

A study on estimating piecewise linear trend model using the simple moving average of differenced time series

Okyoung Na^{1,a}

^aDepartment of Applied Statistics, Kyonggi University

Abstract

In a piecewise linear trend model, the change points coincide with the mean change points of the first differenced time series. Therefore, by detecting the mean change points of the first differenced time series, one can estimate the change points of the piecewise linear trend model. In this paper, based on this fact, a method is proposed for detecting change points of the piecewise linear trend model using the simple moving average of the first differenced time series rather than estimates of the slope or residuals. Our Monte Carlo simulation experiments show that the proposed method performs well in estimating the number of change points not only when the error terms in the piecewise linear trend model are independent but also when they are serially correlated.

Keywords: piecewise linear trend model, change-point, simple moving average

1. 서론

추세가 존재하는 시계열 자료를 분석할 때 $y_t = g(t) + \epsilon_t$ 처럼 결정적 추세(deterministic trend) $g(t)$ 와 정상 시계열(stationary process) $\{\epsilon_t\}$ 의 합으로 표현되는 추세-정상 시계열(trend-stationary process)을 많이 사용한다. 결정적 추세 $g(t)$ 는 시간 t 에 대한 함수로 선형 추세 $g(t) = \beta_0 + \beta_1 t$ 를 가장 많이 사용하지만, 현실적으로 선형 함수만을 사용한 추세 분석에는 한계가 존재하며 이차 추세 $g(t) = \beta_0 + \beta_1 t + \beta_2 t^2$ 와 같은 다항 추세 $g(t) = \beta_0 + \beta_1 t + \dots + \beta_p t^p$ 를 이용하여 분석하기도 한다. 다항 추세 이외에도 조각별 선형 추세(piecewise linear trend)를 이용하여 비선형적 형태의 추세를 분석할 수 있으며, 최근 Campra와 Morales (2016), Tomal과 Rahman (2021), Jiang 등 (2022)은 기후 자료나 주택 가격, COVID-19 자료를 분석할 때 조각별 선형 추세 모형이나 이를 확장한 모형을 사용하였다.

조각별 선형 추세는 이름에서 알 수 있듯이 부분적으로는 선형 추세를 갖지만 기울기가 어떤 시점을 경계로 변하는 모형이며, 만약 기울기가 변하는 지점을 알고 있다면 선형 추세와 마찬가지로 회귀분석 기법을 이용하여 어렵지 않게 추정할 수 있다. 그러나 현실적으로 변화점의 위치는 물론 변화점의 유무 및 개수가 알려진 경우는 거의 없으며, 조각별 선형 추세 모형을 적합시키기 위해서는 변화점을 탐색하는 단계가 반드시 필요하다. 하지만 변화점의 개수와 위치를 추정하는 것은 간단한 문제가 아니며, 이 부분이 조각별 선형 추세 모형을 적합시킬 때 가장 어려운 부분일 것이다. 조각별 선형 회귀 모형이나 변화점 탐지에 대한 연구들은 Yau와 Zhao (2016), Eichinger와 Kirch (2018), Park 등 (2019), Kim 등 (2022), Muggeo (2003, 2016), Wang 등 (2020), Tomal과 Rahman (2021)처럼 최근까지도 다양한 분야에서 진행되고 있으며, R 패키지도 개발된 상태다. Zeileis 등 (2002)의 strucchange, Muggeo (2008, 2020)의 segmented, Fong 등 (2017)의 chngpt 패키지가 대표적이다.

¹Department of Applied Statistics, Kyonggi University, 154-42 Gwanggyosan-Ro, Yeongtong-Gu, Suwon-Si, Gyeonggi-Do 16227, Korea. E-mail: okna@kyonggi.ac.kr

조각별 선형 회귀 모형에 대한 대부분의 연구들은 기울기에 대한 추정값이나 잔차를 사용한다. 그러나 조각별 선형 추세를 이용하여 시계열 자료를 분석하는 경우에는 원 시계열 자료 대신 1차 차분한 시계열을 사용하면 좀 더 쉽게 변화점을 추정할 수 있다. 왜냐하면 조각별 선형 추세 모형에서의 기울기의 변화점이 1차 차분한 시계열의 평균 변화점과 일치하기 때문이다. 따라서 본 논문에서는 1차 차분한 시계열의 평균 변화점 탐색을 기초로 조각별 선형 추세 모형의 변화점을 추정, 모형을 적합시키는 방법을 제안하고자 한다. 1차 차분한 시계열의 평균 변화점을 탐색하는 방법으로 Eichinger와 Kirch (2018)처럼 단순히동평균값의 차이를 검정통계량으로 사용하였으며, 이에 대한 구체적인 설명은 본문의 2장에 기술하였다.

본 논문은 다음처럼 구성한다. 2장에 조각별 선형 추세 모형을 만족하는 시계열과 1차 차분한 시계열의 관계를 기술하고, 이를 바탕으로 변화점을 탐색하고 조각별 선형 추세 모형을 적합시키는 방법을 소개한다. 3장에서는 본 논문에서 제안한 방법과 Muggeo (2020)이 제안한 방법에 대한 비교 실험을 진행, 그 결과를 설명하고, 4장에서 자료 분석을 실시한다. 마지막으로 5장에 본 논문의 방법이 가지는 특징을 간단히 정리한다.

2. 조각별 선형 추세 모형 적합

2.1. 모형

시계열 $\{y_t\}$ 가 관측된 구간 내에서 $(k+1)$ 개의 조각을 가지는 조각별 선형 추세 모형을 따른다고 가정하자. 다시 말해서, 관측값 y_1, \dots, y_n 이 다음 모형식을 만족한다:

$$y_t = \alpha + \sum_{j=0}^k \beta_{j+1} (t - \psi_j)_+ + \epsilon_t, \quad (2.1)$$

여기서 함수 $(x)_+$ 는 실수 x 와 0 중 최대값을 표시하며, k 는 0 이상의 정수, α, β_1 은 실수, $\beta_2, \dots, \beta_{k+1}$ 은 0이 아닌 실수이다. 그리고 $\psi_0, \psi_1, \dots, \psi_k$ 는 각 조각의 왼쪽 경계값으로 $\psi_0 = 0$ 이고 부등식 $1 < \psi_1 < \psi_2 < \dots < \psi_k < n$ 을 만족한다. 관측 시점 t 가 첫 번째 구간 $(\psi_0, \psi_1]$ 에 속할 때, $y_t = \alpha + \beta_1 t + \epsilon_t$ 을 만족하며, 일반적으로 다음이 성립한다:

$$y_t = (\alpha - \beta_2 \psi_1 - \dots - \beta_j \psi_{j-1}) + (\beta_1 + \dots + \beta_j) t + \epsilon_t, \quad \psi_{j-1} < t \leq \psi_j, \quad j = 1, 2, \dots, k+1, \quad \psi_{k+1} = n.$$

그러므로 α 와 β_1 은 첫 번째 구간에서의 선형 추세의 절편과 기울기를 각각 뜻하고, $\beta_2, \dots, \beta_{k+1}$ 은 기울기의 변화량, ψ_1, \dots, ψ_k 는 기울기가 변하는 시점, k 는 변화점의 개수라고 해석할 수 있다.

모형식 (2.1)에 포함된 오차항 $\{\epsilon_t\}$ 는 다음 조건을 만족한다고 가정한다.

(E1) $\{\epsilon_t\}$ 는 평균이 0인 정상 가우스 과정이다.

(E2) 시차 h 에서의 자기공분산을 $\gamma_h = E(\epsilon_t \epsilon_{t+h})$ 이라고 할 때, $\gamma_0 > 0$ 이고 $\sum_{h=0}^{\infty} \gamma_h^2 < \infty$ 이 성립한다.

조각별 선형 회귀 모형에 대한 대부분의 연구에서는 오차항 $\{\epsilon_t\}$ 가 서로 독립이고 동일한 경우만을 고려한다. 그러나 일반적으로 조각별 선형 추세 모형을 이용하여 시계열 자료를 분석하는 경우에는 오차항들이 독립성을 만족하기 어렵다. 그러므로 본 논문에서는 오차항들이 서로 독립인 경우뿐만 아니라 자기상관이 존재하는 경우까지 포함하기 위해 (E1), (E2)처럼 $\{\epsilon_t\}$ 가 평균이 0인 정상시계열이라고 가정하였다. 그리고 변화점을 탐색할 때 사용하는 검정통계량이 $\max_{1 \leq t \leq n} |\epsilon_t|$ 와 관련되어 있는데, 이에 대한 이론적 성질을 얻기 위해 정상성에 더불어 추가로 $\{\epsilon_t\}$ 가 가우스 과정이라고 가정하였다.

2.2. 차분과 평균변화

시계열 $\{y_t\}$ 가 조각별 선형 추세 모형, 즉 모형식 (2.1)을 만족하는 경우, $\{y_t\}$ 를 1차 차분하여 구한 시계열 $\{\nabla y_t\}$ 는 다음 식을 만족한다:

$$\begin{aligned} \nabla y_t &= y_t - y_{t-1}, \quad \nabla \epsilon_t = \epsilon_t - \epsilon_{t-1}, \\ \nabla y_t &= \beta_1 + \dots + \beta_{j+1} + \nabla \epsilon_t, \quad \psi_j < t \leq \psi_{j+1}, \quad j = 0, 1, \dots, k. \end{aligned} \tag{2.2}$$

이로부터 우리는 1차 차분한 시계열의 평균이 $t = \psi_j$ 에서 β_{j+1} 만큼 변한다는 것을 알 수 있다. 그러므로 1차 차분한 시계열 $\{\nabla y_t\}$ 의 평균 변화점을 탐지하면, k 와 ψ_1, \dots, ψ_k 을 추정할 수 있다.

평균의 변화점을 탐지하는 방법에는 CUSUM 절차 (Aue와 Horváth, 2013), MOSUM 절차 (Eichinger와 Kirch, 2018), 가능도비를 이용한 방법 (Aue와 Horváth, 2013; Yau와 Zhao, 2016) 등 다양한 방식이 존재하는데, CUSUM과 MOSUM 절차는 둘 다 비모수적인 방법으로 계산이 쉽다. 특히 MOSUM 절차는 변화점의 존재 여부를 검정할 뿐만 아니라 위치와 개수를 시각적으로 쉽게 알아볼 수 있다는 장점을 가진다. 그러므로 본 논문에서는 MOSUM 절차처럼 단순이동평균의 차이를 이용하여 평균 변화를 탐지하는 방식을 선택하였다. 하지만 1차 차분한 시계열의 오차항 $\{\nabla \epsilon_t\}$ 는 원 시계열의 오차항 $\{\epsilon_t\}$ 가 서로 독립인 정상시계열인 경우라도 장기간 분산이 양수라는 보장이 없으며 Eichinger와 Kirch (2018)에 기술된 조건을 만족하지 못 한다. 따라서 MOSUM 절차와 유사한 통계량을 사용하더라도 임계값 계산시 Eichinger와 Kirch (2018)의 결과를 직접적으로 사용할 수 없으며, 이에 대한 연구가 필요하다.

2.3. 변화점 탐지

1차 차분한 시계열의 평균 변화점을 탐지하기 위한 검정통계량으로

$$D_s = \frac{1}{m} \sum_{t=s+1}^{s+m} \nabla y_t - \frac{1}{m} \sum_{t=s-m+1}^s \nabla y_t, \quad s = m + 1, \dots, n - m \tag{2.3}$$

을 사용한다. D_s 는 이동평균기간이 m 인 단순이동평균값의 차이를 이용하여 계산할 수 있으며, 이 때 사용한 이동평균기간이

$$\min_{1 \leq j \leq k+1} |\psi_j - \psi_{j-1}| > 2m \tag{2.4}$$

을 만족하면, D_s 는 $B_s = m^{-1} \sum_{j=1}^k \beta_{j+1} (m - |s - \psi_j|)_+ I(k > 0)$ 와 $E_s = m^{-1} (\epsilon_{s+m} - 2\epsilon_s + \epsilon_{s-m})$ 의 합으로 나누어 쓸 수 있다. 단, $I(\cdot)$ 는 지시함수를 나타낸다.

