

# 변환-역변환을 통한 자기회귀이동평균모형에서의 예측값 추정\*

여인권<sup>1)</sup> 조혜민<sup>2)</sup>

## 요약

시계열자료 분석에 있어 주요 목적 중에 하나는 미래에 대한 예측값을 추정하는 것이다. 이 논문에서는 정상자기회귀이동평균 모형에서 변환-역변환 방법을 이용하여 예측값을 구하는 과정에서 발생하는 문제에 대해 알아보고 회귀분석에서 제안되었던 smearing 추정방법을 시계열분석에서 사용할 수 있도록 붓스트랩을 이용하여 수정한 추정법을 소개한다. Yeo-Johnson 변환 (2000)을 이용한 KOSDAQ지수의 수익률 실증분석을 통해 기존에 사용되고 있는 방법의 문제점과 제안된 방법의 적절성에 대해 고찰해 보았다.

주요용어: 붓스트랩, 최소평균제곱오차, smearing 추정, Yeo-Johnson 변환.

## 1. 서론

금융시계열자료에 대한 분석은 일반적으로 자료가 정규분포 또는 대칭적인 분포를 따른다는 가정 하에서 이루어진다. 그러나 경제적 상황은 호재(good news)보다 악재(bad news)에 더 영향을 받으며 이에 따라 많은 금융자료들이 음의 왜도를 가지는 것으로 Pagan과 Schwert (1990)와 Engle과 Ng (1993)등에 의해 보고되었다. 이런 문제를 해결하기 위한 연구가 현재 까지 계속 진행되어 오고 있는데 주요 방법론은 다음과 같다. 첫 번째 방법은 왜도를 설명할 수 있는 부분을 모형에 추가한 것으로 Glosten 등 (1993)의 GJR모형이나 Zakoian (1994)의 Threshold GARCH와 같은 비대칭형 변동성모형(asymmetric volatility model)에서 많이 사용되고 있다. 또 다른 방법으로는 오차항에 대한 분포를 대칭인 정규분포나  $t$ -분포가 아닌 비대칭적 구조를 가지는 것으로 가정한 것으로 Nelson (1991)의 일반화된 오차분포 등이 있다. 이 두 방법은 가정한 모형이나 분포가 자료에 얼마나 일치하는가에 따라 분석결과의 타당성이 결정되기 때문에 자료에 대한 모형의 적합성을 추가적으로 확인하는 과정이 필요할 때도 있다. 이들 방법보다 쉽게 적용할 수 있는 방법은 모형의 가정에 근사하도록 자료를 변환시키는 것으로 앞의 방법들보다 상대적으로 쉽게 적용시킬 수 있다는 장점이 있다.

시계열자료 분석의 주요 목적 중에 하나는 과거의 패턴이 미래에도 계속 유지된다는 가정 하에 수집된 자료를 분석한 결과를 바탕으로 미래의 값을 예측하는 것이다. 예측은 일반적으

\* 본 연구는 숙명여자대학교 2007년도 교내연구비 지원에 의해 수행되었음.

이 논문은 2007년 두뇌한국 21 사업의 지원을 받아 연구되었음.

1) (140-742) 교신저자. 서울특별시 용산구 청파동 2가, 숙명여자대학교 이과대학 통계학과, 부교수.

E-mail: inkwon@sookmyung.ac.kr

2) (140-742) 서울특별시 용산구 청파동 2가, 숙명여자대학교 이과대학 통계학과, 석사과정.

E-mail: coolmin98@hanmail.net

로 점추정값인 예측값과 변동성을 고려한 예측구간으로 제시된다. 조혜민 등 (2007)은 Yeo-Johnson 변환 (2000)을 통해 변환된 자료를 이용하여 자기회귀이동평균모형을 적합시켜 얻은 예측구간을 역변환을 통해 본래 척도에서의 예측구간을 구하는 방법에 대해 연구하였다. 이 연구에 의하면 자료가 비대칭적인 구조를 가지는 경우, 단측예측구간의 포함확률이 변환을 사용하지 않은 것보다 사용하는 것이 명목수준에 가까운 것으로 나타났다. 시계열분석에서는 시점  $t$ 에서의 시계열 자료  $Z_t$ 에 대한 예측값으로 평균제곱오차(mean square error: MSE)인  $E\{(Z_t - \hat{Z}_t)^2\}$ 를 최소로 만드는  $\hat{Z}_t$ 을 사용하는데 이를 최소평균제곱오차(minimum MSE: MMSE) 예측값이라고 한다. 시점  $n$ 으로부터  $l$ -시차 후의 시계열  $Z_{n+l}$ 에 대한 MMSE 예측값을  $Z_n(l)$ 이라고 하면,  $Z_n(l)$ 은 다음과 같이 조건부 기대값으로 구할 수 있다.

