Communications for Statistical Applications and Methods
2023, Vol. 30, No. 2, 149–162

https://doi.org/10.29220/CSAM.2023.30.2.149
Print ISSN 2287-7843 / Online ISSN 2383-4757

Bias-reduced ℓ_1-trend filtering

Donghyeon Yu[a], Johan Lim[b], Won Son[1,c]

[a]Department of Statistics, Inha University, Korea;
[b]Department of Statistics, Seoul National University, Korea;
[c]Department of Information Statistics, Dankook University, Korea

Abstract

The ℓ_1-trend filtering method is one of the most widely used methods for extracting underlying trends from noisy observations. Contrary to the Hodrick-Prescott filtering, the ℓ_1-trend filtering gives piecewise linear trends. One of the advantages of the ℓ_1-trend filtering is that it can be used for identifying change points in piecewise linear trends. However, since the ℓ_1-trend filtering employs total variation as a penalty term, estimated piecewise linear trends tend to be biased. In this study, we demonstrate the biasedness of the ℓ_1-trend filtering in trend level estimation and propose a two-stage bias-reduction procedure. The newly suggested estimator is based on the estimated change points of the ℓ_1-trend filtering. Numerical examples illustrate that the proposed method yields less biased estimates for piecewise linear trends.

Keywords: ℓ_1-trend filtering, piecewise linear trend, shrinkage method, total variation denoising

1. Introduction

One of the main interests in sequential data analysis is to extract trends from temporary and unpredictable noises. Suppose that observations y_i $(i = 1, 2, \dots, n)$ can be expressed as an additive model

$$y_i = \theta_i + \epsilon_i, \tag{1.1}$$

where θ_i indicates the trend level at the i^{th} time and ϵ_i represents the corresponding random noise satisfying $\mathbb{E}[\epsilon_i] = 0$ and $\mathrm{Cov}(\epsilon_i, \epsilon_j) = 0$ where $i \neq j$. We further assume that random noises ϵ_is follow a centered Gaussian distribution with the variance σ^2.

Regularization methods are frequently used to estimate underlying trends θ_i $(i = 1, 2, \dots, n)$ from noisy observations. The H-P (Hodrick-Prescott) filtering (Hodrick and Prescott, 1997) is one of the famous regularization methods for trend filtering. The H-P filtering extracts trend estimates by solving the following regularization problem

$$\hat{\boldsymbol{\theta}}^{\mathrm{HP}}(\lambda) = \operatorname*{arg\,min}_{\boldsymbol{\theta}\in\mathbb{R}^n} \left\{ \frac{1}{2} \sum_{i=1}^{n} (y_i - \theta_i)^2 + \lambda \sum_{i=2}^{n-1} (\theta_{i-1} - 2\theta_i + \theta_{i+1})^2 \right\}, \tag{1.2}$$

This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT) (No. 2022R1A5A7033499 (Donghyeon Yu), No. 2021R1A2C1 010786 (Johan Lim), and No. 2020R1F1A1A01 051039 (Won Son)).

[1] Corresponding author: Department of Information Statistics, Dankook University, 152 Jukjeon-ro, Suji-gu, Yongin-si, Gyeonggi-do 16890, Korea. E-mail: son.won@dankook.ac.kr

Published 31 March 2023 / journal homepage: http://csam.or.kr

where $\lambda \geq 0$. Since $\theta_i - \theta_{i-1}$ can be regarded as the slope between the i^{th} and $(i-1)^{th}$ points, $\theta_{i-1} - 2\theta_i + \theta_{i+1}$ in the objective function (1.2) is equal to the difference between two neighboring slopes $\theta_i - \theta_{i-1}$ and $\theta_{i+1} - \theta_i$. Therefore, a smaller λ induces more drastic changes in slopes while a larger λ yields gradual changes.

Consider a piecewise linear trend model satisfying the equation

$$\theta_{i+1} - \theta_i = \theta_i - \theta_{i-1}$$

for $i = 2, \ldots, n-1$ except for a few change points $j_1, j_2, \ldots, j_J$. Here, the number of change points J is relatively small compared to the number of observations n. Since the H-P filtering employs the squared ℓ_2-norm as a penalty term, the neighboring slopes of the H-P filtered estimates cannot be exactly equal to each other. Therefore, the H-P filtering is inadequate for finding a piecewise linear solution.

Instead of the ℓ_2-penalty term in the H-P filtering, the ℓ_1-trend filtering (Kim *et al.*, 2009) employs ℓ_1-penalty term as

$$\hat{\boldsymbol{\theta}}^{\mathrm{L1}}(\lambda) = \underset{\boldsymbol{\theta} \in \mathbb{R}^n}{\arg\min} \left\{ \frac{1}{2} \sum_{i=1}^{n} (y_i - \theta_i)^2 + \lambda \sum_{i=2}^{n-1} |\theta_{i-1} - 2\theta_i + \theta_{i+1}| \right\}, \tag{1.3}$$

and implements sparsity in slope changes. The ℓ_1-trend filtering is widely used for various applications such as image denoising (Selvin *et al.*, 2016) and economics (Yamada and Jin, 2013). In addition, there are many studies for properties and possible extensions of the ℓ_1-trend filtering (Guntuboyina *et al.*, 2020; Rojas and Wahlberg, 2015; Tibshirani, 2014; Yu *et al.*, 2022).

