α 를 자료로부터 추정해야 하는데 이때 추정치들은 다음 식 (1.2)을 최소화하는 방법으로 추정되어 진다고 말할 수 있다.

$$\mathbf{S}(\alpha) = \sum_{t=1}^n L(\epsilon_t) = \sum_{t=1}^n L(X_t - \alpha' \mathbf{Y}_{t-1}) \quad (1.2)$$

여기서 $L(\cdot)$ 은 오차항에 대한 대칭함수(symmetric function)이며 손실함수(loss function)라고 불리운다. 위의 식 (1.2)를 최소화 하는 식은 다음과 같이 표현할 수 있고 모수의 추정치는 이 식의 해(solution)이기도 하다.

$$\sum_{t=1}^n \mathbf{g}(\mathbf{Y}_{t-1}) \varphi(X_t - \alpha' \mathbf{Y}_{t-1}) = 0 \quad (1.3)$$

여기서, $\varphi(\cdot) = L'(\cdot)$ 은 $t\varphi(t) > 0$ 인 함수이고 일반적으로 $\varphi'(0) = 1$ 임을 가정하며 $\mathbf{g}(\cdot)$ 도 또한 어떤 함수임을 가정한다. 각각의 모수 추정방법은 손실함수 $L(\cdot)$ 의 선택에 따라 결정되며 이때 추정치 $\hat{\alpha}$ 의 의미도 또한 달라지게 된다.

최대 우도 추정법(maximum likelihood estimation)은 전통적으로 가장 많이 사용하는 추정법으로 최소 제곱 추정법(least squares estimation)과 더불어 통계학의 여러 분야에서 가장 유용하게 널리 사용되고 있는 일반적인 방법이다. 그러나, 최대 우도 추정법은 항상 모형속의 오차항에 대한 분포를 가정해야하기 때문에 실용적이지 못하다는 단점이 있으나 점근적인 정규분포성(normality)과 유효성(efficiency) 이라는 좋은 성질을 가지고 있다(Brockwell and Davis(1991)).

그림 1은 1차의 자기 회귀모형 (AR(1))을 따르는 시계열 $\{X_t\}$ 에 대하여 임의의 시점 t 에서 x -축은 X_{t-1} 값을 y -축은 X_t 값을 나타내도록 그린 그림으로써 그림에서 실선(-)은 원래 함수식이고 점선(···)은 최소 제곱 방법에 의한 적합식이며 (- · -)은 최소 절대 편차 추정법에 의한 적합식이다. 그림에서 보는 바와 같이 분산이 정규분포보다 큰 이중 지수 분포(double exponential distribution)인 경우에는 최소 제곱 추정법에 의한 적합식보다 최소 절대 편차 추정법에 의한 적합식이 원래 함수식에 근접하여 최소 절대 편차 추정법이 최소 제곱 추정법보다 좋은 추정방법임을 알 수 있다. 그림 2는 AR(1) 모형에서 오차항이 오염된 정규분포(contaminated normal distribution)를 따르는 경우를 나타낸 것으로서 실선(-)은 원래 함수식이고 점선(···)은 최소 제곱 방법에 의한 적합식이며 (- · -)은 로버스트한 추정법의 하나인 GM-추정법(generalized M-estimation)에 의한 적합식이다. 그림에서 보는 바와 같이 관측치에 이상치(outlier)가 많이 포함되어 있는 경우에는 최소 제곱 추정법에 의한 적합식보다 로버스트한 추정법에 의한 적합식이 원래 함수식에 근접하여 로버스트한 추정방법이 최소 제곱 추정방법보다 좋은 추정방법임을 알 수 있다. 따라서 비정규성이라든가 이상치에 의하여 영향을 받지않는 로버스트한 모수 추정방법이 시계열 자료 분석에서 필요하게 됨을 우리는 알 수 있다.

본 논문의 목적은 회귀분석등의 영역에서 주로 사용되어 왔던 최소 절대 편차 추정

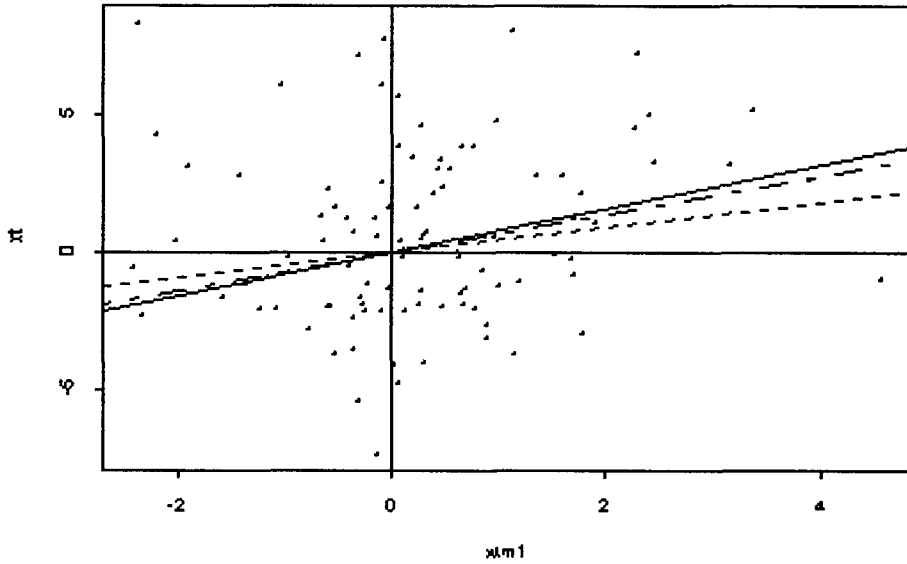


그림 1: 오차항이 이중 지수 분포를 따르는 경우

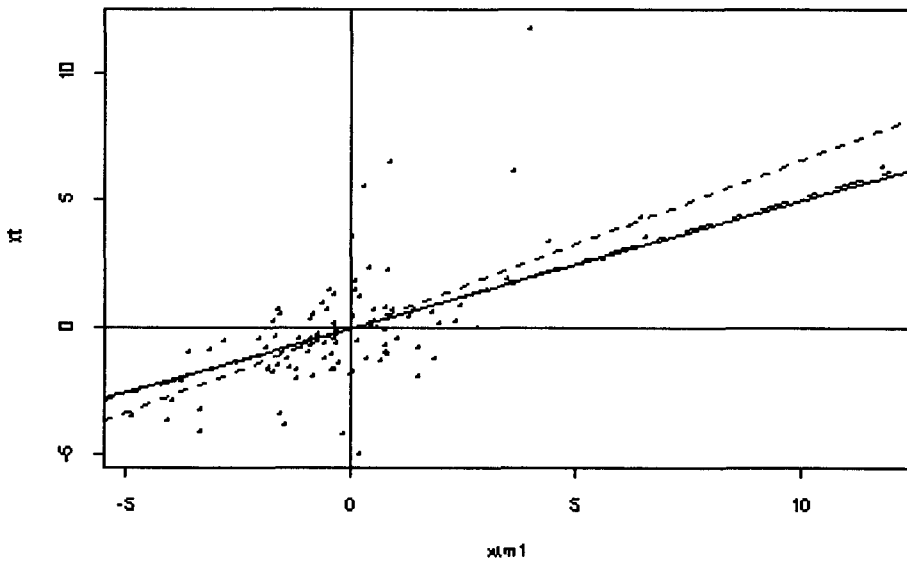


그림 2: 오차항이 오염된 정규 분포를 따르는 경우

법(least absolute deviation estimation) 과 로버스트한 추정법(robust estimation)의 개념을 시계열 모형중의 하나인 자기 회귀모형에서의 모수 추정방법으로 적용시켜 보고 이러한 로버스트한 추정 방법들과 기존의 많이 사용되었던 최대 우도 추정법(maximum likelihood estimation), 최소제곱 추정법(least squares estimation)등을 모의 실험을 통하여 예측의 관점에서 비교·평가하여 이들 로버스트한 추정방법들의 유효성을 확인하는데 있다.

