

이상 트래픽 탐지를 위한 로버스트 추정 방법 비교 연구

정재윤^a, 김삼용^{1,a}

^a중앙대학교 통계학과

요약

본 연구는 이상치가 존재하는 자료에 적용될 수 있는 방법을 비교한 연구로서, 이분산 시계열 모형 하에서 로버스트 추정 방법의 효용성을 보이고자 한다. GARCH 모형 하에서 이상치 탐지 기법과 GARCH 모형을 기반한 로버스트 추정방법의 성능을 비교하였다. 실제 인터넷 트래픽 자료에 두 방법을 적용했을 때, 로버스트 추정방법이 이상치 탐지 기법에 비해 덜 복잡하고 성능이 우수함을 입증하였다.

주요용어: 로버스트 추정, 이상치 탐지, 트래픽 자료, GARCH 모형.

1. 서론

시계열 자료의 분석에서 가장 널리 쓰이는 통계적 기법으로는 Box와 Jenkins (1976)가 제안한 ARIMA(Autoregressive Integrated Moving Average) 모형이 있고, 오차의 분산이 자기 회귀적으로 변하는 경우 적절한 모형으로는 Engle (1982)에 의해 제시된 ARCH 모형과 ARCH 모형의 일반화된 모형으로 Bollerslev (1986)에 의해 제시된 GARCH 모형 등 여러 가지 이분산성 시계열 모형도 있다. 이런 모형들에 적용되는 분석과 예측 방법으로는 우도함수, 최소제곱법, 적률추정법을 이용한 추정과 검정 방법이 있다. 이런 방법들의 성능은 알려진 바와 같이 오차분포 가정에 대해서 민감하게 작용한다. 또한 자료의 이상치 존재 여부 따라 추정과 예측에 매우 심각한 문제를 일으킬 수 있다. 이러한 문제를 해결하기 위해 많은 연구자들이 다양한 통계적 방법을 적용하여 이상치를 탐지하는 방법을 연구하고 있다. Huber (1981)는 독립적인 관측치의 경우에 표준적인 통계적 절차의 로버스트한 버전을 논의하였다. Denby와 Martin (1979), Martin (1980, 1982)은 AR 모형에 대한 로버스트 적률추정량의 효용성을 연구하였다. Martin과 Yohai (1985)는 AR과 MA 모형의 로버스트 추정량에 대하여 연구하였다. 한편 Basawa 등 (1985)는 시계열 모형에 대한 로버스트 스코어 함수 및 로버스트 추정량을 제시하고 이 방법으로부터 유도된 검정의 효용성을 논의하였다. 경제 시계열 자료에서 나타나는 이상치 문제를 해결하기 위하여 Franses와 Ghijsels (1999)은 GARCH 모형을 이용한 이상치의 탐지 방법을 제안하였고, Charles와 Darne (2005)는 재무 시계열 자료에서 GARCH 모형을 이용한 이상치 탐지 기법의 효용성을 보였다.

한편, Shao (2008)는 PAR(Periodic autoregressive) 시계열에 대한 Basawa 등 (1985)의 방법을 근거로 하여 로버스트 추정방법을 제시하였고, 그 결과를 실제 데이터 분석에 적용하였다. 하명호와 김삼용 (2010)은 KOSPI자료의 이상치를 탐지하는데 있어 GARCH 모형 하에서 Charles와 Darne (2005)이 제시한 이상치 탐지 기법의 효용성을 보였다.

이 논문에서는 Basawa 등 (1985)에서 제시된 로버스트 추정량을 이용해 모형 적합한 결과와 Charles와 Darne (2005)에서 제시된 방법을 이용하여 인터넷 트래픽 자료에 적합시켜 로버스트 방법의 성능을 비교하고자 한다.

본 연구는 2009년도 중앙대학교 연구장학기금 지원에 의한 것임.

