

# Re-Transformation of Power Transformation for ARMA( $p, q$ ) Model - Simulation Study

Jun-Hoon Kang<sup>a</sup> · Key-Il Shin<sup>a,1</sup>

<sup>a</sup>Department of Statistics, Hankuk University of Foreign Studies

(Received March 20, 2015; Revised May 19, 2015; Accepted May 26, 2015)

---

## Abstract

For time series analysis, power transformation (especially log-transformation) is widely used for variance stabilization or normalization for stationary ARMA( $p, q$ ) model. A simple and naive back transformed forecast is obtained by taking the inverse function of expectation. However, this back transformed forecast has a bias. Under the assumption that the log-transformed data is normally distributed. The unbiased back transformed forecast can be obtained by the expectation of log-normal distribution; consequently, the property of this back transformation was studied by Granger and Newbold (1976). We investigate the sensitivity of back transformed forecasts under several different underlying distributions using simulation studies.

Keywords: power transformation, moment, stationary time series, power-normal distribution

---

## 1. 서론

시계열 분석에서는 정규성 또는 대칭성이 만족되지 않는 경우 변환을 사용하고 변환된 자료를 기초로 분석을 실시한 후 재변환(re-transformation)하여 원자료의 예측에 사용한다. 주로 사용하는 변환은 Box-Cox (1964)가 제안한 멱변환(power transformation)이다. 멱변환은 자료가 양수인 경우에 사용되거나 자료가 음수인 경우에는 임의의 양수를 더하여 자료를 양수로 만든 후 사용하기도 한다. 최근 Yeo와 Johnson (2000)은 멱변환의 정규성과 대칭성을 향상할 수 있는 새로운 멱변환 족(power transformation family)를 제안하였다.

국내 연구에서 Box-Cox 멱변환을 사용하여 시계열 분석을 실시한 사례는 매우 흔하다. Ha와 Kim (2008)은 인터넷 트래픽 예측 기법 연구에 멱변환 중에서 로그 변환을 이용하였다. Hur 등 (2009)는 다변량 시계열 모형을 이용한 항공 수요 예측 분석을 위해 역시 로그 변환을 실시하였다. Lee 등 (2013)도 한국전력거래소의 일별최대전력수요자료 분석에 로그 변환을 사용하였다. 특히 Kim과 Seong (2011)은 개입모형을 이용한 한국의 입출국자 수 분석에서 로그 변환을 실시하고, 이후 재변환을 이용하여 원자료의 예측을 실시하였다. 이때 사용한 재변환 방법은 예측값의 지수함수를 이용한 단순 재변환 방법이다. 잘 알려진 것처럼 흔히 사용하는 재변환을 이용하여 예측값을 얻는 경우에는 편향이 발

This research was supported by Hankuk University of Foreign Studies research fund (2015).

<sup>1</sup>Corresponding author: Department of Statistics, Hankuk University of Foreign Studies, Gyeonggi-do 449-791, Korea. E-mail: [keyshin@hufs.ac.kr](mailto:keyshin@hufs.ac.kr)

생한다. 이러한 편향을 제거하기 위해 Granger과 Newbold (1976)는 로그 변환의 경우 로그-정규분포의 기댓값을 재변환 예측값으로 사용할 것을 제안하였다. 또한 Freeman과 Modarres (2006)은 power-normal 분포를 가정하였을 경우의 Box-Cox 재변환에 관한 이론을 연구하였다. 최근 Helmut와 Fang (2012)은 단순 재변환과 로그-정규분포를 이용한 재변환 결과를 자료분석을 통해 비교하였다. 이에 본 연구에서는 기댓값의 역함수를 이용한 단순 재변환과 역함수의 기댓값을 이용한 재변환의 특징과 성능을 모의실험을 통하여 비교하였다.

본 연구의 구성은 다음과 같다. 먼저 2절에서는 역변환과 재변환의 편향에 관하여 살펴보았다. 3절에서는 모의실험을 통하여 두 방법의 성능을 비교하였다. 이때 로그변환과 제곱근 변환이 사용되었으며 오차는 정규분포,  $t$ -분포, 이중지수분포를 가정하였다. 4절의 사례분석에서는 Wei (1990)에 수록된 Lynx Pelts 자료와 검정과리 자료가 사용되었다. 5절에 결론이 있다.

## 2. 역변환

### 2.1. 역변환(power transformation)

**2.1.1. Box-Cox 역변환** 시계열 분석에서 Box-Cox 역변환(power transformation)은 분산 안정화(variance stablization) 또는 정규화를 위하여 매우 유용하게 사용되고 있다. Box-Cox (1964)가 제안한 역변환은 다음과 같다.

$$Y = \begin{cases} \frac{X^\lambda - 1}{\lambda}, & \text{for } \lambda > 0, \\ \log(X), & \text{for } \lambda = 0. \end{cases} \quad (2.1)$$

Box-Cox 역변환에서 자료가 음수일 경우에는 각각의 자료에 양의 상수  $c > 0$ 를 더하여 모든 자료를 양수로 만든 후 Box-Cox 변환을 실시한다. 본 논문에서  $\log$ 는 자연로그를 의미한다.

**2.1.2. Yeo-Johnson 변환** Yeo-Johnson (2000)은 Box-Cox 역변환 시 큰 음수 자료가 포함된 경우를 해결하기 위해 다음의 변환을 제안하였다.

$$x^{(\lambda)} = \begin{cases} \frac{(x+1)^\lambda - 1}{\lambda}, & x \geq 0, \lambda \neq 0, \\ \log(x+1), & x \geq 0, \lambda = 0, \\ -\frac{(-x+1)^{2-\lambda} - 1}{2-\lambda}, & x < 0, \lambda \neq 2, \\ -\log(-x+1), & x < 0, \lambda = 2. \end{cases} \quad (2.2)$$

### 2.2. 재변환의 편향

여러 역변환 중에서 가장 흔히 사용되는 변환이 로그 변환(log-transformation)이며 특히 경제 시계열 자료 분석에서는 기본적으로 사용되고 있다. 변환된 자료는 일반적으로 대칭성을 유지하거나 또는 정규분포와 유사한 분포를 따르게 된다. 따라서 모든 분석은 변환된 자료가 사용된다. 이후 모형 식별과 모수 추정 등이 끝나면 예측을 실시하는데 예측은 변환된 자료를 이용하여 이루어진다. 그러나 최종적인 원자료의 예측을 얻기 위해서는 재변환을 실시하여야 한다. 잘 알려진 사실은 흔히 사용하는 재변환이 불편추정량이 아니라는 것이다. 이는 간단히 Jensen의 부등식(Jensen's inequality)에서 쉽게 알 수 있다. 다음이 Jensen의 부등식이다.

$$E(X^\lambda) \leq E(X)^\lambda, \quad \text{for } 0 \leq \lambda \leq 1. \quad (2.3)$$

**Table 2.1.** Means and variances of power-normal distribution

$\lambda$	$E(X)$	$\text{Var}(X)$
0	$\exp(\mu + \frac{1}{2}\sigma^2)$	$\exp(2\mu + \sigma^2)(\exp(\sigma^2) - 1)$
$\frac{1}{4}$	$(\frac{1}{4}\mu + 1)^4 + \frac{3}{8}\sigma^2(\frac{1}{4}\mu + 1)^2 + \frac{3}{256}\sigma^4$	$\frac{8}{2048}\sigma^8 + \frac{3}{32}\sigma^6(\frac{1}{4}\mu + 1)^2 + \frac{21}{32}\sigma^4(\frac{1}{4}\mu + 1)^4 + \sigma^2(\frac{1}{4}\mu + 1)^6$
$\frac{1}{3}$	$(\frac{1}{3}\mu + 1)^3 + \frac{1}{3}\sigma^2(\frac{1}{3}\mu + 1)$	$\frac{5}{243}\sigma^6 + \frac{4}{9}\sigma^6(\frac{1}{3}\mu + 1)^2 + \sigma^2(\frac{1}{3}\mu + 1)^4$
$\frac{1}{2}$	$(\frac{1}{2}\mu + 1)^2 + \frac{1}{4}\sigma^2$	$\frac{1}{8}\sigma^4 + \sigma^2(\frac{1}{2}\mu + 1)^2$