시계열 모형에서의 모수 추정방법으로 최소 절대 편차 추정량에 대한 성질과 계산방법(algorithm)등에 관한 대표적인 연구 성과로는 Bloomfield and Steiger(1983)등이 있으며 로버스트한 GM-추정법을 시계열 분석에 적용하고자 하는 대표적인 이론 성과로는 Bustos(1982)등이 있다. 최근의 연구 결과로써는 Koul and Zhu(1995)가 발표한 일반적인 $AR(p)$ 모형에 대한 최소 절대 편차 추정량과 GM-추정량에 대한 강한 근사식(strong representation) 등이 있으며, 국내 학회에서는 Kang and Shin(1998)이 발표한 1차의 $AR(1)$ 모형에서의 LAD 추정량에 대한 강한 근사식과 차경엽 등(1999)이 발표한 M-추정량에 대한 극한 분포(Asymptotic distribution) 등이 있다.

제 2절에서는 자기회귀모형을 위하여 사용되어질 수 있는 로버스트한 모수 추정방법들과 이러한 추정 방법들을 비교하여 평가할 기준들에 대해 살펴 보고 제 3절에서는 이러한 모수 추정 방법들의 비교를 위한 모의 실험의 설계와 절차에 대해 설명하고 또한 이를 통한 결과들을 분석하고자 한다.

2. 로버스트한 모수 추정 방법들

이 절에서는 자기 회귀 모형을 위한 로버스트한 모수 추정방법들에 대해 알아 보고 이러한 방법에 대한 이론적 성과를 살펴 보고자 한다.

2.1 최소 절대 편차 추정법(Least Absolute Deviation Estimation)

최소 절대 편차 추정법은 오차항의 절대값의 합을 최소로 하는 $\hat{\alpha}_{LAD}$ 값을 찾는 방법으로써 앞서의 식 (1.2)와 식 (1.3)에서 $L(t) = |t|$ 이며 $\mathbf{g}(t) = t$, $\varphi(t) = \text{sign}(t)$ 인 경우이다. 여기서, $\text{sign}(t)$ 은 $t < 0$ 이면 -1 이거나 $t > 0$ 이면 1 값을 가지는 함수이다. 다시 말해서, $\hat{\alpha}_{LAD}$ 는 다음과 같은 식 (2.1)를 만족시키는 해이다.

$$\sum_{t=1}^n |X_t - \hat{\alpha}'_{LAD} \mathbf{Y}_{t-1}| = 0. \quad (2.1)$$

이러한 최소 절대 편차 추정방법에 대한 개념은 최소 제곱 방법보다 훨씬 이전에 제안되었으나 추정량들의 비확정성(non-explicit form)문제로 인한 계산상의 어려움 때문에 최근까지 거의 관심을 끌지 못하여 왔다.

최근 컴퓨터의 계산 능력이 향상됨에 따라 정규분포의 가정이 맞지 않는 현상에 대한 통계적 추정방법이 다시 관심을 받기 시작한 바, 이러한 시도의 하나가 자기 회귀모형에서

의 오차항에 대한 비정규성 가정을 하게 되는 경우이다. 만일 오차항의 분포가 이중 지수 분포일 경우에는 최소 절대 편차 추정법은 최대 우도 추정법과 일치하게 되어 최소 절대 편차 추정법이 최대 우도 추정치의 점근적 정규분포성과 유효성의 좋은 성질을 가지게 된다. 따라서 오차항이 이중 지수 분포인 경우에는 최소 절대 편차 추정법이 최적의 추정방법이라 할 수 있어 최소제곱 추정법이 항상 최적의 추정방법이 아니라는 것을 알 수 있다.

p 차의 자기 회귀모형에서의 최소 절대 편차 추정량에 대한 연구의 대표적인 이론 성과는 Bloomfield and Steiger(1983)등이 발표한 LAD 추정량에 대한 알고리즘 연구 결과와 Koul and Zhu(1995) 등이 발표한 최소 절대 편차 추정량에 대한 강한 근사식(strong representation)인 다음의 정리 2.1이 있으나 아직까지 LAD 추정량에 대표본 분포는 해결되지 않고 있는 실정이다.

정리 2.1 (Koul and Zhu(1995)) 추정량 $\hat{\alpha}_{LAD}$ 에 대한 강한 근사식은, 다음의 가정들 (1) 분포함수 F 는 0에서 유일한 중앙값(median)을 가진다. (2) 유계(bounded)인 확률밀도함수 f 는 0의 주위에서 연속이다. (3) 어떤 $\tau > 8$ 에 대하여 $E(|\epsilon_t|^\tau) < \infty$ 을 만족하면, 확률 1로써

$$-2f(0)n^{-1/2} \sum_{t=1}^n \mathbf{Y}_{t-1} \mathbf{Y}'_{t-1} (\hat{\alpha}_{LAD} - \alpha) = -n^{-1/2} \sum_{t=1}^n \mathbf{Y}_{t-1} \text{sign}(\epsilon_t) + R_n$$

이 성립한다. 여기서 $\|R_n\| = O(n^{-1/4}(\ln n)^{3/4})$ 이다.

2.2 로버스트한 추정법(Robust Estimation)

로버스트 추정 방법의 기본 개념은 오차항의 분포에 대한 가정이나 이상치에 의한 영향을 받지 않도록 하는 것으로 이 추정법은 M-추정법과 이를 좀더 일반화한 GM-추정법(Generalized M-estimation)으로 나누어 생각할 수 있다. 이러한 로버스트 추정방법은 계산량이 많아 추정치의 계산이 쉽지 않고 시계열 자료의 고유의 성질인 관측치들간의 종속성 때문에 시계열 자료 분석에서는 별로 사용되지 않았던 방법이다.

2.2.1 M-추정법(Minimized Estimation)

Huber의 M-추정법은 식 (1.3)에서 $\mathbf{g}(t) = t$ 이고 $\varphi(t)$ 는 적당히 유계(bounded)인 함수인 경우로써 다음의 식 (2.2)을 만족하는 방정식의 해 $\hat{\alpha}_M$ 가 모형 (1.1)에 있는 모수 α 의 M-추정량이 된다.

$$\sum_{t=1}^n \mathbf{Y}_{t-1} \varphi(X_t - \hat{\alpha}'_M \mathbf{Y}_{t-1}) = 0 \tag{2.2}$$

여기서 $\varphi(\cdot)$ 함수는 적당히 유계인 함수로써 여러가지 형태의 함수를 생각할 수 있으며 그 역할은 굉장히 큰 이상치의 영향력을 조절하는 기능을 한다. 그러나 이 추정법은 제1절의 그림 1 또는 그림 2에서 y -축방향으로만의 이상치(vertical outlier)의 영향력을 제한하고 있

으나 x -축방향으로의 지렛대점(leverage point)의 영향력은 제한하지 못한다는 약점을 가지고 있기도 하다(Staudte and Sheather(1990)).

p 차의 자기 회귀모형에서 M -추정량 $\hat{\alpha}_M$ 에 대한 이론적인 성과로는 Denby and Martin(1979)와 Martin(1982)등이 있으며 이들은 각각 유한 분산의 조건하에서 $AR(1)$ 모형과 $AR(p)$ 모형에서의 $\hat{\alpha}_M$ 의 대표본 분포를 구해 내었다.