$$Z_n(l) = E(Z_{n+l} | Z_n, Z_{n-1}, \dots).$$

단조함수에 의해 변환된 분포에서의 위수를 역변환시킨 값은 본래 척도에서의 위수와 같으므로 조혜민 등 (2007)의 예측구간 상하한값 추정에서는 문제가 발생하지 않는다. 그러나 변환  $h$ 가 단조이더라도 선형이 아니면  $E\{h(Y)\}$ 와  $h\{E(Y)\}$ 가 같지 않기 때문에 기대값으로 표시되는 MMSE 예측값을 구할 때 변환된 자료로부터 예측값을 구하고 이를 역변환하여 본래 척도에서의 예측값을 구하면 편향이 발생하는 문제가 있다. 이 논문에서는 이러한 문제를 해결하는 방안을 제시하고 KOSDAQ지수의 수익률분석을 통해 제안하고자 하는 방법의 타당성을 확인한다.

## 2. 변환된 자기회귀이동평균모형에서의 예측값 추정

회귀분석에서 변환-역변환 방법에 대한 연구는 Duan (1983)와 Taylor (1986) 등에 의해 연구되었는데 이 절에서는 Duan (1983)의 smearing 추정법에 대해 간략하게 알아보고 이 방법을 자기회귀이동평균모형에 적용하기 위해 붓스트랩 기법을 이용하여 수정된 추정방법을 소개한다.

### 2.1. Smearing 추정

선형회귀모형에서 원래 척도에서의 반응변수를  $Y_i$ 라 하고 미분 가능한 단조함수  $h$ 를 통해 변환된 값을  $Z_i$ 라 하자. 이 때 변환된 반응변수  $Z$ 와 설명변수  $x$ 간에 다음과 같은 관계가 성립한다고 가정한다.

$$Z_i = h(Y_i) = \mathbf{x}_i^T \boldsymbol{\beta} + \varepsilon_i, \quad i = 1, 2, \dots, n,$$

여기서 Duan (1983)은 오차항  $\varepsilon$ 은 독립이고 평균이 0이고 분산이  $\sigma^2$ 인 임의의 분포  $F$ 를 따른다고 가정하였다. 설명변수  $x$ 들의 설계행렬(design matrix)을  $\mathbf{X}$ 라고 하면  $\boldsymbol{\beta}$ 의 최소제곱추정량은  $\hat{\boldsymbol{\beta}} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{Z}$ 가 된다. 임의의  $x_0$ 에 대해, 변환된 반응변수의 기대값  $E(Z|x_0) = \mathbf{x}_0^T \boldsymbol{\beta}$ 는  $\mathbf{x}_0^T \hat{\boldsymbol{\beta}}$ 으로 추정하고 본래 척도에서의 반응변수 기대값  $E(Y|x_0)$ 에 대한 추정값으로  $h^{-1}(\mathbf{x}_0^T \hat{\boldsymbol{\beta}})$ 를 흔히 사용한다. 그러나 함수  $h$ 가 선형이 아니면, 추정값  $h^{-1}(\mathbf{x}_0^T \hat{\boldsymbol{\beta}})$ 는  $E(Y|x_0)$ 에 대해 비편향성(unbiasedness)뿐만 아니라 일치성(consistency)을 만족하지 않을

수 있으며 모수  $\beta$ 의 참값을 안다고 하더라도  $h^{-1}(\mathbf{x}_0^T \beta)$ 가  $E(Y|\mathbf{x}_0)$ 의 적절한 추정값이라 보기 힘들다. 왜냐하면 다음의 관계가 성립하기 때문이다.

$$E(Y|\mathbf{x}_0) = E\{h^{-1}(Z_0)\} = E\{h^{-1}(\mathbf{x}_0^T \beta + \varepsilon_0)\} \neq h^{-1}(\mathbf{x}_0 \beta).$$

위와 같은 문제를 해결하기 위한 방안으로 Duan (1983)은 다음과 같은 추정법을 제안하였다. 본래척도에서의 반응변수 기대값

$$E(Y|\mathbf{x}_0) = E\{h^{-1}(\mathbf{x}_0^T \beta + \varepsilon_0)\} = \int h^{-1}(\mathbf{x}_0 \beta + \varepsilon_0) dF(\varepsilon_0) \quad (2.1)$$

를 추정하기 위해  $F$ 를 다음과 같은 추정된 잔차의 경험적 분포함수(empirical distribution function)로 추정하는 방법을 고려하였다.

$$\hat{F}_n(e) = \frac{1}{n} \sum_{i=1}^n I(\hat{\varepsilon}_i \leq e),$$

여기서 추정된 잔차는  $\hat{\varepsilon}_i = z_i - \mathbf{x}_i \hat{\beta}$ 이다. 미지의 모수  $\beta$ 를  $\hat{\beta}$ 로 대체하여  $E(Y|\mathbf{x}_0)$ 의 추정값을 다음과 같이 제안하였으며

$$\hat{E}(Y|\mathbf{x}_0) = \int h^{-1}(\mathbf{x}_0 \hat{\beta} + \varepsilon_0) d\hat{F}_n(\varepsilon_0) = \frac{1}{n} \sum_{i=1}^n h^{-1}(\mathbf{x}_0 \hat{\beta} + \hat{\varepsilon}_i)$$

이를 smearing 추정값이라고 불렀다.