However, the ℓ_1-trend filtering is a shrinkage estimator and is prone to be biased in estimating trend levels. In this paper, we assume a piecewise linear model having a relatively small number of changes in slopes. Under this assumption, we demonstrate that the ℓ_1-trend filtered estimates are biased and suggest a bias-reduction procedure. Based on the change points obtained by ℓ_1-trend filtering, the proposed bias-reduced trend estimator yields better estimates of the trend level.

This article is composed as follows. In Section 2, we show that the ℓ_1-trend filter is biased for trend level estimation. As a remedy, we propose a bias-reduced estimator for a piecewise linear trend in Section 3. In Section 4, we numerically illustrate that the proposed estimator outperforms the ℓ_1-trend filter in trend level estimation. Finally, in Section 5, we wrap up with a summary and discussion of possible extension.

2. Biasedness of the ℓ_1-trend filtering

The ℓ_1-trend filtered estimator $\hat{\theta}_i^{\mathrm{L1}}$ is biased in a sense that there exists some points satisfying

$$\mathbb{E}\left[\hat{\theta}_i^{\mathrm{L1}}\right] \neq \theta_i = \mathbb{E}\left[y_i\right].$$

For an observation vector $\mathbf{y} = (y_1, y_2, \ldots, y_n)^\intercal$, an underlying trend vector $\boldsymbol{\theta} = (\theta_1, \theta_2, \ldots, \theta_n)^\intercal$ and the $(n-2) \times n$ matrix $\mathbf{D}$ defined as

$$\mathbf{D} = \begin{bmatrix} -1 & 2 & -1 & 0 & \cdots & 0 & 0 & 0 \\ 0 & -1 & 2 & -1 & \cdots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \cdots & -1 & 2 & -1 \end{bmatrix},$$

the optimization problem (1.3) can be re-expressed in a matrix-vector representation as

$$\underset{\boldsymbol{\theta}\in\mathbf{R}^n}{\text{minimize}} \left\{ \frac{1}{2} \|\mathbf{y} - \boldsymbol{\theta}\|_2^2 + \lambda \|\mathbf{D}\boldsymbol{\theta}\|_1 \right\}, \tag{2.1}$$

where $\|\mathbf{y} - \boldsymbol{\theta}\|_2^2 = \sum_{i=1}^n (y_i - \theta_i)^2$ and $\|\mathbf{D}\boldsymbol{\theta}\|_1 = \sum_{i=2}^{n-1} |\theta_{i-1} - 2\theta_i + \theta_{i+1}|$. The solution of (2.1) is given as

$$\hat{\boldsymbol{\theta}}_\lambda = \mathbf{y} - \mathbf{D}^\mathsf{T}\hat{\mathbf{u}}_\lambda, \tag{2.2}$$

where elements of the vector $\hat{\mathbf{u}}_\lambda = (\hat{u}_2, \hat{u}_3, \ldots, \hat{u}_{n-1})$ are defined as

$$\hat{u}_i \in \begin{cases} \{+\lambda\}, & \text{if } (\mathbf{D}\hat{\boldsymbol{\theta}}_\lambda)_i > 0, \\ \{-\lambda\}, & \text{if } (\mathbf{D}\hat{\boldsymbol{\theta}}_\lambda)_i < 0, \\ [-\lambda, +\lambda], & \text{if } (\mathbf{D}\hat{\boldsymbol{\theta}}_\lambda)_i = 0, \end{cases} \tag{2.3}$$

where $(\mathbf{D}\hat{\boldsymbol{\theta}}_\lambda)_i = -\hat{\theta}_{i-1} + 2\hat{\theta}_i - \hat{\theta}_{i+1}$. In addition, for the boundary set $\mathcal{B}_\lambda = \{i : (\mathbf{D}\hat{\boldsymbol{\theta}}_\lambda)_i \neq 0\}$ and the corresponding sign vector $\mathbf{s}_{\mathcal{B}_\lambda}$, $\hat{\mathbf{u}}_\lambda$ are given as

$$\begin{aligned} \hat{\mathbf{u}}_{\lambda,\mathcal{B}_\lambda} &= \lambda\, \mathbf{s}_{\mathcal{B}_\lambda}, \\ \hat{\mathbf{u}}_{\lambda,\mathcal{B}_\lambda^c} &= \left(\mathbf{D}_{\mathcal{B}_\lambda^c} \left(\mathbf{D}_{\mathcal{B}_\lambda^c}\right)^\mathsf{T}\right)^{-1} \mathbf{D}_{\mathcal{B}_\lambda^c} \left(\mathbf{y}_{\mathcal{B}_\lambda} - \lambda \mathbf{D}_{\mathcal{B}_\lambda}{}^\mathsf{T} \mathbf{s}_{\mathcal{B}_\lambda}\right), \end{aligned}$$

where $\mathbf{D}_A$ and $\mathbf{s}_A$ indicate a submatrix of the matrix $\mathbf{D}$ and a subvector of the vector $\mathbf{s}$ corresponding to the index set A. One can refer to Tibshirani and Taylor (2011) for more details. From the equation