2.2.2 일반화 M -추정법(Generalized M -estimation)

GM-추정법은 일반적인 M -추정법의 형태로 M -추정법이 가지고 있는 단점을 극복하기 위하여 y -축방향의 이상치의 영향력 뿐만 아니라 x -축 방향으로의 지렛대점(leverage point)의 영향력도 제한하는 로버스트한 모수 추정 방법이다. 이러한 GM-추정법은 식 (1.2)에 있는 $L(\cdot)$ 함수의 형태에 따라 Mallows type과 Schweppe type(Staudte and Sheather(1990))으로 나누어 생각할 수 있는데 본 논문에서는 Mallows type을 사용하기로 한다. 따라서 식 (2.2)은 다음의 식 (2.3)이 되며 이 방정식의 해를 식 (1.1)에 있는 모수 α 의 GM-추정량 $\hat{\alpha}_{GM}$ 이라 한다.

$$\sum_{t=1}^n \mathbf{g}(\mathbf{Y}_{t-1})\varphi(X_t - \hat{\alpha}'_{GM}\mathbf{Y}_{t-1}) = 0. \quad (2.3)$$

여기서 $\mathbf{g}(\mathbf{Y}_{t-1})$ 은 x -축방향의 영향력을 제한하는 기능을 가지고 있는 유계인 함수이다. 또한, GM-추정법은 위의 식 (2.3)에서의 함수 $\varphi(\cdot)$ 의 선택에 따라 Huber-type 추정법과 Biweight-type 추정법등의 여러가지 방법으로 구분되어지는데 본 논문에서는 다음과 같은 Huber-type 함수와 Tukey의 Biweight 함수만을 고려하기로 한다.

· Huber-type 함수

$$\varphi_H(t) = \begin{cases} t, & |t| \leq c_h, \\ c_h \cdot \text{sign}(t), & |t| > c_h. \end{cases}$$

· Tukey의 Biweight 함수

$$\varphi_B(t) = \begin{cases} t(1 - (t/c_b)^2)^2, & |t| \leq c_b, \\ 0, & |t| > c_b. \end{cases}$$

여기서 c_h 와 c_b 는 조절상수(tuning constant)라 불리우는 값이며 이 상수값들은 GM-추정량의 효율성(efficiency)를 결정하는데 중요한 역할을 한다. 본 논문에서는 Denby and Martin(1979)에서 얻은 $AR(1)$ 모형을 위한 GM-추정량의 최소 제곱 추정치에 대한 이론적 상대 효율성(relative efficiency)을 유지하기 위하여 $\mathbf{g}(t)$ 와 $\varphi(t)$ 는 같은 형식의 함수를 사용하기로 하였다.

M -추정량이나 GM-추정량은 반복적인 가중 최소 제곱(iterated weighted least squares (IWLS))방법으로 계산되어지며 이러한 방법은 본 논문의 3절에 있는 Monte Carlo 모의 실험을 위하여 사용되어진 S-Plus의 Window Ver 3.2에서 구현되고 있다.

p 차의 자기회귀모형에서의 GM-추정량에 대한 연구 성과로는 Bustos(1982) 와 다음의 정리 2.2의 GM-추정량에 대한 강한 근사식등이 있다.

정리 2.2 (Koul and Zhu(1995)) 추정량 $\hat{\alpha}_{GM}$ 에 대한 강한 근사식(Strong representation)은, 가정 (1) $E(|\epsilon_t|^\tau) < \infty$ for some $\tau > 4$ 을 만족하면, 확률 1로써

$$\lambda(0)n^{-1/2} \sum_{t=1}^n \mathbf{g}(\mathbf{Y}_{t-1})\mathbf{Y}'_{t-1}(\hat{\alpha}_{GM} - \alpha) = -n^{-1/2} \sum_{t=1}^n \mathbf{g}(\mathbf{Y}_{t-1})\varphi(\epsilon_t) + R_n$$

이 성립한다. 여기서

$$\|R_n\| = O(n^{-1/4}(\ln n)^{3/4}) \quad , \quad \lambda(x) = E(\psi(\epsilon_t - x)) \quad ,$$

$$\lambda(x) = E[g(\epsilon_t - x)] \quad , \quad \dot{\lambda}(x) = \frac{d}{dx}\lambda(x) \Big|_{x=0} \quad \text{이다.}$$

시계열 자료 분석에 있어서의 중요한 목적중의 하나가 미래 시점에서의 관측값에 대한 정확한 예측에 있으므로 앞서의 모수 추정 방법들에 따른 각각의 예측 능력을 평가하여 예측 오차를 최소화시키는 추정 방법을 최적의 모수 추정 방법으로 선택하는 것이 바람직하다고 말할 수가 있겠다. 따라서 본 논문에서는 예측 오차들을 바탕으로한 평균 제곱 오차(mean square error:MSE), 평균 절대 오차(mean absolute error:MAE) 와 같은 평가 기준 통계량을 사용하여 각각의 모수 추정법을 비교·평가하고자 한다.

3. 모의 실험 및 결론

이 절에서는 앞에서 언급한 최소 절대 편차 추정법, 로버스트 추정법들과 기존의 시계열 자료 분석에서 많이 사용하는 최소 제곱 추정법을 여러 조건의 모의 실험을 통하여 비교하여 자기회귀모형을 위한 모수 추정 방법으로써 로버스트한 추정법이 유효하게 사용될 수 있음을 확인해 보기로 한다. 이러한 목적을 위하여 우리는 오차항의 분포, 모수의 크기, 예측(forecasting) 값의 갯수등에 변화를 주어가며 모의 실험하는 다음과 같은 일련의 실험 계획을 마련하였다.

3.1 모의 실험 설계

본 논문에서는 자기회귀모형중에서 자주 사용되는 다음과 같은 AR(1) 모형과 AR(2) 모형을 고려하였다.

$$AR(1) \quad : \quad X_t = \alpha_1 X_{t-1} + \epsilon_t, \quad t = 1, \dots, n. \quad (3.1)$$

$$AR(2) \quad : \quad X_t = \alpha_1 X_{t-1} + \alpha_2 X_{t-2} + \epsilon_t, \quad t = 1, \dots, n. \quad (3.2)$$

여기서 $\{\epsilon_t\}$ 는 *i.i.d.*이며 평균이 0이고 분산이 σ_ϵ^2 인 백색 잡음 과정임을 가정한다. 위의 식 (3.1)과 (3.2)에서 설정한 AR(1) 모형과 AR(2) 모형이 정상성의 조건을 만족하도록 하기 위

하여 다음과 같은 모수들을 가정하기로 한다.

$$AR(1) : \{\alpha_1\} = \{0.2, 0.4, 0.6, 0.8\}.$$

$$AR(2) : \{\alpha_1, \alpha_2\} = \{(0.2, 0.2), (-0.6, 0.3), (0.8, -0.4), (-1.2, -0.5)\}.$$

자료의 분포나 이상치의 영향력에 대한 모수 추정 방법들의 유효성을 살펴보기 위하여 오차항의 분포에 대해서는 다음과 같은 3가지 경우를 고려하였다.

