

추세 시계열 자료의 부트스트랩 적용

박진수* · 김윤배**† · 송기범***

Applying Bootstrap to Time Series Data Having Trend

Jinsoo Park* · Yun Bae Kim** · Kiburm Song***

■ Abstract ■

In the simulation output analysis, bootstrap method is an applicable resampling technique to insufficient data which are not significant statistically. The moving block bootstrap, the stationary bootstrap, and the threshold bootstrap are typical bootstrap methods to be used for autocorrelated time series data. They are nonparametric methods for stationary time series data, which correctly describe the original data. In the simulation output analysis, however, we may not use them because of the non-stationarity in the data set caused by the trend such as increasing or decreasing. In these cases, we can get rid of the trend by differencing the data, which guarantees the stationarity. We can get the bootstrapped data from the differenced stationary data. Taking a reverse transform to the bootstrapped data, finally, we get the pseudo-samples for the original data. In this paper, we introduce the applicability of bootstrap methods to the time series data having trend, and then verify it through the statistical analyses.

Keyword : Bootstrap Methods, Trended Time Series Data, Simulation Output Analysis

1. 서 론

전통적인 시뮬레이션 출력 분석 방법 중의 하나

는 독립반복법(independent replication method)으로, 여러 번의 시뮬레이션을 독립적으로 수행하여 신뢰성 있는 통계량을 구하는 방법이다. 그러나 경

논문접수일 : 2012년 12월 28일 논문게재확정일 : 2013년 02월 21일

논문수정일(1차 : 2013년 02월 05일)

* 용인대학교 경영정보학과

** 성균관대학교 시스템경영공학과

*** 한양대학교 산업공학과

† 교신저자, kimyb@skku.edu

우에 따라 시뮬레이션을 수행하는데 많은 시간과 비용이 발생하게 되는데 제한된 시간과 비용 하에서는 유효한 통계적 결과를 나타낼 만한 데이터를 충분히 얻지 못하게 된다. 이러한 경우 적은 양의 데이터를 재추출(resampling)하여 원하는 통계량을 추정하는 방법 중 하나가 부트스트랩[9] 방법이다. 부트스트랩 방법은 본래 독립적이며 동일한 분포를 따르는 (independent and identically distributed, IID) 자료를 재추출하는데 적용되도록 개발되었으나 현재는 종속적인 데이터에 대해서도 부트스트랩 기법을 적용할 수 있는 여러 가지 방법들이 제안되었다[7, 13, 14, 15, 18]. 이 중 대표적인 방법이 이동 블록 부트스트랩(moving block bootstrap, MBB)[14, 15], 정상 부트스트랩(stationary bootstrap, SB) [18], 그리고 임계값 부트스트랩(threshold bootstrap, TB) [1, 13]이다. 추세가 없는 시계열 자료 중 자기상관이 존재하는 정상 데이터(stationary data)의 부트스트랩 적용에 있어 많이 사용되는 기법은 MBB와 TB이며 이들은 계절성을 가진 시계열 데이터에 대해서도 적용이 가능하다[4]. 또한 부트스트랩은 다양한 분야에서 비모수적 분석 기법으로 활용되고 있다[2, 3].

시뮬레이션 출력 혹은 실세계에서 발생하는 시계열 자료들은 추세를 포함하는 경우가 많다. 데이터가 증가 혹은 감소 등의 추세를 띠게 되면 기존의 방법을 그대로 적용하여 원본 데이터(original data)를 묘사하는 것이 어려워진다. 또한 계절성이 존재하는 경우도 마찬가지이다. 이는 정상성의 가정이 깨지기 때문인데 이를 극복하기 위한 일반적인 방법은 비정상 시계열 자료를 차분한 후 단위근 검정(unit root test)[8, 17, 19]을 통하여 정상화가 되었는지 확인하는 것이다. 본 논문에서 제안하는 방법은 차분을 통해 정상화된 데이터에 부트스트랩을 적용하는 것이다. 결국, 재추출된 정상 데이터를 역변환 함으로써 원본 데이터의 묘사가 가능해진다. 본 논문에서는 편의상 계절성도 일종의 추세로 보아 그 내용을 전개하였다.

본 논문의 구성은 다음과 같다. 제 2장에서는 비

정상 시계열의 특징, 비정상 시계열 판별법, 시계열 모수 추정법에 대한 기존 연구를 고찰한다. 제 3장에서는 추세 자료에 대한 MBB 및 TB 적용 방법을 제시한다. 제 4장에서는 시뮬레이션으로부터 얻은 추세 데이터에 부트스트랩을 적용한 결과를 나열하고 분석한다. 제 5장에서는 결론 및 추후 연구 과제를 기술한다.