**Table 3.1.** The comparison results of the methods of AR(1) - Normal

자료수	$\phi$	분산	평균	통계량					
				MAE		Bias		RMSE	
				M1	M2	M1	M2	M1	M2
210	0.3	$\sigma_a^2 = 0.5$	$\mu = 4$	215.0	228.5	99.3	2.8	347.4	333.7
			$\mu = 6$	3745.1	3981.7	1772.2	37.0	6060.8	5812.0
		$\sigma_a^2 = 1$	$\mu = 4$	366.7	414.1	221.2	-0.1	739.5	707.8
			$\mu = 6$	6385.3	7216.9	3895.5	-18.4	12889.6	12326.9
	0.5	$\sigma_a^2 = 0.5$	$\mu = 4$	2397.2	2556.9	1325.3	57.9	4086.7	3881.3
			$\mu = 6$	131358.0	141691.0	77752.7	-3474.9	224934.0	212174.0
		$\sigma_a^2 = 1$	$\mu = 4$	4249.1	4885.3	2860.8	3.5	9378.8	8970.9
			$\mu = 6$	232295.0	270964.0	161590.0	-9300.0	513751.0	490429.0

따라서 오른쪽으로 꼬리가 치우친 분포에서 주로 사용하는 멱변환의 경우 편향이 발생한다.

**2.2.1. 로그 변환**  $\lambda = 0$ 인 로그 변환의 경우  $E(\log(X)) \leq \log(E(X))$ 이 된다. 따라서 단순 재변환을 실시하게 되면 예측값은 과소추정 된다. 이를 해결하기 위해 Granger과 Newbold (1976)은 로그-정규분포의 기댓값을 사용할 것을 제안하였다. 즉  $\log(X) \sim N(\mu, \sigma^2)$ 인 로그-정규분포를 고려하면 기댓값은 다음과 같다.

$$E(X) = \exp(\mu + 0.5\sigma^2). \tag{2.4}$$

단순 재변환인  $\exp(E(\log(x)))$ 를 사용할 경우  $e^\mu(e^{0.5\sigma^2} - 1)$ 의 편향이 발생한다. 따라서 단순 재변환을 사용하지 않고 위의 결과를 사용할 것을 제안하였다.

**2.2.2. 제곱근 및 4승근 변환** 시계열 분석에서 로그 변환보다는 사용 빈도가 적지만 많이 사용하는 변환이 제곱근 변환이며 이 변환도 식 (2.3)의 결과에 의해 편향이 발생한다. 이러한 편향을 제거하기 위한 방법의 하나로 다음을 고려할 수 있다.

즉  $Y = \sqrt{X} \sim N(\mu, \sigma^2)$ 이면  $E(X) = \mu^2 + \sigma^2$ 이 된다. 따라서 단순히 재변환 값으로  $\mu^2$ 을 사용하는 것에 비해  $\sigma^2$ 에 해당되는 편향을 제거할 수 있다. 또한  $Y = X^{1/4} \sim N(\mu, \sigma^2)$ 인 경우를 살펴보면  $E(X) = \mu^4 + 6\mu^2\sigma^2 + 3\sigma^4$ 이 된다. 따라서  $6\mu^2\sigma^2 + 3\sigma^4$ 의 편향을 줄일 수 있다. 또한 Box-Cox의 재변환식인  $X = (\lambda Y + 1)^{1/\lambda}$ 을 위해 Freeman과 Modarres (2006)는  $\lambda = 0, 1/4, 1/3, 1/2$ 의  $E(X)$ 와  $\text{Var}(X)$  결과를 구하였으며 이를 표로 작성하면 다음과 같다.

### 3. 모의 실험

#### 3.1. 로그 변환

이 절에서는 로그 변환과 제곱근 변환을 사용한 후 재변환을 실시할 때 단순 재변환과 편향을 제거한

**Table 3.2.** The comparison results of the methods of MA(1) - Normal

자료수	$\theta$	분산	평균	통계량					
				MAE		Bias		RMSE	
				M1	M2	M1	M2	M1	M2
210	0.3	$\sigma_a^2 = 0.5$	$\mu = 4$	38.6	41.2	16.9	-0.1	62.0	59.8
			$\mu = 6$	285.0	304.2	124.9	-0.7	457.9	441.5
	$\sigma_a^2 = 1$	$\mu = 4$	65.7	74.6	38.7	-0.8	131.6	126.1	
		$\mu = 6$	485.6	550.9	286.1	-5.7	972.6	932.0	
	0.5	$\sigma_a^2 = 0.5$	$\mu = 4$	42.2	45.4	19.5	-0.2	70.0	67.3
			$\mu = 6$	311.6	335.8	144.1	-1.5	517.0	497.6
$\sigma_a^2 = 1$	$\mu = 4$	73.9	85.3	45.7	-1.3	156.9	150.5		
	$\mu = 6$	546.2	630.5	337.7	-9.4	1159.6	1112.1		

**Table 3.3.** The comparison results of the methods of ARMA(1,1) - Normal

자료수	$\phi$	$\theta$	분산	평균	통계량					
					MAE		Bias		RMSE	
					M1	M2	M1	M2	M1	M2
210	0.3	0.5	$\sigma_a^2 = 0.5$	$\mu = 4$	290.0	315.6	160.6	7.3	521.6	497.8
				$\mu = 6$	5052.7	5481.9	2880.4	172.1	9109.0	8668.2
	$\sigma_a^2 = 1$	$\mu = 4$	547.9	653.9	383.4	-6.3	1333.1	1281.7		
		$\mu = 6$	9540.6	11348.0	6768.0	-26.1	23241.0	22315.0		
	0.5	0.3	$\sigma_a^2 = 0.5$	$\mu = 4$	3031.5	3270.3	1900.7	168.3	5640.6	5333.2
				$\mu = 6$	166263	180405.0	110546.0	3851.1	310417.0	291373.0
$\sigma_a^2 = 1$	$\mu = 4$	5861.8	6999.5	4369.2	45.8	15018.0	14445.0			
	$\mu = 6$	320569	385774	246183	-4610.0	822435.0	789362.0			

재변환을 모의실험을 통하여 우수성을 비교하였다. 모의실험에 사용된 모형은 AR(1), MA(1) 그리고 ARMA(1,1) 모형으로 정의는 다음과 같다.

$$\text{AR}(1) : X_t - \mu = \phi(X_{t-1} - \mu) + a_t,$$

$$\text{MA}(1) : X_t - \mu = a_t - \theta a_{t-1},$$

$$\text{ARMA}(1,1) : X_t - \mu = \phi(X_{t-1} - \mu) + a_t - \theta a_{t-1},$$

여기서  $a_t \stackrel{iid}{\sim} (0, \sigma_a^2)$ 이다. 이 모형에서 사용된 모수는 AR(1) 모형에서  $\mu = 4, 6$ ,  $\sigma_a^2 = 0.5, 1$  그리고  $\phi = 0.3, 0.5$ 이고 MA(1) 모형에서는  $\mu = 4, 6$ ,  $\sigma_a^2 = 0.5, 1$  그리고  $\theta = 0.3, 0.5$ 이다. 그리고 ARMA(1,1) 모형에서는  $(\phi, \theta) = (0.3, 0.5), (0.5, 0.3)$ 이 사용되었다. 이제 각각 110개와 210개의 자료를 생성한 후, 마지막 10개는 예측의 정확성 비교를 위해 남겨 놓고 100개 또는 200개를 이용하여 모형의 모수를 추정한 후 10개 값을 예측하였다. 자료의 수를 110개와 210개를 사용한 이유는 자료의 수에 따라 추정의 정확성이 달라지므로 추정의 정확성으로 인해 결과에 차이가 있는지 확인하기 위해서이다. 그러나 모의실험 결과가 매우 유사하여 본문에는 210개의 모의실험 결과만 수록하였다. 다음으로 얻어진 예측값은 재변환을 실시한 후 참 값인  $Y_t = \exp(X_t)$ 와 비교하였다. 재변환은 다음의 두 방법이 사용되었다.