Case 1: 오차항이 정규분포를 따르는 경우.

Case 2: 오차항이 이중 지수 분포를 따르는 경우.

Case 3: 오차항이 정규분포보다 꼬리가 두꺼운(heavy-tailed) 경우.

실제 모의실험에 있어서는 Case 1을 위하여 오차항은 평균이 0이고 분산이 1인 표준 정규 분포로부터 독립적으로 생성되었음을 가정하였고, Case 2을 위하여는 오차항은 평균이 0 이고 분산이 2인 이중 지수 분포로부터 독립적으로 생성되었음을 가정하였으며, Case 3을 위해서는 오차항은 25% 오염된 정규분포, 즉, $\epsilon_t \sim CN(\cdot | 0.25, 3^2) = 0.75 \cdot N(\cdot | 0, 1) + 0.25 \cdot N(\cdot | 0, 3^2)$ 으로부터 독립적으로 생성되었음을 가정하였다. 또한 예측의 갯수에 따른 각 모수 추정 방법의 예측의 정확도를 알아 보기 위하여 예측값의 갯수를 5개와 10개인 경우를 고려하였다.

이러한 모의 실험을 행하기 위해 S-Plus의 Window Version 3.2d 패키지를 사용하였고 표본의 크기(sample size)는 300이며 이러한 실험을 500번 반복 실행하였다. 또한 실험의 동일 반복성을 위하여 초기값(seed number)를 사용하여 오차항에 대한 난수들을 생성하였다.

3.2 모의 실험 절차

앞서 설명한 여러가지 가상의 조건을 다음의 절차를 통하여 실시하였다.

Step 1 : 앞절의 Case 1~3과 같은 각 분포에 따른 오차항(ϵ_t)을 생성한다. 여기서, 오염된 정규분포 자료를 생성하기 위해서 균등분포 $U(0, 1)$ 로부터 생성한 자료가 0.25보다 작으면 $N(0, 3^2)$ 로부터 자료를 생성하고 0.25보다 크면 $N(0, 1)$ 로부터 자료를 생성한다. 이중 지수분포 자료를 생성하기 위해서는 평균이 1인 지수분포를 사용하였다. S-Plus 프로그램에서는 오염된 정규분포 자료를 생성하기 위해서는 함수 `rnorm()` 와 `runif()`을 이용하였고 이중 지수분포 자료를 생성하기 위해서는 `rexp()` 함수를 이용하였다.

Step 2 : 생성한 각 오차항을 가지고 각 모형 $AR(1)$, $AR(2)$ 에 따른 모의 시계열 자료 X_t 를 300개 생성한다. 본 논문에서는 자료의 초기값 X_0 값으로 0을 사용하였다.

Step 3 : 생성한 모의 시계열 자료에 앞에서 소개한 4가지 모수 추정방법을 각각 적용하여 모수를 추정한다 (예 : $\hat{\alpha}_{LS}, \hat{\alpha}_{LAD}, \hat{\alpha}_{GM-H}, \hat{\alpha}_{GM-B}$). 이때, Step 2에서 생성된 300개의 자료중에서 처음 50개의 자료는 모수 추정에서 초기치의 영향력을 배제하기 위하여 잘라 버리기로 하고 나머지 자료 중 앞부분의 200개의 자료를 이용하여 모수 추정을 하기로 한다(Bustos and Yohai(1986)). 나머지 50개의 자료는 후에 예측 오차를 계산하기 위하여

사용되었으며 또한 각 추정치는 S-plus에 내장된 lsfit, l1fit, ar.gm 모듈을 이용하여 계산하였다. 여기서, 오차항이 정규분포를 따르는 $AR(1)$ 모형의 경우에 있어서의 최소 제곱 추정량에 대한 GM-추정량의 상대적 효율성을 81.5%로 유지하기 위하여 2.2.2절에 있는 조절상수로 각각 $c_h = 1.0$, $c_b = 3.9$ 를 사용하였다(Denby and Martin(1979)).

Step 4 : 추정된 모수를 통해 완성된 확률 모형을 이용하여 선시차 l 만큼의 예측(forecasting)을 시도한다. 이를 위하여 S-Plus의 `arima.forecast()` 함수를 이용하였다. 여기서 l 값은 5 또는 10 이다.

Step 5 : 계산된 예측값과 Step 2에서 생성된 뒷부분의 50개의 자료중 해당되는 실제의 자료값과의 차를 구하여 예측 오차를 계산한다.

Step 6 : 이렇게 계산된 예측 오차를 이용하여 2절에서 나오는 평가 기준 통계량(MSE, MAE)를 계산한다.

Step 7 : 앞서의 Step 1에서 Step 6까지의 과정을 500번 반복하여 얻은 각각의 모수 평가 기준 통계량의 평균값과 표준편차를 이용하여 각각의 가정된 $AR(1)$, $AR(2)$ 모형에서의 각 모수 추정방법들의 유효성을 판별한다.

3.3 모의 실험을 통한 비교·분석

앞서의 모의 실험결과들을 바탕으로, 오차항의 분포, 모수의 크기, 예측의 갯수의 변화에 따른 각 추정방법의 유효성을 예측의 관점에서 비교하여 보기로 하자. 이러한 비교·평가를 위하여 500번 반복 실험을 통하여 얻은 MSE, MAE 평가 기준 통계량의 평균과 표준편차를 다음의 표 1에서부터 표 4까지 정리하여 놓았다. 참고로 표 1에서부터 표 4까지의 숫자값들에서 (*)는 각각의 가정된 모수 값과 오차항의 분포하에서 4가지 모수 추정 방법들을 MSE 기준 통계량으로 비교하였을 때 평균이 가장 작은 숫자값을 의미하며, (**)는 MAE 기준 통계량으로 비교하였을 때 평균이 가장 작은 숫자값을 의미한다.

먼저 오차항이 정규분포를 따르는 경우를 살펴보자. 모형이 $AR(1)$ 인 경우에는 가정된 모수값의 크기에 상관없이 전반적으로 최소 제곱 방법이 가장 적은 MSE, MAE 값을 나타내어 예측력이 예상한대로 다른 방법보다 우수함을 알 수 있다. 이러한 현상은 예측의 개수가 5개인 경우의 MSE와 MAE 기준 통계량을 살펴보면 잘 나타나 있는데, 예측의 개수가 5개에서 10개로 늘어나는 경우에는 모수의 크기가 커짐에 따라 최소 제곱 추정방법보다 최소 절대 편차 추정방법이 약간 우수한 예측력도 보이고 있어 Box와 Jenkins가 제안한 AR 모형은 단기 예측에 더 잘 들어 맞는 모형임을 여기서 확인할 수가 있다. 모형이 $AR(2)$ 인 경우에도 MSE나 MAE 기준 통계량을 살펴보면 예측의 개수가 5개인 경우에는 최소 제곱 추정법이 우수한 예측력을 보이고 있으나 10개로 늘어나면 GM-추정법이 오히려 예측을 잘하는 현상을 보이고 있어 AR 모형의 한계를 보여준다고 말할 수가 있다.

