

변환된 자기회귀이동평균 모형에서의 예측구간추정*

조혜민¹⁾ 오승언²⁾ 여인권³⁾

요약

시계열자료를 분석하는데 있어 중요한 목적 중에 하나가 미래값에 대한 예측이다. 일반적으로 자기회귀이동평균모형에서는 백색잡음이 정규분포를 따른다는 가정 하에서 모수의 추론과 예측 및 예측구간의 추정이 이루어지고 있다. 그러나 자료가 이러한 가정을 만족하지 않는 경우, 자료를 가정에 맞게 변환시킨 후 분석하는 방법을 생각해 볼 수 있다. 이 논문에서는 변환된 자료를 분석하여 얻은 결과를 이용하여 본래의 척도에서의 미래값에 대한 예측구간을 추정하는 문제에 대해 알아본다. 제안하는 방법에서는 먼저 적절한 변환을 이용하여 자료를 정규가정을 만족하도록 변환시키고 변환된 자료를 이용하여 미래값에 대한 예측구간을 추정한 후, 역변환을 이용하여 예측구간을 추정한다. 이 논문에서는 시계열분석에서 모델링이 상대적으로 어려운 왜도의 문제를 해결하기 위해 Yeo-Johnson 변환을 중심으로 한 방법론을 소개한다. 모의실험 결과 제안된 방법의 의한 단측예측구간의 포함확률이 변환을 사용하지 않은 구간보다 명목수준에 가까운 것을 확인하였다.

주요용어: 자기회귀이동평균모형, 포함확률, Yeo-Johnson 변환.

1. 서론

경제시계열자료에 대한 고전적 분석모형에서는 일반적으로 자료가 정규분포를 따른다고 가정한다. 그러나 Fama (1965), Li (1999), McCulloch (1996), McDonald (1996) 등이 실증분석에서 보인 것처럼 실제 자료들은 정규분포에 비해 양쪽 꼬리부분이 두텁거나 한쪽으로 꼬리가 길게 뻗은 형태를 가지는 것으로 알려져 있다. 이러한 형태는 정규분포에 비해 꼬리부분에서 관측되는 자료의 빈도가 높으며 이상점이 발생할 가능성이 높다는 것을 의미한다. 정규분포 가정 하에서의 추론은 평균과 분산을 중심으로 이루어지고 이들 값은 이상점에 로버스트하지 않다. 그렇기 때문에 이상점이 존재하는 경우 정규가정 하에서 얻어진 결과는 신뢰성이 떨어질 수 있는데 이런 문제는 크게 두 가지 방법으로 해결할 수 있다.

* 본 연구는 숙명여자대학교 2006년도 교내연구비 지원에 의해 수행되었음.

이 논문은 2006년 두뇌한국21 사업의 지원을 받아 연구되었음.

- 1) (140-742) 서울특별시 용산구 청파동 2가, 숙명여자대학교 통계학과, 석사과정
E-mail: coolmin98@hanmail.net
- 2) (140-742) 서울특별시 용산구 청파동 2가, 숙명여자대학교 통계학과, 석사과정
E-mail: ruma0819@sookmyung.ac.kr
- 3) (교신저자)(140-742) 서울특별시 용산구 청파동 2가, 숙명여자대학교 통계학과, 부교수
E-mail: inkwon@sookmyung.ac.kr

첫 번째 방법은 오차항에 대한 분포로 정규분포 대신 t -분포나 분산이 다른 정규분포의 혼합분포 등을 가정하는 것이다. 이 방법은 일반적으로 대칭이지만 양쪽 꼬리부분이 두터운 형태를 가지는 자료 분석에 많이 활용되고 있으며 가정한 분포가 자료에 적합하지에 따라 분석결과와 적절성이 결정된다. 자료의 분포가 비대칭적 구조를 가지는 경우에는 Nelson (1991)의 일반화된 오차분포나 비대칭 t -분포 등을 이용할 수 있으나 대칭인 경우에 비해 분포가 복잡하여 통계적 추론과정과 해석이 쉽지 않다는 단점이 있다. 두 번째 방법은 모형의 가정에 근사하도록 자료를 변환시키는 것으로 앞의 방법에 비해, 특히 비대칭인 경우, 쉽게 적용시킬 수 있다는 장점이 있다. 변환을 사용하는데 있어 주의해야 할 점은 얻어진 통계추론의 결과는 변환된 자료에 근거한 것으로 본래의 척도에 대한 결과분석 및 해석을 하기 위해서 역변환이 필요한 경우가 있다는 것이다.