회귀모형에서 각각의 관측값이 서로 독립이므로 기대값은 식 (2.1)에서와 같이 단순히  $\varepsilon_0$ 에 대해서만 기대값을 취하여 계산할 수 있다. 자기회귀이동평균모형에서는 어떤 시점의 관측값은 그 시점의 오차뿐만 아니라 과거의 자료 및 오차들로 표시되기 때문에 기대값을 계산하는데 있어 추가적인 작업이 필요하다.

## 2.2. 붓스트랩 추정

시계열 자료  $Y_t$ 를 정규성 가정을 만족 또는 근사하도록 변환을 취하여 얻은 자료를  $Z_t = h(Y_t)$ 라고 하고  $Z_t$ 는 다음과 같이 평균이  $\mu$ 인 정상자기회귀이동평균 과정을 따른다고 가정한다.

$$\Phi(B)(Z_t - \mu) = \Theta(B)\varepsilon_t, \quad t = 1, 2, \dots, n,$$

여기서  $\Phi(B) = 1 - \phi_1 B - \dots - \phi_p B^p$ 는 자기회귀작용소(AR operator)를 나타내고  $\Theta(B) = 1 - \theta_1 B - \dots - \theta_q B^q$ 는 이동평균작용소(MA operator)를 나타낸다. 정상시계열은 Wold정리에 의해 다음과 같은 무한이동평균모형으로 표시할 수 있다.

$$Z_t = \mu + \sum_{j=0}^{\infty} \psi_j \varepsilon_{t-j},$$

여기서  $\psi_0 = 1$ 이고  $\psi_j$ 는  $\Psi(B) = \Phi(B)^{-1}\Theta(B)$ 를 전개할 때  $B^j$ 에 해당되는 계수를 나타낸다.

시점  $n$ 으로부터  $l$ -시차 후의 시계열  $Z_{n+l}$ 에 대한 MMSE 예측 값을  $Z_n(l)$ 이라고 하면, 자기회귀이동평균모형에서의  $Z_n(l)$ 은 다음과 같이 조건부 기대값으로 구할 수 있다.

$$\begin{aligned} Z_n(l) &= E(Z_{n+l}|Z_n, Z_{n-1}, \dots) = \mu + \sum_{j=l}^{\infty} \psi_j \varepsilon_{n+l-j} \\ &= \begin{cases} \mu + \sum_{j=1}^p \phi_j (Z_n(l-j) - \mu) - \sum_{j=l}^q \theta_j \varepsilon_{n+l-j}, & l \leq q, \\ \mu + \sum_{j=1}^p \phi_j (Z_n(l-j) - \mu), & l > q, \end{cases} \end{aligned}$$

여기서,  $k \leq 0$ 에 대해,  $Z_n(k) = Z_{n+k}$ 가 된다. 위의 기대값의 계산과정은,  $j > 0$ 에 대해,  $E(\varepsilon_{n+j}|Z_n, Z_{n-1}, \dots) = 0$ 이 되어 비교적 간단해지는 반면  $h$ 가 비선형인 경우 본래척도에서의 MMSE 예측식은 이것을 사용할 수 없기 때문에 다음과 같이  $n$  시점 이후의 오차항이 그대로 예측식에 포함되어 있다.

$$\begin{aligned} Y_n(l) &= E(Y_{n+l}|Y_n, Y_{n-1}, \dots) = E\{h^{-1}(Z_{n+l})|Z_n, Z_{n-1}, \dots\} \\ &= E\left\{h^{-1}\left(\mu + \sum_{j=0}^{\infty} \psi_j \varepsilon_{n+l-j}\right) \middle| Z_n, Z_{n-1}, \dots\right\} \\ &= E\left\{h^{-1}\left(\mu + \sum_{j=0}^{l-1} \psi_j \varepsilon_{n+l-j} + \sum_{j=l}^{\infty} \psi_j \varepsilon_{n+l-j}\right)\right\}. \end{aligned} \quad (2.2)$$

식 (2.2)에서  $\varepsilon_n, \varepsilon_{n-1}, \dots$ 는  $Z_n, Z_{n-1}, \dots$ 에 의해 주어진 값이고 기대값은  $\varepsilon_{n+1}, \dots, \varepsilon_{n+l}$ 에 대해서만 취한다. 위의 식에서 AR의 차수  $p$ 가 2 이상이 되면  $\psi$ 의 계산이 쉽지 않기 때문에 모형에 AR항이 포함되어 있으면  $Z_{n+l}$ 을  $n+l$ 시점 이전의 시계열로 반복적으로 전개하여  $Z_n, Z_{n-1}, \dots$ 로 모두 표현되도록 하고  $Y_n(l)$ 를 구하여야 한다.