$$\begin{aligned} \mathbb{E}\left[\hat{\boldsymbol{\theta}}_\lambda \mid \mathcal{B}_\lambda\right] &= \mathbb{E}\left[\mathbf{y} - \mathbf{D}^\mathsf{T}\hat{\mathbf{u}}_\lambda \mid \mathcal{B}_\lambda\right] \\ &= \mathbb{E}\left[\mathbf{y} \mid \mathcal{B}_\lambda\right] - \mathbb{E}\left[\mathbf{D}^\mathsf{T}\hat{\mathbf{u}}_\lambda \mid \mathcal{B}_\lambda\right] = \mathbb{E}\left[\mathbf{y} \mid \mathcal{B}_\lambda\right] - \mathbf{D}^\mathsf{T}\mathbb{E}\left[\hat{\mathbf{u}}_\lambda \mid \mathcal{B}_\lambda\right], \end{aligned}$$

if the ℓ_1-trend filtered estimates were unbiased, the equation

$$\mathbb{E}\left[\hat{u}_{\lambda,2} \mid \mathcal{B}_\lambda\right] = \mathbb{E}\left[\hat{u}_{\lambda,3} \mid \mathcal{B}_\lambda\right] = \cdots = \mathbb{E}\left[\hat{u}_{\lambda,n-1} \mid \mathcal{B}_\lambda\right]$$

should hold. However, if the boundary set $\mathcal{B}_\lambda$ is nonempty,

$$\mathbb{E}\left[\hat{u}_{\lambda,i} \mid \mathcal{B}_\lambda\right] = \lambda \quad \text{or} \quad \mathbb{E}\left[\hat{u}_{\lambda,i} \mid \mathcal{B}_\lambda\right] = -\lambda$$

implies that all the ℓ_1-trend filtered estimates are local maximum or minimum. Since this result is implausible, we can conclude that the ℓ_1-trend filter is a biased estimator for trend level estimation.

We introduce a new interpretation to understand the biasedness of the ℓ_1-trend filter. To do so, we represent the ℓ_1-trend filter via piecewise linear formula. Suppose that a block partition $\mathcal{B}_\lambda$ consists of blocks $B_0 = [1, j_1), B_1 = [j_1, j_2), \ldots, B_J = [j_J, n]$ with change points $j_1, j_2, \ldots, j_J$. Here, the block partitions and change points depend on the tuning parameter λ. For notational convenience, we denote $j_0 = 1$ and $j_{J+1} = n + 1$. For the points in the interval $B_k = [j_k, j_{k+1})$, the ℓ_1-trend filtered estimates $\hat{\theta}_i$ has a common slope $\hat{b}_k$ as

$$\hat{b}_k = \hat{\theta}_{j_k+1} - \hat{\theta}_{j_k} = \hat{\theta}_{j_k+2} - \hat{\theta}_{j_k+1} = \cdots = \hat{\theta}_{j_{k+1}} - \hat{\theta}_{j_{k+1}-1}.$$

Hence, for $i \in B_k = [j_k, j_{k+1})$, the ℓ_1-trend filtered estimates can be represented as

$$\hat{\theta}_i = \hat{a}_k + \hat{b}_k i$$

for an intercept $\hat{a}_k$. In addition, since the two neighboring line segments in blocks B_k and B_{k-1} are connected at the boundary point j_k, we have $\hat{a}_{k-1} + \hat{b}_{k-1} j_k = \hat{a}_k + \hat{b}_k j_k$ and

$$(\hat{a}_k - \hat{a}_{k-1}) + \left(\hat{b}_k - \hat{b}_{k-1}\right) j_k = 0$$

for each j_k $(k = 1, 2, \ldots, J)$.

With this representation, the objective function in (2.1) can be re-expressed as

$$\begin{aligned} Q(\mathbf{a}, \mathbf{b}) &= \frac{1}{2}\sum_{i=1}^{n}(y_i - \theta_i)^2 + \lambda \sum_{i=2}^{n-1} |\theta_{i-1} - 2\theta_i + \theta_{i+1}| \\ &= \frac{1}{2}\sum_{k=0}^{J}\sum_{i \in B_k}(y_i - a_k - b_k i)^2 + \lambda \sum_{k=1}^{J} |b_k - b_{k-1}|, \end{aligned} \tag{2.4}$$

where J is the number of change points. Given a block partition $\mathcal{B}_\lambda$, vectors of parameters $\mathbf{a} = (a_0, a_1, \ldots, a_J)$ and $\mathbf{b} = (b_0, b_1, \ldots, b_J)$ can be found as the minimizers of the objective function (2.4). $\hat{\mathbf{a}}$ and $\hat{\mathbf{b}}$, the minimizers of the objective function $Q(\mathbf{a}, \mathbf{b})$, are given as the solutions of the $2(J + 1)$ equations

$$\frac{\partial Q(\mathbf{a}, \mathbf{b})}{\partial a_k} = \sum_{i=j_k}^{j_{k+1}-1}(a_k + b_k i - y_i) = 0, \tag{2.5}$$

$$\frac{\partial Q(\mathbf{a}, \mathbf{b})}{\partial b_k} = \begin{cases} \sum_{i=1}^{j_1-1} i\ (a_0 + b_0 i - y_i) + \lambda\, \mathrm{sgn}(b_0 - b_1) = 0 & \text{if } k = 0, \\ \sum_{i=j_k}^{j_{k+1}-1} i\ (a_k + b_k i - y_i) + \lambda\ (\mathrm{sgn}(b_k - b_{k-1}) + \mathrm{sgn}(b_k - b_{k+1})) = 0 & \text{if } k = 1, \ldots, J-1, \\ \sum_{i=j_J}^{n} i\ (a_J + b_J i - y_i) + \lambda\ (\mathrm{sgn}(b_J - b_{J-1})) = 0 & \text{if } k = J \end{cases} \tag{2.6}$$