시계열자료 분석에 있어 가장 중요한 목적 중에 하나가 미래값에 대한 예측이다. 미래값에 대한 예측은 기존 자료의 정보를 이용하여 이루어지기 때문에 결과는 항상 변동성을 내포하고 있어 예측값 만큼 변동성을 포함한 예측구간이 자료분석에 있어 중요한 역할을 한다. 또한 위험관리 분야에서 많이 사용되고 있는 위험측도 VaR (Value at Risk)의 경우, 수익률 분포에서의 중심위치보다 1%, 2.5%, 5% 등의 위수가 더욱 중요한 값으로 사용되고 있다. 이러한 값은 분포의 형태에 영향을 많이 받는다. 일반적인 자기회귀이동평균모형은 백색잡음이 정규분포를 따른다는 가정 하에 모수의 추론과 예측 및 예측구간 추정이 이루어지고 있다. 이 논문에서는 자기회귀이동평균모형에서 시계열 자료가 정규가정이 만족되지 않는 경우, 적절한 변환을 통하여 자료를 정규분포에 근사하도록 만든 후 변환된 자료를 분석하여 얻은 결과를 바탕으로 모수를 추정하고 예측구간을 구한 후 다시 역변환을 취하여 본래 척도에서의 미래 값에 대한 예측구간을 구하는 방법에 대해 알아본다. 모의실험을 통해 제안된 방법이 변환을 이용하지 않았을 때와 어떤 차이가 있고 얼마나 예측성능이 개선되었는지를 알아본다.

2. 변환선택

서론에서 언급한 것과 같이 자료가 대칭인 경우에는 기존의 방법으로도 쉽게 모형화할 수 있기 때문에 이 논문에서는 비대칭적인 구조에 적합한 변환을 중심으로 방법을 소개한다. 비대칭적 자료를 정규분포에 근사시킬 때 일반적으로 Box-Cox (1964) 변환이 많이 사용되고 있다. Box-Cox 변환은 $x > 0$ 와 변환모수 λ 에 대해 다음과 같이 정의된다.

$$h(x, \lambda) = \begin{cases} (x^\lambda - 1) / \lambda, & \lambda \neq 0, \\ \log(x), & \lambda = 0. \end{cases}$$

Box-Cox 변환은 $\lambda < 1$ 일 때 오목 (concave) 함수가 되고 $\lambda > 1$ 일 때 볼록 (convex) 함수가 되어, von Zwet (1964)의 상대적 왜도의 정의에 의하면, 자료의 왜도를 조정하는데 적절한 변환이 된다. 문제는 양의 관측값에 대해서만 정의할 수 있기 때문에 수익률과 같은 금융 시계열자료 분석에는 적용하기 어렵다는 것이다. 변환에 아래와 같은 하한모수 τ 를 추가하

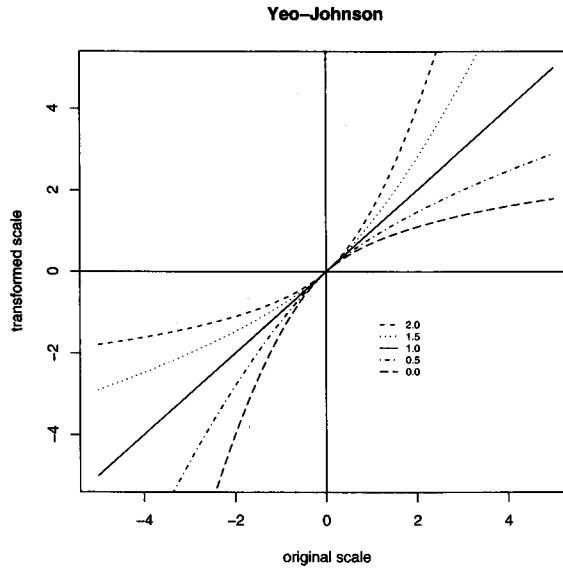


그림 2.1: 변환모수에 따른 Yeo-Johnson 변환의 비교

면, $-\tau$ 보다 큰 음수인 자료에 대해서는 정의할 수 있다.