다음으로 오차항이 이중지수분포를 따르는 경우를 살펴보자. 모형이 $AR(1)$ 인 경우에는 예상한 바와 같이 가정된 모수의 크기에 상관없이 예측의 개수가 5개이거나 10인 경우

표 1: 예측오차에 근거한 추정방법들의 비교 : $AR(1)$, 예측갯수 = 5개인 경우

AR(5)		α_1	0.2	0.4	0.6	0.8
N O R M A L	LS	MSE의 평균 (표준편차)	*1.033602 (0.606473)	*1.139483 (0.726232)	*1.377026 (1.026132)	*1.885750 (1.721898)
		MAE의 평균 (표준편차)	**0.816745 (0.263417)	**0.859438 (0.298458)	**0.944737 (0.375441)	**1.098476 (0.521102)
	LAD	MSE의 평균 (표준편차)	1.035335 (0.606983)	1.141313 (0.727912)	1.378030 (1.027269)	1.891492 (1.712611)
		MAE의 평균 (표준편차)	0.817656 (0.263594)	0.859938 (0.298686)	0.944850 (0.374516)	1.099867 (0.520897)
	GM-H	MSE의 평균 (표준편차)	1.033861 (0.606159)	1.139839 (0.726894)	1.377276 (1.028715)	1.887261 (1.726354)
		MAE의 평균 (표준편차)	0.816875 (0.263136)	0.859680 (0.298083)	0.944952 (0.374897)	1.098801 (0.521501)
	GM-B	MSE의 평균 (표준편차)	1.034064 (0.606346)	1.140230 (0.727386)	1.377757 (1.029735)	1.887369 (1.726324)
		MAE의 평균 (표준편차)	0.816913 (0.263040)	0.859795 (0.297986)	0.945239 (0.374910)	1.099084 (0.521172)
D · E	LS	MSE의 평균 (표준편차)	2.058215 (1.968626)	2.272268 (2.286899)	2.745112 (3.011695)	3.743599 (4.551373)
		MAE의 평균 (표준편차)	1.031728 (0.482642)	1.106420 (0.549065)	1.234919 (0.662565)	1.455081 (0.865997)
	LAD	MSE의 평균 (표준편차)	*2.053854 (1.965835)	*2.264327 (2.278607)	*2.729115 (2.987375)	*3.707874 (4.503717)
		MAE의 평균 (표준편차)	**1.030692 (0.481659)	**1.104182 (0.547793)	**1.230308 (0.659579)	**1.448699 (0.859581)
	GM-H	MSE의 평균 (표준편차)	2.056659 (1.965778)	2.266790 (2.278733)	2.731761 (2.987621)	3.719072 (4.507026)
		MAE의 평균 (표준편차)	1.031738 (0.481782)	1.105289 (0.547918)	1.231246 (0.660060)	1.451239 (0.860658)
	GM-B	MSE의 평균 (표준편차)	2.056409 (1.965422)	2.265715 (2.276215)	2.72943 (2.979423)	3.713833 (4.482833)
		MAE의 평균 (표준편차)	1.031624 (0.481320)	1.104904 (0.547098)	1.230401 (0.658793)	1.450864 (0.857781)
C · N	LS	MSE의 평균 (표준편차)	3.026306 (3.490080)	3.293912 (3.788998)	*3.911221 (4.590946)	*5.246364 (6.514837)
		MAE의 평균 (표준편차)	1.209585 (0.598176)	1.292862 (0.670151)	**1.440016 (0.803601)	1.689987 (1.040219)
	LAD	MSE의 평균 (표준편차)	*3.022981 (3.492504)	*3.293153 (3.798890)	3.923387 (4.621446)	5.284396 (6.611291)
		MAE의 평균 (표준편차)	**1.207936 (0.598743)	**1.291686 (0.670708)	1.442695 (0.804393)	1.696909 (1.042646)
	GM-H	MSE의 평균 (표준편차)	3.025581 (3.491401)	3.294883 (3.792784)	3.913739 (4.599455)	5.246908 (6.547506)
		MAE의 평균 (표준편차)	1.209023 (0.598260)	1.292695 (0.670700)	1.440447 (0.804527)	**1.688859 (1.041415)
	GM-B	MSE의 평균 (표준편차)	3.025255 (3.491043)	3.295100 (3.793937)	3.914182 (4.601678)	5.249390 (6.553407)
		MAE의 평균 (표준편차)	1.208851 (0.598404)	1.292555 (0.670956)	1.440542 (0.804406)	1.690492 (1.039965)

표 2: 예측오차에 근거한 추정방법들의 비교 : AR(1), 예측갯수=10개인 경우

AR1(10)		α_1	0.2	0.4	0.6	0.8
N O R M A L	LS	MSE의 평균 (표준편차)	*1.137649 (0.467370)	*1.301314 (0.587831)	1.692525 (0.920957)	2.761679 (2.046679)
		MAE의 평균 (표준편차)	**0.894122 (0.193669)	**0.958810 (0.229249)	1.094829 (0.314252)	1.387699 (0.539769)
	LAD	MSE의 평균 (표준편차)	1.138651 (0.466852)	1.302371 (0.587355)	*1.691674 (0.918449)	*2.759534 (2.038452)
		MAE의 평균 (표준편차)	0.894774 (0.193812)	0.959206 (0.229243)	**1.094189 (0.313168)	**1.386735 (0.538332)
	GM-H	MSE의 평균 (표준편차)	1.138037 (0.467525)	1.301823 (0.588556)	1.692875 (0.924042)	2.763715 (2.062737)
		MAE의 평균 (표준편차)	0.894381 (0.193700)	0.958995 (0.229349)	1.094830 (0.314707)	1.387608 (0.541582)
	GM-B	MSE의 평균 (표준편차)	1.138098 (0.467502)	1.302003 (0.588543)	1.693331 (0.924720)	2.765465 (2.068338)
		MAE의 평균 (표준편차)	0.894395 (0.193580)	0.959058 (0.229244)	1.095087 (0.314836)	1.388008 (0.542423)
D · E	LS	MSE의 평균 (표준편차)	2.267141 (1.512016)	2.578285 (1.823917)	3.327667 (2.692517)	5.344293 (5.634882)
		MAE의 평균 (표준편차)	1.142416 (0.363412)	1.245882 (0.430576)	1.440669 (0.563487)	1.837373 (0.893861)
	LAD	MSE의 평균 (표준편차)	*2.264349 (1.510712)	*2.572975 (1.821243)	*3.314195 (2.677192)	*5.302480 (5.552921)
		MAE의 평균 (표준편차)	**1.141392 (0.362794)	**1.244498 (0.429735)	**1.437668 (0.560455)	**1.831696 (0.883455)
	GM-H	MSE의 평균 (표준편차)	2.266190 (1.51088)	2.574889 (1.820643)	3.318191 (2.678266)	5.320144 (5.591703)
		MAE의 평균 (표준편차)	1.14199 (0.362878)	1.245067 (0.429722)	1.438592 (0.561053)	1.834691 (0.888580)
	GM-B	MSE의 평균 (표준편차)	2.266318 (1.510875)	2.574705 (1.819943)	3.316988 (2.674220)	5.317695 (5.577675)
		MAE의 평균 (표준편차)	1.141979 (0.362670)	1.244949 (0.429321)	1.438287 (0.560148)	1.835029 (0.886624)
C · N	LS	MSE의 평균 (표준편차)	3.328249 (2.518777)	3.804124 (2.936218)	4.943023 (4.030649)	8.029973 (7.545005)
		MAE의 평균 (표준편차)	1.335366 (0.435631)	1.467548 (0.518832)	1.723508 (0.678435)	2.242894 (1.056832)
	LAD	MSE의 평균 (표준편차)	3.326022 (2.518091)	3.803818 (2.935753)	4.949341 (4.035183)	8.045924 (7.577629)
		MAE의 평균 (표준편차)	**1.334433 (0.435160)	1.467297 (0.517784)	1.725779 (0.677441)	2.247534 (1.057698)
	GM-H	MSE의 평균 (표준편차)	3.326466 (2.516937)	3.802783 (2.933646)	4.939537 (4.024491)	8.005105 (7.511603)
		MAE의 평균 (표준편차)	1.334937 (0.435056)	1.467367 (0.518247)	1.723549 (0.677205)	2.240257 (1.052643)
	GM-B	MSE의 평균 (표준편차)	*3.325763 (2.515595)	*3.801792 (2.932514)	*4.937508 (4.024530)	*7.995715 (7.510285)
		MAE의 평균 (표준편차)	1.334772 (0.434781)	**1.467116 (0.517994)	**1.723338 (0.677118)	**2.239673 (1.051742)