$$M1 = \exp(\text{forecast}),$$

$$M2 = \exp\left(\text{forecast} + \frac{1}{2} \hat{\sigma}_{fore}^2\right),$$

**Table 3.4.** The comparison results of the methods of AR(1) - Double Exponential

자료수	$\phi$	분산	평균	통계량					
				MAE		Bias		RMSE	
				M1	M2	M1	M2	M1	M2
210	0.3	$\sigma_a^2 = 0.5$	$\mu = 4$	399.0	467.2	266.1	42.1	1552.6	1531.9
			$\mu = 6$	6945.9	8142.5	4677.1	718.1	27041.0	26672.0
		$\sigma_a^2 = 1$	$\mu = 4$	1140.6	1411.6	976.9	355.0	10352.0	10315.0
			$\mu = 6$	19857.0	24595.0	17053.0	6153.0	180265.0	179613.0
	0.5	$\sigma_a^2 = 0.5$	$\mu = 4$	4623.0	5495.7	3314.8	400.6	17596.0	17323.0
			$\mu = 6$	252604.0	304789.0	186471.0	12486.0	961725.0	945527.0
		$\sigma_a^2 = 1$	$\mu = 4$	13917.0	17872.0	12222.0	3530.7	117624.0	117235.0
			$\mu = 6$	759707.0	988190.0	673003.0	173561.0	6.42E+07	6.39E+07

**Table 3.5.** The comparison results of the methods of MA(1) - Double Exponential

자료수	$\theta$	분산	평균	통계량					
				MAE		Bias		RMSE	
				M1	M2	M1	M2	M1	M2
210	0.3	$\sigma_a^2 = 0.5$	$\mu = 4$	71.4	83.9	46.9	7.0	277.6	274.0
			$\mu = 6$	527.6	620.3	346.3	51.7	2051.2	2024.6
		$\sigma_a^2 = 1$	$\mu = 4$	203.9	252.8	174.0	63.1	1837.7	1831.1
			$\mu = 6$	1506.7	1867.9	1285.8	466.3	13579.0	13529.0
	0.5	$\sigma_a^2 = 0.5$	$\mu = 4$	80.1	95.6	53.9	6.4	300.8	296.4
			$\mu = 6$	592.2	706.1	398.3	47.0	2222.3	2190.0
		$\sigma_a^2 = 1$	$\mu = 4$	235.3	299.7	203.1	62.5	1955.1	1946.5
			$\mu = 6$	1738.5	2214.3	1500.6	462.1	14446.0	14382.0

여기서  $\hat{\sigma}_{fore}^2$ 는 예측오차의 분산으로 예측 시점에 따라 달라지게 된다. 물론 이 값은 이론적으로 이미 각 모형별로 구해져 있으며 SAS/ARIMA 모형에서는 예측시 이 값을 출력해 준다. 따라서 M2 방법을 사용하는 것은 어려움이 없다. 본 모의실험에서는 이제 재변환에 의해서 얻어진  $r$ 번째 반복의 예측값을  $\hat{Y}_l^{(r)}$ 이라 하고, 실제값을  $Y_l^{(r)}$ 라 하면 다음의 비교 통계량을 이용하여 두 방법의 우수성을 비교하였다. 사용된 비교통계량의 정의는 다음과 같다. 여기서 반복수  $R = 1,000$ 을 사용하였고  $n = 100, 200$ 을 사용하였다.

$$\begin{aligned}
 \text{R.M.S.E} &= \sqrt{\frac{1}{10R} \sum_{r=1}^R \sum_{l=n+1}^{n+10} \left( \hat{Y}_l^{(r)} - Y_l^{(r)} \right)^2}, \\
 \text{M.A.E} &= \frac{1}{10R} \sum_{r=1}^R \sum_{l=n+1}^{n+10} \left| \hat{Y}_l^{(r)} - Y_l^{(r)} \right|, \\
 \text{Bias} &= \frac{1}{10R} \sum_{r=1}^R \sum_{l=n+1}^{n+10} \left( \hat{Y}_l^{(r)} - Y_l^{(r)} \right).
 \end{aligned}$$

변환 된 자료에는 정규분포 뿐만아니라 이중지수분포와 자유도가 5와 15인  $t$  분포를 따르는 경우도 살펴보았으며 절단 이중지수분포와 절단  $t$  분포도 함께 살펴보았다. 절단된 분포의 경우 오차의 절대값이 2.58이상이면 제거하였다. Table 3.1에서 Table 3.15에 결과를 수록하였다.

Table 3.1에서 Table 3.3는 정규분포에서 생성된 자료를 기반으로 얻어진 결과로 AR(1) 모형과 MA(1)

**Table 3.6.** The comparison results of the methods of ARMA(1, 1) - Double Exponential

자료수	$\phi$	$\theta$	분산	평균	통계량					
					MAE		Bias		RMSE	
					M1	M2	M1	M2	M1	M2
210	0.3	0.5	$\sigma_a^2 = 0.5$	$\mu = 4$	554.0	663.8	388.7	-2.1	1380.0	1320.5
				$\mu = 6$	9647.6	11519.0	6859.2	45.3	24062.0	22997.0
			$\sigma_a^2 = 1$	$\mu = 4$	1458.0	2042.7	1250.4	-97.5	6193.1	6008.7
	$\mu = 6$	25388.0		35421.0	21873.0	-1477.2	107876.0	104650.0		
	0.5	0.3	$\sigma_a^2 = 0.5$	$\mu = 4$	5948.1	7113.7	4454.0	125.6	15702.0	15024.0
				$\mu = 6$	325255.0	391943.0	250756.0	-356.0	860116.0	820691.0
$\sigma_a^2 = 1$			$\mu = 4$	16556.0	23418.0	14590.0	-988.1	78827.0	77183.0	
	$\mu = 6$	904308.0	1288953.0	805244.0	-70260	4306758.0	4216031.0			

**Table 3.7.** The comparison results of the methods of AR(1) -  $t$  Distribution

자료수	$\phi$	$df$	평균	통계량					
				MAE		Bias		RMSE	
				M1	M2	M1	M2	M1	M2
210	0.3	$\nu = 5$	$\mu = 4$	4674.7	4847.9	4517.1	4050.6	277869.0	277864.0
			$\mu = 6$	116324.0	119247.0	113637.0	105423.0	6845252.0	6845117.0
		$\nu = 15$	$\mu = 4$	447.2	518.9	292.9	22.1	1246.7	1213.9
	$\mu = 6$		7298.8	8483.5	4733.5	-15.8	17465.0	16849.0	
	0.5	$\nu = 5$	$\mu = 4$	108678.0	110977.0	107132.0	100881.0	5674956.0	5674839.0
			$\mu = 6$	1528037.0	1664394.0	1447768.0	1087107.0	5.5E+07	5.5E+07
$\nu = 15$		$\mu = 4$	5296.9	6312.0	3792.8	215.8	17825.0	17486.0	
	$\mu = 6$	309485.0	361685.0	236807.0	28754.0	1192934.0	1170018.0		

**Table 3.8.** The comparison results of the methods of MA(1) -  $t$  Distribution

자료수	$\theta$	$df$	평균	통계량					
				MAE		Bias		RMSE	
				M1	M2	M1	M2	M1	M2
210	0.3	$\nu = 5$	$\mu = 4$	314.8	345.6	284.5	201.6	12888.0	12886.0
			$\mu = 6$	1829.7	2035.9	1616.8	1010.7	41640.0	41621.0
		$\nu = 15$	$\mu = 4$	76.2	88.4	48.8	0.6	205.8	200.2
	$\mu = 6$		568.3	662.0	363.6	3.4	1458.5	1414.9	
	0.5	$\nu = 5$	$\mu = 4$	381.0	421.4	350.6	242.9	12540.0	12536.0
			$\mu = 6$	1796.2	2092.5	1564.6	789.7	31878.0	31851.0
$\nu = 15$		$\mu = 4$	88.2	103.5	59.9	2.1	288.8	282.8	
	$\mu = 6$	662.7	777.6	449.1	21.3	1987.9	1939.9		