$$h(x, \lambda, \tau) = \begin{cases} \{(x + \tau)^\lambda - 1\} / \lambda, & \lambda \neq 0, \\ \log(x + \tau), & \lambda = 0 \end{cases}$$

하한모수 τ 를 모르는 경우, λ 와 함께 추정해야 하는데 Atkinson과 Pericchi (1991)에 의하면 λ 의 통계적 추론이 τ 에 영향을 많이 받으며 분포의 지지 (support)가 모수에 영향을 받고 있어 고전적 점근이론을 적용할 수 없다.

변환을 선택하는데 있어 기준이 되는 것은 자료의 형태이다. 비정규 분포의 형태를 여러 가지로 표시할 수 있겠지만 Pearson (1930)이 언급했던 것처럼 왜도 (skewness)와 첨도 (kurtosis)가 비정규성을 나타내는 측도로 사용될 수 있다. 왜도는 비대칭 형태의 정도를 나타내는 측도로 경기지표 변화율의 경우 경기가 상승할 때보다 하락할 때 영향을 더 많이 받기 때문에 일반적으로 음의 왜도를 가지는 경향이 있다. 즉, 분포의 왼쪽 꼬리부분이 오른쪽 보다 길게 뻗어 있는 형태이다. 왜도의 문제를 해결하기 위해서는 von Zwet (1964)가 언급한 것과 같이 변환의 형태는 전체적으로 볼록이거나 오목해야 한다. 이러한 형태를 가지면서 실수 전체구간에서 정의되는 변환으로는 Yeo-Johnson (2000)변환이 있다.

$$h(x, \lambda) = \begin{cases} \{(x + 1)^\lambda - 1\} / \lambda, & \lambda \neq 0, x \geq 0, \\ \log(x + 1), & \lambda = 0, x \geq 0, \\ -\{(-x + 1)^{2-\lambda} - 1\} / (2 - \lambda), & \lambda \neq 2, x < 0, \\ -\log(-x + 1), & \lambda = 2, x < 0. \end{cases}$$

그림 2.1에서 보는 것과 같이 Yeo-Johnson 변환의 경우에는 $\lambda > 1$ 이면 전체적으로 불룩하기 때문에 음의 왜도 문제를 해결할 수 있고 $\lambda < 1$ 이면 오목한 형태를 보이고 있으므로 양의 왜도를 줄일 수 있다.

3. ARMA 모형에서의 예측구간추정

시계열 분석의 가장 중요한 목적 중의 하나는 과거의 패턴이 미래에도 계속 유지된다는 가정 하에 수집된 자료를 분석한 결과를 바탕으로 미래의 값을 예측하는 것이다. 자료를 관측하고 모형을 적합 시키는 과정에서 오차가 수반되며 오차의 확률분포에 근거하여 예측값을 구하게 되는데 가능한 오차를 작게 하는 예측값을 최적의 값으로 생각할 수 있다. 이러한 관점에서 일반적으로 많이 사용되는 기준이 최소평균제곱오차 (Minimum Mean Square Error, MMSE) 예측이다. MMSE 예측은 시점 t 에서의 시계열 자료 Z_t 에 대한 예측값으로 평균제곱오차 $E[Z_t - \hat{Z}_t]^2$ 를 최소로 만드는 \hat{Z}_t 를 사용한다.