예제 2.1: 시계열 자료가 AR(1)과정을 따른다고 하면, 미래 예측값은 최종 관측자료  $Z_n$ 에 영향을 받으며

$$Y_n(l) = E\left\{h^{-1}\left(\delta \sum_{j=0}^{l-1} \phi^j + \phi^l Z_n + \sum_{j=1}^l \phi^{l-j} \varepsilon_{n+j}\right)\right\}$$

가 되고 만약 MA(1)과정을 따른다면, 예측값은 다음과 같이 구할 수 있다.

$$Y_n(l) = E\{h^{-1}(\mu + \varepsilon_{n+l} - \theta \varepsilon_{n+l-1})\},$$

여기서  $\delta = (1 - \phi)\mu$ 이고 기대값은  $\varepsilon_{n+1}, \dots, \varepsilon_{n+l}$ 에 대해 취한다.

기대값 (2.2)의 계산은 비록  $\varepsilon$ 이 정규분포를 따른다고 가정하여도 쉬운 작업이 아니다. 이 논문에서는 이 기대값을 추정하기 위해 붓스트랩 방법을 제안한다. 시점  $t \leq n$ 에서의  $Z_t$ 의 예측값을  $\hat{Z}_t$ 이라고 하면 예측오차  $\varepsilon_t$ 는 잔차  $\hat{\varepsilon}_t = Z_t - \hat{Z}_t$ 로 대체할 수 있다. 예측오차

$\varepsilon_{n+1}, \dots, \varepsilon_{n+l}$ 에 대한 조건부 기대값의 추정값은 난수를 반복적으로 발생시켜 다음과 같이 표본평균을 계산하여 구할 수 있다.

$$\hat{Y}_n(l) = \frac{1}{B} \sum_{b=1}^B h^{-1} \left( \hat{\mu} + \sum_{j=1}^p \hat{\phi}_j (Z_{n+l-j}^* - \hat{\mu}) + \sum_{j=0}^q \hat{\theta}_j \varepsilon_{n+l-j}^* \right),$$

여기서  $\varepsilon_t^*$ 는  $t \leq n$ 일 때 기존 자료로부터 계산된 잔차  $\hat{\varepsilon}_t$ 이고  $t > n$ 일 때는 난수  $\varepsilon_t^{(b)}$ 가 된다. 시계열  $Z_t^*$ 는  $Z_n, Z_{n-1}, \dots$ 와  $\varepsilon_t^*, \varepsilon_{t-1}^*, \dots$ 로 표현되도록 반복적으로 전개하여 구한다.

예측 추정값  $\hat{Y}_n(l)$ 를 계산하기 위해서는 난수  $\varepsilon^{(b)}$ 가 필요하며 이를 모수적인 방법과 비모수적인 방법으로 표집할 수 있다. 변환된 시계열의 오차항이 근사적으로 정규분포를 따르게 만드므로 모수적인 방법에서는 평균이 0이고 분산이  $\hat{\sigma}^2$ 인 정규분포에서 난수를 발생시켜 표본  $\varepsilon_j^{(b)}$ 를 얻는다. 비모수적인 방법에서는 잔차  $\hat{\varepsilon}_1, \dots, \hat{\varepsilon}_n$ 에서  $B$ 개를 무작위로 복원추출하여 재표집한다. 비록 변환을 통해 시계열이 정규분포에 근사하도록 만들었지만, 변환된 시계열이 정규성을 따른다고 확신하지 못할 수 있다. Smearing 추정에서 Duan (1983)은 잘못된 정규성 가정하에서 일치성을 만족하지 않는 예측이 나타날 수 있다는 것을 보였으며 이 때문에 앞에서 언급한 것과 같이 정규분포 대신에 일반적인 분포  $F$ 를 가정하였다. 경제시계열의 경우 외부의 큰 충격에 의해 예상보다 큰 이상점이 발생하는 경우가 종종 있으며 이들 자료를 변환하더라도 여전히 이상점으로 남아있을 가능성이 크다. 이로 인하여 자료의 정규성은 개선이 되지만 만족하지 않을 수 있으므로 극한 이상점이 많은 시계열자료에서 예측값을 구하고자 할 때에는 비모수적인 방법을 추천한다.