표 3: 예측오차에 근거한 추정방법들의 비교 : AR(2), 예측갯수 = 5개인 경우

AR2(5)		α_1 α_2	0.2 0.2	-0.6 0.3	0.8 -0.4	-1.2 -0.5
N O R M A L	LS	MSE의 평균 (표준편차)	1.073726 (0.652065)	*1.803622 (1.808092)	1.539196 (1.140460)	*2.874851 (2.953133)
		MAE의 평균 (표준편차)	0.834004 (0.279683)	**1.056850 (0.546101)	0.995696 (0.389990)	**1.326478 (0.703170)
	LAD	MSE의 평균 (표준편차)	1.076339 (0.652100)	1.824947 (1.855835)	1.549119 (1.142436)	2.928279 (3.030850)
		MAE의 평균 (표준편차)	0.835580 (0.279168)	1.059362 (0.552503)	1.000096 (0.389200)	1.335949 (0.714725)
	GM-H	MSE의 평균 (표준편차)	*1.073307 (0.651778)	1.805273 (1.799238)	*1.537862 (1.139881)	2.885395 (2.949173)
		MAE의 평균 (표준편차)	0.834013 (0.278891)	1.057891 (0.543945)	**0.995598 (0.388837)	1.329801 (0.702639)
	GM-B	MSE의 평균 (표준편차)	1.073369 (0.651819)	1.803679 (1.795354)	1.539011 (1.139686)	2.890146 (2.953963)
		MAE의 평균 (표준편차)	**0.834000 (0.278760)	1.057299 (0.543927)	0.996160 (0.388571)	1.330854 (0.703420)
D · E	LS	MSE의 평균 (표준편차)	2.141976 (2.078906)	3.590885 (4.718641)	3.062918 (3.320076)	5.721910 (7.946306)
		MAE의 평균 (표준편차)	1.065569 (0.507754)	1.409078 (0.866825)	1.304060 (0.688266)	1.762315 (1.097188)
	LAD	MSE의 평균 (표준편차)	*2.130169 (2.089727)	*3.569440 (4.753258)	3.059965 (3.345614)	5.685721 (8.041769)
		MAE의 평균 (표준편차)	**1.060192 (0.509707)	**1.403805 (0.863088)	1.301104 (0.691062)	**1.753523 (1.098971)
	GM-H	MSE의 평균 (표준편차)	2.134098 (2.076102)	3.580551 (4.729432)	*3.057303 (3.324341)	5.675745 (7.925634)
		MAE의 평균 (표준편차)	1.062540 (0.507254)	1.406514 (0.865683)	1.301874 (0.687882)	1.756142 (1.092253)
	GM-B	MSE의 평균 (표준편차)	2.131752 (2.073336)	3.581693 (4.743426)	3.057601 (3.330842)	*5.665398 (7.919587)
		MAE의 평균 (표준편차)	1.061939 (0.506478)	1.406534 (0.866812)	**1.300781 (0.688332)	1.754304 (1.092274)
C · N	LS	MSE의 평균 (표준편차)	3.126955 (3.570858)	5.149104 (7.346501)	4.424526 (5.274824)	8.491202 (12.55515)
		MAE의 평균 (표준편차)	1.245880 (0.623688)	1.631617 (1.086127)	1.523564 (0.847136)	2.093361 (1.421445)
	LAD	MSE의 평균 (표준편차)	3.130255 (3.588020)	5.128558 (7.329751)	4.401199 (5.278786)	*8.350498 (12.34313)
		MAE의 평균 (표준편차)	1.245653 (0.624970)	1.627735 (1.087865)	1.518209 (0.844985)	2.076041 (1.404113)
	GM-H	MSE의 평균 (표준편차)	*3.122854 (3.580178)	*5.125683 (7.319842)	4.404075 (5.275365)	8.373485 (12.50721)
		MAE의 평균 (표준편차)	1.244239 (0.624299)	**1.625650 (1.084883)	1.517589 (0.847126)	2.076782 (1.409103)
	GM-B	MSE의 평균 (표준편차)	3.122996 (3.584139)	5.131974 (7.320017)	*4.400157 (5.279181)	8.363930 (12.57306)
		MAE의 평균 (표준편차)	**1.244226 (0.624646)	1.627516 (1.085238)	**1.515942 (0.847840)	**2.075165 (1.407304)