변화점이 적어도 하나 존재하는 경우, $|B_s|$ 는 $\psi_j - m < s < \psi_j + m$ 일 때만 0이 아닌 값을 가지며, 그래프의 모양이 삼각형 모양이고, $s = \psi_j$ 에서 극대값 $|\beta_{j+1}|$ 을 가진다. 그리고 Deo (1972, 1975)에 의하면, $\{\epsilon_t\}$ 가 (E1), (E2)를 만족할 때

$$\max_{m < s \leq n-m} |E_s| \leq \frac{4}{m} \max_{1 \leq t \leq n} |\epsilon_t| = \frac{4}{m} \left(\sqrt{2 \log n} + o_P(1) \right) \quad \text{as } n \rightarrow \infty$$

이 성립한다. 따라서 이동평균기간 m 이 (2.4)를 만족하면, 다음이 성립함을 알 수 있다.

- 만약 $\min_{1 \leq j \leq k} |s - \psi_j| \geq m$ 이면, $B_s = 0$ 이므로

$$|D_s| = |E_s| \leq \frac{4}{m} \max_{1 \leq t \leq n} |\epsilon_t|$$

을 만족한다.

- 만약 어떤 j 에 대하여 $m > |s - \psi_j| > 8 \max_{1 \leq t \leq n} |\epsilon_t| / |\beta_{j+1}|$ 이면, $B_s = |\beta_{j+1}|(1 - |s - \psi_j|/m)$ 이므로 다음 부등식이 성립한다:

$$\begin{aligned} |D_s| &\leq |B_s| + |E_s| \\ &< |\beta_{j+1}| - \frac{8 \max_{1 \leq t \leq n} |\epsilon_t|}{m} + \frac{4 \max_{1 \leq t \leq n} |\epsilon_t|}{m} \\ &\leq |B_{\psi_j}| - |E_{\psi_j}| \leq |D_{\psi_j}|. \end{aligned}$$

위 사실로부터 이동평균기간 m 이 조건 (2.4)와 더불어

$$\lim_{n \rightarrow \infty} \frac{\sqrt{\log n}}{m \min_{1 \leq j \leq k} |\beta_{j+1}|} = 0 \quad (2.5)$$

을 만족하면, $|D_s|$ 의 최대점이 변화점 $\psi_j, j = 1, \dots, k$ 근처에 존재함을 알 수 있다. 따라서 $|D_s|$ 의 최대점 또는 국소적 최대점을 이용하여 변화점 ψ_1, \dots, ψ_k 를 추정할 수 있다. 추가로 본 논문에서는 변화점의 개수 k 와 기울기의 변화량 $\beta_2, \dots, \beta_{k+1}$ 이 모두 유한한 상수값을 가지고, 변화점의 위치가 $\psi_j = c_j n$ 처럼 표현되는 경우를 주로 고려한다. 여기서 c_j 는 $0 < c_1 < \dots < c_k < 1$ 을 만족하는 상수이며, n 은 자료의 개수이다. 그러면 $\min_{1 \leq j \leq k} |\beta_{j+1}|$ 가 양의 상수가 되므로, 이동평균기간을 $m = \log n, \log n \sqrt{\log n}, \sqrt{n}$ 처럼 $\lim_{n \rightarrow \infty} (m/n) = 0$ 과 $\lim_{n \rightarrow \infty} (\sqrt{\log n}/m) = 0$ 를 만족하는 값으로 선택하면 된다.

이제 E_s 에 대해서 좀 더 살펴보자. 오차항에 대한 가정 (E1), (E2)로부터 $\{mE_s\}$ 가 다음 조건을 만족하는 가우스 과정임을 유도할 수 있다.

$$\begin{aligned} E(mE_s) &= E(\epsilon_{s+m} - 2\epsilon_s + \epsilon_{s-m}) = 0, \quad \forall s, \\ \text{Var}(mE_s) &= 6\gamma_0 - 8\gamma_m + 2\gamma_{2m}, \quad \forall s, \\ \text{Cov}(mE_s, mE_{s+h}) &= \gamma_{h-2m} - 4\gamma_{h-m} + 6\gamma_h - 4\gamma_{h+m} + \gamma_{h+2m}, \quad \forall s. \end{aligned}$$

그러므로 Deo (1972)의 Corollary에 의해 다음이 성립한다:

$$\lim_{n \rightarrow \infty} P \left(\frac{m}{\sqrt{6\gamma_0 - 8\gamma_m + 2\gamma_{2m}}} \max_{m < s \leq n-m} |E_s| \leq a_{n-2m}x + b_{n-2m} \right) = \exp(-e^{-x}), \quad \forall x \in \mathbb{R}. \quad (2.6)$$

단, $a_{n-2m} = (2 \log(n-2m))^{-1/2}$, $b_{n-2m} = (2 \log(n-2m))^{1/2} - (8 \log(n-2m))^{-1/2} (\log \log(n-2m) + \log(4\pi) - 4)$ 이다. 따라서 유의수준 α 에서

$$\max_{m < s \leq n-m} |D_s| \geq c_\alpha, \quad c_\alpha = m^{-1} (-a_{n-2m} \log(-\log(1-\alpha)) + b_{n-2m}) \sqrt{6\hat{\gamma}_0 - 8\hat{\gamma}_m + 2\hat{\gamma}_{2m}} \quad (2.7)$$

이면 변화점이 존재한다고 판단할 수 있다. 여기서 $\hat{\gamma}_h$ 는 γ_h 에 대한 추정량이다.

2.4. 모형 적합 과정

본 논문에서는 다음 과정에 따라 조각별 선형 추세 모형을 적합시킬 것을 제안한다.

- (S1) 이동평균기간 m 을 선택하고, 주어진 시계열 자료를 이용하여 $D_s, s = m+1, \dots, n-m$ 을 계산한다.
- (S2) 조각의 경계값들의 후보 집합 \mathcal{P} 에 대한 초기치를 $\mathcal{P}_0 = \{1, n\}$ 으로 설정하고, 다음 과정을 멈출 때까지 반복한다.

(2-1) \mathcal{P} 의 원소의 개수가 q 이고, 원소를 순서대로 나열한 값을 $p_1 < p_2 < \dots < p_q$ 라고 하자.

- (2-2) 만약 모든 j 에 대하여 $p_j - p_{j-1} < 2m$ 이면, 계산 과정을 멈춘다.
- (2-3) 만약 $p_j - p_{j-1} \geq 2m$ 를 만족하는 j 가 존재하면, 조건을 만족하는 j 에 대해서만 구간 $[p_{j-1} + m, p_j - m]$ 에서의 최대점 $\arg \max_{p_{j-1} + m \leq s \leq p_j - m} |D_s|$ 을 계산하여 \mathcal{P} 에 추가한다. 만약 구간 내에서의 최대점의 개수가 한 개가 아니라면, 최대점 중 최소값만을 \mathcal{P} 에 추가한다.
- (S3) (S2)의 반복 계산을 통해 구한 \mathcal{P} 의 원소 중 1과 n 을 제외한 나머지를 $\tilde{\psi}_1 < \dots < \tilde{\psi}_k$ 으로 두고, 다음을 수행한다.
 - (3-1) (2.1)의 모형식에서 ψ_1, \dots, ψ_k 을 $\tilde{\psi}_1, \dots, \tilde{\psi}_k$ 으로 대체하고, 최소제곱법을 사용하여 모형식을 추정, 잔차 e_1, \dots, e_n 을 구한다.
 - (3-2) 잔차 e_1, \dots, e_n 의 표본자기공분산 $\gamma_h^e = n^{-1} \sum_{t=1}^{n-h} e_t e_{t+h}$ 를 계산하고, 이를 $\hat{\gamma}_h$ 로 사용한다.
 - (3-3) 유의수준 α 를 선택하고, (2.7)에 주어진 임계값 c_α 를 계산한다.
- (S4) $\tilde{\psi}_1, \dots, \tilde{\psi}_k$ 중 $|D_{\tilde{\psi}_j}| \geq c_\alpha$ 를 만족하는 것만 선택하여, 변화점 ψ_1, \dots, ψ_k 에 대한 추정값 $\hat{\psi}_1, \dots, \hat{\psi}_k$ 으로 사용한다.
- (S5) (2.1)의 모형식에서 ψ_1, \dots, ψ_k 을 $\hat{\psi}_1, \dots, \hat{\psi}_k$ 으로 대체하고, 최소제곱법을 사용하여 모형식을 추정한다.

이동평균기간을 $m < (n - 1)/2$ 을 만족하는 자연수로 선택하면, (S2)에서 \mathcal{P} 에 대한 초기치로 $\mathcal{P}_0 = \{1, n\}$ 을 사용하기 때문에 적어도 한 번 이상 (2-1)~(2-3)의 과정을 반복하게 되며, (S3)의 최종 \mathcal{P} 는 1과 n 을 원소로 가질 뿐만 아니라 적어도 3개 이상의 원소를 가진다. 따라서 k 는 1 이상이며, (S3)를 수행하는데 문제가 없다.

3. 모의실험

본 모의실험에서는 변화점의 개수가 0, 1, 2개인 경우를 고려하였으며, Muggeo (2020)의 방법과 비교하였다.

먼저 본 논문에서 제안한 방법을 사용하기 위해서는 1차 차분한 시계열의 단순이동평균을 계산해야 하며, 이동평균기간으로 $m = \text{floor}(2 \log n \sqrt{2 \log n})$ 을 사용하였다. 여기서 $\text{floor}(x)$ 는 x 이하의 정수 중 최대값을 의미하며, n 은 주어진 시계열 자료의 개수를 나타낸다. 비록 이동평균기간 선택에 대한 자세한 연구는 진행하지 못 하였으나 앞서 2장에서 기술한 m 에 대한 조건을 만족하면서 모의실험에서 고려한 $100 \leq n \leq 1000$ 에 대해 m 의 값이 약 20에서 50 정도가 될 수 있도록 설정한 것이다.

Muggeo (2020)은 BIC를 이용하여 변화점의 개수를 선택하는 방법과 가설 검정을 순차적으로 이용하여 선택하는 방법 두 가지를 제안하였다. 이 중 BIC를 사용하는 방법부터 간단히 소개하면, 우선 변화점의 개수 k 를 고정하고 Muggeo (2003)이 제안한 방식대로 모형식을 추정한다. 즉, s -단계의 변화점 $\psi_1^{(s)}, \dots, \psi_k^{(s)}$ 이 주어졌을 때, 모형식 (2.1)에서 $\beta_1 t + \sum_{j=1}^k \beta_{j+1} (t - \psi_j)_+$ 대신 $\beta_1 t + \sum_{j=1}^k \beta_{j+1} (t - \psi_j^{(s)})_+ - \sum_{j=1}^k \delta_j I(t > \psi_j^{(s)})$ 를 사용하여 모형을 적합시키고, 이 때 구한 추정값 $\hat{\beta}_{j+1}$ 과 $\hat{\delta}_j$ 를 바탕으로 다음 단계의 변화점을 $\psi_j^{(s+1)} = \psi_j^{(s)} + \hat{\delta}_j / \hat{\beta}_{j+1}$ 처럼 계산한다. 이와 같은 작업을 변화점의 값이 수렴할 때까지 반복하여 변화점 ψ_1, \dots, ψ_k 와 $\beta_1, \dots, \beta_{k+1}$ 의 값을 추정한다. 마지막으로 추정된 모형식을 이용하여 잔차 r_1, \dots, r_n 을 구하고, 변화점의 개수가 k 개인 모형에 대한 $\text{BIC} = n \log(\sum_{t=1}^n r_t^2 / n) + (k + 2) \log n$ 를 계산한다. 이와 같은 방식으로 변화점의 개수가 0개인 경우부터 미리 지정한 최대값 K_{max} 인 경우까지 BIC를 계산하여 BIC의 값이 최소인 경우를 선택한다. 또 다른 방법은 $H_0 : k = 0$ vs. $H_1 : k = K_{max}$ 의 가설부터 시작하여 score test와 같은 가설 검정을 실시하고, 검정 결과에 따라 귀무가설에서 고려하는 변화점의 개수를 한 개씩 늘이거나 대립가설에서 고려하는 변화점의 개수를 한 개씩 줄이는 순차적인 방법으로 자세한 설명은 Muggeo (2020)을 참고하기 바란다. 본 실험에서는 Muggeo (2020)가 제안한 두 가지 모두 고려하였으며 segmented 패키지의 selgmented() 함수를 사용하여 결과를 얻었다. selgmented() 함수 사용시 변화점에 대한 최대 개수를 지정해야하는데, BIC의 경우는 10으로 지정하였고, score test의 경우에는 기본값 2를 선택하였다.