시계열 $\{Z_t\}_{t=1}^n$ 이 평균이 μ 인 정상 자기회귀이동평균 (ARMA)과정을 따른다고 하면 모형식은 다음과 같이 쓸 수 있다.

$$\begin{aligned}\Phi(B)(Z_t - \mu) &= \Theta(B)\varepsilon_t, \\ \Phi(B) &= 1 - \phi_1 B - \dots - \phi_p B^p, \\ \Theta(B) &= 1 - \theta_1 B - \dots - \theta_q B^q.\end{aligned}$$

여기서 B 는 후진작용소를 나타낸다. 정상성을 만족하므로 아래와 같이 이동평균 모형의 형태로 다시 표현할 수 있다.

$$Z_t = \mu + \Psi(B)\varepsilon_t = \mu + \sum_{j=0}^{\infty} \psi_j B^j \varepsilon_t.$$

여기서 $\Psi(B) = \Theta(B)/\Phi(B)$ 이고 $\psi_0 = 1$ 이다. 시점 n 으로부터 l -시차 후의 시계열 Z_{n+l} 에 대한 MMSE 예측값을 $Z_n(l)$ 이라고 하면, $Z_n(l)$ 은 조건부 기대값을 이용하여 다음과 같이 구할 수 있다.

$$Z_n(l) = E(Z_{n+l} | Z_n, Z_{n-1}, \dots) = \mu + \sum_{j=l}^{\infty} \psi_j \varepsilon_{n+l-j}.$$

관측값과 예측값의 차인 예측오차는 $e_n(l) = Z_{n+l} - Z_n(l) = \sum_{j=0}^{l-1} \psi_j \varepsilon_{n+l-j}$ 으로 $E(e_n(l)) = 0$ 이고 분산은 다음과 같다.

$$\text{Var}(e_n(l)) = \text{Var}\left(\sum_{j=0}^{l-1} \psi_j \varepsilon_{n+l-j}\right) = \sigma_\varepsilon^2 \sum_{j=0}^{l-1} \psi_j^2.$$

오차 ε_t 들이 독립이고 $N(0, \sigma_\varepsilon^2)$ 를 따른다면, ε_t 의 선형결합인 예측오차는 기대값이 0이고 분산이 $\sigma_\varepsilon^2 \sum_{j=0}^{l-1} \psi_j^2$ 인 정규 분포를 따르게 된다. 또한 $\{Z_t\}_{t=1}^n$ 이 주어졌을 때 Z_{n+l} 의 조건

부 분포는 $N\left(Z_n(l), \sigma_\epsilon^2 \sum_{j=0}^{l-1} \psi_j^2\right)$ 가 되므로 다음과 같은 확률을 구할 수 있다.

$$P\left\{-z_{\alpha/2} \leq \frac{Z_{n+l} - Z_n(l)}{\sigma_\epsilon \sqrt{\sum_{j=0}^{l-1} \psi_j^2}} \leq z_{\alpha/2}\right\} = 1 - \alpha. \quad (3.1)$$

식 (3.1)를 이용하여 Z_{n+l} 에 대한 $100(1 - \alpha)\%$ 양측 예측구간을 다음과 같이 추정할 수 있다.

$$\hat{Z}_n(l) \pm z_{\alpha/2} \hat{\sigma}_\epsilon \sqrt{\sum_{j=0}^{l-1} \hat{\psi}_j^2}. \quad (3.2)$$

여기서, $Z_{\alpha/2}$ 는 표준정규분포의 $(1 - \alpha/2)$ 위수를 의미한다. 정칙조건 하에서 최대가능도 추정량 $\hat{\sigma}_\epsilon$ 과 $\hat{\psi}$ 은 모수 σ_ϵ 과 ψ 의 일치추정량이 되므로 추정된 예측구간 (3.2)은 n 이 커지면 포함확률 (coverage probability)이 명목확률 (nominal probability)과 같아지는 점근적으로 옳은 (asymptotically correct) 예측구간이 된다.