for $k = 0, 1, \ldots, J$ where $\mathrm{sgn}(x) = \mathrm{sign}(x)$ if $x \neq 0$ and $\mathrm{sgn}(x) \in [-1, 1]$ if $x = 0$. Comparing these equations with the normal equations of the ordinary least squares, the difference can be found in the Equation (2.6) where $\mathrm{sgn}(b_k - b_{k-1}) = \mathrm{sgn}(b_k - b_{k+1}) \neq 0$.

In the case of $\mathrm{sgn}(b_k - b_{k-1}) = \mathrm{sgn}(b_k - b_{k+1}) = 1$, i.e., when the slope of the k^{th} interval is larger than those of the neighboring intervals, the normal Equations (2.5) and (2.6) can be re-expressed as

$$\begin{aligned} \hat{a}_k + \hat{b}_k \bar{x}_k - \bar{y}_k &= 0, \\ n_k \hat{a}_k \bar{x}_k + \hat{b}_k s_{xx,k} - s_{xy,k} + 2\lambda &= 0, \end{aligned}$$

where $n_k = j_k - j_{k-1}$, $\bar{x}_k = \sum_{i=j_k}^{j_{k+1}-1} i/n_k$, $\bar{y}_k = \sum_{i=j_k}^{j_{k+1}-1} y_i/n_k$, $s_{xx,k} = \sum_{i=j_k}^{j_{k+1}-1} i^2$ and $s_{xy,k} = \sum_{i=j_k}^{j_{k+1}-1} i y_i$. Therefore, the slope of the k^{th} interval is given as

$$\hat{b}_k = \frac{s_{xy,k} - n_k \bar{x}_k \bar{y}_k - 2\lambda}{s_{xx,k} - n_k \bar{x}_k^2},$$

and $\hat{b}_k$ gets smaller for a larger λ. Similarly, for the interval satisfying $\mathrm{sgn}(\hat{b}_k - \hat{b}_{k-1}) = \mathrm{sgn}(\hat{b}_k - \hat{b}_{k+1}) = -1$, slope $\hat{b}_k$ gets larger for increasing λ as in the equation

$$\hat{b}_k = \frac{s_{xy,k} - n_k \bar{x}_k \bar{y}_k + 2\lambda}{s_{xx,k} - n_k \bar{x}_k^2}.$$

From these results, we can conclude that the slope estimates are biased for non-zero regularization parameter λ. However, from the normal Equation (2.5), we have

$$y_i - \bar{y}_k = \hat{b}_k (i - \bar{x}_k),$$

and at the center of the line segment $(\bar{x}_k, \bar{y}_k)$, the estimate is unbiased.

3. Bias-reduced trend estimator

Considering the nature of regularization methods, the biasedness of the ℓ_1-trend filtering estimator is not surprising. In the ℓ_1-trend filtering estimator, biases are due to the Equation (2.6) which contain differentiated penalty terms. Thus, one can consider replacing the Equation (2.6) to obtain bias-reduced estimates. We propose a bias-reduction procedure based on the ℓ_1-trend filtered estimates of change points.===== However, since two neighboring line segments should be connected at a change point, the proposed trend estimator cannot be entirely bias-free.

3.1. The proposed estimator

The proposed procedure utilizes the change points set $\widehat{\mathcal{J}}(\lambda)$ of the ℓ_1-trend filtered estimates and the Equation (2.5). The Equation (2.5) can be expressed as

$$\hat{a}_k + \hat{b}_k \frac{\sum_{i=j_k}^{j_{k+1}-1} i}{j_k - j_{k-1}} - \frac{\sum_{i=j_k}^{j_{k+1}-1} y_i}{j_k - j_{k-1}} = \hat{a}_k + \hat{b}_k \bar{x}_k - \bar{y}_k = 0 \quad (k = 0, 1, \ldots, J).$$

Therefore, each line segment given by the proposed procedure passes through the point $(\bar{x}_k, \bar{y}_k)$.

For bias reduction, one can consider the equations obtained by removing the differentiated penalty term from the Equation (2.6) as

$$\sum_{i=j_k}^{j_{k+1}-1} i\,(a_h + b_h i - y_i) = 0 \quad (k = 0, 1, \ldots, J).$$

However, the resulting solutions are not guaranteed to be connected at each change point. Thus, we propose bias-reduced trend estimator obtained by minimizing the objective functions

$$\sum_{k=0}^{J} \sum_{i \in B_k} (a_k + b_k i - y_i)^2 \tag{3.1}$$

subject to

$$(a_k - a_{k-1}) + (b_k - b_{k-1})\, j_k = 0, \tag{3.2}$$

$$a_k + b_k \bar{x}_k - \bar{y}_k = 0 \tag{3.3}$$

for $k = 0, 1, \ldots, J$. The additional requirement (3.2) is necessary for the connectedness of the estimates while Equation (3.3) is equivalent to (2.5).