2. 비정상 시계열

본 장에서는 비정상 시계열 자료(non-stationary time series data)와 정상화 방법에 대한 기존 연구를 고찰한다. 먼저 비정상 시계열의 특징과 비정상 시계열을 판별하는 세 가지 방법에 대해 간략히 소개한다. 다음으로 차분을 통해 비정상 데이터를 정상화하는 방법에 대해 알아본다.

2.1 비정상 시계열의 특징 및 판별

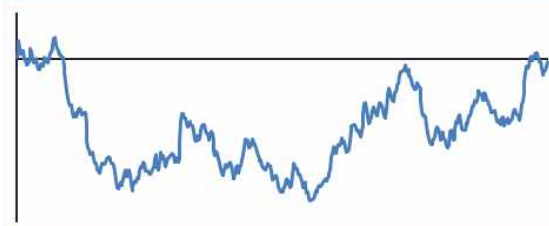
비정상 시계열 자료의 특징은 평균의 비정상성과 분산의 비정상성으로 구분된다. 일반적으로 평균의 비정상성은 추세를 띠는 데이터에서 많이 발견되며 분산의 비정상성은 데이터 내에 계절적 요소가 존재하는 상황에 자주 일어난다. 시계열 자료의 비정상성에 대한 판별은 자기상관함수(autocorrelation function, ACF)와 부분자기상관함수(partial autocorrelation function, PACF)의 그래프를 통해 가능하다. 그러나 ACF와 PACF를 이용해도 비정상성을 판별하기 어려운 경우가 있는데 이러한 경우 사용하는 통계적 검정 방법이 존재한다[8, 17, 19].

현재 비정상 시계열을 판별하기 위한 대표적인 방법으로는 DF(Dickey-Fuller) 검정[8], ADF(augmented DF) 검정[19], PP(Phillips-Perron) 검정[17]의 세 가지가 있다. DF 검정과 ADF 검정 모두 시계열 자료가 AR 과정을 따른다는 가정 하에 전개됐으나 현실적으로 시계열 과정은 누적자기회귀이동평균(autoregressive integrated moving ave-

rage, ARIMA) 과정을 따르는 경우가 더 많다. 이 경우 ARIMA 과정을 AR 과정으로 하여 단위근 검정을 수행함으로써 비정상성을 판별하게 된다. PP 검정은 비모수적인(non-parametric) 방법으로써 다양한 시뮬레이션 결과 데이터 생성과정에서 음의 자기상관이 있는 경우, 특히 작은 크기의 표본에서 심각한 검정의 왜곡현상이 발생한다고 알려져 있다. 이러한 단점 때문에 PP 검정은 ADF 검정보다 신뢰성이 떨어지게 된다.

2.2 차분을 통한 데이터 정상화

추세 시계열 자료는 일반적으로 하향 또는 상향의 형태를 보인다. 물론 주기적인 추세를 띠는 경우도 있는데 이러한 경우 계절성(seasonality)으로 판단하여 처리한다. 본 논문에서 주로 다루는 상·하향 추세 시계열 자료는 ARIMA 모형의 시계열 자료이고, ARIMA 모형의 추세는 모두 백색잡음(white noise)에서 기인한 확률적 추세이므로 시계열 값들의 변동이 크게 나타난다. 또한 시계열 모형 특성상 현재 값이 과거 값의 영향을 받기 때문에 백색잡음과정에서 생성되는 값이 추세를 결정하게 된다. [그림 1]은 이를 설명하기 위한 ARIMA 시계열 모형의 시뮬레이션 결과의 예시이다.



[그림 1] 추세 시계열 데이터의 예

일반적으로 추세를 띠는 비정상 시계열에는 동차 비정상성(homogeneous nonstationary)이 존재한다. 동차 비정상이 나타나는 경우는 차분을 통해 평균을 정상적으로 만들 수 있으며 1계차 또는 2계차 차분을 사용할 수 있다. 보통 1계차 차분을 주로 사용하며 2계차 이상의 차분을 하는 경우는 흔

치 않다. [그림 2]는 [그림 1]에 제시된 ARIMA 모형의 1차 차분 결과이다. 그래프 형태만으로도 시계열 값이 정상적이라는 것을 판단할 수 있다.