¹ 교신저자: (156-756) 서울시 동작구 흑석동 221, 중앙대학교 통계학과, 교수. E-mail: sahm@cau.ac.kr

2. 이분산성 시계열 모형

GARCH 모형은 오차의 분산이 자기 회귀적으로 변하는 조건부 이분산자기회귀모형인 ARCH 모형의 일반화된 모형으로 Bollerslev (1986)에 의해 제시된 모형으로 GARCH(q, p)의 모형은 수식 (2.1)과 같다.

$$e_t = \sqrt{h_t}\eta_t, \quad \eta_t \sim \text{iid } N(0, 1) \quad (2.1)$$

$$\begin{aligned} \text{Var}(e_t | \Psi_{t-1}) = h_t &= \alpha_0 + \sum_{i=1}^q \alpha_i e_{t-i}^2 + \sum_{j=1}^p \beta_j h_{t-j} \\ &= \alpha_0 + A(L)e_t^2 + B(L)h_t, \end{aligned}$$

여기서 $A(L) = \alpha_1 L + \alpha_2 L^2 + \dots + \alpha_q L^q$, $B(L) = \beta_1 L + \beta_2 L^2 + \dots + \beta_p L^p$ 이다. 또, $\alpha_0 > 0$, $\alpha_i \geq 0$ ($i = 1, \dots, q$), $\beta_j \geq 0$ ($j = 1, \dots, p$), $p \geq 0$, $q > 0$ 이다.

ARCH(∞) 모형은 차수가 너무 커서 추정해야 할 모수가 많아져 비효율적이다. 그래서 같은 의미를 가지는 GARCH(p, q) 모형을 이용하면 적은 개수의 모수를 이용하여 설명 가능하다.

일반적으로 GARCH(1, 1) 모형은 시계열 자료의 변동성을 ARCH 모형보다 효율적으로 탐지 할 수 있다고 알려져 있으므로, 본 연구에서는 GARCH(1, 1) 모형을 고려하기로 한다.

GARCH(1, 1)의 모형식은 수식 (2.2)와 같다.

$$e_t = \sqrt{h_t}\eta_t, \quad \eta_t \sim \text{iid } N(0, 1) \quad (2.2)$$

$$h_t = \alpha_0 + \alpha_1 e_{t-1}^2 + \beta_1 h_{t-1}$$

$$\alpha_0 > 0, \quad 0 \leq \alpha_1 + \beta_1 < 1.$$

3. 이상치 탐지기법 및 로버스트 추정량

3.1. 이상치 탐지기법

Charles와 Darne (2005)이 제안한 이상치 탐지 및 대체 기법에 대해서 살펴보겠다. 다음과 같은 절차에 따라 이상치를 탐지하고 이상치로 탐지된 값들은 적절한 값으로 대체하였고, GARCH 모형을 기반으로 하여 이상치를 탐지하였다. 절차는 다음과 같다.

먼저, 잔차에 대해 GARCH(1, 1) 모형의 계수 $\alpha_0, \alpha_1, \beta_1$ 를 추정한다. GARCH(1, 1) 모형식은 수식 (2.2)와 같다. 추정된 계수를 이용하여 \hat{v}_t 를 추정하고, \hat{v}_t 는 수식 (3.1)과 같이 구할 수 있다.

$$\hat{v}_t = e_t^2 - \hat{h}_t. \quad (3.1)$$

두 번째 단계는 $\tau = 1, 2, \dots, n$ 에 대해 $\hat{\zeta}(\tau)$ 를 계산한다. $\hat{\zeta}(\tau)$ 는 수식 (3.2)와 같이 구할 수 있다.

$$\hat{\zeta}(\tau) = \frac{\sum_{t=\tau}^n x_t \hat{v}_t}{\sum_{t=\tau}^n x_t^2}, \quad (3.2)$$

여기서 $x_t = \begin{cases} 0, & t < \tau \\ 1, & t = \tau \end{cases}$ 이고, $t > \tau$ 일때는 $x_{t+k} = -\pi_k$ 이다.