3.2. Estimation

The optimization problem for the proposed method can be easily solved. Since the midpoint of each line segment is fixed as $(\bar{x}_k, \bar{y}_k)$, if another point on the line segment is specified, the entire line segment can be completely determined. Therefore, the left and right end points of the line segment of the k^{th} block, B_k, say (j_k, v_k) and (j_{k+1}, v_{k+1}), satisfying

$$\bar{x}_k = \frac{j_k + j_{k+1}}{2} \quad \text{and} \quad \bar{y}_k = \frac{v_k + v_{k+1}}{2}$$

can be determined. Consequently, the starting point of the k^{th} block B_k can be found as

$$j_k = 2\bar{x}_k - j_{k+1} \quad \text{and} \quad v_k = 2\bar{y}_k - v_{k+1}$$

and similarly the ending point of the k^{th} block, (u_{k+1}, v_{k+1}), can be identified. In this sequential procedure, the entire estimates can be determined if any one point on the piecewise linear estimates is specified.

Therefore, we can think of the objective function (3.1) as a function of v_0, where v_0 is the y-coordinate of the starting point of the first block, $(u_0, v_0) = (1, v_0) = (1, \hat{\theta}_1)$. Let $f(i|v_0, \mathcal{B})$ be a piecewise linear trend estimate given $\hat{\theta}_1 = v_0$ and the ℓ_1-trend filtered change points estimate $\mathcal{B}$. We represent the optimization problem as

$$\begin{aligned} &\underset{v_0}{\text{minimize}}\ \mathcal{L}(v_0 \mid \mathcal{B}), \\ &\text{where } \mathcal{L}(v_0 \mid \mathcal{B}) = \sum_{i=1}^{n} \{y_i - f(i \mid v_0, \mathcal{B})\}^2. \end{aligned} \tag{3.4}$$

Since the objective function (3.4) is quadratic in v_0, we can find the numerical solution easily.

4. Numerical illustration

To examine the effectiveness of the proposed bias-reduction procedure, we consider the following four scenarios of length 50 ($i = 1, 2, \ldots, 50$) for the piecewise linear structures containing one to four change points.

1. Scenario 1 :

$$\theta_i^* = \begin{cases} -i & \text{if } i = 1, \ldots, 25, \\ i - 50 & \text{if } i = 26, \ldots, 50. \end{cases}$$

2. Scenario 2 :

$$\theta_i^* = \begin{cases} -i & \text{if } i = 1, \ldots, 12, \\ i - 24 & \text{if } i = 13, \ldots, 38, \\ -i + 52 & \text{if } i = 39, \ldots, 50. \end{cases}$$

3. Scenario 3 :

$$\theta_i^* = \begin{cases} -i & \text{if } i = 1, \ldots, 12, \\ i - 24 & \text{if } i = 13, \ldots, 25, \\ -i + 26 & \text{if } i = 26, \ldots, 38, \\ i - 50 & \text{if } i = 39, \ldots, 50. \end{cases}$$

4. Scenario 4 :

$$\theta_i^* = \begin{cases} -i & \text{if } i = 1, \ldots, 10, \\ i - 20 & \text{if } i = 11, \ldots, 20, \\ -i + 20 & \text{if } i = 21, \ldots, 30, \\ i - 40 & \text{if } i = 31, \ldots, 40, \\ -i + 40 & \text{if } i = 41, \ldots, 50. \end{cases}$$

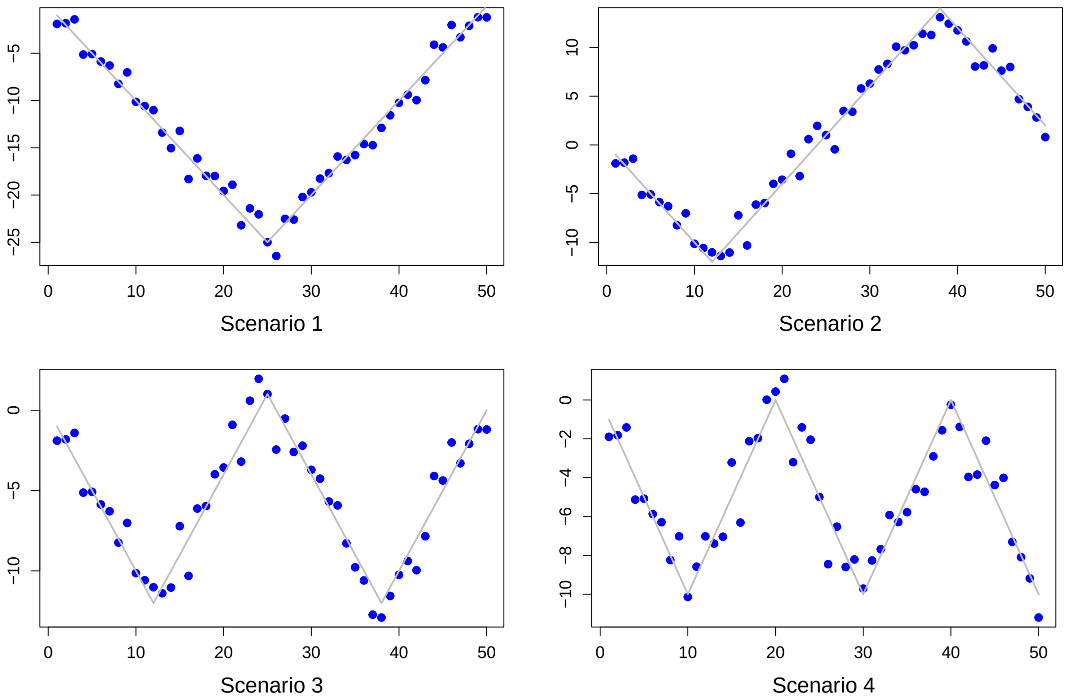


Figure 1: *Underlying piecewise linear trends (grey solid line) and observations (blue dots) for* $\epsilon_i \sim N(0, 1)$.