[그림 2] 1차 차분 결과의 예

앞서 설명한 바와 같이 분산이 비정상적인 경우는 계절성이 존재하는 경우가 많다. 계절성이 존재하는 시계열 모형은 다양한 사례에 적용이 가능하다[5, 6]. 계절성 시계열 데이터의 경우 계절 주기로 차분한 계절차분을 이용하여 데이터를 정상화 할 수 있다. 또한 분산이 비정상적이면 평균도 비정상적인 경우가 빈번하다. 이 경우 변수변환을 통해 분산을 안정화시키고 이를 차분하여 평균을 정상적으로 만들어야 한다.

정상화된 시계열은 보통 ARIMA 모형으로 표현이 가능해진다. 일반적으로 ARIMA 모형의 모수 추정을 위해서는 Gardner 등[10]이 제안한 방법이 사용된다. 이는 칼만 필터(Kalman filter)[12]를 이용하여 추정하는 방법으로서 $ARMA(p, q)$ 의 모수 집합($\phi_1, \dots, \phi_p, \theta_1, \dots, \theta_q$)을 추정하는 방법이다. 본 논문에서는 칼만 필터를 이용하여 ARMA 모형의 모수를 추정하는 기법이 내장된 오픈 소스 통계 소프트웨어 "R"을 이용하여 부트스트랩 결과의 모수를 추정하였다.

3. 추세 시계열 데이터의 재추출

원본 시계열 자료가 추세 또는 계절성을 띠는 경우 이를 정상화 하는 일반적인 방법으로 차분을 사용한다. 앞서 언급한 바와 같이 보통 최대 2계차 차분까지 수행하면 추세 자료들은 정상성을 띠게 된다. 또한 계절성이 존재하는 경우 이를 제거하기

위해서는 주기만큼의 간격으로 차분을 수행하면 된다. 원본 시계열 자료 $\{Z_1, Z_2, \dots, Z_n\}$ 에 대해 간격 d 인 차분으로 일반화하면 다음과 같이 정의할 수 있다.

$$\nabla_t = Z_{t+d} - Z_t, \quad (t=1, \dots, n-d) \quad (1)$$

여기서 $d=1$ 이면 추세 차분이며 $d \geq 2$ 인 경우 계절 주기가 d 인 계절 차분이 된다. 식 (1) 으로부터 유도된 1차 차분 시계열 자료 $\{\nabla_1, \dots, \nabla_{n-d}\}$ 의 정상성을 확인하였다면 이에 대한 부트스트랩 적용이 가능해진다.

3.1 부트스트랩 방법의 적용

본 논문에서 제안 및 적용하고자 하는 부트스트랩 방법은 MBB와 TB이다. 본 절에서는 이들에 대한 간략한 이론과 추세 시계열 자료에의 직접적인 적용방법을 소개한다. 본 논문에서 제안하는 방법은 먼저 차분을 통해 정상화된 데이터를 부트스트랩 방법으로 재추출 한 후, 이를 역변환하여 원 시데이터를 묘사하는 시계열을 생성하는 방법이다.

MBB는 상관관계가 존재하는 데이터를 위한 비모수적 재추출 방법으로 Künsch[14]과 Liu and Singh[15]에 의해 독립적으로 제안되었다. 식 (1)에 의해 정의된 간격 d 의 차분 시계열 자료 $\{\nabla_1, \dots, \nabla_m\}$ ($m=n-d$)가 정상 시계열이면 재추출 단위인 이동블록 B_i 는 다음과 같이 정의된다.

$$B_i = \{\nabla_i, \dots, \nabla_{i+b-1}\}, \quad (i=1, \dots, m-b+1) \quad (2)$$

여기서 b 는 이동 블록의 크기를 나타내는데 이는 MBB의 성능에 영향을 미치는 가장 중요한 요소이다. b 를 결정하는 이론적인 방법은 존재하지 않으며 경험적인 방법이 Hall[11]에 의해 제안되었다. 이는 여러 개의 블록크기 후보 중에서 평균제곱오차(mean squared error, MSE)를 최소화하는 블록크기를 최적의 블록크기로 결정하는 방법이다.

TB는 시뮬레이션 출력 분석을 위한 비모수적 표본 재추출 방법으로 Kim et al.[13]에 의해 제안

되었다. TB는 자료의 자기상관관계를 묘사하기 위해 재추출 기본단위를 주기(cycle)로 설정한다. 주기는 평균 등의 임계값을 기준으로 상위흐름(high run)과 하위흐름(low run)을 구성하고 한 쌍씩 묶은 단위이다. TB는 이러한 일정 수의 주기를 묶은 청크(chunk) 단위로 재추출을 수행하게 된다. 임의의 임계값이 R 개의 주기를 만든다고 가정하면 주기 C_i 는 다음과 같다.