자료가 오차항의 정규성 가정을 만족하지 않는 경우, (3.2)은 점근적으로도 옳은 예측구간이라는 보장이 없다. 이런 경우 적절한 변환을 통해 자료를 정규분포에 근사하도록 변환시킨 후 이를 바탕으로 예측구간을 추정하는 것이 보다 옳은 예측구간이 될 것이다. 변환된 확률변수 $h(Z_1, \lambda)$, $h(Z_2, \lambda), \dots, h(Z_n, \lambda)$ 들은 평균이 ν 인 정상 ARMA 모형을 따른다고 가정하면, 모형식은 다음과 같이 쓸 수 있다.

$$\Phi(B) \{h(Z_t, \lambda) - \nu\} = \Theta(B)\epsilon_t.$$

오차항 ϵ_t 들은 독립이고 $N(0, \sigma_\epsilon^2)$ 이라고 하고 추정해야 하는 모수를 $\xi = (\lambda, \nu, \phi, \theta, \sigma_\epsilon^2)$ 라고 하면 로그 가능도 함수 $L(\xi)$ 는 다음과 같다.

$$L(\xi) = -\frac{n}{2} \log(\sigma_\epsilon^2) - \frac{1}{2\sigma_\epsilon^2} \sum_{t=1}^n \left[\Phi(B) \{h(Z_t, \lambda) - \nu\} + \sum_{j=1}^q B^j \epsilon_t \right]^2 + \sum_{t=1}^n \log\{J(Z_t, \lambda)\}. \quad (3.3)$$

여기서 $J(Z_t, \lambda)$ 는 변환된 확률변수들에 대한 자코비안으로 Yeo-Johnson 변환인 경우 다음과 같이 계산된다.

$$J(x, \lambda) = \frac{\partial}{\partial x} h(x, \lambda) = (|x| + 1)^{\text{sign}(x)(\lambda-1)}.$$

위의 로그가능도 함수를 최대로 만드는 추정량을 구하기 위해서 λ 를 고정시킨다. 고정된 λ 에 대해 $L(\xi)$ 를 최대로 만드는 $\hat{\nu}(\lambda)$, $\hat{\phi}(\lambda)$, $\hat{\theta}(\lambda)$, $\hat{\sigma}_\epsilon^2(\lambda)$ 를 구한 후 이를 (3.3)에 대입하면 λ 의 프로파일 로그가능도함수가 만들어 진다. 이 함수를 최대로 만드는 $\hat{\lambda}$ 를 구하여 (3.3)에 다시 대입하면 최대가능도추정량 $\hat{\xi} = (\hat{\lambda}, \hat{\nu}(\hat{\lambda}), \hat{\phi}(\hat{\lambda}), \hat{\theta}(\hat{\lambda}), \hat{\sigma}_\epsilon^2(\hat{\lambda}))$ 을 얻을 수 있다. 이렇게 구한 모수의 추정 값을 예측구간 (3.2)에 대입하여 변환된 척도에서의 $100(1 - \alpha)\%$ 예측구간을 다음과 같이 구할 수 있다.

$$\hat{h}(Z_n, \hat{\lambda})(l) \pm z_{\alpha/2} \hat{\sigma}_\epsilon(\hat{\lambda}) \sqrt{\sum_{j=0}^{l-1} \hat{\psi}_j^2(\hat{\lambda})}.$$

변환된 척도에서의 추정된 예측구간의 경계점을 (L, U) 이라고 할 때 다음과 같은 역변환을 이용하여 본래 척도에서의 예측구간을 추정할 수 있다.

$$h^{-1}(x, \lambda) = \begin{cases} (\lambda x + 1)^{1/\lambda} - 1, & x \geq 0, \lambda \neq 0, \\ \exp(x) - 1, & x \geq 0, \lambda = 0, \\ 1 - \{(\lambda - 2)x + 1\}^{1/(2-\lambda)}, & x < 0, \lambda \neq 2, \\ 1 - \exp(-x), & x < 0, \lambda = 2. \end{cases}$$

즉, 본래 척도에서의 추정된 예측구간 (L^*, U^*) 는 $L^* = h^{-1}(L, \hat{\lambda})$, $U^* = h^{-1}(U, \hat{\lambda})$ 가 된다.

Remark 3.1: Yeo-Johnson 변환의 경우, $\lambda < 0$ 일 때 변환된 값은 항상 $-\lambda^{-1}$ 보다 작고 $\lambda > 2$ 일 때 항상 $(2 - \lambda)^{-1}$ 보다 큰 값을 가진다. 그러므로 $\hat{\lambda}$ 가 0보다 작거나 2보다 큰 경우, 변환된 척도에서의 예측구간 L 또는 U 에 따라 역변환을 적용할 수 없을 수도 있다. 이런 경우가 발생하면 아주 크거나 작은 값으로 포함확률에 영향을 주지 않는 범위내에서 대체 값을 정하여 예측구간의 경계를 구해야 한다.