모형 그리고 ARMA(1, 1) 모형에서의 결과가 유사하다. 먼저 RMSE를 기준으로 M1와 M2 결과를 비교하면 큰 차이를 보이고 있지 않지만 M2 방법이 우수한 결과를 주고 있다. 본 연구에서 중점적으로 고려한 편향(Bias)의 경우에는 M2 방법이 매우 우수한 결과를 주고 있다. 반면 MAE를 기준으로 본다면 M1 방법이 우수한 결과를 주고 있다. 결국 RMSE와 편향(Bias)을 기준으로 재변환을 실시한다면 M2 방법이 우수한 결과를 준다. 그러나 MAE를 기준으로 한다면 M1 방법의 사용도 고려할 수 있다. 이러한 현상은 Table 3.4에서 Table 3.9에 수록된 이중지수분포와  $t$  분포에서도 같은 결과를 주고 있다. 또한 절단 이중지수분포와 절단  $t$  분포의 결과에서도 유사한 현상을 보이고 있다. 결론적으로 RMSE와

**Table 3.9.** The comparison results of the methods of ARMA(1, 1) -  $t$  Distribution

자료수	$\phi$	$\theta$	df	평균	통계량							
					MAE		Bias		RMSE			
					M1	M2	M1	M2	M1	M2		
210	0.3	0.5	$\nu = 5$	$\mu = 4$	7463.4	7935.3	7274.2	6265.1	376450.0	376431.0		
				$\mu = 6$	54728.0	63106.0	51200.0	33231.0	1363300.0	1364096.0		
			$\nu = 15$	$\mu = 4$	738.2	889.3	567.2	67.2	3371.6	3319.8		
				$\mu = 6$	12716.0	15390.0	9839.2	1114.8	58226.0	57436.0		
			0.5	0.3	$\nu = 5$	$\mu = 4$	74448.0	80298.0	72713.0	60452.0	2767344.0	2766960.0
						$\mu = 6$	1.2E+08	1.2E+08	1.2E+08	1.2E+08	1.1E+10	1.1E+10
	$\nu = 15$	$\mu = 4$			7639.0	9415.7	6002.0	305.0	26579.0	25958.0		
		$\mu = 6$			430999.0	525599.0	352817.0	39523.0	1917245.0	1890865.0		

**Table 3.10.** The comparison results of the methods of AR(1) - Truncated Double Exponential

자료수	$\phi$	분산	평균	통계량						
				MAE		Bias		RMSE		
				M1	M2	M1	M2	M1	M2	
210	0.3	$\sigma_a^2 = 0.5$	$\mu = 4$	280.8	314.0	156.5	6.8	542.5	520.8	
			$\mu = 6$	4890.5	5473.0	2769.0	106.0	9457.8	9068.2	
		$\sigma_a^2 = 1$	$\mu = 4$	371.0	431.7	224.8	-8.0	725.8	692.2	
			$\mu = 6$	6460.2	7522.9	3957.6	-153.2	12651.0	12053.0	
		0.5	$\sigma_a^2 = 0.5$	$\mu = 4$	3228.0	3628.6	2034.9	101.2	6654.8	6364.0
				$\mu = 6$	176615.0	201214.0	116535.0	-2132.2	365016.0	347602.0
	$\sigma_a^2 = 1$		$\mu = 4$	4355.5	5151.5	2933.3	-79.9	9591.5	9168.7	
			$\mu = 6$	238015.0	285843.0	165628.0	-13860	525384.0	501112.0	

**Table 3.11.** The comparison results of the methods of AR(1) - Truncated Double Exponential

자료수	$\theta$	분산	평균	통계량						
				MAE		Bias		RMSE		
				M1	M2	M1	M2	M1	M2	
210	0.3	$\sigma_a^2 = 0.5$	$\mu = 4$	50.3	56.5	27.1	0.5	96.7	93.1	
			$\mu = 6$	371.5	417.3	200.2	3.8	714.8	687.8	
		$\sigma_a^2 = 1$	$\mu = 4$	66.3	77.5	39.3	-2.2	129.3	123.6	
			$\mu = 6$	489.9	572.3	290.2	-16.6	955.7	913.1	
		0.5	$\sigma_a^2 = 0.5$	$\mu = 4$	56.2	63.7	31.5	0.3	112.2	108.0
				$\mu = 6$	415.3	470.3	233.1	2.6	829.2	797.7
	$\sigma_a^2 = 1$		$\mu = 4$	75.2	89.0	46.6	-2.9	158.6	151.9	
			$\mu = 6$	555.6	657.7	344.5	-21.4	1171.6	1122.6	

편향을 고려한다면 비록 정규분포를 따르지 않더라도 M2 방법을 사용한다면 예측의 정확성은 향상될 수 있다.

**3.2. 제곱근 변환**

3.1절의 로그 변환과 같이 AR(1) 모형과 MA(1) 모형 그리고 ARMA(1, 1) 모형이 사용되었다. 제곱근 변환에 의해 얻어진 자료는 항상 양수가 되기 때문에 자료에 음수가 있을 경우가 발생하지 않도록 하기 위해 상대적으로 큰 값인  $\mu = 10, 20$ 를 선택하였으며  $\sigma_a^2 = 0.5, 1$ 인 경우를 살펴보았다.  $t$  분포의 경우

**Table 3.12.** The comparison results of the methods of ARMA(1, 1) - Truncated Doble Exponential

자료수	$\phi$	$\theta$	분산	평균	통계량					
					MAE		Bias		RMSE	
					M1	M2	M1	M2	M1	M2
210	0.3	0.5	$\sigma_a^2 = 0.5$	$\mu = 4$	409.4	470.0	264.8	14.6	943.8	909.1
				$\mu = 6$	7130.1	8154.9	4702.8	321.9	16461.0	15826.0
		$\sigma_a^2 = 1$	$\mu = 4$	571.9	699.0	403.4	-11.1	1472.7	1419.9	
			$\mu = 6$	9956.7	12124.0	7118.9	-101.1	25670.0	24719.0	
	0.5	0.3	$\sigma_a^2 = 0.5$	$\mu = 4$	4347.1	4965.4	3052.2	272.8	10546.0	10154.0
				$\mu = 6$	237978.0	273549.0	174013.0	9285.9	578243.0	554477.0
			$\sigma_a^2 = 1$	$\mu = 4$	6173.5	7558.3	4631.1	27.0	17628.0	17061.0
				$\mu = 6$	337466.0	416298.0	260641.0	-5405.7	964733.0	931888.0

**Table 3.13.** The comparison results of the methods of AR(1) - Truncated  $t$  Distribution

자료수	$\phi$	$df$	평균	통계량					
				MAE		Bias		RMSE	
				M1	M2	M1	M2	M1	M2
210	0.3	$\nu = 5$	$\mu = 4$	371.5	424.1	221.8	-11.3	687.4	653.0
			$\mu = 6$	6520.1	7448.8	4017.4	-87.4	12131.0	11491.0
		$\nu = 15$	$\mu = 4$	352.2	401.2	204.9	-9.8	647.1	615.6
			$\mu = 6$	6103.4	6968.2	3587.0	-217.2	11273.0	10725.0
	0.5	$\nu = 5$	$\mu = 4$	4411.3	5150.0	2955.6	-76.8	9107.7	8653.1
			$\mu = 6$	236108.0	281309.0	162788.0	-17031.0	494445.0	469634.0
		$\nu = 15$	$\mu = 4$	4306.4	4834.0	2995.6	229.1	8754.9	8238.0
			$\mu = 6$	234263.0	267546.0	165279.0	-677.7	491209.0	461403.0