세 번째 단계는 $\hat{\tau}$ 를 계산하게 되는데 $\hat{\tau}$ 는 두 번째 단계에서 구한 $\hat{\zeta}(\tau)$ 를 이용하여 계산할 수 있다. $\hat{\tau}$ 는 수식 (3.3)과 같이 구할 수 있다.

$$\hat{\tau} = \frac{[\hat{\zeta}(\tau)/\hat{\sigma}_a]}{[\sum_{t=\tau}^n x_t^2]^2}, \quad (3.3)$$

$\hat{\tau}$ 값 중 최대값을 $t_{\max}(\hat{\tau})$ 라 하고, $t_{\max}(\hat{\tau}) > C$ 이면, 그 관측치는 이상치라 할 수 있다.

C 는 특정 상수이고 보통 4, 실제 데이터에서는 10을 사용한다. 이상치로 판명된 값의 시점을 ξ 라고 할 때, 다음 단계를 이용하여 e_t 를 e_t^* 값을 대체하고, e_t^* 값은 수식 (3.4)와 같이 계산할 수 있다.

$$e_t^* = \begin{cases} e_t, & t \neq \xi, \\ \text{sign}(e_\xi) \sqrt{e_\xi^{2*}}, & t = \xi, \end{cases} \quad (3.4)$$

여기서 $e_\xi^{2*} = e_\xi^2 - \hat{\zeta}(\xi)$ 이다.

대체된 시계열 자료 e_t^* 을 가지고 단계 1부터 $t_{\max}(\hat{\tau})$ 가 특정상수 C 를 넘지 않을 때까지 반복한다.

3.2. 로버스트 추정 방법

Basawa 등 (1985)에서 제안한 로버스트 추정량에 대해서 살펴보겠다. $X_{(n)}$ 은 n 개 관측치 (X_1, \dots, X_n) 의 벡터로 쓰는데 이 관측치는 $k \times 1$ 모수 θ 벡터에 의존하는 시계열 모형으로부터 얻은 관측치이다. 로버스트 추정량을 추정하기 위해서는 먼저 로버스트 스코어 함수 $\tilde{S}_n(\theta)$ 를 구하여야 한다. 추정방정식 $\tilde{S}_n(\theta) = 0$ 의 해가 로버스트 추정량이다. 여기서 $\tilde{S}_n(\theta)$ 는 $k \times 1$ 벡터인데 각 원소는 $\tilde{S}_n^{(j)}(\theta)$ ($1 \leq j \leq k$)로 $X_{(n)}$ 과 θ 의 함수이다.

θ 의 로버스트 추정량을 $\tilde{\theta}_n$ 라 하고 수식 (3.5)와 같이 정의한다.

$$\tilde{\theta}_n = t_n + n^{-1}G^{-1}(t_n)S_n(t_n), \quad (3.5)$$

여기서 t_n 은 θ 에 대한 적절한 초기 추정량이다.

로버스트 추정 방정식을 구하기 위해서 η_t 를 $\tilde{\eta}_t$ 로 대체한다. 여기서 $\tilde{\eta}_t$ 는 유한한 분산을 가지는 유계 기함수이다. 이같은 함수의 예로는 Huber의 ψ 함수가 있다. Basawa 등 (1985)에서도 추정방정식을 로버스트화 하기위해서 Huber의 ψ 함수를 사용하였다. 본 연구에서도 Huber의 ψ 함수를 사용하여 추정방정식을 로버스트화 하였다. Huber의 ψ 함수는 수식 (3.6)와 같다.

$$\tilde{\eta}_t = \psi(\eta_t) = \begin{cases} \eta_t, & |\eta_t| < k, \\ k \times \text{sign}(\eta_t), & |\eta_t| \geq k, \end{cases} \quad (3.6)$$

여기서 k 는 양의 실수이고, $\text{sign}(x)$ 는 x 가 양수이면 1, 음수이면 -1, 0이면 0을 갖는 함수이다.