Remark 3.2: 변환-역변환 방법을 이용하여 기대값을 추정하는데 있어 주의해야 할 것은 h 가 선형이 아니면 $E\{h(Z, \lambda)\}$ 와 $h\{E(Z), \lambda\}$ 가 같지 않다는 것이다. 그러므로 MMSE 예측을 이용하여 미래값을 예측할 때에는 단순히 변환-역변환을 적용하면 편향의 문제가 발생할 수 있다. 이러한 문제는 Duan (1983)의 smearing 추정법을 사용하여 해결할 수 있다.

4. 모의실험

변환을 통해 예측구간의 신뢰도가 얼마나 향상 되는지 알아보기 위해 AR(1)과 MA(1) 모형을 이용하였다. AR(1)과 MA(1)에 대한 MMSE 예측값과 분산은 다음과 같다.

- AR(1)

$$\begin{aligned} Z_n(l) &= \mu + \phi^l(Z_n - \mu), \\ \text{Var}(e_n(l)) &= \sigma_\varepsilon^2 \left(\frac{1 - \phi^{2l}}{1 - \phi^2} \right). \end{aligned}$$

- MA(1)

$$\begin{aligned} Z_n(l) &= \begin{cases} \mu - \theta\varepsilon_n, & l = 1, \\ \mu, & l \geq 2, \end{cases} \\ \text{Var}(e_n(l)) &= (1 + \theta^2) \sigma_\varepsilon^2. \end{aligned}$$

모의실험을 위한 자료의 생성, 변환모수를 포함한 모수의 추정, 예측구간의 계산은 Matlab 프로그램을 이용하였다. 구체적 모의실험과정은 다음과 같다. Matlab에서 지원하고 있는 garchsim 함수를 이용하여 오차항이 표준정규백색잡음이고 $\phi = 0.5$ 인 AR(1)과 $\theta = 0.5$ 인

표 4.1: Yeo-Johnson 변환을 통해 얻은 자료에 의한 포함확률

모형	λ	변환사용								
		변환사용하지 않음			정규분포위수			t -분포위수		
		L	R	B	L	R	B	L	R	B
AR(1) $n = 50$	0.0	0.995	0.928	0.937	0.937	0.934	0.925	0.941	0.937	0.932
	0.5	0.974	0.916	0.930	0.933	0.933	0.923	0.936	0.937	0.930
	1.0	0.935	0.931	0.922	0.933	0.930	0.920	0.937	0.935	0.926
	1.5	0.918	0.978	0.934	0.935	0.935	0.926	0.939	0.938	0.932
	2.0	0.921	0.994	0.930	0.930	0.932	0.925	0.934	0.935	0.932
AR(1) $n = 100$	0.0	0.999	0.934	0.947	0.951	0.940	0.939	0.952	0.942	0.942
	0.5	0.984	0.912	0.934	0.945	0.933	0.933	0.946	0.934	0.936
	1.0	0.944	0.943	0.934	0.944	0.942	0.935	0.945	0.943	0.939
	1.5	0.918	0.984	0.939	0.939	0.948	0.937	0.940	0.949	0.939
	2.0	0.939	0.998	0.949	0.943	0.945	0.938	0.944	0.947	0.941
MA(1) $n = 50$	0.0	0.999	0.931	0.943	0.947	0.939	0.938	0.951	0.942	0.943
	0.5	0.980	0.927	0.944	0.945	0.944	0.938	0.949	0.947	0.943
	1.0	0.948	0.946	0.941	0.946	0.942	0.937	0.948	0.945	0.942
	1.5	0.927	0.980	0.940	0.938	0.942	0.933	0.942	0.945	0.939
	2.0	0.930	0.999	0.943	0.938	0.943	0.932	0.942	0.947	0.939
MA(1) $n = 100$	0.0	0.999	0.942	0.953	0.950	0.946	0.943	0.952	0.948	0.946
	0.5	0.985	0.931	0.949	0.949	0.948	0.943	0.9504	0.950	0.945
	1.0	0.948	0.951	0.947	0.947	0.948	0.944	0.949	0.949	0.947
	1.5	0.931	0.986	0.947	0.946	0.945	0.944	0.948	0.947	0.948
	2.0	0.937	0.998	0.947	0.942	0.950	0.943	0.944	0.951	0.946