**Table 3.14.** The comparison results of the methods of MA(1) - Truncated  $t$  Distribution

자료수	$\theta$	$df$	평균	통계량					
				MAE		Bias		RMSE	
				M1	M2	M1	M2	M1	M2
210	0.3	$\nu = 5$	$\mu = 4$	66.7	76.3	39.3	-2.1	124.5	118.4
			$\mu = 6$	501.1	567.0	303.7	-2.7	927.6	878.3
		$\nu = 15$	$\mu = 4$	63.3	72.2	35.7	-2.5	117.1	111.8
			$\mu = 6$	460.4	527.2	255.9	-24.5	836.2	798.3
	0.5	$\nu = 5$	$\mu = 4$	74.7	87.0	46.1	-3.5	151.1	144.4
			$\mu = 6$	565.5	657.8	355.2	-9.2	1147.2	1094.8
		$\nu = 15$	$\mu = 4$	72.5	83.0	45.0	-0.3	143.5	136.6
			$\mu = 6$	535.5	617.8	325.4	-10.3	1028.9	977.2

에는 자유도  $\nu = 5, 15$ 를 사용하였다. 또한 이중지수분포의 경우 분산  $\sigma_a^2 = 0.5, 1$ 를 사용하였다. 절단 분포를 위해서는 생성된 오차의 절대값이 2.58이상이면 이를 제거한 후 새로운 난수를 생성하여 오차를 만들었다. 또한 재변환을 위하여 다음의 두 재변환 방법이 사용되었다.

$$M1 = \text{forecast}^2,$$

$$M2 = \text{forecast}^2 + \hat{\sigma}_{fore}^2,$$



**Table 3.15.** The comparison results of the methods of ARMA(1, 1) - Truncated  $t$  Distribution

자료수	$\phi$	$\theta$	df	평균	통계량					
					MAE		Bias		RMSE	
					M1	M2	M1	M2	M1	M2
210	0.3	0.5	$\nu = 5$	$\mu = 4$	588.1	709.4	418.2	3.7	1499.1	1442.1
				$\mu = 6$	10374.0	12364.0	7531.6	278.8	25503.0	24403.0
			$\nu = 15$	$\mu = 4$	547.5	644.0	384.8	13.6	1281.9	1226.3
	$\mu = 6$	9666.8		11331.0	6900.8	334.1	23251.0	22248.0		
	0.5	0.3	$\nu = 5$	$\mu = 4$	6164.4	7450.3	4649.9	80.3	15442.0	14778.0
				$\mu = 6$	360667.0	432305.0	283569.0	17530.0	966468.0	927511.0
$\nu = 15$			$\mu = 4$	5895.2	6859.2	4457.9	283.2	14597.0	13832.0	
	$\mu = 6$	333678.0	391428.0	262575.0	21566.0	845675.0	807512.0			

**Table 3.16.** The comparison results of the methods of AR(1) - Normal

자료수	$\phi$	분산	평균	통계량					
				MAE		Bias		RMSE	
				M1	M2	M1	M2	M1	M2
210	0.3	$\sigma_a^2 = 0.5$	$\mu = 10$	16.9	16.9	1.9	1.2	21.2	21.1
			$\mu = 20$	34.0	33.9	5.6	4.6	42.5	42.4
		$\sigma_a^2 = 1$	$\mu = 10$	23.9	23.9	2.5	1.3	29.9	29.9
	$\mu = 20$		47.9	47.8	6.2	4.8	60.0	59.8	
	0.5	$\sigma_a^2 = 0.5$	$\mu = 10$	26.4	26.2	8.1	6.8	33.2	32.9
			$\mu = 20$	57.0	55.9	30.3	27.1	71.3	69.9
$\sigma_a^2 = 1$		$\mu = 10$	36.9	36.6	8.7	6.8	46.3	46.0	
	$\mu = 20$	76.7	75.8	30.9	27.1	96.2	95.0		

**Table 3.17.** The comparison results of the methods of MA(1) - Normal

자료수	$\theta$	분산	평균	통계량					
				MAE		Bias		RMSE	
				M1	M2	M1	M2	M1	M2
210	0.3	$\sigma_a^2 = 0.5$	$\mu = 10$	11.8	11.8	0.6	0.1	14.7	14.7
			$\mu = 20$	23.5	23.5	0.7	0.2	29.4	29.4
		$\sigma_a^2 = 1$	$\mu = 10$	16.6	16.7	1.2	0.1	20.9	20.8
	$\mu = 20$		33.3	33.3	1.3	0.3	41.6	41.6	
	0.5	$\sigma_a^2 = 0.5$	$\mu = 10$	12.5	12.5	0.7	0.1	15.7	15.7
			$\mu = 20$	25.0	25.0	0.8	0.2	31.3	31.3
$\sigma_a^2 = 1$		$\mu = 10$	17.7	17.7	1.4	0.1	22.2	22.2	
	$\mu = 20$	35.4	35.4	1.5	0.3	44.3	44.3		

여기서  $\sigma_{fore}^2$ 도 예측오차의 분산 추정량이며 이 값도 SAS/ARIMA 분석에서 값이 출력되므로 이 값을 사용하면 된다. 또한 비교 통계량은 3.1절에서 사용한 통계량을 사용하였으며 결과를 Table 3.16에서 Table 3.30에 수록하였다.

제공근 변환의 재변환 결과인 Table 3.16에서 Table 3.30을 살펴보면 모든 Table에서 RMSE와 MAE를 기준으로 하였을 때 M1과 M2는 유사한 결과를 주고 있다. 그러나 편향을 기준으로 한다면 M2가 우수한 것을 확인할 수 있다. 따라서 제공근 변환의 재변환은 M2 방법을 사용할 필요가 있다.

**Table 3.18.** The comparison results of the methods of ARMA(1, 1) - Normal

자료수	$\phi$	$\theta$	분산	평균	통계량					
					MAE		Bias		RMSE	
					M1	M2	M1	M2	M1	M2
210	0.3	0.5	$\sigma_a^2 = 0.5$	$\mu = 10$	20.7	20.6	3.2	2.3	26.0	25.9
				$\mu = 20$	41.7	41.6	8.2	7.0	52.3	52.1
			$\sigma_a^2 = 1$	$\mu = 10$	29.2	29.1	4.2	2.5	36.8	36.6
				$\mu = 20$	58.7	58.6	10.0	8.0	73.7	73.4
	0.5	0.3	$\sigma_a^2 = 0.5$	$\mu = 10$	30.9	30.6	10.2	8.7	38.8	38.4
				$\mu = 20$	66.0	64.8	34.0	30.5	82.7	81.3
			$\sigma_a^2 = 1$	$\mu = 10$	43.0	42.7	11.8	9.4	54.2	53.7
				$\mu = 20$	89.4	88.3	36.6	32.2	112.3	110.9

**Table 3.19.** The comparison results of the methods of AR(1) - Double Exponential

자료수	$\phi$	분산	평균	통계량					
				MAE		Bias		RMSE	
				M1	M2	M1	M2	M1	M2
210	0.3	$\sigma_a^2 = 0.5$	$\mu = 10$	21.8	21.8	2.1	0.9	29.9	29.9
			$\mu = 20$	43.8	43.7	5.6	4.1	59.9	59.8
		$\sigma_a^2 = 1$	$\mu = 10$	30.8	30.9	3.1	0.9	42.4	42.3
			$\mu = 20$	61.8	61.7	6.5	4.0	84.6	84.4
	0.5	$\sigma_a^2 = 0.5$	$\mu = 10$	34.6	34.3	8.2	6.2	46.2	45.9
			$\mu = 20$	72.4	71.3	29.9	26.0	95.7	94.6
		$\sigma_a^2 = 1$	$\mu = 10$	48.5	48.3	9.3	6.1	65.0	64.6
			$\mu = 20$	99.3	98.3	30.8	25.6	132.2	131.1