### 3. 실증분석

제안된 방법의 타당성을 확인하기 위해 기존 방법과 제안된 방법을 실증분석을 통해 비교하였다. 실증분석에는 기준지수가 1000으로 상향된 2004년 1월 26일부터의 KOSDAQ 종합지수 수익률이 사용되었다.  $P_t$ 를  $t$  시점의 KOSDAQ종합지수라고 하면, 수익률을 다음과 같이 정의하였다.

$$Y_t = 100 \times \log \left( \frac{P_t}{P_{t-1}} \right).$$

그림 3.1은 수익률에 대한 시계열그림으로 음의 큰 값이 양의 큰 값보다 상대적으로 더 많은 것을 볼 수 있다. 이것은 서론에서 언급한 것과 같이 대부분의 금융시계열이 가지는 특징이며 이런 음의 왜도를 해결할 수 있는 변환선택이 자료분석에 있어 중요한 시발점이 될 것이다. 왜도를 조절하는 대표적인 변환은 Box-Cox변환 (1964)이다. 그러나 수익률은 음의 값을 가질 수 있으므로 Box-Cox변환 (1964)과 같이 자료의 하한 값에 대한 제약이 있는 변환은 사용하는데 문제가 있다. 이런 문제가 없으면서 왜도를 조절할 수 있는 변환으로 아래와 같은

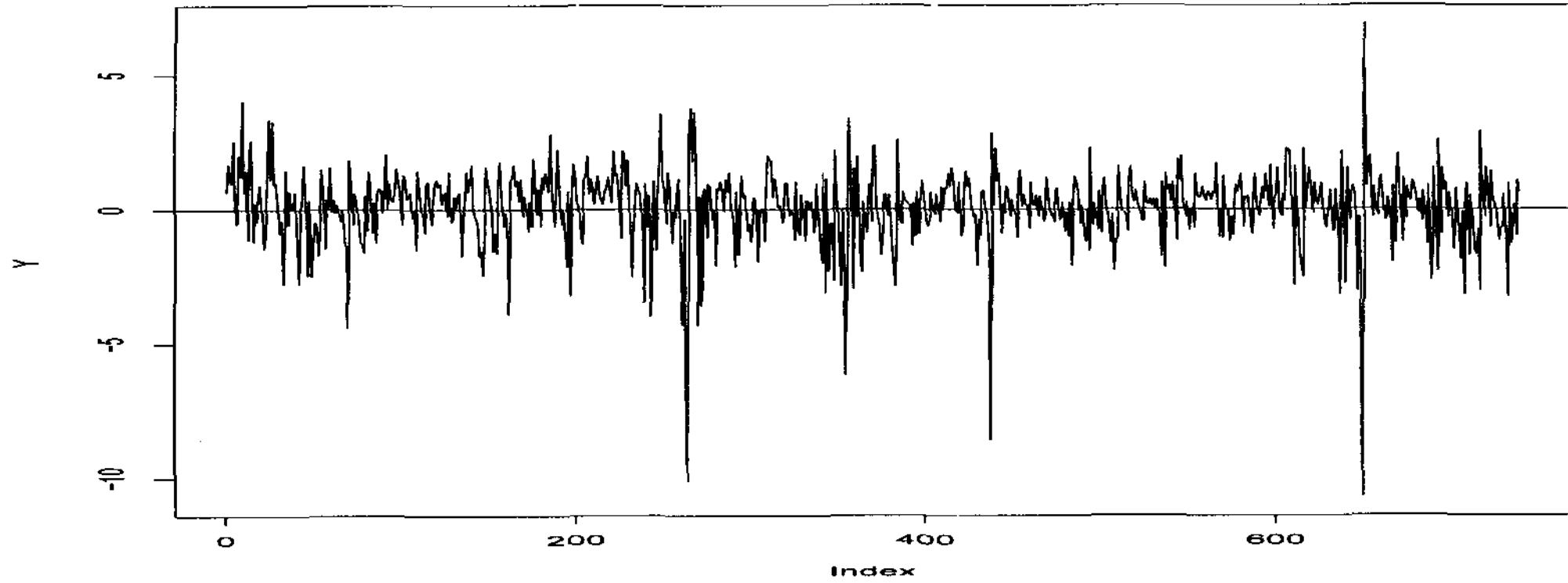


그림 3.1: 2006년 1월 17일부터의 KOSDAQ 수익률 시계열그림

Yeo-Johnson 변환 (2000)이 있다.

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

Yeo-Johnson 변환 (2000)은  $\lambda > 1$ 일 때 음의 왜도를 줄여주고  $\lambda < 1$ 일 때 양의 왜도를 줄여준다. 이 변환의 변환모수를 추정하는 이론적 내용은 조혜민 등 (2007)을 참고하면 된다.

이 논문에서 비교를 위해 사용된 방법은 다음과 같다.