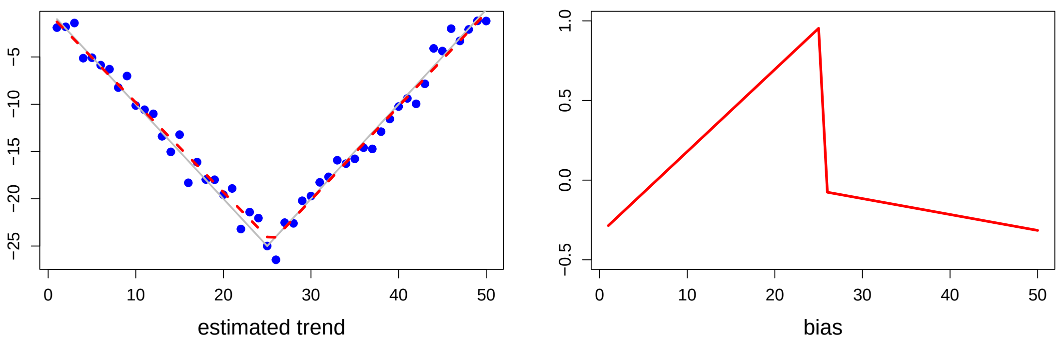


Figure 2: ℓ_1*-trend filtered estimates (left panel) and bias (right panel) for scenario 1.*

For the above four mean structures, noise levels $\sigma = 0.1, 0.2, 0.5, 1.0$ are considered. Figure 1 shows the four true piecewise linear trends and observations whose error terms ϵ follow standard normal distributions.

4.1. Biasedness of the ℓ_1-trend filtering

Under the above assumptions, we illustrate the biasedness of the ℓ_1-trend filtering for scenario 1. For practical applications, the solution path of the ℓ_1-trend filtering can be found by generalized lasso algorithm (Tibshirani and Taylor, 2011). The R-package "genlasso" provides functions implementing the generalized lasso algorithm. In this paper, we use the "`trendfilter`" function to obtain a solution path of the ℓ_1-trend filtering.

In Figure 1, a sample observation from scenario 1 is displayed. For this sample, there is no true change point set in the solution path of the ℓ_1-trend filtering and the minimal change points set which contains the true change points, and the least change points can be found for the tuning parameter $\lambda = 24.978$. The left panel of Figure 2 shows the estimated trend $\hat{\theta}_i$ obtained by the ℓ_1-trend filtering and the right panel depicts the difference between the trend estimates $\hat{\theta}_i$ and the true trend θ_i^*. As we can see in the subfigures, there is systematic discrepancy, i.e., unignorable bias between the estimates and the true trend level. The estimated trend tends to underestimate the local extremes.

We estimate average biases $\hat{\mathbb{E}}(\hat{\theta}_i) - \theta_i^*$ $(i = 1, 2, \ldots, n)$ of the ℓ_1-trend filtering numerically. We generate $R = 1000$ random samples and find empirical expectations of the ℓ_1-trend filtered trend as

$$\hat{\mathbb{E}}\left[\hat{\theta}_i(\lambda)\right] = \frac{1}{R}\sum_{r=1}^{R} \hat{\theta}_{i,r}(\lambda),$$

where $\hat{\theta}_{i,r}$ indicates the i^{th} observation in the r^{th} sample. Using this empirical expectation, we can estimate the bias of the i^{th} observation as $\hat{\mathbb{E}}[\hat{\theta}_i(\lambda)] - \theta_i^*$. Note that the empirical expectation and the estimated bias depend on the tuning parameter λ.

For the four scenarios and various λ levels, the estimated biases are displayed in Figure 3. The horizontal axes indicate the indices of the observations while the vertical axes show the levels of biases. Each subfigure contains four lines that exhibit the levels of biases for different values of the tuning parameter, $\lambda = 1, 10, 20$, and 50.

All the subfigures show that there exist structural patterns of biases. We can see that biases tend to exhibit sequentially increasing or decreasing patterns. In addition, absolute values of biases get larger near the peaks and troughs of the true trends. Larger values of the tuning parameter λ give larger biases in absolute values. When λ is as small as 1, the values of biases concentrate near zero, except in the case of indices being near peaks and troughs, while larger values of λ lead to wider gaps from the zero level. The noise level σ does not affect the patterns of biases much. However, for the higher noise levels and smaller tuning parameter values, absolute values of biases are inclined to get larger, especially near peaks and troughs.