**Table 3.20.** The comparison results of the methods of MA(1) - Double Exponential

자료수	$\theta$	분산	평균	통계량					
				MAE		Bias		RMSE	
				M1	M2	M1	M2	M1	M2
210	0.3	$\sigma_a^2 = 0.5$	$\mu = 10$	15.2	15.2	1.0	-0.1	20.9	20.9
			$\mu = 20$	30.4	30.4	0.9	-0.2	41.6	41.6
		$\sigma_a^2 = 1$	$\mu = 10$	21.4	21.6	2.0	-0.2	29.7	29.6
			$\mu = 20$	42.9	43.0	1.9	-0.3	58.9	58.9
	0.5	$\sigma_a^2 = 0.5$	$\mu = 10$	16.4	16.5	1.1	-0.1	22.2	22.2
			$\mu = 20$	32.9	32.9	1.0	-0.2	44.3	44.3
		$\sigma_a^2 = 1$	$\mu = 10$	23.2	23.3	2.3	-0.2	31.6	31.5
			$\mu = 20$	46.5	46.5	2.2	-0.3	62.7	62.7

## 4. 실제자료분석

### 4.1. 로그 변환 실제자료분석

자료분석에는 Andrews와 Herzberg (1985)에 보고된 자료로 캐나다 Hudson Bay company의 1857년부터 1911년까지의 연간 Lynx Pelts의 판매 수 자료가 사용되었으며 이 자료는 Wei (1990)의 부록에도 수록되어 있다. 총 55개 자료 중 5개는 예측 비교를 위해 남겨 놓았으며 따라서 50개의 자료가 모형식별 및 추정에 사용되었다.

**Table 3.21.** The comparison results of the methods of ARMA(1, 1) - Double Exponential

자료수	$\phi$	$\theta$	분산	평균	통계량					
					MAE		Bias		RMSE	
					M1	M2	M1	M2	M1	M2
210	0.3	0.5	$\sigma_a^2 = 0.5$	$\mu = 10$	29.1	29.1	3.8	2.0	36.7	36.5
				$\mu = 20$	58.5	58.4	9.1	7.0	73.5	73.3
		$\sigma_a^2 = 1$	$\mu = 10$	41.1	41.1	5.4	2.0	52.0	51.7	
			$\mu = 20$	82.5	82.3	11.4	7.7	103.8	103.4	
	0.5	0.3	$\sigma_a^2 = 0.5$	$\mu = 10$	42.8	42.5	11.1	8.7	54.0	53.6
				$\mu = 20$	88.5	87.5	35.4	31.0	111.7	110.3
			$\sigma_a^2 = 1$	$\mu = 10$	60.1	59.7	13.2	9.1	76.0	75.4
				$\mu = 20$	122.5	121.5	39.2	33.1	154.6	153.21

**Table 3.22.** The comparison results of the methods of AR(1) -  $t$  Distribution

자료수	$\phi$	$df$	평균	통계량					
				MAE		Bias		RMSE	
				M1	M2	M1	M2	M1	M2
210	0.3	$\nu = 5$	$\mu = 10$	29.2	29.2	3.7	1.8	39.2	39.1
			$\mu = 20$	57.6	57.5	5.9	3.7	78.6	78.4
		$\nu = 15$	$\mu = 10$	25.3	25.3	2.4	1.1	32.5	32.4
			$\mu = 20$	50.6	50.5	7.3	5.7	65.0	64.8
	0.5	$\nu = 5$	$\mu = 10$	44.0	43.8	7.9	5.1	58.5	58.2
			$\mu = 20$	92.1	91.2	26.2	21.5	120.8	119.8
		$\nu = 15$	$\mu = 10$	39.4	39.2	8.6	6.5	50.2	49.8
			$\mu = 20$	80.5	79.6	30.6	26.5	102.5	101.3

**Table 3.23.** The comparison results of the methods of MA(1) -  $t$  Distribution

자료수	$\theta$	$df$	평균	통계량					
				MAE		Bias		RMSE	
				M1	M2	M1	M2	M1	M2
210	0.3	$\nu = 5$	$\mu = 10$	19.9	19.9	2.5	0.7	26.9	26.8
			$\mu = 20$	40.0	40.1	1.3	-0.5	54.5	54.5
		$\nu = 15$	$\mu = 10$	17.7	17.8	1.4	0.1	22.6	22.6
			$\mu = 20$	35.4	35.4	0.4	-0.8	45.0	45.0
	0.5	$\nu = 5$	$\mu = 10$	21.1	21.2	2.1	0.1	28.3	28.3
			$\mu = 20$	42.5	42.5	1.1	-1.0	56.7	56.8
		$\nu = 15$	$\mu = 10$	18.7	18.8	1.3	-0.1	23.9	23.9
			$\mu = 20$	38.2	38.2	0.9	-0.5	48.3	48.3

Figure 4.1을 살펴보면 로그 변환이 타당한 것으로 판단되어 로그 변환을 실시하였으며 로그 변환을 실시한 결과 그래프는 Figure 4.2이다. 여기서 Figure 4.2는 뒤의 5개 자료를 제거한 후의 그래프이며 분석 후에 얻어진 5개의 예측값을 M1, M2 두 방법을 이용하여 재변환하였다. 로그변환과 선택된 최종 모형은 Wei (1990)의 분석 결과와 같다.

Table 4.1은 모수 추정결과를 나타낸 표이며 이를 이용한 최종 모형식은 다음과 같다.

$$(1 - 0.92762B + 0.55419B^3)(\log(Z_t) - 9.81311) = a_t.$$

**Table 3.24.** The comparison results of the methods of ARMA(1, 1) -  $t$  Distribution

자료수	$\phi$	$\theta$	df	평균	통계량					
					MAE		Bias		RMSE	
					M1	M2	M1	M2	M1	M2
210	0.3	0.5	$\nu = 5$	$\mu = 10$	36.3	36.3	4.3	1.5	48.2	48.0
				$\mu = 20$	72.8	72.7	8.6	5.5	99.1	98.8
			$\nu = 15$	$\mu = 10$	31.1	31.0	3.5	1.6	40.1	39.9
	$\mu = 20$	62.6		62.4	8.9	6.6	80.0	79.7		
	0.5	0.3	$\nu = 5$	$\mu = 10$	52.5	52.2	11.5	7.9	70.4	69.9
				$\mu = 20$	111.2	110.0	40.9	35.4	145.8	144.3
$\nu = 15$			$\mu = 10$	46.1	45.6	14.9	12.2	59.2	58.6	
	$\mu = 20$	94.4	93.3	36.5	31.9	119.9	118.5			

**Table 3.25.** The comparison results of the methods of AR(1) - Truncated Double Exponential

자료수	$\phi$	분산	평균	통계량					
				MAE		Bias		RMSE	
				M1	M2	M1	M2	M1	M2
210	0.3	$\sigma_a^2 = 0.5$	$\mu = 10$	19.6	19.6	2.1	1.2	25.8	25.7
			$\mu = 20$	39.4	39.3	5.8	4.6	51.7	51.6
		$\sigma_a^2 = 1$	$\mu = 10$	23.9	23.8	2.0	0.8	30.5	30.5
	$\mu = 20$		47.9	47.8	5.4	3.9	61.1	61.0	
	0.5	$\sigma_a^2 = 0.5$	$\mu = 10$	31.0	30.8	8.3	6.7	40.1	39.8
			$\mu = 20$	65.9	64.8	30.5	26.9	84.4	83.1
$\sigma_a^2 = 1$		$\mu = 10$	37.1	36.9	8.0	6.1	47.3	47.0	
	$\mu = 20$	77.1	76.2	29.5	25.6	97.8	96.6		

**Table 3.26.** The comparison results of the methods of AR(1) - Truncated Double Exponential

자료수	$\theta$	분산	평균	통계량					
				MAE		Bias		RMSE	
				M1	M2	M1	M2	M1	M2
210	0.3	$\sigma_a^2 = 0.5$	$\mu = 10$	16.6	16.6	0.9	-0.2	21.3	21.2
			$\mu = 20$	27.3	27.3	0.9	0.1	35.9	35.8
		$\sigma_a^2 = 1$	$\mu = 10$	13.6	13.6	0.8	0.1	18.0	17.9
	$\mu = 20$		33.2	33.2	0.8	-0.4	42.4	42.4	
	0.5	$\sigma_a^2 = 0.5$	$\mu = 10$	14.7	14.7	1.0	0.1	19.1	19.1
			$\mu = 20$	29.4	29.4	1.0	0.1	38.2	38.2
$\sigma_a^2 = 1$		$\mu = 10$	17.8	17.8	1.0	-0.2	22.6	22.6	
	$\mu = 20$	35.6	35.6	0.9	-0.4	45.1	45.1		

최종 모형식을 이용하여 5개의 시점을 예측한 후 3절에서 설명한 M1, M2 방법으로 재변환을 실시하였다. MAE, Bias, RMSE 세 가지의 비교통계량이 계산되었으며 결과는 Table 4.2와 같다.