Table 1: Frequencies of $\hat{k} = 0$ when $k = 0$

| ϕ | $n = 100$ | | | $n = 300$ | | | $n = 500$ | | | $n = 1000$ | | |
|--------|-----------|------|------|-----------|------|------|-----------|------|------|------------|------|------|
| | MA | BIC | ST | MA | BIC | ST | MA | BIC | ST | MA | BIC | ST |
| -0.9 | 1000 | 1000 | 1000 | 1000 | 1000 | 1000 | 999 | 1000 | 1000 | 999 | 1000 | 1000 |
| -0.8 | 1000 | 999 | 1000 | 998 | 1000 | 1000 | 999 | 1000 | 1000 | 997 | 1000 | 1000 |
| -0.7 | 1000 | 1000 | 1000 | 998 | 1000 | 1000 | 997 | 1000 | 1000 | 991 | 1000 | 1000 |
| -0.6 | 1000 | 999 | 999 | 998 | 1000 | 1000 | 992 | 1000 | 1000 | 998 | 1000 | 1000 |
| -0.5 | 999 | 999 | 997 | 997 | 1000 | 1000 | 993 | 1000 | 999 | 992 | 1000 | 999 |
| -0.4 | 1000 | 1000 | 997 | 996 | 1000 | 996 | 992 | 1000 | 998 | 991 | 1000 | 997 |
| -0.3 | 1000 | 999 | 993 | 994 | 1000 | 991 | 992 | 1000 | 986 | 992 | 1000 | 994 |
| -0.2 | 1000 | 998 | 981 | 996 | 1000 | 979 | 996 | 1000 | 984 | 991 | 999 | 987 |
| -0.1 | 999 | 993 | 967 | 996 | 999 | 970 | 994 | 1000 | 967 | 995 | 1000 | 963 |
| 0.0 | 997 | 990 | 941 | 995 | 995 | 944 | 989 | 995 | 936 | 992 | 1000 | 932 |
| 0.1 | 995 | 971 | 906 | 997 | 992 | 892 | 991 | 999 | 911 | 989 | 996 | 911 |
| 0.2 | 996 | 945 | 877 | 989 | 974 | 855 | 990 | 981 | 842 | 994 | 987 | 845 |
| 0.3 | 999 | 903 | 825 | 992 | 938 | 794 | 992 | 960 | 806 | 993 | 970 | 794 |
| 0.4 | 995 | 806 | 718 | 990 | 854 | 705 | 991 | 914 | 702 | 986 | 936 | 737 |
| 0.5 | 996 | 633 | 623 | 990 | 769 | 640 | 991 | 804 | 643 | 987 | 848 | 660 |
| 0.6 | 991 | 416 | 528 | 992 | 525 | 519 | 989 | 627 | 531 | 989 | 697 | 557 |
| 0.7 | 984 | 194 | 416 | 981 | 249 | 412 | 982 | 295 | 423 | 979 | 411 | 420 |
| 0.8 | 939 | 49 | 269 | 964 | 62 | 292 | 964 | 70 | 281 | 966 | 85 | 286 |
| 0.9 | 873 | 7 | 153 | 880 | 1 | 138 | 857 | 8 | 151 | 880 | 11 | 148 |

Tables 1–8까지 실험 결과를 표로 정리하였으며, 각 표에서 MA는 본 논문에서 제안한 방법, BIC는 Mugge (2020)의 방법 중 BIC를 사용하는 방법, ST는 score test를 사용하여 변화점의 개수를 선택하는 방법을 각각 의미한다.

3.1. 변화점이 1개 이하인 경우

변화점이 존재하지 않거나 한 개 존재하는 경우에 대한 모의실험을 수행하기 위하여 다음과 같은 모형을 고려하였다.

$$y_t = \alpha + \beta_1 t + \beta_2 (t - \psi_1)_+ + \epsilon_t, \quad t = 1, 2, \dots, n$$

모수의 값으로 $\alpha = 0$, $\beta_1 = 1$, $\beta_2 = 0, 0.2, 0.5, 1$ 의 값을 사용하였으며, $\beta_2 = 0$ 인 경우는 변화가 없는 경우, $\beta_2 > 0$ 인 경우는 변화점의 개수가 한 개인 경우를 나타낸다. 자료의 개수는 $n = 100, 300, 500, 1000$ 으로 설정하였으며, 변화점의 위치는 $\psi_1 = 0.5n$ 으로 하였다.

본 논문에서 고려한 오차항은 (E1), (E2)를 만족하는 정상시계열로, 서로 독립인 경우뿐만 아니라 자기상관이 존재하는 경우까지 포함한다. 그러므로 모의실험에서는 1차 자기회귀모형을 이용하여 오차항을 생성하였으며, 구체적인 모형식은 다음과 같다:

$$\begin{aligned} \epsilon_t &= \sqrt{1 - \phi^2} a_t, \\ a_t &= \phi a_{t-1} + b_t, \quad b_t \sim_{iid} N(0, 1). \end{aligned}$$

1차 자기상관계수인 ϕ 의 값으로 $0, \pm 0.1, \dots, \pm 0.9$ 을 고려하였고, $\{a_t\}$ 생성 시 `arma.sim()` 함수를 사용하였다. 여기서 $\phi = 0$ 이면, 오차항들은 서로 독립이 된다. 그리고 $\phi \neq 0$ 이면, 오차항들 사이에 자기상관이 존재하고,

Table 2: Frequency distributions of \hat{k} when $k = 1$ and $n = 100$

| β_2 | ϕ | $\hat{k} = 0$ | | | $\hat{k} = 1$ | | | $\hat{k} = 2$ | | | $\hat{k} > 2$ | | |
|-----------|--------|---------------|-----|-----|---------------|------|------|---------------|-----|-----|---------------|-----|----|
| | | MA | BIC | ST | MA | BIC | ST | MA | BIC | ST | MA | BIC | ST |
| 0.2 | -0.9 | 684 | 0 | 0 | 316 | 998 | 1000 | 0 | 2 | 0 | 0 | 0 | 0 |
| | -0.5 | 613 | 0 | 0 | 387 | 1000 | 999 | 0 | 0 | 1 | 0 | 0 | 0 |
| | 0.0 | 577 | 0 | 0 | 423 | 992 | 966 | 0 | 8 | 34 | 0 | 0 | 0 |
| | 0.3 | 565 | 0 | 0 | 435 | 883 | 867 | 0 | 86 | 133 | 0 | 31 | 0 |
| | 0.4 | 525 | 0 | 0 | 475 | 765 | 813 | 0 | 140 | 187 | 0 | 95 | 0 |
| | 0.5 | 520 | 0 | 0 | 480 | 619 | 744 | 0 | 200 | 256 | 0 | 181 | 0 |
| | 0.6 | 525 | 0 | 0 | 475 | 424 | 692 | 0 | 202 | 308 | 0 | 374 | 0 |
| | 0.7 | 459 | 0 | 0 | 540 | 187 | 586 | 1 | 177 | 414 | 0 | 636 | 0 |
| | 0.8 | 397 | 0 | 0 | 600 | 66 | 511 | 3 | 78 | 489 | 0 | 856 | 0 |
| 0.9 | 262 | 0 | 0 | 717 | 24 | 423 | 21 | 50 | 577 | 0 | 926 | 0 | |
| 0.5 | -0.9 | 14 | 0 | 0 | 986 | 1000 | 1000 | 0 | 0 | 0 | 0 | 0 | 0 |
| | -0.5 | 0 | 0 | 0 | 1000 | 1000 | 999 | 0 | 0 | 1 | 0 | 0 | 0 |
| | 0.0 | 0 | 0 | 0 | 1000 | 980 | 941 | 0 | 20 | 59 | 0 | 0 | 0 |
| | 0.3 | 0 | 0 | 0 | 1000 | 868 | 866 | 0 | 93 | 134 | 0 | 39 | 0 |
| | 0.4 | 1 | 0 | 0 | 998 | 776 | 812 | 1 | 137 | 188 | 0 | 87 | 0 |
| | 0.5 | 2 | 0 | 0 | 998 | 600 | 735 | 0 | 190 | 265 | 0 | 210 | 0 |
| | 0.6 | 2 | 0 | 0 | 995 | 396 | 681 | 3 | 203 | 319 | 0 | 401 | 0 |
| | 0.7 | 0 | 0 | 0 | 996 | 180 | 597 | 4 | 160 | 403 | 0 | 660 | 0 |
| | 0.8 | 2 | 0 | 0 | 989 | 55 | 508 | 9 | 74 | 492 | 0 | 871 | 0 |
| 0.9 | 5 | 0 | 0 | 987 | 16 | 382 | 8 | 31 | 618 | 0 | 953 | 0 | |
| 1.0 | -0.9 | 0 | 0 | 0 | 1000 | 1000 | 1000 | 0 | 0 | 0 | 0 | 0 | 0 |
| | -0.5 | 0 | 0 | 0 | 1000 | 1000 | 997 | 0 | 0 | 3 | 0 | 0 | 0 |
| | 0.0 | 0 | 0 | 0 | 1000 | 985 | 947 | 0 | 13 | 53 | 0 | 2 | 0 |
| | 0.3 | 0 | 0 | 0 | 1000 | 901 | 869 | 0 | 79 | 131 | 0 | 20 | 0 |
| | 0.4 | 0 | 0 | 0 | 1000 | 786 | 821 | 0 | 130 | 179 | 0 | 84 | 0 |
| | 0.5 | 0 | 0 | 0 | 1000 | 613 | 737 | 0 | 174 | 263 | 0 | 213 | 0 |
| | 0.6 | 0 | 0 | 0 | 1000 | 423 | 697 | 0 | 196 | 303 | 0 | 381 | 0 |
| | 0.7 | 0 | 0 | 0 | 1000 | 172 | 605 | 0 | 125 | 395 | 0 | 703 | 0 |
| | 0.8 | 0 | 0 | 0 | 1000 | 59 | 508 | 0 | 83 | 492 | 0 | 858 | 0 |
| 0.9 | 0 | 0 | 0 | 999 | 11 | 390 | 1 | 37 | 610 | 0 | 952 | 0 | |

$|\phi|$ 이 커질수록 자기상관이 강하게 존재하게 된다. a_t 의 분산은 $(1 - \phi^2)^{-1}$ 으로 $|\phi|$ 이 클수록, 즉 자기상관이 강하게 존재할수록 분산이 커진다. 그러므로 ϵ_t 대신 a_t 를 오차항으로 사용하면, (2.7)에서 보듯이 변화점 여부를 판단할 때 사용하는 임계값 c_α 도 ϕ 의 영향을 받게 된다. 물론 (2.7)을 자세히 보면 오차항의 분산뿐만 아니라 m 차 자기공분산과 $2m$ 차 자기공분산도 임계값 c_α 에 영향을 준다. 그러나 본 논문에서 제안한 m 은 그 값이 작지 않고 $\sqrt{\log n}$ 보다 큰 값이므로, m 차, $2m$ 차 자기공분산 값은 0에 가까우며 영향이 미비하다. 따라서 본 모의실험에서는 위에 기술한 것처럼 a_t 대신 분산을 1로 조정된 ϵ_t 를 오차항으로 사용하였다.