GARCH(1, 1) 모형 하에서 로버스트 추정방정식은 아래와 같이 유도할 수 있다. GARCH(1, 1) 모형은 수식 (2.2)에서 확인할 수 있다. $S_n(\theta) = \partial/\partial\theta \log L(\theta)$ 은 수식 (3.7)과 같다. 여기서 $\theta^T = (\alpha_0, \alpha_1, \beta_1)$ 이다.

$$S_n(\theta) = -\frac{1}{2} \sum \frac{1}{h_t(\theta)} \frac{\partial}{\partial\theta} h_t(\theta) + \frac{1}{2} \sum \frac{e_t^2}{h_t^2(\theta)} \frac{\partial}{\partial\theta} h_t(\theta). \quad (3.7)$$

또 수식 (3.7)은 수식 (3.8)과 같이 나타낼 수 있다.

$$\begin{pmatrix} S_n(\alpha_0) \\ S_n(\alpha_1) \\ S_n(\beta_1) \end{pmatrix} = \begin{pmatrix} -\frac{1}{2} \sum \frac{1}{h_t(\theta)} + \frac{1}{2} \sum \frac{\eta_t^2}{h_t(\theta)} \\ -\frac{1}{2} \sum \frac{e_{t-1}^2}{h_t(\theta)} + \frac{1}{2} \sum \frac{\eta_t^2 e_{t-1}^2}{h_t(\theta)} \\ -\frac{1}{2} \sum \frac{h_{t-1}(\theta)}{h_t(\theta)} + \frac{1}{2} \sum \frac{\eta_t^2 h_{t-1}(\theta)}{h_t(\theta)} \end{pmatrix}. \quad (3.8)$$

로버스트 추정량을 구하기 위해서는 앞에서 언급한 것처럼 오차항에 Huber의 ψ 함수, 수식 (3.6)을 적용하여 수식 (3.8)을 로버스트화 한다. 로버스트화된 추정방정식 $\tilde{S}_n(\theta)$ 은 수식 (3.9)와 같다. 단, 수식 (3.9)에서 $\tilde{h}(\theta)$ 를 $\tilde{h}(\theta)$ 로 표기하였다.

$$(\tilde{S}_n(\theta)) = \begin{pmatrix} \tilde{S}_n(\alpha_0) \\ \tilde{S}_n(\alpha_1) \\ \tilde{S}_n(\beta_1) \end{pmatrix} = \begin{pmatrix} -\frac{1}{2} \sum \frac{1}{\tilde{h}_t} + \frac{1}{2} \sum \frac{\tilde{\eta}_t^2}{\tilde{h}_t} \\ -\frac{1}{2} \sum \frac{\tilde{e}_{t-1}^2}{\tilde{h}_t} + \frac{1}{2} \sum \frac{\tilde{\eta}_t^2 \tilde{e}_{t-1}^2}{\tilde{h}_t} \\ -\frac{1}{2} \sum \frac{\tilde{h}_{t-1}}{\tilde{h}_t} + \frac{1}{2} \sum \frac{\tilde{\eta}_t^2 \tilde{h}_{t-1}}{\tilde{h}_t} \end{pmatrix}, \quad (3.9)$$

여기서 $\tilde{\eta}_t = \psi(\eta_t)$, $\tilde{e}_t = \sqrt{\tilde{h}_t} \tilde{\eta}_t$, $\tilde{h}_t = \alpha_0 + \alpha_1 \tilde{e}_{t-1}^2 + \beta_1 \tilde{h}_{t-1}$ 이다.