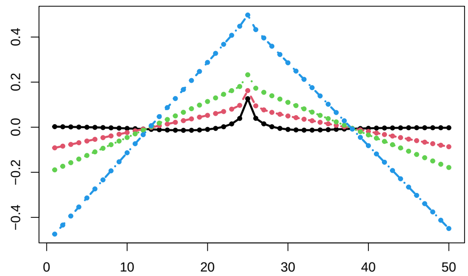
4.2. Performance of the bias-reduced estimator

To examine the effectiveness of our proposed bias-reduction procedure numerically, we fit the bias-reduced estimator $\hat{\theta}_i^{\mathrm{BR}}(\lambda)$ on the randomly generated samples and evaluate empirical biases $\hat{\mathbb{E}}[\hat{\theta}_i^{\mathrm{BR}}(\lambda)] - \theta_i^*$ as in the previous subsection. And then, the absolute values of empirical biases of the two estimators, the ℓ_1-trend filtering estimator $\hat{\theta}^{L1}(\lambda)$ and the proposed bias-reduced estimator $\hat{\theta}^{\mathrm{BR}}(\lambda)$, are averaged as

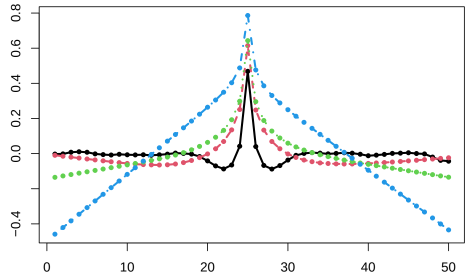
$$\frac{1}{n}\sum_{i=1}^{n}\left|\hat{\mathbb{E}}\left[\hat{\theta}_i(\lambda)\right] - \theta_i^*\right|. \tag{4.1}$$

The values obtained by (4.1) are compared in Table 1 and Figure 4. In addition, the average mean squared errors for the tuning parameter λ are evaluated as

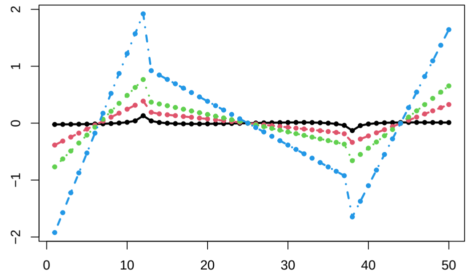
$$\frac{1}{nR}\sum_{r=1}^{R}\sum_{i=1}^{n}\left(\hat{\theta}_{i,r}(\lambda) - \theta_i^*\right)^2,$$



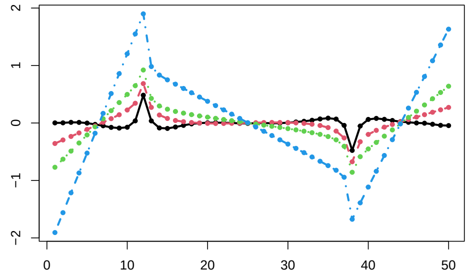
(a) Scenario 1, $\sigma = 0.2$



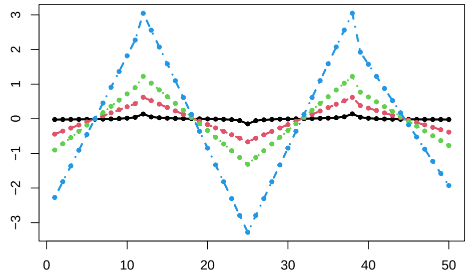
(b) Scenario 1, $\sigma = 1$



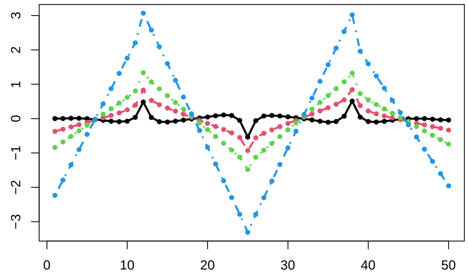
(c) Scenario 2, $\sigma = 0.2$



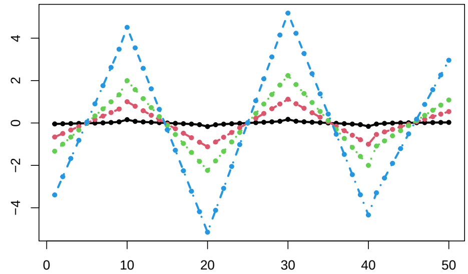
(d) Scenario 2, $\sigma = 1$



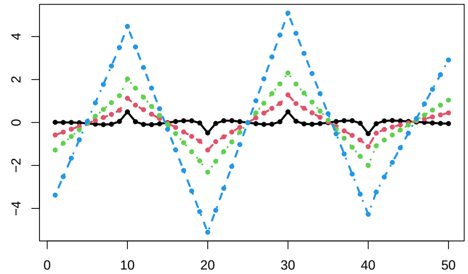
(e) Scenario 3, $\sigma = 0.2$



(f) Scenario 3, $\sigma = 1$



(g) Scenario 4, $\sigma = 0.2$



(h) Scenario 4, $\sigma = 1$

Figure 3: *Empirical biases for $\epsilon_i \sim N(0, \sigma^2)$ where $\sigma = 0.2$ and 1 (black solid line: $\lambda = 1$, red dashed line: $\lambda = 10$, green dotted line: $\lambda = 20$, blue dot-dash line: $\lambda = 50$).*

Table 1: Average absolute biases and MSEs of the ℓ_1-trend filtered (L1) and bias-reduced (BR) estimates