각각의 경우에 대해 1000번씩 반복 실험을 하였으며, 실험 결과는 Tables 1-5에 정리하였다. 먼저 Table 1은 1000번의 반복 실험 중 변화점이 존재하지 않을 때 변화점이 존재하지 않는다고 옳게 판단한 경우, 즉 변화점의 개수에 대한 추정값 \hat{k} 이 0인 사건의 도수를 나타낸다. Table 1의 결과를 보면, $\phi \leq 0$ 인 경우에는 세 가지 방법에 대한 도수가 대부분 950 이상으로 변화점 여부를 잘 판단한다. 그러나 $\phi > 0$ 인 경우에는 세 가지 방법들의 도수에 차이가 발생하며, 이 차이는 ϕ 의 값이 1에 가까워질수록 커진다. 본 논문에서 제안한 방법은 $\phi \leq 0.8$ 일 때 900 이상의 도수를 가지고, $\phi = 0.9$ 일 때 850 이상의 높은 도수를 가진다. 이에 비해 Muggeo

Table 3: Frequency distributions of \hat{k} when $k = 1$ and $\beta_2 = 0.2$

| n | ϕ | $\hat{k} = 0$ | | | $\hat{k} = 1$ | | | $\hat{k} = 2$ | | | $\hat{k} > 2$ | | |
|------|--------|---------------|-----|----|---------------|------|------|---------------|-----|-----|---------------|-----|----|
| | | MA | BIC | ST | MA | BIC | ST | MA | BIC | ST | MA | BIC | ST |
| 300 | -0.9 | 350 | 0 | 0 | 650 | 1000 | 1000 | 0 | 0 | 0 | 0 | 0 | 0 |
| | -0.5 | 209 | 0 | 0 | 788 | 1000 | 999 | 3 | 0 | 1 | 0 | 0 | 0 |
| | 0.0 | 180 | 0 | 0 | 815 | 997 | 947 | 5 | 3 | 53 | 0 | 0 | 0 |
| | 0.3 | 164 | 0 | 0 | 827 | 926 | 853 | 9 | 57 | 147 | 0 | 17 | 0 |
| | 0.4 | 201 | 0 | 0 | 796 | 887 | 828 | 3 | 75 | 172 | 0 | 38 | 0 |
| | 0.5 | 196 | 0 | 0 | 801 | 721 | 747 | 3 | 145 | 253 | 0 | 134 | 0 |
| | 0.6 | 204 | 0 | 0 | 793 | 494 | 660 | 3 | 199 | 340 | 0 | 307 | 0 |
| | 0.7 | 188 | 1 | 0 | 802 | 243 | 591 | 9 | 143 | 409 | 1 | 613 | 0 |
| | 0.8 | 222 | 2 | 0 | 755 | 40 | 483 | 23 | 49 | 517 | 0 | 909 | 0 |
| | 0.9 | 155 | 0 | 0 | 738 | 3 | 334 | 86 | 12 | 666 | 21 | 985 | 0 |
| 500 | -0.9 | 243 | 0 | 0 | 755 | 1000 | 1000 | 2 | 0 | 0 | 0 | 0 | 0 |
| | -0.5 | 54 | 0 | 0 | 946 | 1000 | 998 | 0 | 0 | 2 | 0 | 0 | 0 |
| | 0.0 | 57 | 0 | 0 | 935 | 996 | 951 | 8 | 4 | 49 | 0 | 0 | 0 |
| | 0.3 | 62 | 0 | 0 | 935 | 960 | 866 | 3 | 35 | 134 | 0 | 5 | 0 |
| | 0.4 | 80 | 0 | 0 | 916 | 892 | 811 | 4 | 79 | 189 | 0 | 29 | 0 |
| | 0.5 | 88 | 0 | 0 | 905 | 778 | 758 | 6 | 134 | 242 | 1 | 88 | 0 |
| | 0.6 | 93 | 0 | 0 | 897 | 559 | 649 | 9 | 186 | 351 | 1 | 255 | 0 |
| | 0.7 | 139 | 1 | 0 | 853 | 280 | 563 | 8 | 158 | 437 | 0 | 561 | 0 |
| | 0.8 | 110 | 2 | 0 | 862 | 66 | 478 | 25 | 66 | 522 | 3 | 866 | 0 |
| | 0.9 | 103 | 1 | 0 | 785 | 2 | 339 | 91 | 6 | 661 | 21 | 991 | 0 |
| 1000 | -0.9 | 49 | 0 | 0 | 949 | 1000 | 1000 | 2 | 0 | 0 | 0 | 0 | 0 |
| | -0.5 | 2 | 0 | 0 | 992 | 1000 | 1000 | 6 | 0 | 0 | 0 | 0 | 0 |
| | 0.0 | 0 | 0 | 0 | 995 | 999 | 955 | 5 | 1 | 45 | 0 | 0 | 0 |
| | 0.3 | 3 | 0 | 0 | 990 | 966 | 840 | 7 | 30 | 160 | 0 | 4 | 0 |
| | 0.4 | 6 | 0 | 0 | 986 | 924 | 816 | 8 | 57 | 184 | 0 | 19 | 0 |
| | 0.5 | 8 | 0 | 0 | 987 | 830 | 754 | 5 | 115 | 246 | 0 | 55 | 0 |
| | 0.6 | 13 | 1 | 0 | 979 | 651 | 666 | 8 | 154 | 334 | 0 | 194 | 0 |
| | 0.7 | 22 | 2 | 0 | 964 | 324 | 591 | 13 | 158 | 409 | 1 | 516 | 0 |
| | 0.8 | 30 | 2 | 0 | 944 | 82 | 479 | 23 | 81 | 521 | 3 | 835 | 0 |
| | 0.9 | 32 | 0 | 0 | 876 | 3 | 357 | 76 | 13 | 643 | 16 | 984 | 0 |

가 제안한 방법은 BIC와 score test를 이용하는 두 가지 방법 모두 ϕ 의 값이 증가할수록 $\hat{k} = 0$ 인 도수가 줄어들며, 특히 BIC 경우에 도수가 줄어드는 속도가 크다. 비록 자료의 개수가 증가함에 따라 변화점의 개수를 올바르게 선택한 도수는 증가하지만, $\phi \geq 0.8$ 인 경우 도수가 100보다 작으며 $\phi = 0.6, 0.7$ 일 때도 도수가 높지 않다. $\phi = 0.9$ 일 때는 $n = 1000$ 이라도 도수가 11밖에 되지 않으며, 변화점이 없어도 대부분 변화점이 있다고 판단한다.

Table 2는 변화점의 개수가 1개인 경우 기울기의 변화량 β_2 와 ϕ 의 값에 따른 \hat{k} 의 도수분포를 구한 것이다. 전체 실험 결과 중 특징이 유사하게 나타나는 결과들은 생략하였으며, $n = 100$ 인 결과만을 작성하였다. 먼저 Muggeo의 방법에 대한 결과를 보면, β_2 의 값에 따른 차이는 별로 없다. BIC와 score test를 이용한 두 가지 방법 모두 1000번의 반복 실험 중 단 한 건도 변화점이 없다고 잘못 판단한 경우는 없으며, 모두 변화점이 한 개 이상 있다고 판단하였다. 그리고 $\phi \leq 0.2$ 인 경우에는 $\hat{k} = 1$ 로 변화점의 개수를 올바르게 선택한 비율이 약 90% 이상으로 꽤 높다. $\hat{k} = 1$ 인 비율은 ϕ 의 값이 1에 가까워질수록 줄어들며, BIC를 이용한 방법은 $\phi \geq 0.6$ 일 때, score test를 이용한 방법은 $\phi = 0.9$ 일 때 그 비율이 50% 미만으로 떨어진다. Score test를 이용하여 변화점

Table 4: Five-number summaries of $\hat{\psi}_1$ when $k = 1$ and $\psi_1 = 0.5n$

| | | $n = 100$ | | | | | | | | | | | | | | | |
|-----------|--------|------------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|--|
| β_2 | ϕ | MA | | | | | BIC | | | | | ST | | | | | |
| | | min | Q1 | Q2 | Q3 | max | min | Q1 | Q2 | Q3 | max | min | Q1 | Q2 | Q3 | max | |
| 0.2 | -0.9 | 40.0 | 48.0 | 50.0 | 52.0 | 61.0 | 47.6 | 49.9 | 50.0 | 50.4 | 51.5 | 47.6 | 49.7 | 50.0 | 50.4 | 53.0 | |
| | -0.5 | 34.0 | 48.0 | 50.0 | 53.0 | 64.0 | 46.0 | 49.2 | 50.1 | 50.9 | 53.7 | 45.6 | 49.0 | 50.0 | 51.0 | 53.7 | |
| | 0.0 | 32.0 | 47.0 | 50.0 | 53.0 | 66.0 | 42.0 | 49.0 | 50.2 | 51.2 | 58.0 | 42.0 | 48.8 | 50.0 | 51.5 | 58.0 | |
| | 0.3 | 29.0 | 47.0 | 50.0 | 53.0 | 67.0 | 39.5 | 48.2 | 50.0 | 51.7 | 61.6 | 35.7 | 48.0 | 50.0 | 52.0 | 61.6 | |
| | 0.6 | 32.0 | 45.0 | 49.0 | 54.0 | 73.0 | 35.0 | 47.1 | 50.0 | 52.1 | 67.3 | 34.9 | 47.0 | 50.0 | 52.6 | 71.8 | |
| | 0.9 | 28.0 | 44.0 | 50.0 | 55.0 | 73.0 | 30.7 | 45.2 | 50.3 | 54.6 | 60.9 | 21.6 | 45.1 | 49.8 | 54.5 | 78.4 | |
| 0.5 | -0.9 | 40.0 | 49.0 | 50.0 | 51.0 | 61.0 | 49.3 | 50.0 | 50.0 | 50.2 | 50.8 | 49.3 | 49.9 | 50.0 | 50.2 | 50.8 | |
| | -0.5 | 39.0 | 49.0 | 50.0 | 52.0 | 61.0 | 48.6 | 49.7 | 50.0 | 50.3 | 51.7 | 48.5 | 49.7 | 50.0 | 50.3 | 51.7 | |
| | 0.0 | 39.0 | 48.0 | 50.0 | 52.0 | 62.0 | 47.5 | 49.5 | 50.0 | 50.5 | 52.6 | 47.5 | 49.5 | 50.0 | 50.5 | 52.6 | |
| | 0.3 | 34.0 | 48.0 | 50.0 | 52.0 | 67.0 | 46.1 | 49.4 | 50.1 | 50.8 | 54.5 | 45.7 | 49.4 | 50.0 | 50.8 | 54.6 | |
| | 0.6 | 33.0 | 48.0 | 50.0 | 52.0 | 66.0 | 44.4 | 49.1 | 50.0 | 50.9 | 55.2 | 44.4 | 49.0 | 50.1 | 51.1 | 55.2 | |
| | 0.9 | 32.0 | 48.0 | 50.0 | 51.0 | 68.0 | 46.7 | 48.4 | 50.1 | 51.5 | 53.7 | 41.1 | 48.0 | 49.6 | 51.4 | 58.4 | |
| 1.0 | -0.9 | 46.0 | 49.0 | 50.0 | 51.0 | 55.0 | 49.6 | 50.0 | 50.0 | 50.1 | 50.3 | 49.6 | 49.9 | 50.0 | 50.1 | 50.3 | |
| | -0.5 | 44.0 | 49.0 | 50.0 | 51.0 | 57.0 | 49.2 | 49.9 | 50.0 | 50.2 | 50.9 | 49.2 | 49.8 | 50.0 | 50.2 | 50.9 | |
| | 0.0 | 44.0 | 49.0 | 50.0 | 51.0 | 58.0 | 48.7 | 49.8 | 50.0 | 50.3 | 51.4 | 48.7 | 49.8 | 50.0 | 50.3 | 51.4 | |
| | 0.3 | 41.0 | 49.0 | 50.0 | 51.0 | 58.0 | 48.3 | 49.7 | 50.0 | 50.3 | 51.6 | 48.3 | 49.6 | 50.0 | 50.3 | 51.6 | |
| | 0.6 | 40.0 | 49.0 | 50.0 | 51.0 | 59.0 | 47.3 | 49.5 | 50.0 | 50.5 | 52.3 | 47.5 | 49.5 | 50.0 | 50.5 | 52.3 | |
| | 0.9 | 42.0 | 50.0 | 50.0 | 50.0 | 59.0 | 48.3 | 49.4 | 50.3 | 50.5 | 51.4 | 45.8 | 49.3 | 50.0 | 50.8 | 55.5 | |
| | | $n = 1000$ | | | | | | | | | | | | | | | |
| 0.2 | -0.9 | 475.0 | 497.0 | 500.0 | 504.0 | 528.0 | 499.6 | 499.9 | 500.0 | 500.1 | 500.4 | 499.6 | 499.9 | 500.0 | 500.1 | 500.4 | |
| | -0.5 | 474.0 | 497.0 | 500.0 | 504.0 | 525.0 | 498.9 | 499.7 | 500.0 | 500.2 | 501.1 | 498.9 | 499.7 | 500.0 | 500.2 | 501.1 | |
| | 0.0 | 473.0 | 496.0 | 500.0 | 504.0 | 528.0 | 497.8 | 499.5 | 500.0 | 500.4 | 501.9 | 497.8 | 499.5 | 500.0 | 500.4 | 501.9 | |
| | 0.3 | 469.0 | 496.0 | 500.0 | 505.0 | 524.0 | 496.9 | 499.4 | 500.0 | 500.5 | 502.9 | 496.9 | 499.4 | 500.0 | 500.5 | 502.9 | |
| | 0.6 | 469.0 | 495.0 | 500.0 | 504.5 | 530.0 | 496.1 | 499.2 | 500.1 | 500.8 | 503.9 | 496.1 | 499.3 | 500.1 | 500.9 | 503.9 | |
| | 0.9 | 444.0 | 495.0 | 500.0 | 505.0 | 927.0 | 498.7 | 500.6 | 502.5 | 502.6 | 502.6 | 492.9 | 498.0 | 499.7 | 501.7 | 508.8 | |
| 0.5 | -0.9 | 487.0 | 499.0 | 500.0 | 501.0 | 511.0 | 499.8 | 500.0 | 500.0 | 500.0 | 500.2 | 499.8 | 500.0 | 500.0 | 500.0 | 500.2 | |
| | -0.5 | 487.0 | 499.0 | 500.0 | 502.0 | 513.0 | 499.6 | 499.9 | 500.0 | 500.1 | 500.4 | 499.6 | 499.9 | 500.0 | 500.1 | 500.4 | |
| | 0.0 | 488.0 | 498.0 | 500.0 | 502.0 | 513.0 | 499.3 | 499.8 | 500.0 | 500.2 | 500.9 | 499.3 | 499.8 | 500.0 | 500.2 | 500.9 | |
| | 0.3 | 487.0 | 498.0 | 500.0 | 502.0 | 511.0 | 498.9 | 499.8 | 500.0 | 500.2 | 501.1 | 499.1 | 499.8 | 500.0 | 500.2 | 501.1 | |
| | 0.6 | 487.0 | 498.0 | 500.0 | 502.0 | 514.0 | 498.7 | 499.7 | 500.0 | 500.3 | 501.2 | 498.7 | 499.7 | 500.0 | 500.3 | 501.4 | |
| | 0.9 | 486.0 | 499.0 | 500.0 | 501.0 | 514.0 | 499.0 | 499.0 | 499.0 | 499.0 | 499.0 | 497.1 | 499.2 | 500.0 | 500.7 | 502.8 | |
| 1.0 | -0.9 | 492.0 | 499.0 | 500.0 | 501.0 | 505.0 | 499.9 | 500.0 | 500.0 | 500.0 | 500.1 | 499.9 | 500.0 | 500.0 | 500.0 | 500.1 | |
| | -0.5 | 493.0 | 499.0 | 500.0 | 501.0 | 506.0 | 499.8 | 499.9 | 500.0 | 500.0 | 500.2 | 499.8 | 499.9 | 500.0 | 500.0 | 500.2 | |
| | 0.0 | 491.0 | 499.0 | 500.0 | 501.0 | 509.0 | 499.6 | 499.9 | 500.0 | 500.1 | 500.4 | 499.6 | 499.9 | 500.0 | 500.1 | 500.4 | |
| | 0.3 | 494.0 | 499.0 | 500.0 | 501.0 | 506.0 | 499.5 | 499.9 | 500.0 | 500.1 | 500.6 | 499.5 | 499.9 | 500.0 | 500.1 | 500.6 | |
| | 0.6 | 492.0 | 499.0 | 500.0 | 501.0 | 508.0 | 499.2 | 499.8 | 500.0 | 500.2 | 500.7 | 499.2 | 499.8 | 500.0 | 500.2 | 500.7 | |
| | 0.9 | 493.0 | 500.0 | 500.0 | 500.0 | 506.0 | 499.2 | 499.5 | 499.7 | 500.0 | 500.3 | 498.6 | 499.6 | 500.0 | 500.4 | 501.6 | |