표 4: 예측오차에 근거한 추정방법들의 비교 : AR(2), 예측갯수=10개인 경우

AR2(10)		α_1	0.2	-0.6	0.8	-1.2	
		α_2	0.2	0.3	-0.4	-0.5	
N O R M A L	LS	MSE의 평균 (표준편차)	1.100216 (0.501456)	2.598272 (2.600253)	1.903177 (0.990940)	3.964121 (2.808667)	
		MAE의 평균 (표준편차)	0.840313 (0.202061)	**1.259139 (0.627071)	1.157177 (0.314596)	1.645434 (0.606316)	
	LAD	MSE의 평균 (표준편차)	1.101191 (0.500593)	2.617983 (2.621728)	1.911805 (0.994109)	3.991869 (2.848954)	
		MAE의 평균 (표준편차)	0.840929 (0.201890)	1.263465 (0.629215)	1.159981 (0.314971)	1.649804 (0.608640)	
	GM-H	MSE의 평균 (표준편차)	*1.099896 (0.501780)	2.594538 (2.561346)	1.904130 (0.991586)	3.968763 (2.809461)	
		MAE의 평균 (표준편차)	**0.840260 (0.202239)	1.260283 (0.622358)	1.157286 (0.314592)	1.646896 (0.605399)	
	GM-B	MSE의 평균 (표준편차)	1.099983 (0.501987)	*2.592855 (2.563558)	*1.669391 (0.910440)	*3.381315 (2.630915)	
		MAE의 평균 (표준편차)	0.840267 (0.202191)	1.259265 (0.623102)	**1.035447 (0.296922)	**1.453579 (0.577793)	
	D · E	LS	MSE의 평균 (표준편차)	2.162165 (1.582998)	*4.749543 (5.007206)	3.755388 (2.765637)	7.783584 (6.705739)
			MAE의 평균 (표준편차)	1.073143 (0.382583)	1.654666 (0.869263)	1.532403 (0.555267)	2.209844 (0.932127)
		LAD	MSE의 평균 (표준편차)	*2.155707 (1.587037)	4.755097 (5.059713)	3.755727 (2.769568)	7.775428 (6.727008)
			MAE의 평균 (표준편차)	**1.070380 (0.382644)	**1.654131 (0.869942)	1.531783 (0.555218)	2.206853 (0.928983)
GM-H		MSE의 평균 (표준편차)	2.157692 (1.580918)	4.76032 (5.029393)	3.752659 (2.762762)	7.75997 (6.683775)	
		MAE의 평균 (표준편차)	1.071565 (0.381609)	1.656092 (0.869911)	1.531387 (0.553446)	2.206806 (0.926362)	
GM-B		MSE의 평균 (표준편차)	2.156633 (1.579912)	4.768677 (5.058920)	*3.244281 (2.541645)	*6.541024 (6.232634)	
		MAE의 평균 (표준편차)	1.071288 (0.381100)	1.656918 (0.871262)	**1.357002 (0.521934)	**1.929221 (0.870938)	
C · N		LS	MSE의 평균 (표준편차)	3.224510 (2.582405)	7.191421 (8.064000)	5.606072 (4.615789)	11.75983 (11.10134)
			MAE의 평균 (표준편차)	1.27410 (0.458775)	2.002679 (1.118163)	1.815796 (0.700729)	2.642787 (1.198180)
		LAD	MSE의 평균 (표준편차)	3.226127 (2.589623)	7.148193 (8.077182)	5.590245 (4.599675)	11.66607 (11.02722)
			MAE의 평균 (표준편차)	1.274291 (0.459574)	1.993379 (1.118325)	1.812923 (0.697947)	2.629743 (1.193395)
	GM-H	MSE의 평균 (표준편차)	3.221577 (2.584319)	7.139393 (8.039737)	5.594563 (4.604449)	11.68441 (11.07689)	
		MAE의 평균 (표준편차)	1.273223 (0.458910)	1.992006 (1.115901)	1.812798 (0.699656)	2.631648 (1.195271)	
	GM-B	MSE의 평균 (표준편차)	*3.221319 (2.586654)	*7.135632 (8.062784)	*4.850569 (4.304111)	*9.871495 (10.15299)	
		MAE의 평균 (표준편차)	**1.273169 (0.459321)	**1.990972 (1.117881)	**1.606119 (0.663479)	**2.309343 (1.120232)	

에도 최소 절대 편차 추정법이 가장 작은 MSE, MAE 값들을 나타내어 예측의 관점에서 다른 추정방법보다 더 우수한 방법임을 알 수가 있다. 즉, 정규성의 가정을 할수 없는 상태에서의 최소제곱 추정법의 한계성을 이 경우가 잘 보여 준다고 말할 수 있겠다. $AR(2)$ 모형의 경우에는 예측의 갯수가 5개일때는 위와 같은 현상을 잘 살펴볼 수 있으나 예측의 개수가 10개로 늘어나면서 GM-추정법이 최소 절대 편차 추정법보다 더 유효한 방법인 경우도 보이고 있어 역시 AR 모형의 한계를 보여주고 있다.

마지막으로 오차항이 오염된 정규분포를 따르는 경우를 살펴보자. 여기서도 예상한대로 전반적으로 다른 추정 방법보다 GM-추정법이 예측을 잘 수행한다는 것을 알수가 있다. 즉, 최소 제곱 추정법이나 최소 절대편차 추정법이 GM-추정법보다는 이상치에 의한 영향을 상대적으로 많이 받고 있음을 알 수가 있어 이상치가 있거나 정규 분포의 가정을 할 수가 없는 경우에는 최소 제곱 추정방법이 최적의 방법이 아님을 우리는 알 수가 있다. 특히나 이러한 현상은 $AR(1)$ 모형이나 $AR(2)$ 모형에서 예측의 개수가 5개에서 10개로 늘어날수록 잘 나타나고 있는데 GM-추정법중에서도 Tukey의 Biweight 함수를 이용한 GM-B 추정법이 더 나은 예측력을 가지는 것을 볼 수 있어 AR 모형을 이용하여 예측을 할 경우에는 GM-B 추정법의 사용을 고려하는 것이 좋겠다.

3.4 결론

앞절의 모의 실험 결과들을 바탕으로한 분석을 살펴 볼 때, GM-추정방법은 오차항이 오염된 정규분포를 따르는 경우에 예측의 관점에서 가장 좋은 추정방법이 되지만 오차항이 다른 분포일 경우에도 어느 정도 좋은 예측력을 가지고 있다고 판단된다. 또한 다른 추정방법에 비해 모수의 크기에 따른 영향을 적게 받는다고 할 수 있어 오차항의 분포에 대한 정규성의 가정을 할 수 없고 또한 모수의 크기도 어느 정도인지 알 수 없는 실제 자료분석시에는 다른 방법보다는 GM-추정방법의 사용이 보다 타당하다고 여겨진다.

그림 3과 그림 4는 $AR(1)$ 모형을 따르는 시계열 $\{X_t\}$ 에 대한 시도표(time plot)로써 시점 200 이후 10개의 시점에 대한 예측값들을 나타낸 그림들이다. 그림 3은 오차항이 이중지수분포를 따른다는 가정을 한 경우로써, 시점 200 이후의 실선(-)은 3.2절의 모의 실험 절차상의 Step 2에서 생성된 시계열 자료를 나타낸 것이고 점선(···)은 최소 제곱 추정법을 이용한 적합식에 의해 예측된 값이며 (--)은 최소 절대 편차 추정법을 이용한 적합식에 의해 예측된 값을 나타낸다. 그림에서 보는 바와 같이 최소 절대 편차 방법에 의한 예측값이 최소 제곱 추정방법에 의한 예측값보다 실제 값에 더 근접함을 알 수 있다. 따라서 예측에 있어서도 오차항이 이중 지수 분포를 따를 경우에는 최소 절대 편차 방법이 더 타당하다는 것을 확인할 수가 있다. 그림 4는 오차항이 오염된 정규분포를 따른다는 가정을 한 경우로써, 역시 앞서와 같이 시점 200 이후의 실선(-)은 3.2절의 모의 실험 절차상의 Step 2에서 생성된 시계열 자료를 나타낸 것이고 점선(···)은 최소 제곱 추정법을 이용한 적합식에 의해 예측된 값이며 (--)은 로버스트한 방법중 GM-B 추정방법을 이용한 적합식에 의해 예측