- 방법 I : 변환하지 않고  $Y$ 에 대해 ARMA모형을 적합하고 예측값을 구함
- 방법 II : 자료를 변환하여 예측값을 구한 후 역변환하여 본래척도에서의 예측값을 구함
- 방법 III : 제안된 방법 - 모수적 붓스트랩이용
- 방법 IV : 제안된 방법 - 비모수적 붓스트랩이용

자료는 2004년 1월 27일부터 계산된 수익률 중 앞에서 200개의 자료로 모형을 추정하고 1시차부터 5시차 후까지의 예측값을 각각의 방법으로 추정하여 실제 관측값과 비교하였다. 이러한 방법으로 하루씩 이동하여 개장한 500일 동안의 수익률에 대해 예측오차를 구하였다. 모형 추정에 있어 정확성을 높이기 위해 각 자료마다 잠정모형의 AIC를 비교하여 AIC가 가장 작은 모형을 선택한 후 예측값을 계산하도록 하였다. 분석은 R에서 제공하는 arima함수를 사용하였으며 변환모수  $\lambda$ 는 arima함수의 출력값 loglik에 다음과 같은 야코비안(Jacobian)항을 더

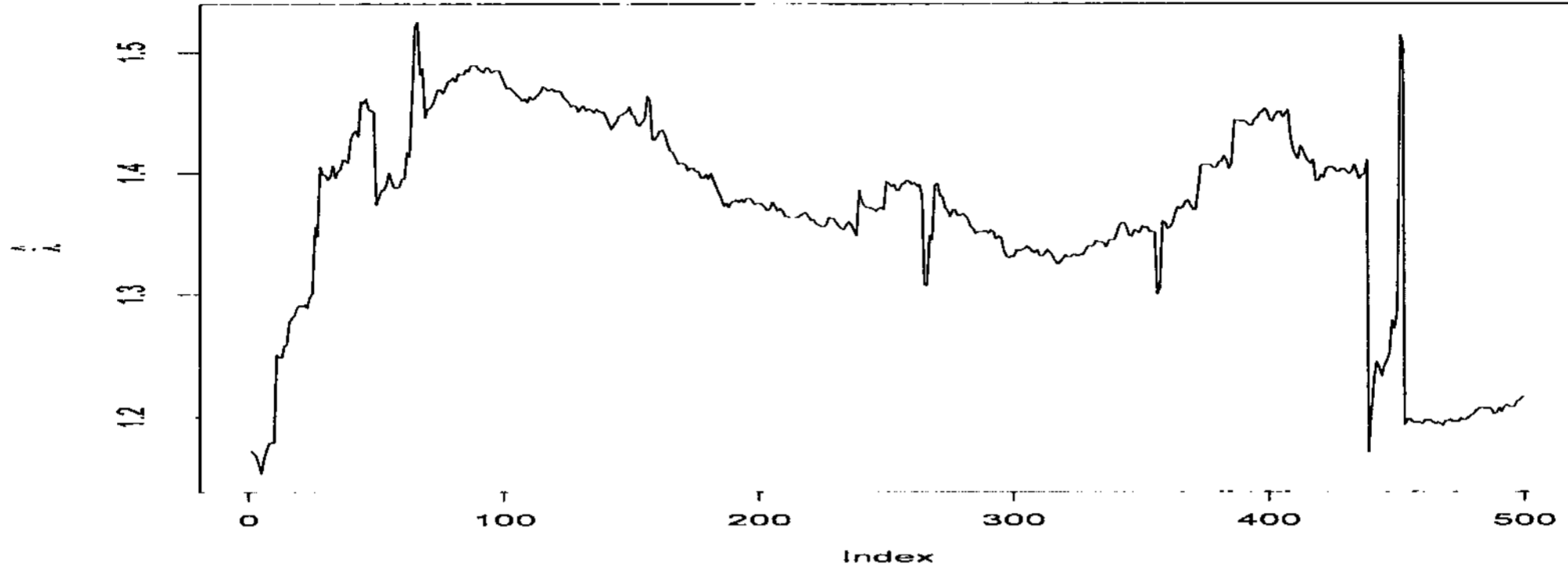


그림 3.2: 시간에 따른 변환모수의 추정값

표 3.1: 예측오차의 평균과 MSE

방법	1-시차		2-시차		3-시차		4-시차		5-시차	
	평균	MSE	평균	MSE	평균	MSE	평균	MSE	평균	MSE
I	0.000	2.287	0.007	2.300	0.014	2.266	0.014	2.266	0.010	2.258
II	0.185	2.262	0.201	2.313	0.205	2.316	0.204	2.314	0.200	2.309
III	-0.007	2.228	0.000	2.278	0.004	2.268	0.004	2.269	-0.002	2.267
IV	0.005	2.227	0.011	2.275	0.016	2.265	0.016	2.267	0.016	2.260

하고  $-1$ 을 곱한 후 optimize 함수로 최적값을 구하였다.

$$(\lambda - 1) \sum_{j=t+1}^{t+200} \text{sign}(y_j) \log(|y_j| + 1),$$

여기서  $\text{sign}(y)$ 은  $y \geq 0$ 일 때  $1$ 이고  $y < 0$ 일 때  $-1$ 이다.