의 개수를 선택하는 방법에서는 최대 변화점의 개수를 2개로 설정하였기 때문에 Table 2에서 보듯이 $\hat{k} > 2$ 인 경우가 존재하지 않는다. 이에 비해 Muggeo의 방법 중 BIC를 기초로 변화점의 개수를 선택하는 방법에서는 최대 변화점의 개수를 10개로 설정하였으며, 이로 인해 $\hat{k} > 2$ 인 경우도 나타날 수 있다. 그러나 $\phi \leq 0$ 일 때는 $\hat{k} > 2$ 인 경우가 거의 나타나지 않으며, $\phi > 0$ 일 때 주로 나타난다. 게다가 ϕ 의 값이 1에 가까워질수록 그 빈도가 급격하게 늘어난다. 이와 같은 현상은 앞선 변화점이 존재하지 않는 경우에도 발생했으며, 이로부터 BIC를 이용한 Muggeo의 방법은 오차항들 사이에 강한 양의 자기상관이 존재할 때 변화점이 참값보다 더 많이 존재한다고 잘못 판단하는 경향이 있음을 알 수 있다.

본 논문에서 제안한 방법은 β_2 의 값이 증가함에 따라 $\hat{k} = 1$ 인 비율이 증가하고, $\beta_2 = 0.2$ 를 제외한 나머지 경우에는 그 비율이 1에 가깝게 나타난다. $\beta_2 = 0.2$ 는 자료의 수 n 이 작은 경우 (2.7)에 주어진 임계값과 큰 차이가 나지 않기 때문에 $\hat{k} = 0$ 인 비율이 꽤 높게 나온다. 그러나 Table 3에서 보듯이 β_2 의 값이 작더라도 자료의

Table 5: Five-number summaries of $\hat{\psi}_1^*$ when $k = 1$ and $\psi_1 = 0.5n$

| | | $n = 100$ | | | | | | | | | | | | | | | |
|-----------|--------|------------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|--|
| β_2 | ϕ | MA | | | | | | BIC | | | | | | ST | | | |
| | | min | Q1 | Q2 | Q3 | max | min | Q1 | Q2 | Q3 | max | min | Q1 | Q2 | Q3 | max | |
| 0.2 | 0.6 | 32.0 | 45.0 | 49.0 | 54.0 | 73.0 | 28.4 | 46.4 | 50.0 | 53.0 | 75.2 | 28.0 | 46.6 | 50.0 | 53.1 | 77.9 | |
| | 0.9 | 28.0 | 44.0 | 50.0 | 56.0 | 73.0 | 21.0 | 47.1 | 50.0 | 53.0 | 75.0 | 18.0 | 44.7 | 49.8 | 54.6 | 84.3 | |
| 0.5 | 0.6 | 33.0 | 48.0 | 50.0 | 52.0 | 66.0 | 40.0 | 48.6 | 50.0 | 51.4 | 60.0 | 40.0 | 48.7 | 50.0 | 51.3 | 59.7 | |
| | 0.9 | 32.0 | 48.0 | 50.0 | 51.0 | 68.0 | 39.1 | 48.7 | 50.0 | 51.3 | 59.7 | 36.4 | 48.1 | 50.0 | 51.8 | 61.3 | |
| 1.0 | 0.6 | 40.0 | 49.0 | 50.0 | 51.0 | 59.0 | 43.0 | 49.3 | 50.0 | 50.7 | 55.0 | 43.0 | 49.4 | 50.0 | 50.6 | 54.6 | |
| | 0.9 | 42.0 | 50.0 | 50.0 | 50.0 | 59.0 | 44.6 | 49.5 | 50.0 | 50.5 | 57.7 | 42.2 | 49.2 | 49.9 | 50.7 | 57.7 | |
| | | $n = 1000$ | | | | | | | | | | | | | | | |
| 0.2 | 0.6 | 469.0 | 495.0 | 500.0 | 504.5 | 530.0 | 479.0 | 499.0 | 500.1 | 501.1 | 513.1 | 487.0 | 499.0 | 500.0 | 501.2 | 518.4 | |
| | 0.9 | 444.0 | 495.0 | 500.0 | 505.0 | 927.0 | 469.6 | 495.6 | 500.0 | 504.5 | 528.0 | 475.6 | 497.2 | 499.8 | 502.9 | 528.5 | |
| 0.5 | 0.6 | 487.0 | 498.0 | 500.0 | 502.0 | 514.0 | 492.0 | 499.6 | 500.0 | 500.4 | 507.5 | 492.6 | 499.6 | 500.0 | 500.5 | 507.6 | |
| | 0.9 | 486.0 | 499.0 | 500.0 | 501.0 | 514.0 | 490.0 | 498.2 | 500.0 | 501.6 | 508.7 | 486.6 | 498.9 | 500.0 | 501.2 | 509.8 | |
| 1.0 | 0.6 | 492.0 | 499.0 | 500.0 | 501.0 | 508.0 | 496.4 | 499.8 | 500.0 | 500.2 | 502.2 | 495.7 | 499.8 | 500.0 | 500.2 | 504.5 | |
| | 0.9 | 493.0 | 500.0 | 500.0 | 500.0 | 506.0 | 494.1 | 499.2 | 500.0 | 500.9 | 508.8 | 494.1 | 499.5 | 500.0 | 500.6 | 506.4 | |

개수 n 이 증가하면 $\hat{k} = 0$ 의 비율은 줄어들고, $\hat{k} = 1$ 의 비율은 늘어나게 된다. 변화점의 개수를 1개로 옳게 추정할 비율은 $n = 100$ 일 때 40~50%에 불과하였으나 $n = 300$ 일 때는 약 80%, $n = 500$ 일 때는 약 90%, $n = 1000$ 일 때는 대부분 95% 이상으로 증가한 것을 확인할 수 있다. Table 2에서는 $\hat{k} > 1$ 인 경우가 거의 나타나지 않는 것에 비해 Table 3에서는 그 도수가 크지는 않으나 $\hat{k} = 2$ 인 경우와 $\hat{k} > 2$ 인 경우도 나타나는 것을 확인할 수 있다. 이는 2.4에 소개된 적합 과정을 살펴보면 그 이유를 알 수 있다. $n = 100$ 인 경우 사용한 $m = 27$ 의 값이 상대적으로 크기 때문에 처음에 찾은 $|D_s|$ 의 최대점이 참 변화점인 $\psi_1 = 50$ 근처인 $45 < s < 55$ 의 범위에 존재하면 변화점을 탐색하는 (S2) 단계를 멈추게 된다. 또한 $s \leq 45$ 또는 $s \geq 55$ 의 구간에서 $|D_s|$ 의 최대점이 존재하더라도 (S2)를 반복하는 횟수는 그리 많지 않다. 게다가 β_2 의 값이 큰 경우에는 Table 4에서 보듯이 $|D_s|$ 의 최대점이 참값 $\psi_1 = 50$ 근처에 대부분 존재한다. 그러므로 $n = 100$ 인 경우에는 $\hat{k} > 1$ 인 비율이 0에 가까운 것이다. 그러나 $n = 300, 500, 1000$ 인 경우에는 모의실험에서 사용한 이동평균기간이 각각 $m = 38, 43, 51$ 으로 그 값이 크지 않고, (S2) 단계를 여러번 반복하게 된다. 따라서 Table 3처럼 $n \geq 300$ 일 때는 $\hat{k} \geq 2$ 인 경우도 나타나게 된다. 하지만 여기서 눈여겨볼 것은 그 비율이 Muggeo의 방법처럼 높지 않으며, 비율이 채 10%를 넘지 않는다는 것이다.