$\tilde{S}_n(\theta)/\sqrt{n}$ 는 수식 (3.10)과 같이 정규분포로 분포수렴하고, 여기서 평균행렬은 영행렬, 분산 공분산 행렬은 F 이다. F 는 수식 (3.11)과 같이 계산할 수 있다.

$$\frac{\tilde{S}_n(\theta)}{\sqrt{n}} \xrightarrow{D} N_3(0, F) \quad (3.10)$$

$$F = \lim \frac{1}{n} E(\tilde{S}_n(\theta) \tilde{S}_n^T(\theta)). \quad (3.11)$$

다음, 3×3 행렬 $G_n(\theta)$ 를 수식 (3.12)와 같이 정의하자.

$$G_n(\theta) = -\frac{1}{n} \frac{\partial \tilde{S}_n(\theta)}{\partial \theta} \quad (3.12)$$

그리고 $G_n(\theta)$ 는 수식 (3.13)과 같이 확률 수렴한다.

$$G_n(\theta) \xrightarrow{P} G \quad (3.13)$$

테일러 전개를 이용하여 수식 (3.14)로부터 수식 (3.15)를 유도할 수 있다.

$$\mathbf{0} = \frac{1}{\sqrt{n}} \tilde{S}_n(\theta_0) + \frac{1}{n} \frac{\partial \tilde{S}_n(\theta_0)}{\partial \theta} \sqrt{n} (\tilde{\theta}_n - \theta_0) \quad (3.14)$$

수식 (3.14)에서 수식 (3.11)과 수식 (3.13)의 결과에 따라 수식 (3.15)를 따른다.

$$n^{\frac{1}{2}} (\tilde{\theta}_n - \theta_0) \rightarrow N_k \left[0, \{G^T(\theta_0) F^{-1}(\theta_0) G(\theta_0)\}^{-1} \right]. \quad (3.15)$$

4. 트래픽 자료 분석결과

본 연구에서 사용한 자료는 2004년 5월 1일부터 2004년 9월 10일까지 약 5개월 동안 미국 텍사스 주립대학교에서 수집된 자료이다. 12개 도시에서 주고받은 인터넷 트래픽 자료로서 5분 단위로 트래픽 양을 측정하였다. 실제 분석에는 일요일부터 토요일까지 7일간 측정된 자료를 이용하였다. 이 자료는 (<http://www.cs.utexas.edu/users/yzhang/research/AbileneTM>)에서 이용할 수 있다. 다음 그림 1은 원자료의 시계열 그림이다. 시계열 자료가 정상성으로 보이지 않기 때문에 차분을 통하여 정상 시계열로 변환시켜 준다.