$$C_i = \{\nabla_{i,1}, \dots, \nabla_{i,m_i}\}, \quad \left(\sum_{i=1}^R m_i = m \right) \quad (3)$$

여기서 m_i 는 i 번째 주기의 크기로서 그 값이 일정치 않음에 주의해야 한다. 이렇게 생성된 주기를 c 개씩 묶으면 총 $\lceil R/c \rceil$ 개의 청크가 생성된다. 먼저 $\lceil R/c \rceil$ 개의 청크에 대해 i 번째 청크 H_i 는 다음과 같이 구성된다.

$$H_i = \{C_{\alpha-c+1}, \dots, C_{\alpha}\}, \quad (i=1, \dots, \lceil R/c \rceil) \quad (4)$$

만일 R 이 c 개로 나누어지지 않는 경우 남은 주기를 묶어 식 (5)와 같이 마지막 청크로 구성한다.

$$H_i = \{C_{\alpha-c+1}, \dots, C_R\}, \quad (i= \lceil R/c \rceil) \quad (5)$$

이는 원본 자료의 손실을 막기 위함이다. 결국 TB의 성능을 좌우하는 것은 청크별 주기 수인 c 값이다. 최적의 c 를 결정하기 위해서는 MSE를 최소화하는 경험적인 방법을 사용할 수 있다[16].

3.2 역변환에 의한 원본 데이터의 묘사

부트스트랩 방법에 의해 재추출된 자료는 원래의 추세 또는 계절성을 반영 할 수 있도록 역변환되어야 한다. 원본 자료 $\{Z_1, \dots, Z_n\}$ 에 대해, 차분 등에 의해 정상화된 자료집합을 $\{\nabla_1, \dots, \nabla_m\}$ 라 하고 이로부터 부트스트랩 방법으로 재추출된 표본을 $\{\nabla_1^*, \dots, \nabla_m^*\}$ 이라 하자. 차분자료의 재추출 표본은 다음과 같은 역변환을 통해 원본 자료의 추세 또는 계절성을 반영할 수 있다.

$$Z_i^* = \begin{cases} Z_i^*, & (i \leq d) \\ Z_{i-d}^* + \nabla_{i-d}^*, & (i = d+1, \dots, n) \end{cases} \quad (6)$$

여기서 우리가 주목해야 할 것은 Z_1^* 부터 Z_d^* 의 결정이다. 이들의 결정 방법은 원본의 해당 시점 자료인 Z_1 부터 Z_d 를 그대로 이용하는 방법, 임의의 Z_1 부터 Z_{i+d-1} 을 취하는 방법, 마지막으로 ∇_i^* 을 구성한 원본 차분 자료가 ∇_k 라면 이에 해당하는 Z_k 를 사용하는 방법이다. 모수추정과 같은 단순한 통계적 추론의 경우 위 세 가지 방법은 비슷한 성능을 보인다. 하지만 원본 시계열 자료의 추세를 가장 잘 반영하는 방법은 첫 번째 방법임을 실험을 통해 확인하였다. 특히 다음절에서 설명하는 시계열 예측에 응용하는 경우 Z_1 부터 Z_d 를 그대로 사용하여 역변환하는 방법의 예측력이 가장 우수하였다.

3.3 응용 및 활용범위

식 (1)에 의해 표현되는 차분 방법은 여러 단계로 적용이 가능하다. 특히 계절성과 추세가 동시에 존재하는 데이터의 경우 추세 차분과 계절 차분의 두 단계를 거쳐 정상화가 가능하다. 즉 원본 데이터를 $d=1$ 로 차분하고 이를 다시 계절 주기로 차분($d \geq 2$)하여 정상화 할 수 있다. 본 논문에서는 계절성 ARIMA(seasonal ARIMA, SARIMA) 모형에 대한 실험을 수행하여 그 응용 가능성을 검증하였다.

본 논문에서 제시하는 알고리즘은 시계열 예측에도 활용이 가능하다. 일반적으로 부트스트랩 방법은 n 개의 원본 데이터에 대해 n 개의 재추출 데이터를 생성한다. 그러나 차분 자료의 정상성을 고려하면 $(n+1)$ 개의 데이터 재추출이 가능하다. 즉 m 개의 차분 데이터로부터 $(m+1)$ 개를 재추출하여 이를 역변환하면 $(n+1)$ 번째 시계열 데이터의 예측이 가능하다. 본 논문에서는 NASDAQ 주가지수에 이를 적용한 결과를 제시하여 그 응용 가능성을 타진하였다.