그림 3.2는 시간에 따른 변환모수  $\lambda$ 의 추정값을 그린 것으로 모든 추정값이 1보다 큰 것으로 나타났다. 이들 추정값의 평균은 1.370이었고 표준편차는 0.083인 것으로 나타났는데 이것은 원 자료는 음의 왜도 형태를 가지고 있으며 정규근사를 위해 변환이 필요한 것을 의미한다. 변환을 사용하지 않은 경우 500번 중 174번의 분석에서 AR(1) 모형을 나머지분석에서 MA(1) 모형이 선택되었다. 변환을 사용한 경우 165번의 분석에서 AR(1)을 나머지에서 MA(1) 모형이 선택되었는데 변환을 하지 않은 경우의 MA(1) 모형은 변환한 경우에도 모두 MA(1) 모형인 것으로 나타났다.

표 3.1는 1시차에서 5시차 후까지의 예측오차에 대한 평균과 MSE를 정리한 것이다. 표 3.1에서 주의해서 볼 것은 방법 II의 예측오차평균이 다른 것이 비해 확연하게 크다는 것으로 이것은 방법 II에 의해 계산된 예측값이 편향되어 있다는 것으로 의미한다. 방법 III의 경우 대체로 편향이 적은 것으로 나타났으며 방법 IV는 MSE가 작은 것으로 나타났다. 특히 1시차 후 예측에서 다른 두 방법보다 방법 III과 IV가 MSE 측면에서 매우 우수한 것으로 나타났다. 이와 같이 방법 III과 IV가 유사한 결과를 보이는 것은 변환을 통해 자료가 정규분포에 잘 근

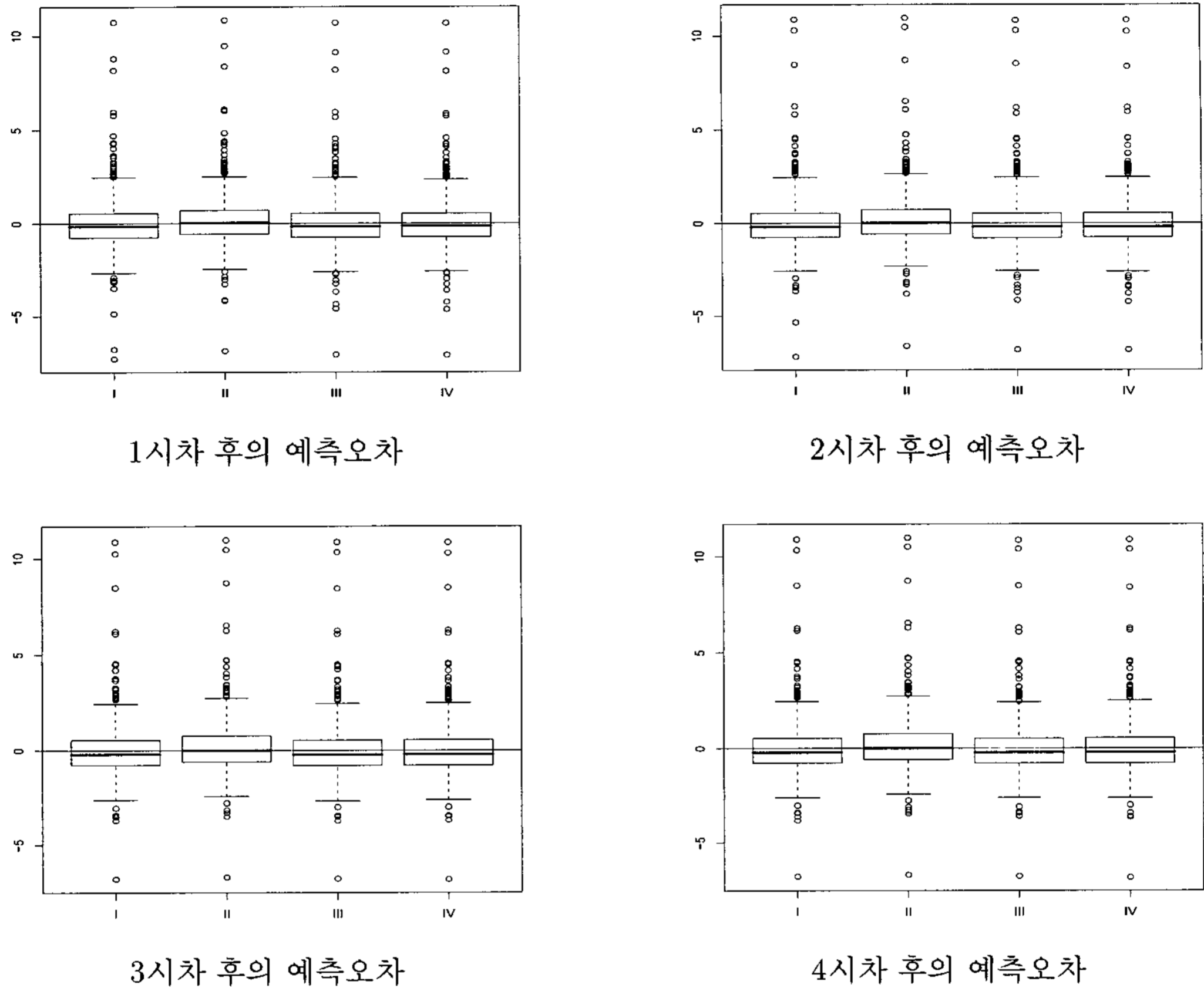


그림 3.3: 예측오차에 대한 상자그림

사하는 것으로 유추할 수 있다.