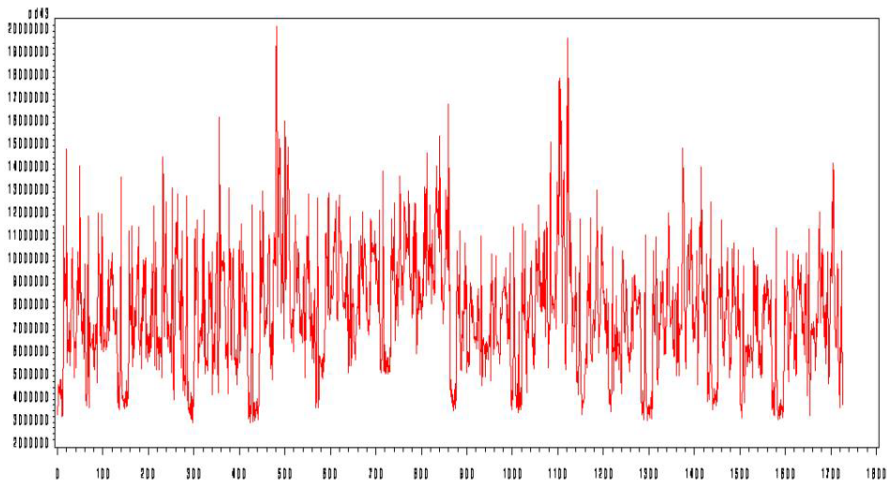


그림 1: 원자료의 시계열 그림

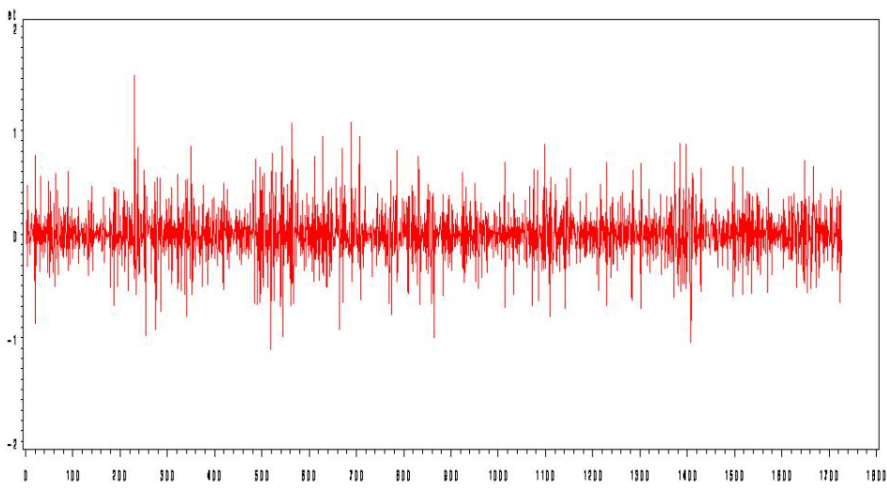


그림 2: 변환된 자료의 시계열 그림

원자료에 로그변환 및 차분을 통해 변환된 자료의 시계열도 그림 2에 나타나 있다. 변환한 시계열 자료의 시계열 그림을 보면 이분산성이 존재할 것으로 예상할 수 있는데, 분산의 군집화가 나타남을 확인할 수 있다. 실제 트래픽 자료가 이상치가 없는 자료로 간주하고, 이 자료에 5개의 이상치 값을 랜덤하게 부여한다. 이상치는 243, 275, 500, 923, 1145 시점에 부여되었고 이 시점들은 랜덤하게 선택되었다. 이상치 값은 2.5로 부여되었다. 이상치 값이 부여된 후의 그림은 그림 3에 나타나 있다.

이상치가 부여된 자료를 이용하여 앞에서 설명한 두 가지 방법에 대해서 시뮬레이션 하였다. 먼저 이상치를 탐지하기 전의 모형 적합 결과는 표 1과 같다.

Charles와 Darne (2005)이 제안한 방법에 따른 이상치 탐지 결과는 표 2에 요약되어 있다. 시점 243, 923, 1145, 500, 275 순서로 이상치로 생각되는 값들 5개가 탐지 되었으며, 각각의 t_{max} 값은

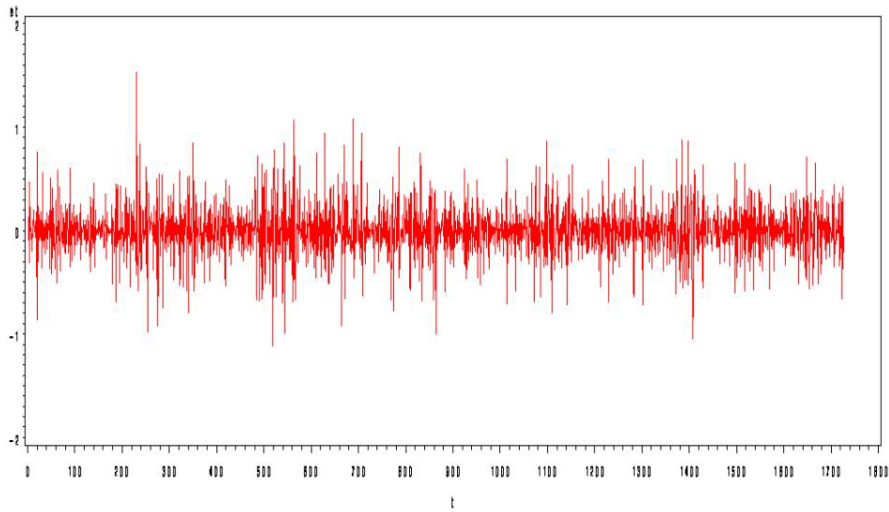


그림 3: 이상치 부여 후 시계열 그림

표 1: 이상치 부여 후 모형 적합 결과

모수	추정량	표준편차	유의확률
α_0	0.0207	0.0011	<.0001
α_1	0.4370	0.0273	<.0001
β_1	0.4266	0.0261	<.0001