4. 실험 및 결과 분석

본 장에서는 제 3장에서 제안한 재추출 방법의 적용 및 응용에 대한 타당성을 검증한다. 이를 위해 다양한 시계열 데이터에 제안한 부트스트랩 방법을 적용하여 그 결과를 분석한다. 기본적으로 재추출 과정은 <표 1>과 같은 과정을 통해 진행된다.

<표 1> 추세 데이터의 재추출 절차

단계	실험 내용
1	n 개의 원본 시계열 확보
2	원본 시계열 차분하여 m 개의 차분 시계열 생성
3	차분 시계열의 최적 블록크기 및 청크크기를 결정
4	MBB와 TB를 사용하여 차분 시계열 재추출
5	MBB 및 TB로 재추출한 시계열을 역변환
6	단계 4와 단계 5를 b 회 반복

4.1 실험 계획 및 적용 모형

본 논문에서는 3가지 종류의 실험 및 결과를 제시한다. 먼저 알고리즘의 성능을 확인하기 위해 이론적 모형의 시뮬레이션 결과를 이용하여 실험을 수행한다. 다음으로 추세와 계절성이 동시에 존재하는 경우에 대한 응용 실험을 수행하였다. 마지막으로 앞 절에서 언급한 예측방법을 적용하여 그 결과를 보여준다.

첫 번째 실험에 사용된 원본 시계열 데이터는 기존의 이론적 모형에 의한 시뮬레이션 결과이다. 시뮬레이션 적용 모형은 $ARIMA(0, 1, 1)$ 모형으로서 $IMA(1, 1)$ 모형이라고도 하며 시점 t 의 시계열 z_t 는 다음과 같이 정의된다.

$$(1-B)z_t = (1-\theta B)a_t, \quad (-1 < \theta < 1) \quad (7)$$

여기서 B 는 백시프트 연산자(backshift operator)이며 a_t 는 평균이 0, 분산이 σ_a^2 인 백색잡음을 나타낸다. 본 논문에서는 $ARIMA(0, 1, 1)$ 모형의 모수 $\theta = -0.4$ 와 $\theta = 0.4$ 인 경우에 대해 각각의 실험을 수행하였다. 단계 1의 원본 데이터 크기는 100, 500, 1000의 3가지로 구분하였고 단계 2에서는 $d=1$ 인 차분

결과($m=n-1$)를 사용한다. 이 과정을 100회 독립 반복(independent replication, IR) 수행하여 이항 신뢰구간을 산출하였다. 따라서 수행된 실험의 총 회수는 6회(2가지 모형×3가지 크기)이다.

두 번째 실험에 사용된 데이터는 SARIMA 모형의 시뮬레이션 결과이다. 본 논문에서 사용된 SARIMA 모형은 $SARIMA(1, 1, 1) \times (0, 1, 0)_4$ 모형으로서 다음과 같이 정의된다.

$$(1-\phi B)(1-B)(1-B^4)z_t = (1-\theta B)a_t \quad (8)$$

본 실험에서는 $\phi=0.8$, $\theta=0.5$ 로 설정하여 사용하였다. 이 모형의 경우 먼저 $d=1$ 로 추세 차분한 후 이를 다시 $d=4$ 로 계절차분하여 데이터를 정상화하였다. 실험 결과의 다양성을 위해 추정된 모수와 ACF를 분석하였다.

마지막 실험에 사용된 데이터는 실측 자료로서 2011년 4월 1일부터 2012년 5월 18일까지의 NASDAQ 지수이다. 이 시계열 데이터는 $d=1$ 로 1차 차분하여 정상성을 확보하였다. 2012년 5월 12일까지의 282개 지수 데이터를 이용하여 단계 4에서 ($m+5$)개의 데이터를 재추출하면 287개의 부트스트랩 자료를 생성할 수 있다. 마지막 5개의 데이터는 2012년 5월 14일부터 18일까지 지수의 예측치로 간주할 수 있다.

첫 번째 실험의 최적 블록크기 및 청크크기는 표 2에 나열하였다. <표 2>에 표기된 최적 청크크기는 그 평균크기를 나타낸다. 두 번째 실험의 경우 최적 블록크기는 42이며, 최적 청크크기는 51.05이다. 세 번째 실험의 경우 최적 블록크기는 3, 최적 청크크기는 4.53이다.