그림 3.3는 1시차에서 4시차 후까지의 예측오차를 표시한 것으로 방법 I에 비해 방법 III과 IV는 음의 예측오차부분에서 개선되는 것으로 볼 수 있다. 흥미로운 것은 방법 II에 의한 예측오차의 중앙값이 거의 0에 가까이 있는 것으로 자료의 중앙값을 변환한 값과 변환한 자료의 중앙값이 같기 때문에 나타나는 현상으로 방법 II로는 기대값에 관련된 정보를 구하기에는 부족하다는 것을 보여준다.

#### 4. 결론

이 논문에서는 정상자기회귀이동평균모형에서 변환-역변환을 이용하여 예측값을 추정할 때 고려해야 할 문제와 해결방안에 대해 알아보았다. 변환된 자료를 이용하여 MMSE 예측값을 구하고 그 예측값을 역변환하여 본래척도에서의 예측값을 구하는 방법을 흔히들 사용하는데 이러한 방법에 의해 계산된 예측값은 편향된다는 사실을 실증분석을 통해 확인하였다. 이



러한 편향을 줄여주기 위해 이 논문에서는 역변환의 기대값을 계산하는 과정에서 나타나는 미래의 오차항을 모수적 붓스트랩과 비모수적 붓스트랩을 통해 표집하여 표본평균으로 기대값을 추정하는 방법을 제안하였다. 또한 KOSDAQ 지수의 수익률 자료를 분석하여 제안된 방법이 편향을 줄여주고 MSE 측면에서도 우수한 것을 확인하였다.

## 참고문헌

- 조혜민, 오승언, 여인권 (2007). 변환된 자기회귀이동평균 모형에서의 예측구간추정, <응용 통계연구>, **20**, 541-550.
- Box, G. E. P. and Cox, C. R. (1964). An analysis of transformations, *Journal of the Royal Statistical Society, Series B*, **26**, 211-252.
- Duan, N. (1983). Smearing estimate: A nonparametric retransformation method, *Journal of the American Statistical Association*, **78**, 605-610.
- Engle, R. F. and Ng, V. K. (1993). Measuring and testing the impact of news on volatility, *The Journal of Finance*, **48**, 1749-1778.
- Glosten, L. R., Jagannathan, R. and Runkle, D. (1993). On the relation between the expected value and the volatility of nominal excess return on stocks, *The Journal of Finance*, **48**, 1779-1801.
- Nelson, D. B. (1991). Conditional heteroskedasticity in asset returns: A new approach, *Econometrica*, **59**, 347-370.
- Pagan, A. R. and Schwert, G. W. (1990). Alternative models for conditional stock volatility, *Journal of Econometrics*, **45**, 267-290.
- Taylor, J. M. G. (1986). The retransformed mean after a fitted power transformation, *Journal of the American Statistical Association*, **81**, 144-118.
- Yeo, I. K. and Johnson, R. A. (2000). A new family of power transformations to improve normality or symmetry, *Biometrika*, **87**, 954-959.
- Zakoian, J. M. (1994). Threshold heteroskedastic models, *Journal of Economic Dynamics and Control*, **18**, 931-955.

[ 2008년 2월 접수, 2008년 4월 채택 ]

## Estimation of Prediction Values in ARMA Models via the Transformation and Back-Transformation Method\*

In-Kwon Yeo<sup>1)</sup> Hyemin Cho<sup>2)</sup>

### ABSTRACT

One of main goals of time series analysis is to estimate prediction of future values. In this paper, we investigate the bias problem when the transformation and back-transformation approach is applied in ARMA models and introduce a modified smearing estimation to reduce the bias. An empirical study on the returns of KOSDAQ index via Yeo-Johnson transformation was executed to compare the performance of existing methods and proposed methods and showed that proposed approaches provide a bias-reduced estimation of the prediction value.

*Keywords:* Bootstrap, minimum mean square error, smearing estimation, Yeo-Johnson transformation.

---

\* This research was supported by the Sookmyung Women's University Research Grants 2007.

This work was supported by Brain Korea 21 project in 2007.

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

E-mail: inkwon@sookmyung.ac.kr

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

E-mail: coolmin98@hanmail.net