표 2: 이상치 탐지 결과

탐지 순서	시점	tmax	대체값
1	243	18.378	0.280
2	923	18.962	0.321
3	1145	19.377	0.469
4	500	19.970	0.496
5	275	20.213	0.634
6		7.834	

표 3: 이상치 탐지 및 대체 후 모형 적합 결과

모수	추정량	표준편차	유의확률
α_0	0.0136	0.0016	<.0001
α_1	0.2521	0.0246	<.0001
β_1	0.5612	0.0340	<.0001

18.378, 18.962, 19.377, 19.970, 20.213으로 모두 10보다 큰 것을 알 수 있다. 탐지된 이상치 값들은 $e_{243}^* = 0.28$, $e_{923}^* = 0.321$, $e_{1145}^* = 0.469$, $e_{500}^* = 0.496$, $e_{275}^* = 0.634$ 로 대체되었다. 5개의 이상치 값을 탐지하고 대체된 후 계산한 tmax 값은 7.834로 10보다 작은 것을 알 수 있다. 그러므로 반복을 멈추게 된다. 앞서 부여되었던 이상치 5개를 모두 정확히 탐지해낸 것을 확인할 수 있었고, 또 tmax 값이 7.834로 10보다 작기 때문에 이상치가 없다고 판단된다. 이상치 대체 후 모형적합 결과는 표 3과 같다.

Basawa 등 (1985)가 제안한 Huber의 ψ 함수를 이용하여 추정된 로버스트 추정량 추정하기로 한다. Huber의 ψ 함수에서 양의 실수 k 값은 0.9로 하기로 하였다. 추정된 로버스트 추정량은 표 4와 같다.

표 4: Huber ψ 함수를 이용한 로버스트 추정량 결과

모수	추정량	표준편차	유의확률
α_0	0.0134	0.0016	<.0001
α_1	0.2334	0.0279	<.0001
β_1	0.5759	0.0386	<.0001

표 5: 각각의 방법에 따른 RMSE

	이상치 처리 전	이상치 탐지 및 대체	로버스트 추정량
RMSE	0.2930	0.2581	0.2575

RMSE(Root Mean Square Error)를 통해서 앞에서 살펴본 두가지 방법을 비교하기로 한다. RMSE는 수식 (4.1)과 같이 계산된다.

$$RMSE = \sqrt{\frac{1}{n} \sum_{t=1}^n (X_t - \hat{X}_t)^2}. \tag{4.1}$$

계산된 RMSE 값은 표 5와 같다. 그 결과 아무 방법도 적용하지 않고 모형 적합 후 계산된 RMSE는 0.2930으로 RMSE가 가장 큰 것을 알 수 있다. 또한 앞에서 살펴본 두 가지 방법을 적용하였을 때 RMSE가 상당히 줄어드는 것을 알 수 있다. Huber ψ 함수를 이용한 로버스트 추정량 적합결과에 대한 RMSE가 이상치 탐지 및 대체 후 적합한 결과의 RMSE보다 다소 작은 것을 알 수 있다.