결과를 살펴보면 Bias와 RMSE의 경우 M1 방법이 우수하고 MAE는 M2가 우수한 것을 확인 할 수 있다. 이 결과는 모의실험 결과와는 상반된 결과이나 예측에 사용된 자료를 살펴보면 실제값이 추세에 비해 작기 때문에 작은 값으로 예측하는 M1 방법이 우수한 결과를 주는 것으로 판단된다. 그러나 이 결과를 일반화하는 것은 문제가 있다.

**Table 3.27.** The comparison results of the methods of ARMA(1, 1) - Truncated Doble Exponential

자료수	$\phi$	$\theta$	분산	평균	통계량					
					MAE		Bias		RMSE	
					M1	M2	M1	M2	M1	M2
210	0.3	0.5	$\sigma_a^2 = 0.5$	$\mu = 10$	24.7	24.6	3.7	2.4	31.7	31.6
				$\mu = 20$	49.7	49.5	9.1	7.5	63.7	63.5
		$\sigma_a^2 = 1$	$\mu = 10$	29.5	29.4	3.7	1.9	37.5	37.4	
			$\mu = 20$	59.2	59.1	8.9	6.8	75.2	74.9	
	0.5	0.3	$\sigma_a^2 = 0.5$	$\mu = 10$	36.8	36.5	11.1	9.1	47.2	46.7
				$\mu = 20$	77.3	76.2	35.3	31.4	98.6	97.2
		$\sigma_a^2 = 1$	$\mu = 10$	43.5	43.2	11.1	8.6	55.4	55.0	
			$\mu = 20$	90.1	89.1	35.2	30.8	114.4	113.0	

**Table 3.28.** The comparison results of the methods of AR(1) - Truncated  $t$  Distribution

자료수	$\phi$	df	평균	통계량					
				MAE		Bias		RMSE	
				M1	M2	M1	M2	M1	M2
210	0.3	$\nu = 5$	$\mu = 10$	23.6	23.6	2.2	1.1	29.5	29.4
			$\mu = 20$	49.1	49.1	6.2	4.7	61.5	61.3
		$\nu = 15$	$\mu = 10$	24.6	24.6	2.0	0.8	30.8	30.7
			$\mu = 20$	47.9	47.8	5.9	4.4	59.7	59.6
	0.5	$\nu = 5$	$\mu = 10$	37.5	37.3	8.1	6.2	47.3	47.0
			$\mu = 20$	78.0	77.2	27.9	23.9	98.2	97.2
		$\nu = 15$	$\mu = 10$	36.6	36.4	7.3	5.4	45.6	45.3
			$\mu = 20$	76.5	75.5	31.5	27.6	95.3	94.0

**Table 3.29.** The comparison results of the methods of MA(1) - Truncated  $t$  Distribution

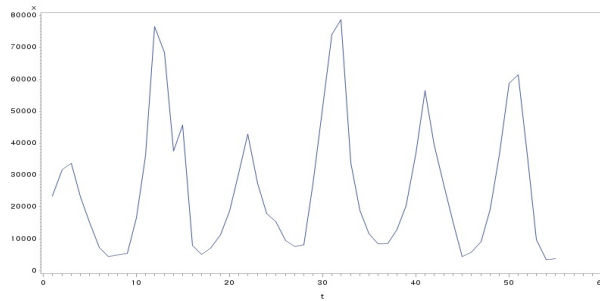
자료수	$\theta$	df	평균	통계량					
				MAE		Bias		RMSE	
				M1	M2	M1	M2	M1	M2
210	0.3	$\nu = 5$	$\mu = 10$	17.1	17.1	1.3	0.2	21.4	21.4
			$\mu = 20$	34.3	34.3	1.0	-0.1	42.8	42.8
		$\nu = 15$	$\mu = 10$	16.8	16.8	1.3	0.2	21.0	21.0
			$\mu = 20$	33.4	33.4	1.5	0.4	41.3	41.3
	0.5	$\nu = 5$	$\mu = 10$	18.3	18.3	1.4	0.1	22.8	22.7
			$\mu = 20$	36.2	36.2	0.7	-0.6	45.2	45.2
		$\nu = 15$	$\mu = 10$	17.7	17.7	1.7	0.5	22.1	22.0
			$\mu = 20$	35.4	35.4	1.4	0.2	44.0	44.0

**4.2. 제공된 변환 실제자료분석**

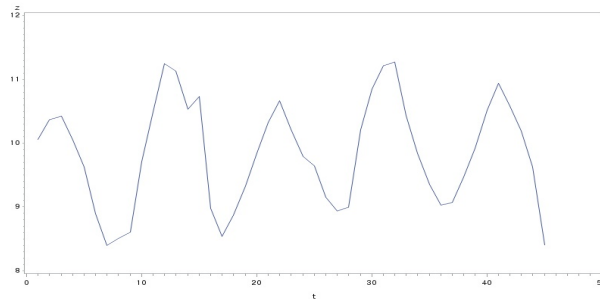
사용된 자료는 Wei (1990)에 수록된 자료로 균등 성비로 구성된 검정파리(Blowfly)를 일정 공간의 실험 상자에 넣고 매일 일정한 양의 먹이를 주었을 때 얻어진 검정파리의 총 수를 일 간격으로 대략 2년 동안 수집한 것이며 Nicholson (1950)으로부터 얻은 데이터이다. 이 데이터 중 Tong (1983)이 고려한 두 개의 부분 시계열 중 Blowfly B를 이용하여 자료를 분석하였다. 총 82개 자료 중 10개는 예측 비교를 위해 남겨 놓았으며 따라서 72개의 자료가 모형식별 및 추정에 사용되었다.

**Table 3.30.** The comparison results of the methods of ARMA(1, 1) - Truncated  $t$  Distribution

자료수	$\phi$	$\theta$	df	통계량						
				평균	MAE		Bias		RMSE	
					M1	M2	M1	M2	M1	M2
210	0.3	0.5	$\nu = 5$	$\mu = 10$	30.1	30.0	4.3	2.5	37.8	37.7
				$\mu = 20$	60.9	60.8	10.1	8.0	76.5	76.2
			$\nu = 15$	$\mu = 10$	28.7	28.6	4.0	2.3	36.1	35.9
				$\mu = 20$	58.4	58.3	8.4	6.4	73.3	73.1
	0.5	0.3	$\nu = 5$	$\mu = 10$	44.0	43.7	10.3	7.8	55.5	55.1
				$\mu = 20$	91.3	90.2	36.4	31.9	115.0	113.6
		$\nu = 15$	$\mu = 10$	42.9	42.5	13.5	11.2	54.2	53.6	
			$\mu = 20$	88.3	87.3	34.8	30.5	111.2	109.9	



**Figure 4.1.** Raw data graph of Lynx data.



**Figure 4.2.** Log transformed data graph of Lynx data.

**Table 4.1.** Parameter estimation results of log transformation

Parameter	Estimate	Standard Error	P-value	Lag
$\mu$	9.81311	0.08924	<0.001	0
$\phi_1$	0.92762	0.07259	<0.001	1
$\phi_3$	-0.55419	0.07259	<0.001	3

**Table 4.2.** Comparison results of Lynx data

MAE		Bias		RMSE	
M1	M2	M1	M2	M1	M2
6001.4	5963.0	812.4	3506.8	6411.8	6719.8

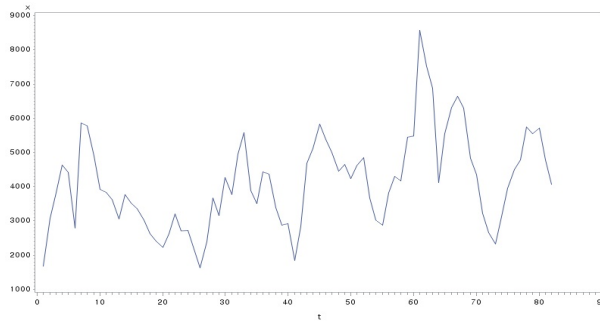


Figure 4.3. Raw data graph of Blowfly data.