Table 4와 Table 5는 변화점의 위치를 추정할 결과를 정리한 것이다. Table 4는 변화점의 개수를 1개로 옳게 추정할 경우에 구한 $\hat{\psi}_1$ 값들에 대한 요약통계량이다. 변화점의 참값은 $\psi_1 = 0.5n$ 으로 $n = 100, 1000$ 일 때 각각 50, 500이며, 대부분의 경우 중앙값이 이와 일치한다. 세 가지 방법 모두 β_2 의 값이 증가할수록 사분위수범위와 범위가 줄어들며, 본 논문에서 제안한 방법보다 Muggeo가 제안한 방법들의 사분위수범위와 범위가 더 작다. 비록 Muggeo의 방법들이 본 논문에서 제안한 방법보다 변화점의 위치를 잘 추정하지만, 우리의 방법도 $\beta_2 \geq 0.5$ 인 경우에는 사분위수 범위가 0~4로 아주 크지 않은 편이다. Table 2와 Table 3의 결과에서 보듯이 오차항들 사이에 비교적 강한 양의 자기상관이 존재할 때 Muggeo의 방법들은 변화점의 개수를 크게 추정한다. 그러므로 Table 5에서는 변화점의 개수가 1 이상인 경우에 구한 변화점에 대한 추정값들 중에서 참값 ψ_1 에 가장 가까운 값을 찾아 추정값 $\hat{\psi}_1^*$ 로 사용하였으며, 이들에 대한 요약통계량을 정리하였다. 그 결과 본 논문에서 제안한 방법은 $\hat{k} > 1$ 인 도수가 크지 않으므로 Table 4와 Table 5의 결과에 차이가 거의 없다. 이에 비해 Muggeo의 방법들은 $\hat{k} > 1$ 인 도수가 크기 때문에 Table 5의 사분위수범위와 범위가 Table 4보다 증가하였으며, 본 논문의 방법과 그 차이가 줄어들었다. 그러나 여전히 Muggeo의 방법이 본 논문에서 제안한 방법보다 변화점의 위치를 좀 더 정확하게 추정한다.

Table 6: Frequency distributions of \hat{k} when $k = 2$ and $\psi_1 = 0.3n$

| ψ_2 | n | ϕ | $\hat{k} = 0$ | | | $\hat{k} = 1$ | | | $\hat{k} = 2$ | | | $\hat{k} > 2$ | | |
|----------|-----|--------|---------------|-----|----|---------------|-----|----|---------------|------|------|---------------|-----|----|
| | | | MA | BIC | ST | MA | BIC | ST | MA | BIC | ST | MA | BIC | ST |
| 0.5n | 100 | -0.9 | 57 | 0 | 0 | 810 | 0 | 0 | 133 | 1000 | 1000 | 0 | 0 | 0 |
| | | -0.5 | 33 | 0 | 0 | 867 | 0 | 1 | 100 | 1000 | 999 | 0 | 0 | 0 |
| | | 0.0 | 49 | 0 | 0 | 869 | 0 | 1 | 82 | 974 | 999 | 0 | 26 | 0 |
| | | 0.3 | 40 | 0 | 0 | 877 | 2 | 12 | 83 | 811 | 988 | 0 | 187 | 0 |
| | | 0.4 | 34 | 0 | 0 | 863 | 2 | 22 | 103 | 678 | 978 | 0 | 320 | 0 |
| | | 0.5 | 32 | 0 | 0 | 846 | 1 | 36 | 122 | 558 | 964 | 0 | 441 | 0 |
| | | 0.6 | 22 | 0 | 0 | 865 | 0 | 42 | 113 | 340 | 958 | 0 | 660 | 0 |
| | | 0.7 | 38 | 0 | 0 | 832 | 0 | 83 | 130 | 163 | 917 | 0 | 837 | 0 |
| | | 0.8 | 29 | 0 | 0 | 802 | 0 | 91 | 169 | 92 | 909 | 0 | 908 | 0 |
| 0.9 | 36 | 0 | 0 | 733 | 0 | 80 | 231 | 24 | 920 | 0 | 976 | 0 | | |
| 0.5n | 300 | -0.9 | 0 | 0 | 0 | 0 | 0 | 0 | 1000 | 1000 | 1000 | 0 | 0 | 0 |
| | | -0.5 | 0 | 0 | 0 | 0 | 0 | 0 | 1000 | 1000 | 1000 | 0 | 0 | 0 |
| | | 0.0 | 0 | 0 | 0 | 0 | 0 | 0 | 998 | 998 | 1000 | 2 | 2 | 0 |
| | | 0.3 | 0 | 0 | 0 | 0 | 0 | 0 | 998 | 928 | 1000 | 2 | 72 | 0 |
| | | 0.4 | 0 | 0 | 0 | 0 | 0 | 0 | 994 | 862 | 1000 | 6 | 138 | 0 |
| | | 0.5 | 0 | 0 | 0 | 0 | 0 | 0 | 996 | 683 | 1000 | 4 | 317 | 0 |
| | | 0.6 | 0 | 0 | 0 | 1 | 0 | 0 | 996 | 510 | 1000 | 3 | 490 | 0 |
| | | 0.7 | 0 | 1 | 0 | 0 | 0 | 0 | 997 | 233 | 1000 | 3 | 766 | 0 |
| | | 0.8 | 0 | 1 | 0 | 0 | 0 | 0 | 991 | 53 | 1000 | 9 | 946 | 0 |
| 0.9 | 0 | 2 | 0 | 1 | 0 | 0 | 948 | 8 | 1000 | 51 | 990 | 0 | | |
| 0.7n | 100 | -0.9 | 11 | 0 | 0 | 27 | 0 | 0 | 962 | 1000 | 1000 | 0 | 0 | 0 |
| | | -0.5 | 1 | 0 | 0 | 2 | 0 | 0 | 997 | 1000 | 1000 | 0 | 0 | 0 |
| | | 0.0 | 1 | 0 | 0 | 5 | 0 | 0 | 994 | 980 | 1000 | 0 | 20 | 0 |
| | | 0.3 | 5 | 0 | 0 | 6 | 0 | 0 | 989 | 871 | 1000 | 0 | 129 | 0 |
| | | 0.4 | 3 | 0 | 0 | 6 | 0 | 0 | 991 | 737 | 1000 | 0 | 263 | 0 |
| | | 0.5 | 6 | 0 | 0 | 13 | 0 | 0 | 981 | 603 | 1000 | 0 | 397 | 0 |
| | | 0.6 | 9 | 0 | 0 | 8 | 0 | 0 | 983 | 400 | 1000 | 0 | 600 | 0 |
| | | 0.7 | 3 | 0 | 0 | 17 | 0 | 0 | 980 | 203 | 1000 | 0 | 797 | 0 |
| | | 0.8 | 3 | 0 | 0 | 19 | 0 | 0 | 978 | 93 | 1000 | 0 | 907 | 0 |
| 0.9 | 8 | 0 | 0 | 6 | 0 | 0 | 986 | 27 | 1000 | 0 | 973 | 0 | | |

3.2. 변화점이 2개인 경우

변화점이 2개인 모형으로 다음 모형을 고려하였다:

$$y_t = \alpha + \beta_1 t + \beta_2 (t - \psi_1)_+ + \beta_3 (t - \psi_2)_+ + \epsilon_t, \quad t = 1, 2, \dots, n.$$

오차항은 앞서 변화점의 개수가 1개 이하인 경우와 같은 방식으로 생성하였고, $\alpha = 0, \beta_1 = 1, \beta_2 = 0.5, \beta_3 = 0.5, \psi_1 = 0.3n, \psi_2 = 0.5n, 0.7n, n = 100, 300, 500, 1000$ 으로 설정하였다.

Table 6는 변화점의 개수에 대한 추정량 \hat{k} 에 대한 결과를 정리한 것이며, $n \geq 500$ 인 경우는 결과가 가지는 특징이 $n = 300$ 인 경우와 유사하므로 표에서 제외하였다. 본 논문의 방법은 $n = 300$ 이거나 $\psi_2 = 0.7n$ 일 때 변화점의 개수를 2개로 옳게 추정된 비율이 95% 이상이었으나, $n = 100$ 이고 $\psi_2 = 0.5n$ 인 경우에는 변화점의 개수를 1개 또는 0개로 작게 추정한 비율이 90% 근처였다. 이에 비해 Muggeo의 방법 중 score test를 이용한 방법은 모든 경우에 $\hat{k} = 2$ 인 비율이 90% 이상이며, $n = 300$ 또는 $\psi_2 = 0.7n$ 일 때는 100%인 것으로 나타났다. 그리고 BIC를 이용하여 변화점의 개수를 선택한 방법에 대해서는 $\hat{k} \leq 1$ 의 비율이 거의 0%이었으나, 오차항들 사이에 양의 자기상관이 존재하면서 그 정도가 커질 때 $\hat{k} = 2$ 의 비율이 감소하고 $\hat{k} > 2$ 의 비율이 증가하였다. 특히 $\phi \geq 0.8$ 로 강한 자기상관이 존재하는 경우에는 올바르게 변화점의 개수를 추정한 비율, 즉 $\hat{k} = 2$ 인 비율이 약 10%도 넘지 않았다. 앞서 변화점의 개수가 1개인 경우에서 설명했듯이 score test를 수행할 때 최대 변화점의 개수를 2개로 지정하였기 때문에 BIC를 이용한 방법과는 다르게 score test를 이용한 방법에서 $\hat{k} = 2$

Table 7: Five-number summaries of $\hat{\psi}_1$ and $\hat{\psi}_2$ when $\psi_1 = 0.3n$ and $\psi_2 = 0.5n$