MA(1) 과정의 시계열 $\{X_t\}_{t=1}^{n+10}$ 를 생성한다. Yeo-Johnson 변환의 역함수를 이용하여 시계열 자료 $Z_t = h^{-1}(X_t, \lambda)$ 를 얻는다. 이렇게 하면, $h(Z_t, \lambda)$ 는 정규성을 만족하는 시계열이 된다.

변환을 사용하지 않고 예측구간을 구하는 경우, 시계열 $\{Z_t\}_{t=1}^n$ 를 가지고 garchfit 함수를 이용하여 모수, 예측값, 예측오차의 분산을 추정하여 예측구간을 구한다. 추가로 생성된 10개의 자료를 가지고 예측구간에 포함되어 있는지를 확인한다. 변환을 사용하는 경우에는 함수 garchfit에서 제공하는 로그가능도에 자코비안 항을 추가하여 프로파일 로그가능도함수 값을 계산하고 fminbnd 함수를 이용하여 $[-3, 5]$ 사이에서 프로파일 로그가능도 함수를 가장 크게 만드는 $\hat{\lambda}$ 을 구한다. 추정된 $\hat{\lambda}$ 를 이용하여 자료를 변환시킨 후 garchfit 함수를 이용하여 예측구간의 경계를 계산한 후 이 경계 값들을 역변환을 이용하여 본래 척도에서의 예측구간을 구한다. 이렇게 구한 예측구간에 추가로 생성된 10개의 자료가 포함되

어 있는지를 계산한다.

이 논문에서는 이러한 작업을 1,000번 반복하여 10,000개의 자료에 대한 예측구간을 구하여 포함확률을 추정하였다. 표 4.1는 $n = 50, 100$ 이고 예측명목수준을 95%로 했을 때 양측(B) 및 단측(L,R) 예측구간의 포함확률을 추정한 것이다. 변환을 이용하여 예측구간을 추정한 경우, 정규분포와 자유도가 $n - 4$ 인 t -분포로부터 구한 위수를 각각 사용하였다. 표 4.1의 결과를 정리하면 다음과 같다.

- 변환을 사용하지 않은 경우, 양측구간에 대한 포함확률에서는 대부분 큰 문제가 발생하지 않았으나 단측구간의 경우 과대포함 또는 과소포함의 현상을 보이고 있다. 양의 왜도가 심한 경우, AR과 MA 과정 모두에서 단측구간의 하한값이 상대적으로 너무 작아 예측구간에 미래값의 대부분이 포함되고 있으며 n 이 클수록 더 뚜렷한 형태를 보이고 있다. 음의 왜도를 가지는 경우, 상한값에 대해 같은 현상을 일어나는 것을 볼 수 있다. AR(1)과정에서 왜도가 심하지 않은 $\lambda = 0.5$ 또는 $\lambda = 1.5$ 의 경우, 반대 현상이 일어나 포함확률이 0.92보다 작은 값을 가지는 경우가 있는 것으로 나타났다.
- 변환을 사용한 경우, 좌측예측구간과 우측예측구간의 포함확률이 비슷하게 나오고 있으며 n 커질수록 명목수준에 근접하는 것을 볼 수 있다. 양측구간의 포함확률은 명목수준에 비해 전반적으로 낮은 것을 나타냈으며 특히 n 이 작은 AR 과정에서 상당히 작은 것으로 나타났다.