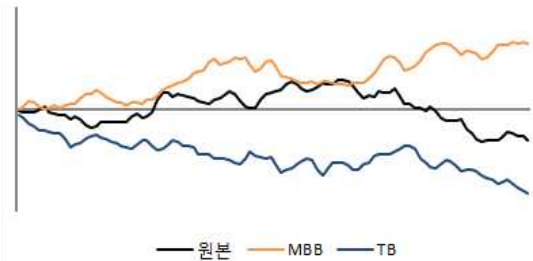
<표 2> 최적의 블록크기와 청크크기

모수	표본크기	최적 블록 및 청크크기	
		MBB	TB
$\theta = -0.4$	100	8	3.23
	500	24	28.60
	1,000	16	32.21
$\theta = 0.4$	100	4	5.09
	500	8	5.17
	1,000	18	5.12

4.2 결과 및 분석

본 절에서는 각 실험의 결과를 다양한 각도에서 분석한다. [그림 3]은 첫 번째 실험의 단계 5에 의한 결과 중 하나의 표본 집합을 그림으로 나타낸 것이다. 이 그림만으로는 추세 반영에 대한 결과를 정확히 판단 할 수 없으므로 모수 추정 결과와 원본 자료의 결과를 비교하는 과정을 거쳐야 한다. 먼저 1000개의 부트스트랩 표본 집합을 이용하여 추정된 모수의 상·하위 90% 백분위 구간을 계산한다. 다음으로 100회의 독립 반복에 대해 계산된 백분위 구간의 참값 포함 회수를 계산한다. 표 3은 이에 대한 결과를 요약하여 나타낸 것이다. 표 3의 결과를 이용하여 이항 신뢰구간을 구하게 되는데 이를 구하는 공식은 다음과 같다.

$$\left(\hat{p} \pm z_{1-\alpha/2} \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \right) \quad (9)$$



[그림 3] 재추출된 표본집합의 예($\theta = 0.4$, 표본크기 100)

<표 3> 상·하위 90% 백분위 구간의 참값 포함 회수

모형	재추출방법	표본크기	포함회수
$\theta = -0.4$	MBB	100	91
		500	93
		1,000	76
	TB	100	88
		500	90
		1,000	87
$\theta = 0.4$	MBB	100	83
		500	63
		1,000	81
	TB	100	87
		500	89
		1,000	87

식 (9)에서 \hat{p} 는 베르누이 시행의 성공확률로서 <표 3>의 참값 포함 회수를 전체 반복시행 회수인 100으로 나눈 결과이다. n 은 본래 표본의 크기이지만 본 실험에서는 반복 회수인 100을 나타낸다. 이로부터 90% 이항 신뢰구간을 계산하여 정리하면 <표 4>와 같다. <표 4>의 하한과 상한 사이에 $\hat{p} = 0.9$ 가 포함되어야 최종적으로 추정된 모수의 타당성을 입증할 수 있다. <표 4>에서는 이에 해당하지 않는 실험 결과를 음영으로 처리하여 부트스트랩이 최종적으로 실패한 경우를 표시하였다. TB의 경우 모든 실험에 대해 성공적이었으나 MBB의 경우 전체의 절반 정도는 실패하고 있다. 이는 MBB의 성능이 TB에 비해 상대적으로 떨어지는 것을 보여 주며 이는 이미 Park and Willemain[16]에 의해 보고된 바 있다.

<표 4> 모수 추정치의 90% 이항 신뢰구간

모수	재추출방법	표본크기	하한값	상한값
$\theta = -0.4$	MBB	100	0.86	0.96
		500	0.89	0.97
		1,000	0.69	0.83
	TB	100	0.83	0.93
		500	0.85	0.95
		1000	0.81	0.93
$\theta = 0.4$	MBB	100	0.77	0.89
		500	0.55	0.71
		1,000	0.75	0.87
	TB	100	0.81	0.93
		500	0.84	0.94
		1000	0.81	0.93

두 번째 실험의 경우 먼저 각 모수의 90% 백분위 신뢰구간을 산출하여 <표 5>에 정리하였다. <표 5>에서 IR은 100회 독립반복한 시계열 데이터에 의한 결과이다. MBB와 TB의 결과는 100개 중 하나의 시계열 데이터만 이용한 결과임에 주의해야 한다. 즉 IR이 100개의 독립 원본 시계열을 사용한 반면 부트스트랩은 하나의 시계열만 사용한 것이다. <표 5>로 부터 각 방법을 통해 추정된 모수의 신뢰구간이 매우 유사함을 확인할 수 있다. 이는 부트스트랩 방법에 의해 추정된 모수의 신뢰성을 확인할

수 있는 결과이다.