5. 결론

본 연구에서는 이상치 탐지 방법과 로버스트 추정량에 대해서 논의 해 보았다. 이 두 가지 방법은 이상치가 존재할 경우 추정 시 생길 수 있는 문제에 대한 대처 방안으로 생각할 수 있다. 이상치가 존재하는 자료에 대해서 일반적인 모형추정과 앞에서 살펴본 두 가지 방법을 이용한 모형추정을 적합해 보았다. 시뮬레이션 결과 살펴본 두 가지 방법 모두 RMSE 값이 일반적인 모형추정의 RMSE보다 작게 추정되는 것을 확인할 수 있었다. 일반적인 추정방법은 이상치의 존재 유무에 따라 민감한 것이 알려져 있다. 이를 통해서 이상치 탐지 및 대체 하는 방법과 로버스트 추정량의 방법은 이상치가 존재하는 자료에서 좋은 효율성을 가진다고 할 수 있다. 다만 이상치 탐지 및 대체하는 방법은 그 과정 자체가 로버스트 추정량 보다는 좀 더 복잡하고, 시뮬레이션 결과 로버스트 추정량의 RMSE가 다소 작은 것으로 보아 Huber ψ 함수를 이용한 로버스트 추정량이 더 간편하고 좋은 효율성을 가진다고 사료된다.

참고 문헌

하명호, 김삼용 (2010). 재무 시계열 자료의 이상치 탐지 기법 연구, <응용통계연구>, **23**, 41-47.
 Basawa, I. V., Huggins, R. M. and Staudte, R. G. (1985). Robust test for time series with an application to first-order autoregressive processes, *Biometrika*, **72**, 559-571.
 Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity, *Journal of Econometrics*, **31**, 307-327.
 Box, G. E. P. and Jenkins, G. M. (1976). *Time Series Analysis: Forecasting and Control*, Holden Day, San Francisco.
 Charles, A. and Darne, O. (2005). Outliers and GARCH models in financial data, *Journal of Economics Letters*, **86**, 347-352.
 Denby, L. and Martin, R. D. (1979). Robust estimation of the first order autoregressive parameter, *Journal of the American Statistical Association*, **74**, 140-146.

- Engle, R. F. (1982). Autoregressive conditional heteroskedasticity with estimates of the variance of UK inflation, *Econometrica*, **50**, 987–1007.
- Fox, A. J. (1972). Outliers in time series, *Journal of Royal Statistical Society B*, **34**, 350–363.
- Franses, P. H. and Ghijssels, H. (1999). Additive outliers, GARCH and forecasting volatility, *International Journal of Forecasting*, **15**, 1–9.
- Huber, P. (1981). *Robust Statistics*, Wiley, New York.
- Martin, R. D. (1980). *Robust Estimation of Autoregressive Models (with discussion)*. *Directions in Time Series*, The Institute of Mathematical Statistics Publishers, Hayward, 228–262.
- Martin, R. D. (1982). The Cramer-Rao bound and robust M-estimates for autoregressions, *Biometrika*, **69**, 437–442.
- Martin, R. D. and Yohai, V. J. (1985). *Robustness in Time Series and Estimating ARMA Models*, Handbook of Statistics, 4, Elsevier, New York.
- Shao, Q. (2008). Robust estimation for periodic autoregressive time series, *Journal of Time Series Analysis*, **29**, 251–263.

2011년 2월 접수; 2011년 3월 채택

A Comparative Study of a Robust Estimate Method for Abnormal Traffic Detection

Jae-Yoon Jung^a, Sahm Kim^{1,a}

^aDepartment of Statistics, Chung-Ang University

Abstract

This paper shows the performance evaluation of a robust estimator based on the GARCH model. We first introduce the method of a robust estimate in the GARCH model and the method of an outlier detection in the GARCH model. The results of the real internet traffic data show the out-performance of the robust estimator over the outlier detection method in the GARCH model. In addition, the method of the robust estimate is less complex than the method of the outlier detection method in the GARCH model.

Keywords: Robust estimation, outlier detection, traffic data, GARCH model.

This research was supported by the research fund of ChungAng University in 2009.

¹ Corresponding author: Professor, Department of Statistics, Chung-Ang University, 221 Heukseok-Dong, Dongjack-Gu, Seoul 156-756, Korea. E-mail: sahm@cau.ac.kr