5. 결론

이 논문에서는 변환된 자기회귀이동평균모형에서 미래값에 대한 예측구간 추정에서의 문제를 살펴보았다. 모의실험을 통해, 자료가 왜도가 심한 경우 단측구간에 대한 포함확률에서 과대 또는 과소 포함되는 문제가 발생하는 것을 확인하였으며 제안된 방법은 이런 문제를 해결해 주는 것으로 나타났다. 이는 경제시계열과 관련된 자료에서 흔히 접하는 현상으로 VaR과 같은 위험관리에서 쉽게 활용될 수 있을 것으로 생각된다. 또한 양측구간의 경우 작은 n 에 대해서는 제안한 방법에 개선이 필요하다는 것을 인식하였다. 개선방안으로 신뢰수준에 대한 임계값으로 표준정규분포나 t -분포의 위수가 아닌 Cho와 Loh (2005)에서 언급된 표준화된 잔차로부터 구한 표본위수를 이용하는 방법등을 고려할 수 있다.

참고문헌

- Atkinson, A. C. and Pericchi, L. R. (1991). Grouped likelihood for the shifted power transformation, *Journal of the Royal Statistical Society, Ser. B*, **53**, 473-482.
- Box, G. E. P. and Cox, D. R. (1964). An analysis of transformations, *Journal of the Royal Statistical Society, Ser. B*, **26**, 211-252.
- Cho, K and Loh, W.-Y. (2005). Bias and convergence rate of the coverage probability of prediction intervals in Box-Cox transformed linear models, *Journal of Statistical Planning and Inference*, **136**, 3614-3624.

- Duan, N. (1983). Smearing estimate: a nonparametric retransformation method, *Journal of the American Statistical Association*, **78**, 605–610.
- Fama, E. F. (1965). The behavior of stock-market prices, *The Journal of Business*, **38**, 34–105.
- Li, D. (1999). Value at risk based on the volatility skewness and kurtosis, RiskMetrics Group.
- McCulloch, J. H. (1996). Financial applications of stable distribution, *Handbook of Statistics*, **14**, 393–425.
- McDonald, J. B. (1996). Probability distribution for financial models, *Handbook of Statistics*, **14**, 427–461.
- Nelson, D. B. (1991). Conditional heteroskedasticity in asset returns: a new approach, *Econometrica*, **59**, 347–370.
- Pearson, E. S. (1930). A further development of tests for normality, *Biometrika*, **22**, 239–249.
- van Zwet, W. R. (1964). *Convex Transformations of Random Variables*, Mathematisch Centrum, Amsterdam.
- Yeo, I. K. and Johnson, R. A. (2000). A new family of power transformation to improve normality or symmetry, *Biometrika*, **87**, 954–959.

[2007년 4월 접수, 2007년 6월 채택]

Prediction Interval Estimation in Transformed ARMA Models*

Hyemin Cho¹⁾ Sungun Oh²⁾ In-Kwon Yeo³⁾

ABSTRACT

One of main aspects of time series analysis is to forecast future values of series based on values up to a given time. The prediction interval for future values is usually obtained under the normality assumption. When the assumption is seriously violated, a transformation of data may permit the valid use of the normal theory. We investigate the prediction problem for future values in the original scale when transformations are applied in ARMA models. In this paper, we introduce the methodology based on Yeo-Johnson transformation to solve the problem of skewed data whose modelling is relatively difficult in the analysis of time series. Simulation studies show that the coverage probabilities of proposed intervals are closer to the nominal level than those of usual intervals.

Keywords: ARMA models, coverage probability, Yeo-Johnson transformation.

* This Research was supported by the Sookmyung Women's University Research Grants 2006.

This work was supported by Brain Korea 21 project in 2006.

1) Graduate student, Department of Statistics, Sookmyung Women's University, Chungpa-dong 2-ga, Yongsan-gu, Seoul 140-742, Korea

E-mail: coolmin98@hanmail.net

2) Graduate student, Department of Statistics, Sookmyung Women's University, Chungpa-dong 2-ga, Yongsan-gu, Seoul 140-742, Korea

E-mail: ruma0819@sookmyung.ac.kr

3) (Corresponding author) Associate Professor, Department of Statistics, Sookmyung Women's University, Chungpa-dong 2-ga, Yongsan-gu, Seoul 140-742, Korea

E-mail: inkwon@sookmyung.ac.kr