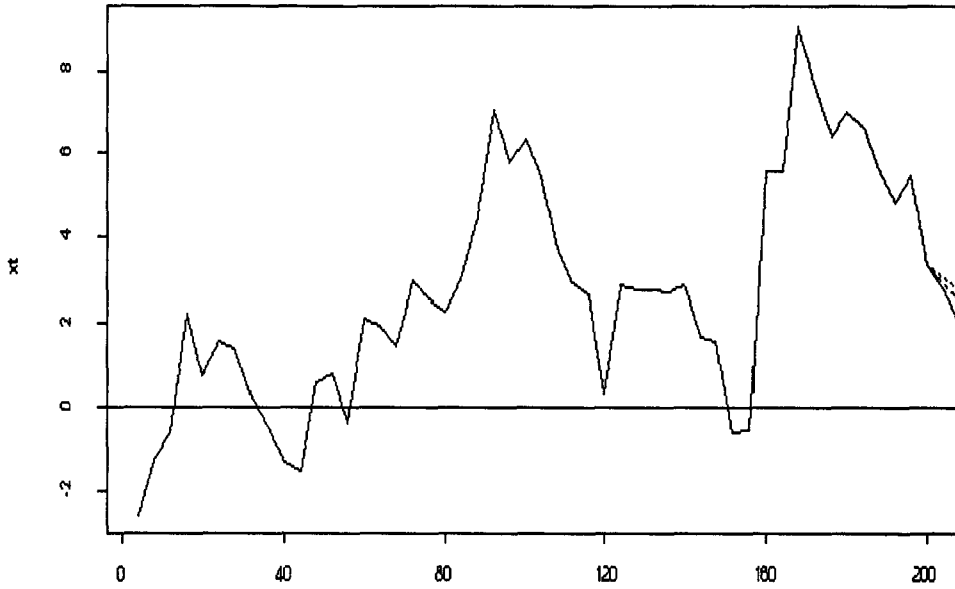


그림 3: 오차항이 이중 지수 분포를 따르는 경우의 선시차 $l=10$ 의 예측

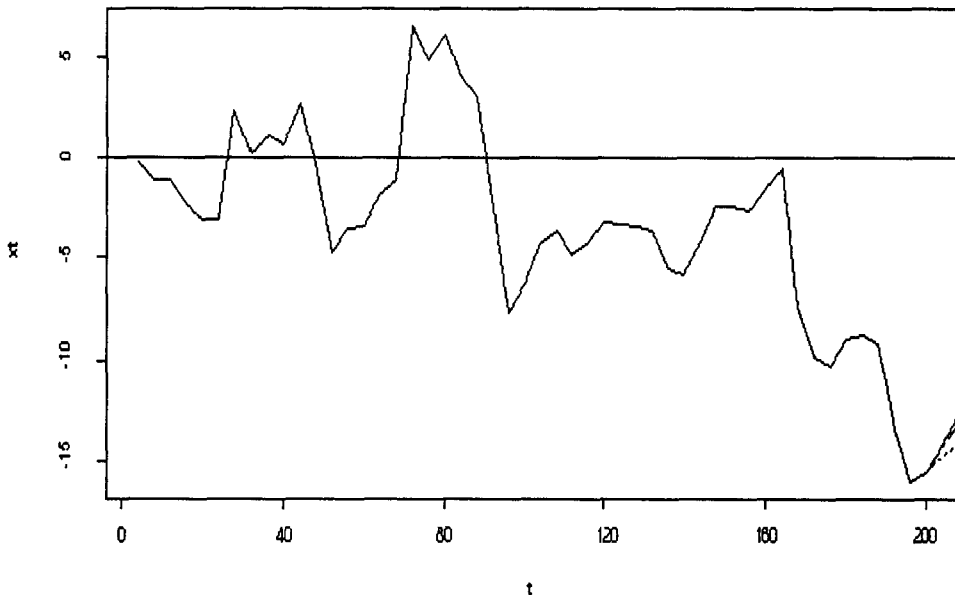


그림 4: 오차항이 오염된 정규 분포를 따르는 경우의 선시차 $l=10$ 의 예측

된 값을 나타낸다. 그림에서 보는 바와 같이 GM-B 방법에 의한 예측값이 최소 제곱 추정 방법에 의한 예측값보다 실제 값에 더 근접함을 알 수 있다. 따라서 예측에 있어서도 오차항에 이상치가 포함되어 있는 경우에는 최소 제곱 추정방법보다는 로버스트한 방법이 더 타당하다는 것을 확인할 수가 있다.

따라서 앞서의 모의 실험결과들을 바탕으로 우리는 다음과 같은 결론들을 얻을 수가 있다.

첫째, 자료의 분포에 이상치가 발견되거나 오차항의 분포가 정규분포임을 확신할 수 없는 상황에서는 최소 절대 편차 추정법이나 GM-추정방법이 보통의 최소 제곱 추정방법보다는 보다 정확한 예측을 할 수 있다.

둘째, 모수의 크기도 모르고 오차항에 대한 정규분포의 가정을 할 수 없으며 이상치에 의한 영향력을 무시할 수 없는 실제의 자료분석의 경우에는 다른 모수 추정방법보다는 GM-B 추정방법이 보다 정확한 예측을 하므로 이 방법의 사용을 제안한다.

셋째, 예측의 관점에서 볼때 Box와 Jenkins가 제안한 자기 회귀 모형(AutoRegressive Models)은 단기 예측의 경우에 더 정확한 예측을 하므로 중·장기적인 예측을 할 경우에는 다른 모형을 고려해야 할 것이다.

참 고 문 헌

1. 차경엽, 김삼용, 이성덕 (1999). 시계열모형에서 추정함수를 이용한 로버스트 추론방법, 응용통계연구, 제12권 2호, 479-490.
2. Bloomfield, P., and Steiger, W. L. (1983). *Least Absolute Deviations : Theory, Applications and Algorithms*, Birkhauser, Boston.
3. Brockwell, P. J., and Davis, R. A. (1991). *Time Series : Theory and Methods*, 2nd ed., Springer-Verlag, New York.
4. Bustos, O. H. (1982). General M-estimates for Contaminated p -th Order Autoregressive Processes : Consistency and Asymptotic Normality, *Z. Wahrscheinlichkeitstheorie verw. Gebiete*, 59, 491-504.
5. Bustos, O. H., and Yohai, V. J. (1986). Robust Estimates for ARMA Models, *Journal of the American Statistical Association*, 81, 155-168.
6. Denby, L., and Martin, R. D. (1979). Robust Estimation of the First-Order Autoregressive Parameter, *Journal of the American Statistical Association*, 74, 140-146.
7. Kang, H. J., and Shin, K. I. (1998). Strong Representations for LAD Estimators in AR(1) Models, *Journal of the Korean Statistical Society*, 27, 349-358.

8. Koul, H. L., and Zhu, Z. (1995). Bahadur-Kiefer Representations for GM-estimators in Autoregression Models, *Stochastic Processes and their Applications*, 57, 167-189.
9. Martin, R. D. (1982). The Cramer-Rao Bound and Robust M-estimates for Autoregression, *Biometrika*, 69, 437-442.
10. Staudte, R. G., and Sheather, S. J. (1990). *Robust Estimation and Testing*, John Wiley & Sons, New York.

A Comparison of Robust Parameter Estimations for Autoregressive Models ⁴

Hee-Jeong Kang⁵ and Soon-Young Kim⁶

Abstract

In this paper, we study several parameter estimation methods used for autoregressive processes and compare them in view of forecasting. The least square estimation, least absolute deviation estimation, robust estimation are compared through Monte Carlo simulations.

Key Words and Phrases: Autoregressive Processes, Least Squares Estimator, LAD(least absolute deviation) Estimator, Generalized M-estimator, Forecasting Error.

⁴The authors wish to acknowledge the partial financial support of the Korea Research Foundation (1998-015-D00048) made in the program year of 1998.

⁵Assistant Professor, Dept. of Statistics, Chonbuk National University, Chonju, Chonbuk.

⁶Dept. of Statistics, Chonbuk National University, honju, Chonbuk 561-756, Korea.