<표 5> 모수의 90% 백분위 신뢰구간

모수	IR	MBB	TB
$\phi = 0.8$	0.7998±0.0624	0.8158±0.0606	0.8201±0.0533
$\theta = 0.5$	0.5026±0.0938	0.5315±0.0898	0.5247±0.0635

본 논문에서 제시하는 재추출 방법이 원본 데이터의 상관관계 구조를 얼마나 잘 반영하는지 확인하기 위해 ACF를 확인해 보자. <표 6>은 $w_t = (1-B)(1-B^4)z_t$ 의 각 방법에 의한 ACF 추정치의 90% 백분위 신뢰구간을 계산하여 정리한 것이다. 이는 w_t 가 $ARMA(1, 1)$ 로 표현되어 ACF의 이론값을 구할 수 있기 때문이다. <표 6>에서 보는 바와 같이 부트스트랩 방법의 결과는 IR의 결과와 유사하며 모두 이론값을 포함하는 것을 확인할 수 있다.

<표 6> w_t 의 ACF에 대한 90% 백분위 신뢰구간

k	이론값	IR	MBB	TB
1	0.4000	0.3931±0.0614	0.3571±0.0606	0.3762±0.0674
2	0.3200	0.3152±0.0770	0.3214±0.0898	0.3339±0.0634
3	0.2560	0.2502±0.0669	0.2226±0.0653	0.2454±0.0585
4	0.2048	0.1993±0.0696	0.1376±0.0722	0.1539±0.0728
5	0.1638	0.1591±0.0662	0.1512±0.0729	0.1706±0.0746

마지막 실험은 본 재추출 방법을 적용하여 예측을 수행하는 것이다. 2011년 4월 1일부터 2012년 5월 12일의 NASDAQ 지수 자료는 1차 차분하면 정상화됨을 확인하였다. 또한 이에 부합하여 다음과 같은 $ARIMA(1, 1, 1)$ 모형에 적합함을 알 수 있다.

$$(1 + 0.87B)(1 - B)z_t = 0.96 + (1 + 0.78B)a_t \quad (10)$$

따라서 본 논문에서는 ARIMA 모형에 의한 5일치의 예측 결과와 재추출 방법에 의한 예측 결과를 비교하였다. <표 7>은 실측값과 각 방법에 의한 예측값들을 정리하여 비교한 결과이다. <표 7>에서 보는 바와 같이 부트스트랩 방법을 적용한 결과가 예측력이 매우 우수한 것을 확인할 수 있다.

<표 7> NASDAQ 지수 예측결과
(2012. 05. 14.~2012. 05. 18)

	05.14.	05.15.	05.16.	05.17.	05.18.
실측	2902.58	2893.76	2874.04	2813.69	2778.79
ARIMA	2932.95 (1.05)*	2934.66 (1.41)	2934.13 (2.09)	2935.55 (4.33)	2935.28 (5.63)
MBB	2925.42 (0.79)	2924.51 (1.06)	2927.76 (1.87)	2929.56 (4.12)	2929.57 (5.42)
TB	2902.29 (0.01)	2903.16 (0.32)	2904.29 (1.05)	2906.15 (3.29)	2905.51 (4.56)

주) *괄호안의 숫자는 상대백분율오차를 나타내며 단위는 %임.

본 실험들로부터 다양한 시계열 자료에 부트스트랩을 적용한 결과, 다음과 같은 결론을 도출할 수 있다. 먼저 추세 및 계절성을 그대로 반영할 수 있다는 점이 그 첫 번째 이다. 다음으로 추정된 모수의 정확성을 들 수 있다. 적은 양의 자료를 이용하여 정확도 높은 분석을 수행할 수 있다는 것은 부트스트랩 방법의 큰 장점이다. 마지막으로 상관관계 구조의 반영에 대한 정확성을 확인할 수 있다. 부트스트랩을 이용하는 가장 큰 이유가 바로 상관관계 구조의 반영에 있으므로 그 장점을 그대로 가지고 있다는 것이다. 추세 시계열에 존재하는 상관관계 구조를 그대로 반영한다는 것은 부트스트랩 방법을 적용한 결과가 매우 우수한 성능을 보임을 판단할 수 있는 근거이다.

5. 결 론

본 논문에서는 추세 및 계절성을 띠며 자기상관관계가 있는 시계열 자료를 비모수 부트스트랩을 이용하여 재추출하는 방법을 제안하고 검증해 보았다. 추세를 동반한 원본 시계열 자료를 차분하여 정상 시계열 자료로 변환하고 이로부터 부트스트랩 기법을 적용하여 재추출하게 된다. 마지막으로 재추출된 자료를 다시 역변환함으로써 원본 자료와 유사한 표본을 얻을 수 있다. 본 논문에서 적용한 부트스트랩 방법은 이동 블록 부트스트랩과 임계값 부트스트랩 방법이다. 제안한 방법을 검증하기 위

해 추세 및 계절성을 동반한 이론적 모형의 시뮬레이션 결과를 원본 시계열 자료로 활용하여 실험 및 결과를 분석하였다. 또한 그 응용 가능성을 타진하기 위해 실제 시계열 자료인 NASDAQ 지수를 이용, 이를 예측하는 방법을 간략히 소개하였다.