| n | ϕ | Method | Freq. of $\hat{k} = 2$ | $\hat{\psi}_1$ | | | | | $\hat{\psi}_2$ | | | | | |
|-----|--------|--------|------------------------|----------------|------|------|------|-------|----------------|-------|-------|-------|-------|-------|
| | | | | min | Q1 | Q2 | Q3 | max | min | Q1 | Q2 | Q3 | max | |
| 100 | -0.9 | MA | 133 | 28.0 | 30.0 | 31.0 | 33.0 | 39.0 | 56.0 | 58.0 | 59.0 | 61.0 | 67.0 | |
| | | BIC | 1000 | 29.0 | 30.0 | 30.0 | 30.4 | 31.9 | 49.0 | 49.9 | 50.0 | 50.2 | 51.2 | |
| | | ST | 1000 | 29.0 | 29.8 | 30.0 | 30.2 | 31.0 | 49.0 | 49.8 | 50.0 | 50.2 | 51.0 | |
| | -0.5 | MA | 100 | 28.0 | 30.0 | 31.0 | 33.0 | 39.0 | 56.0 | 59.0 | 60.0 | 62.0 | 67.0 | |
| | | BIC | 1000 | 27.2 | 29.6 | 30.0 | 30.7 | 32.7 | 48.0 | 49.6 | 50.0 | 50.5 | 53.2 | |
| | | ST | 999 | 26.0 | 29.4 | 30.0 | 30.5 | 32.7 | 47.4 | 49.5 | 50.0 | 50.4 | 52.8 | |
| | 0.0 | MA | 82 | 28.0 | 30.0 | 31.0 | 33.0 | 38.0 | 56.0 | 59.0 | 60.0 | 62.8 | 66.0 | |
| | | BIC | 974 | 23.6 | 29.3 | 30.0 | 31.0 | 37.7 | 46.3 | 49.5 | 50.1 | 51.0 | 62.2 | |
| | | ST | 999 | 25.0 | 29.1 | 30.0 | 30.9 | 34.7 | 45.7 | 49.4 | 50.0 | 50.8 | 56.1 | |
| | 0.3 | MA | 83 | 28.0 | 30.5 | 32.0 | 34.0 | 40.0 | 56.0 | 59.0 | 61.0 | 63.0 | 69.0 | |
| | | BIC | 811 | 22.3 | 29.1 | 30.2 | 31.2 | 39.8 | 45.3 | 49.3 | 50.2 | 51.5 | 63.0 | |
| | | ST | 988 | 22.8 | 29.0 | 30.0 | 31.1 | 37.6 | 45.2 | 49.2 | 50.1 | 51.3 | 62.7 | |
| | 0.6 | MA | 113 | 28.0 | 30.0 | 31.0 | 33.0 | 41.0 | 55.0 | 59.0 | 61.0 | 63.0 | 69.0 | |
| | | BIC | 340 | 21.5 | 28.7 | 30.2 | 31.8 | 39.6 | 44.8 | 49.2 | 50.4 | 51.9 | 63.4 | |
| | | ST | 958 | 13.0 | 28.4 | 30.0 | 31.6 | 41.5 | 43.2 | 49.0 | 50.3 | 51.9 | 86.2 | |
| | 0.9 | MA | 231 | 28.0 | 30.0 | 31.0 | 34.0 | 43.0 | 56.0 | 60.0 | 62.0 | 65.0 | 73.0 | |
| | | BIC | 24 | 24.9 | 28.7 | 29.7 | 30.8 | 38.0 | 42.9 | 49.3 | 50.9 | 52.1 | 62.6 | |
| | | ST | 920 | 15.6 | 28.3 | 29.9 | 31.5 | 48.5 | 40.6 | 49.0 | 50.5 | 52.6 | 89.0 | |
| | 300 | -0.9 | MA | 1000 | 78.0 | 89.0 | 90.0 | 91.0 | 102.0 | 138.0 | 149.0 | 150.0 | 151.0 | 161.0 |
| | | | BIC | 1000 | 89.4 | 90.0 | 90.0 | 90.1 | 90.6 | 149.5 | 150.0 | 150.0 | 150.1 | 150.6 |
| | | | ST | 1000 | 89.4 | 89.9 | 90.0 | 90.1 | 90.6 | 149.5 | 149.9 | 150.0 | 150.1 | 150.6 |
| | | -0.5 | MA | 1000 | 77.0 | 88.0 | 90.0 | 92.0 | 101.0 | 140.0 | 149.0 | 150.0 | 152.0 | 161.0 |
| | | | BIC | 1000 | 88.6 | 89.7 | 90.0 | 90.3 | 91.3 | 149.0 | 149.8 | 150.0 | 150.3 | 151.2 |
| | | | ST | 1000 | 88.6 | 89.7 | 90.0 | 90.3 | 91.3 | 149.0 | 149.8 | 150.0 | 150.3 | 151.2 |
| 0.0 | | MA | 998 | 76.0 | 88.0 | 90.0 | 92.0 | 106.0 | 136.0 | 148.0 | 150.0 | 152.0 | 162.0 | |
| | | BIC | 998 | 87.3 | 89.5 | 90.0 | 90.5 | 92.2 | 148.0 | 149.6 | 150.0 | 150.4 | 151.8 | |
| | | ST | 1000 | 87.3 | 89.5 | 90.0 | 90.5 | 92.2 | 148.0 | 149.6 | 150.0 | 150.4 | 151.8 | |
| 0.3 | | MA | 998 | 74.0 | 88.0 | 90.0 | 92.0 | 107.0 | 138.0 | 148.0 | 150.0 | 152.0 | 163.0 | |
| | | BIC | 928 | 86.6 | 89.5 | 90.0 | 90.7 | 92.7 | 147.3 | 149.4 | 150.0 | 150.5 | 153.3 | |
| | | ST | 1000 | 86.6 | 89.5 | 90.0 | 90.6 | 92.7 | 147.3 | 149.4 | 150.0 | 150.5 | 153.3 | |
| 0.6 | | MA | 996 | 76.0 | 88.0 | 90.0 | 92.0 | 104.0 | 134.0 | 148.0 | 150.0 | 152.0 | 164.0 | |
| | | BIC | 510 | 86.1 | 89.0 | 89.9 | 90.8 | 94.0 | 145.6 | 149.2 | 150.0 | 150.8 | 155.3 | |
| | | ST | 1000 | 85.9 | 89.0 | 89.9 | 90.8 | 94.0 | 145.6 | 149.2 | 150.0 | 150.8 | 155.3 | |
| 0.9 | | MA | 948 | 72.0 | 89.0 | 90.0 | 91.0 | 107.0 | 135.0 | 148.0 | 150.0 | 151.0 | 165.0 | |
| | | BIC | 8 | 84.7 | 88.8 | 89.5 | 90.4 | 92.4 | 147.4 | 148.6 | 149.8 | 151.0 | 152.7 | |
| | | ST | 1000 | 79.3 | 88.5 | 90.0 | 91.5 | 100.7 | 143.1 | 148.5 | 149.9 | 151.2 | 156.8 | |

인 비율이 매우 높게 나온 것으로 보이며, 만약 최대 변화점의 개수를 늘린다면 BIC와 유사한 양상이 나타날 것으로 예상된다. 본 논문의 방법도 $n = 100$ 인 경우에는 (S2) 단계를 반복하는 수가 많지 않으며 이로 인해 $\hat{k} > 2$ 인 비율이 영향을 받은 것으로 보인다. 그러나 $n = 300$ 으로 자료의 수가 좀 큰 경우에는 $\hat{k} > 2$ 인 경우도 나타났으나 그 비율이 높지 않은 것을 볼 때 본 논문의 방법이 변화점의 개수를 잘 추정한다고 할 수 있다. $n = 100$ 이고 $\psi_2 = 0.5n$ 인 경우는 자료의 수보다 두 변화점 사이의 거리가 $\psi_2 - \psi_1 = 0.2n = 20$ 으로 $m = 27$ 보다 작으며, 조건 (2.4)를 만족하지 못 한다. 그러므로 대부분 변화점의 개수를 1개로 작게 추정한 것이며, 이는 $n = 300$ 의 결과에서 보듯이 자료의 수가 증가하면 저절로 해결이 된다.

Table 7과 Table 8은 $\hat{k} = 2$ 인 경우에 구한 두 개의 변화점 ψ_1, ψ_2 의 추정값에 대한 요약통계량이다. 먼저 Table 7은 $\psi_1 = 0.3n, \psi_2 = 0.5n$ 일 때의 추정 결과로 $n = 300$ 인 경우를 보면 본 논문에서 제안한 방법과 Muggeo의 방법에 대한 사분위수 값들은 서로 비슷하다. 그러나 최대값과 최소값을 보면 본 논문의 방법보다 Muggeo의 방법이 변화점의 위치를 더 잘 추정하는 것을 알 수 있다. $n = 100$ 인 경우를 보면, Muggeo의 방법들은 중앙값이 대부분 참값 $\psi_1 = 30, \psi_2 = 50$ 과 유사하고 사분위수범위가 크지 않으며 ϕ 의 값이 증가할

Table 8: Five-number summaries of $\hat{\psi}_1$ and $\hat{\psi}_2$ when $\psi_1 = 0.3n$ and $\psi_2 = 0.7n$

| n | ϕ | Method | Freq. of $\hat{k} = 2$ | $\hat{\psi}_1$ | | | | | $\hat{\psi}_2$ | | | | |
|-----|--------|--------|------------------------|----------------|------|------|------|------|----------------|------|------|------|------|
| | | | | min | Q1 | Q2 | Q3 | max | min | Q1 | Q2 | Q3 | max |
| 100 | -0.9 | MA | 962 | 28.0 | 29.0 | 30.0 | 31.0 | 42.0 | 61.0 | 69.0 | 70.0 | 71.0 | 73.0 |
| | | BIC | 1000 | 28.7 | 29.9 | 30.0 | 30.2 | 31.3 | 69.0 | 69.8 | 70.0 | 70.1 | 71.0 |
| | | ST | 1000 | 29.0 | 29.8 | 30.0 | 30.2 | 31.0 | 69.0 | 69.8 | 70.0 | 70.2 | 71.0 |
| | -0.5 | MA | 997 | 28.0 | 29.0 | 30.0 | 32.0 | 40.0 | 59.0 | 68.0 | 70.0 | 71.0 | 73.0 |
| | | BIC | 1000 | 27.5 | 29.6 | 30.0 | 30.4 | 32.3 | 67.9 | 69.6 | 70.0 | 70.4 | 72.4 |
| | | ST | 1000 | 27.6 | 29.6 | 30.0 | 30.4 | 32.0 | 68.0 | 69.6 | 70.0 | 70.4 | 72.4 |
| | 0.0 | MA | 994 | 28.0 | 29.0 | 30.0 | 32.0 | 42.0 | 61.0 | 68.0 | 70.0 | 71.0 | 73.0 |
| | | BIC | 980 | 26.5 | 29.4 | 30.0 | 30.7 | 34.2 | 66.2 | 69.3 | 70.0 | 70.7 | 73.9 |
| | | ST | 1000 | 26.3 | 29.4 | 30.0 | 30.6 | 33.3 | 66.2 | 69.4 | 70.0 | 70.7 | 73.6 |
| | 0.3 | MA | 989 | 28.0 | 29.0 | 31.0 | 32.0 | 40.0 | 58.0 | 68.0 | 69.0 | 71.0 | 73.0 |
| | | BIC | 871 | 24.9 | 29.1 | 30.0 | 31.0 | 35.1 | 63.7 | 69.1 | 70.0 | 70.8 | 73.6 |
| | | ST | 1000 | 24.9 | 29.1 | 30.0 | 31.0 | 34.9 | 63.7 | 69.1 | 70.0 | 70.8 | 76.0 |
| | 0.6 | MA | 983 | 28.0 | 29.0 | 30.0 | 33.0 | 42.0 | 59.0 | 68.0 | 70.0 | 71.0 | 73.0 |
| | | BIC | 400 | 24.2 | 28.9 | 30.0 | 31.1 | 36.1 | 64.0 | 69.0 | 70.0 | 71.3 | 78.8 |
| | | ST | 1000 | 21.6 | 28.7 | 30.0 | 31.2 | 39.8 | 60.7 | 69.0 | 70.1 | 71.2 | 78.8 |
| | 0.9 | MA | 986 | 28.0 | 29.0 | 30.0 | 31.0 | 41.0 | 57.0 | 69.0 | 70.0 | 71.0 | 73.0 |
| | | BIC | 27 | 24.8 | 27.7 | 29.0 | 31.7 | 34.5 | 63.0 | 68.1 | 68.8 | 71.4 | 74.9 |
| | | ST | 1000 | 19.4 | 28.4 | 29.9 | 31.5 | 39.6 | 53.2 | 68.7 | 70.1 | 71.6 | 79.4 |

수록 범위가 증가한다. 이에 비해 본 논문의 방법은 첫 번째 변화점에 대한 추정값의 중앙값이 30 근처로 잘 추정하나, 두 번째 변화점은 중앙값이 약 60 정도로 잘 추정하지 못 함을 알 수 있다. 이는 두 변화점 사이의 거리가 $\psi_2 - \psi_1 = 20$ 으로 모의실험에서 사용한 이동평균기간 $m = 27$ 보다 약 10만큼 작기 때문에 나타난 현상이라고 이해할 수 있다. Table 8은 $n = 100$, $\psi_2 = 0.7n$ 인 경우에 대한 결과로 Muggeo의 방법은 $\psi_2 = 0.5n$ 의 경우와 큰 차이가 없다. 하지만 본 논문의 방법은 $\psi_2 = 0.5n$ 의 경우에 비해 변화점의 개수를 올바르게 추정할 비율과 변화점의 위치를 추정한 정확도가 대폭 높아졌으며, 이는 두 변화점 사이의 거리가 $\psi_2 - \psi_1 = 40$ 으로 모의실험에서 사용한 이동평균기간 $m = 27$ 보다 크기 때문이다.

4. 최저기온 자료 분석

Figure 1의 검은색 실선은 기상청 기상자료개방포털 (Open MET Data Portal, 2023)에서 수집한 기온자료에 대한 시계열 그림으로 서울의 일최저기온의 연평균을 나타낸다. 그리고 초록색 실선(OLS)은 최소제곱법을 이용하여 선형추세분석을 실시한 결과로 추정된 모형식은 $\hat{y} = 7.361 + 0.032t$ 이며, 유의수준 5%에서 기울기가 유의한 것으로 나타났다. 그러나 검은색 실선을 살펴보면, 기온이 1980년대 초반까지 완만하게 증가하다가 1990년대에 급격히 증가하고, 다시 2000년대 이후에는 다시 기울기가 완만해졌으며, 초록색 직선과는 증가 패턴이 다른 것을 확인할 수 있다. 그러므로 본 논문에서는 조각별 선형추세모형을 이용하여 연평균 최저기온의 추세를 분석하였으며, Figure 1의 빨간색(MA-OLS)과 파란색 실선(segmented)은 각각 본 논문에서 제안한 방법과 Muggeo (2003)의 방법을 적용한 결과이다.