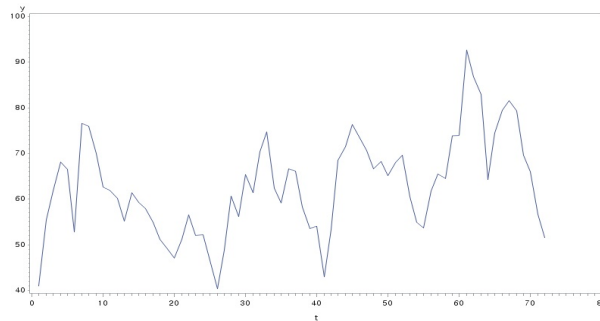


Figure 4.4. Square root transformed data graph of Blowfly data.

Table 4.3. Parameter estimation results of square root transformation

Parameter	Estimate	Standard Error	$P$ -value	Lag
$\mu$	59.01800	3.58320	<0.001	0
$\phi_1$	0.76889	0.08122	<0.001	1

Table 4.4. Comparison results of Blowfly data

MAE		Bias		RMSE	
$M1$	$M2$	$M1$	$M2$	$M1$	$M2$
790.9	783.7	-284.8	-229.3	1037.5	1023.7

분석에서는 Wei (1990)에서와 같이 제곱근 변환 후 AR(1) 모형이 사용되었다. 이때 얻어진 모수추정 값과 표준오차가 Table 4.3에 수록되었다.

Table 4.3의 모수 추정결과를 이용한 최종 모형식은 다음과 같다.

$$(1 - 0.76889B) (\sqrt{Z_t} - 59.018) = a_t.$$

최종 모형식을 이용하여 10개 시점을 예측한 후 3.2절에서 설명한 M1, M2 방법으로 재변환을 실시하였다. MAE, Bias, RMSE 세 가지의 비교통계량이 계산되었으며 결과는 Table 4.4와 같다.

Table 4.4 결과를 살펴보면 모든 비교 통계량에서 M2 방법이 우수한 것을 확인할 수 있으며 특히 편향이 많이 작아진 것을 확인할 수 있다. 이 결과는 모의실험 결과와 일치한다.

## 5. 결론

본 논문에서는 시계열 분석에서 사용하는 멱변환을 연구하였다. 흔히 변환은 자료가 비대칭이거나 분산이 일정하지 않은 경우에 정규성 또는 대칭성을 만족하기 위해 사용된다. 또한 원자료의 예측을 위해 변환된 자료의 예측값을 재변환하여 최종 예측값을 얻게 된다. 흔히 사용하는 재변환을 사용할 경우 편향이 발생하는 것으로 알려져 있다. 이에 본 논문에서는 여러 분포를 가정한 후 재변환을 실시하였을 때 변환에서 얻어진 예측값의 역함수 값을 이용하는 방법과 역함수의 기댓값을 이용하는 방법의 성능을 비교하였다.

모의실험 결과는 역함수의 기댓값을 이용하여 예측을 하는 것이 편향을 줄일 수 있는 것으로 나타났으며 큰 차이는 아닐지 모르지만 MSE 기준에서도 M2 방법이 우수한 것을 확인할 수 있다. 그러나 로그 변환의 경우 MAE를 기준으로 한 결과를 살펴보면 M1 방법이 우수한 것을 확인할 수 있다. 물론 로그 변환을 이용한 실제자료 분석 결과는 모의실험 결과와 약간 다른 결과가 얻어졌다. 이는 하나의 특별한 자료 분석 결과로 이를 일반화하는 것은 문제가 있다. 결론적으로 로그 변환의 재변환을 실시할 경우에는 M2 방법을 사용하는 것이 타당하고 만약 MAE가 중요하다고 판단되면 M1 방법을 고려할 수 있다. 그러나 로그 변환을 제외하면 모두 M2 방법을 사용할 것을 제안한다.

## References

- Andrews, P. E. and Herzberg, A. M. (1985). *The Data: A Collection of Problems from Statistics*, Springer-Verlag, Berlin.
- Box, G. E. P. and Cox, D. R. (1964). An analysis of transformation, *Journal of the Royal Statistical Society, Series, B*, **26**, 211–252.
- Freeman, J. and Modarres, R. (2006). Inverse Box-Cox: The power-normal distribution, *Statistics & Probability Letters*, **76**, 764–772.
- Granger, C. W. J. and Newbold, P. (1976). Forecasting transformed series, *Journal of the Royal Statistical Society, Series, B*, **38**, 189–203.
- Ha, M. H. and Kim, S. (2008). Internet traffic forecasting using power transformation heteroscedastic time series models, *The Korean Journal of Applied Statistics*, **21**, 1037–1044.
- Helmut, L. and Fang, X. (2012). The role of the log transformation in forecasting economic variables, *Empir Econ*, **42**, 619–638.
- Hur, N.-K., Jung, J.-Y., Kim, S. (2009). A Study on air demand forecasting using multivariate time series models, *The Korean Journal of Applied Statistics*, **22**, 1007–1017.
- Kim, S. and Seong, B. (2011). Intervention analysis of Korea tourism data, *The Korean Journal of Applied Statistics*, **24**, 735–743.
- Lee, J.-S., Sohn, H. G. and Kim, S. (2013). Daily peak load forecasting for electricity demand by time series models, *The Korean Journal of Applied Statistics*, **26**, 349–360.
- Nicholson, D. F. (1950). Population oscillations caused by competition for food, *Nature*, London, **165**, 476–477.
- Tong, H. (1983). *Threshold Models in Non-Linear Time Series Analysis*, Springer-verlag, Berlin.
- Wei, W. W. S. (1990). *Time Series Analysis : Univariate and Multivariate Methods*, Pearson.
- Yeo, I. and Johnson, R. A. (2000). A new family of power transformation to improve normality or symmetry, *Biometrika*, **87**, 954–959.



# ARMA( $p, q$ ) 모형에서 멱변환의 재변환에 관한 연구 - 모의실험을 중심으로

강전훈<sup>a</sup> · 신기일<sup>a,1</sup>

<sup>a</sup>한국외국어대학교 통계학과

(2015년 3월 20일 접수, 2015년 5월 19일 수정, 2015년 5월 26일 채택)

---

## 요약

ARMA( $p, q$ ) 모형 분석에서 분산 안정화 또는 정규화를 위해 멱변환(power transformation)이 사용된다. 변환된 자료를 이용하여 분석이 이루어지며 원 자료의 예측을 위해 재변환이 사용된다. 이때 흔히 변환된 자료 분석에서 얻어진 예측값의 역함수 값이 원자료 예측값으로 사용되지만 이는 편향이 있는 것으로 알려져 있다. 이를 해결하기 위해 로그 변환의 경우 Granger과 Newbold (1976)는 로그-정규분포의 기댓값을 이용할 것을 제안하였다. 본 연구에서는 모의실험을 통하여 제곱근 변환과 로그 변환 후 재변환을 사용할 때 예측값으로 기댓값의 역함수를 이용하는 방법과 역함수의 기댓값을 사용하였을 때의 추정의 결과를 모의실험을 통하여 비교하였다.

주요용어: 멱변환, 적률, 정상시계열, 멱-정규분포

---

---

이 연구는 2015년 한국외국어대학교 교내연구비 지원을 받아 수행되었음.

<sup>1</sup>교신저자: (449-791) 경기도 용인시 처인구 모현면 외대로81, 한국외국어대학교 통계학과.

E-mail: keyshin@hufs.ac.kr