본 논문에서는 비정상 데이터 중 증감 추세 또는 계절성이 존재하는 경우에 대한 적용 및 실험 결과만을 제시하였다. 추후 일반적인 모든 비정상 데이터에 재추출 기법의 적용이 가능하도록 확장한다면 다양한 실제 사례에 응용이 가능할 것으로 사료된다.

참 고 문 헌

- [1] 김윤배, “시뮬레이션 출력분석을 위한 임계값 부트스트랩의 성능개선”, 『대한산업공학회지』, 제23권, 제4호(1997), pp.755-767.
- [2] 김윤배, 김재범, “임계값 부트스트랩을 사용한 시뮬레이션 입력 시나리오의 생성”, 『경영과학』, 제22권, 제1호(2005), pp.15-26.
- [3] 민재형, 김진한, “DEA의 효율성 평균 차이에 대한 비모수적 검증: 부트스트랩 접근법”, 『한국경영과학회지』, 제24권, 제2호(1999), pp.53-68.
- [4] 박진수, 김윤배, “계절성 데이터의 부트스트랩 적용에 관한 연구”, 『한국시뮬레이션학회 논문지』, 제19권, 제3호(2010), pp.119-125.
- [5] 서명윤, 이종태, “계절상품 판매매출액 시계열의 계절 조정에 관한 연구”, 『경영과학』, 제20권, 제1호(2003), pp.103-124.
- [6] 전태준, “계절적 시계열 모형화를 위한 VSACF의 확장”, 『한국경영과학회지』, 제16권, 제1호(1991), pp.68-75.
- [7] Carlstein, E., “The use of subseries methods for estimating the variance of a general statistic from a stationary time series,” *Annals of Statistics*, Vol.14(1986), pp.1171-1179.
- [8] Dickey, D.A. and W.A. Fuller, “Distribution of the estimators for autoregressive time

- series with a unit root," *Journal of the American Statistical Association*, Vol.74, No.366 (1979), pp.427-431.
- [9] Efron, B., "Bootstrap methods : another look at the jackknife," *Annals of Statistics*, Vol. 7(1979). pp.1-26.
- [10] Gardner, G., A.C. Harvey, and G.D.A. Phillips, "Algorithm AS 154 : An algorithm for exact maximum likelihood estimation of autoregressive-moving average models by means of Kalman filtering," *Applied Statistics*, Vol. 29, No.3(1980), pp.311-322.
- [11] Hall, P., J. Horowitz, and B. Jing, "On blocking rules for the bootstrap with dependent data," *Biometrika*, Vol.82(1995), pp.561-574.
- [12] Kalman, R.E., "A new approach to linear filtering and prediction problems," *Transactions of the ASME-Journal of Basic Engineering*, Vol.82, Series D(1960), pp.35-45.
- [13] Kim, Y.B., T.R. Willemain, J. Haddock, and G.C. Runger, "The threshold bootstrap : a new approach to simulation output analysis," *Proceedings of the 1993 Winter Simulation Conference*, (1993), pp.498-502.
- [14] Künsch, H.R., "The jackknife and the bootstrap for general stationary observations," *Annals of Statistics*, Vol.17(1989), pp.1217-1241.
- [15] Liu, R. and K. Singh, "Moving blocks jackknife and bootstrap capture weak dependence," In : LePage, R., Billard, L. (Eds.), *Exploring the Limits of Bootstrap*, Wiley, New York, (1992), pp.225-248.
- [16] Park, D. and T.R. Willemain, "The threshold bootstrap and threshold jackknife," *Computational Statistics and Data Analysis*, Vol.31 (1999), pp.187-202.
- [17] Phillips, P.C.B. and P. Perron, "Testing for a unit root in time series regression," *Biometrika*, Vol.75, No.2(1988), pp.335-346.
- [18] Politis, D.N. and J.P. Romano, "The stationary bootstrap," *Journal of American Statistics Association*, Vol.89(1994), pp.1303-1313.
- [19] Said, S.E. and D.A. Dickey, "Testing for unit roots in autoregressive-moving average models of unknown order," *Biometrika*, Vol.71, No. 3(1984), pp.599-607.