우선 본 논문의 방법을 이용하여 분석한 결과부터 살펴보면, $m = 15$ 를 이동평균기간으로 사용하여 변화점을 추정한 결과 변화점의 개수는 2개였으며, 그 위치는 1980년($t = 21$)과 1998년($t = 39$)인 것으로 나타났다. 그리고 추정된 변화점을 바탕으로 최소제곱법을 이용하여 조각별 선형 추세 모형을 추정한 결과 추정식은 $\hat{y} = 7.605 + 0.007t + 0.058(t - 21)_+ - 0.054(t - 39)_+$ 이었으며, 이 식은 Figure 1에서 빨간색 실선으로 표시되었다. 잔차에 대한 추가 분석 결과 잔차들은 백색잡음이며, 회귀계수에 대한 t -test 결과 기울기의 변화량을 나타내는 두 모수 β_2 와 β_3 는 모두 유의수준 5%에서 유의한 것으로 나타났다. 이동평균기간으로 $m = 10$ 도 고려하였으며, 분석 결과 1980년($t = 21$), 1990년($t = 31$), 2007년($t = 48$) 세 개의 변화점이 존재하고 모형식이

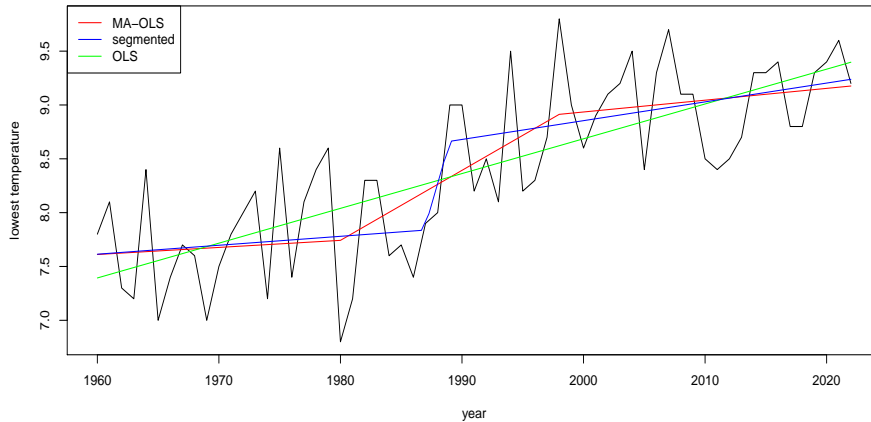


Figure 1: Lowest temperature (1960–2022).

$\hat{y} = 7.603 + 0.007t + 0.065(t - 21)_+ - 0.036(t - 31)_+ - 0.031(t - 48)_+$ 인 것으로 추정되었다. 그러나 기울기의 변화량을 나타내는 모수 중 β_2 만 유의수준 10%에서 유의하고, 나머지 모수 β_3 와 β_4 는 유의하지 않은 것으로 나타났다. 따라서 $m = 10$ 보다 $m = 15$ 의 분석 결과가 기온 자료에는 좀 더 적합하다고 생각한다.

다음으로 Muggeo (2020)의 방법을 적용하여 변화점의 개수를 추정한 결과 변화점이 존재하지 않는 것으로 나타났다. 하지만 앞서 본 논문이 제안한 방법으로 추정한 변화점의 개수가 2개였으므로 변화점의 개수를 2개로 지정하고 Muggeo (2008)의 segmented 함수를 이용하여 추가 분석을 실시한 결과 모형식이 $\hat{y} = 7.606 + 0.008t + 0.383(t - 27.89)_+ - 0.374(t - 30)_+$ 인 것으로 추정되었으며, 이 추정식은 Figure 1의 파란색 실선으로 표시하였다.

5. 결론

본 논문에서는 조각별 선형 추세 모형에서의 변화점이 1차 차분한 시계열의 평균 변화점과 일치한다는 사실에 근거하여 1차 차분한 시계열의 단순이동평균값을 이용하여 조각별 선형 추세 모형의 변화점을 추정, 모형을 적합시키는 방법을 제안하였다.

모의실험 결과에서 보듯이 본 논문에서 제안한 방법은 변화점의 개수를 잘 추정하는 편이다. 비록 변화점의 거리가 단순이동평균을 계산할 때 사용하는 이동평균기간에 비해 짧거나 기울기의 변화량의 크기가 크지 않은 경우에는 변화점의 개수나 위치를 추정할 때 정확도가 떨어지지만, 자료의 수가 충분히 많을 때는 정확도가 상당히 올라간다. 특히 모형에 포함된 오차항들 사이에 강한 양의 자기상관이 존재하는 경우에도 정확도가 많이 떨어지지 않는다. 이에 비해 Muggeo (2020)이 제안한 방법은 변화점의 위치는 잘 추정하지만, 오차항들 사이에 강한 양의 자기상관이 존재하면 변화점의 개수를 참값보다 큰 값으로 추정하는 경향이 있다. 따라서 실제 자료를 분석하는 경우에는 각각의 방법만 사용하는 것보다 변화점의 개수를 본 논문에서 제안한 방식으로 추정하고, 변화점의 위치는 Muggeo (2003)의 방법으로 추정하는 것이 좋지 않을까 생각한다.

앞서 제안한 모형 적합 단계를 살펴보면, 변화점 유무를 판단할 때 사용하는 검정통계량과 임계값 모두 이동평균기간 m 과 관련되어 있다. 만약 이동평균기간을 짧게 잡으면 임계값이 커지기 때문에 기울기의 변화량이 작은 경우 변화점을 거의 찾아내지 못할 수 있다. 반대로 이동평균기간을 길게 잡으면 임계값도 작아지고 오차항이 검정통계량에 주는 영향도 작아져 변화량의 크기가 작은 경우도 잘 찾아낼 수 있다. 그러나 2.4에 기술한 과정에서 (S2)를 반복하는 횟수가 줄어들고 이로 인해 변화점의 후보가 되는 집합이 제대로 구성이

안 될 수 있다. 따라서 변화점을 잘 찾기 위해서는 이동평균기간을 적절한 값으로 선택해야 한다. 2장에서 간단하게 이동평균기간이 만족해야할 조건을 기술하였으나 아직 부족하며, 이동평균기간을 선택하는 방법에 대한 추가 연구가 필요하다.

본 논문에서는 조각별 선형 추세 모형을 적합시키는 방법만을 제안하였으나, 방법의 핵심은 1차 차분한 시계열의 평균 변화점을 찾아내는 것이다. 또한 변화점을 찾아낼 때 사용한 검정통계량도 1차 차분한 시계열의 단순히 이동평균값으로만 이루어져 있기 때문에 모니터링 문제, 즉 새롭게 관측되어 추가되는 시계열 자료에 마지막 구간의 기울기를 지속적으로 적용할 수 있는가를 검토하는 문제로 어렵지 않게 확장 가능하다. 그러나 구체적인 모니터링 절차와 임계값, 이동평균기간의 선택 등에 대한 연구가 필요하며 추후 이에 대한 연구를 진행할 예정이다.

References

- Aue A and Horváth L (2013). Structural breaks in time series, *Journal of Time Series Analysis*, **34**, 1–16.
- Campra P and Morales M (2016). Trend analysis by a piecewise linear regression model applied to surface air temperatures in Southeastern Spain (1973–2014), Retrieved Oct. 18, 2023, [preprint], Available from: <https://doi.org/10.5194/npg-2016-29>.
- Deo CM (1972). Some limit theorems for maxima of absolute values of Gaussian sequences, *Sankhyā: The Indian Journal of Statistics, Series A*, **34**, 289–292.
- Deo CM (1975). Weak convergence of stochastic processes related to the absolute extremes of a gaussian sequence, *Sankhyā: The Indian Journal of Statistics, Series A*, **37**, 334–344.
- Eichinger B and Kirch C (2018). A MOSUM procedure for the estimation of multiple random change points, *Bernoulli*, **24**, 526–564.
- Fong Y, Huang Y, Gilbert PB, and Permar SR (2017). chngpt: Threshold regression model estimation and inference, *BMC Bioinformatics*, **18**, 1–7, Available from: <https://doi.org/10.1186/s12859-017-1863-x>
- Jiang F, Zhao Z, and Shao X (2022). Modelling the COVID-19 infection trajectory: A piecewise linear quantile trend model, *Journal of the Royal Statistical Society Series B: Statistical Methodology*, **84**, 1589–1607.
- Kim J, Oh H, and Cho H (2022). Moving sum procedure for change point detection under piecewise linearity, Retrieved Oct. 18, 2023, Available from: <https://api.semanticscholar.org/CorpusID:251442756>.
- Muggeo VMR (2003). Estimating regression models with unknown break-points, *Statistics in Medicine*, **22**, 3055–3071.
- Muggeo VMR (2008). Segmented: An R package to fit regression models with broken-line relationships, *R News*, **8**, 20–25.
- Muggeo VMR (2016). Testing with a nuisance parameter present only under the alternative: A score-based approach with application to segmented modelling journal of statistical, *Journal of Statistical Computation and Simulation*, **86**, 1–9.
- Muggeo VMR (2020). Selecting number of breakpoints in segmented regression: Implementation in the R package segmented, Retrieved Aug. 1, 2023, Available from: <https://www.researchgate.net/publication/343737604>.
- Open MET Data Portal (2023). Retrieved Oct. 18, 2023, Available from: <https://data.kma.go.kr>
- Park JS, Park CG, and Lee KE (2019). On study for change point regression problems using a difference-based regression model, *Communications for Statistical Applications and Methods*, **26**, 539–556.
- Tomal JH and Rahman H (2021). A Bayesian piecewise linear model for the detection of breakpoints in housing prices, *METRON*, **79**, 361–381.

- Wang Y, Huang G, Yang J, Lai H, Liu S, Chen C, and Xu W (2020). Change point detection with mean shift based on AUC from symmetric sliding windows, *Symmetry*, **12**, 1–17, Available from: <https://doi.org/10.3390/sym12040599>
- Yau CY and Zhao Z (2016). Inference for multiple change points in time series via likelihood ratio scan statistics, *Journal of the Royal Statistical Society. Series B (Statistical Methodology)*, **78**, 895–916.
- Zeileis A, Leisch F, Hornik K, and Kleiber C (2002). Strucchange: An R package for testing for structural change in linear regression models, *Journal of Statistical Software*, **7**, 1–38.

Received September 24, 2023; Revised October 23, 2023; Accepted October 24, 2023

차분한 시계열의 단순이동평균을 이용하여 조각별 선형 추세 모형을 추정하는 방법에 대한 연구

나옥경^{1,a}

^a경기대학교 응용통계학과

요 약

조각별 선형 추세 모형에서의 변화점은 1차 차분한 시계열의 평균 변화점과 일치한다. 그러므로 1차 차분한 시계열의 평균 변화점을 탐색하면 조각별 선형 추세 모형의 변화점을 추정할 수 있다. 본 논문에서는 이와 같은 사실에 근거하여 원 시계열이 아닌 1차 차분한 시계열의 단순이동평균을 이용하여 원 시계열의 기울기가 변하는 변화점을 탐색하는 방법을 제안하고, 이에 대한 모의실험을 수행하였다. 모의실험 결과 본 논문에서 제안한 방법은 오차항들이 서로 독립인 경우뿐만 아니라 오차항들 사이에 강한 양의 자기상관이 존재하는 경우에도 변화점의 개수를 잘 추정하는 것으로 나타났다.

주요용어: 조각별 선형 추세 모형, 변화점, 단순이동평균

¹(16227) 대한민국 경기도 수원시 영통구 광고산로 154-42, 경기대학교 응용통계학과. E-mail: okna@kyonggi.ac.kr