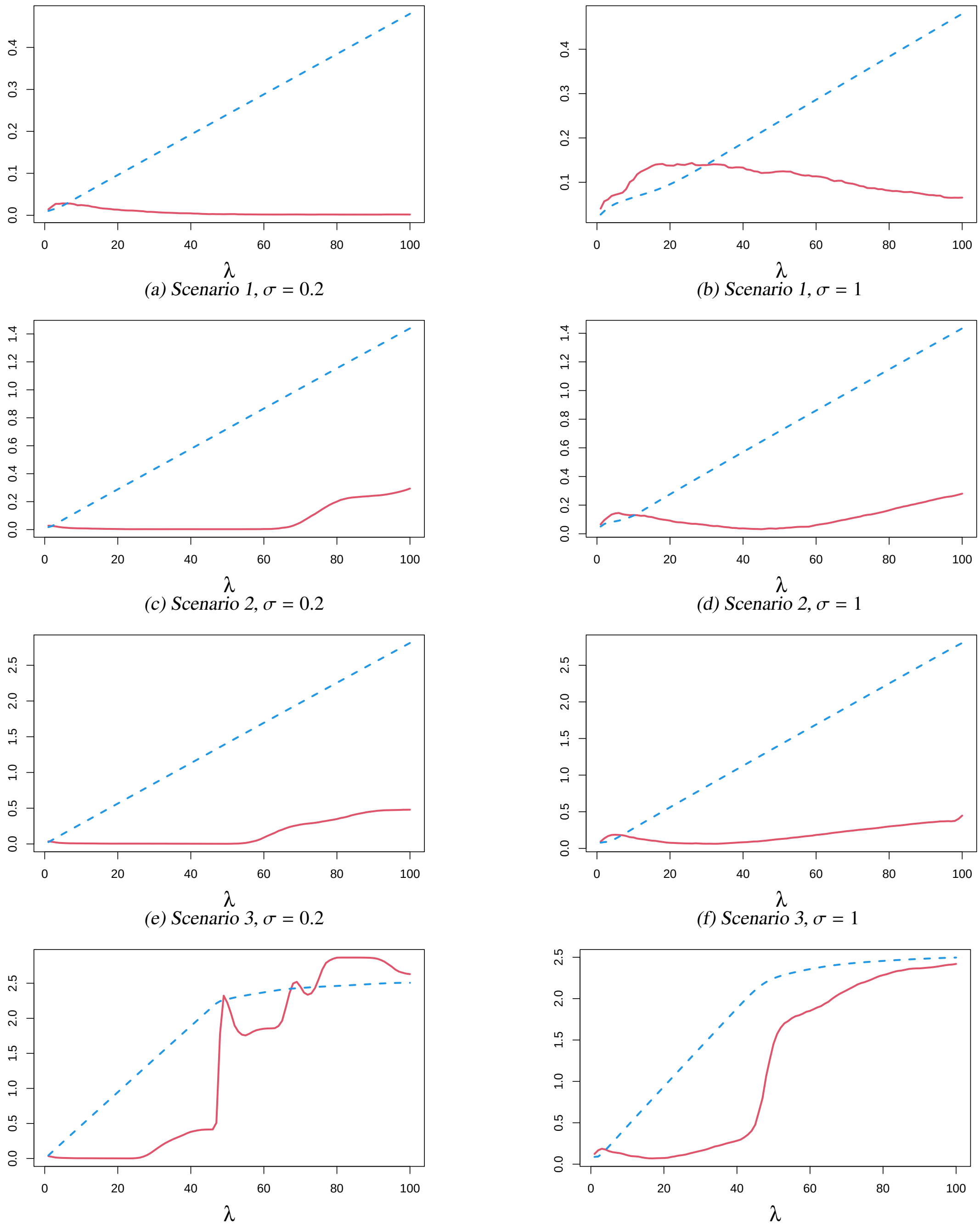
	trend	λ	$\sigma = 0.1$		$\sigma = 0.2$		$\sigma = 0.5$		$\sigma = 1.0$	
			BR	L1	BR	L1	BR	L1	BR	L1
bias	scenario 1	1	0.010	0.007	0.014	0.010	0.028	0.018	0.041	0.027
		10	0.007	0.048	0.024	0.047	0.068	0.048	0.106	0.066
		20	0.002	0.096	0.013	0.096	0.067	0.094	0.138	0.095
		50	0.001	0.240	0.003	0.240	0.034	0.240	0.124	0.237
		opt	0.000	0.185	0.001	0.371	0.002	0.927	0.004	1.600
	scenario 2	1	0.013	0.013	0.028	0.018	0.047	0.035	0.066	0.051
		10	0.002	0.144	0.009	0.143	0.046	0.138	0.131	0.127
		20	0.002	0.289	0.004	0.288	0.025	0.285	0.092	0.276
		50	0.001	0.722	0.003	0.722	0.015	0.720	0.039	0.716
		opt	0.002	0.066	0.004	0.142	0.009	0.360	0.018	0.574
	scenario 3	1	0.015	0.027	0.037	0.026	0.069	0.045	0.093	0.078
		10	0.002	0.283	0.008	0.282	0.039	0.280	0.149	0.273
		20	0.002	0.565	0.005	0.565	0.022	0.565	0.076	0.561
		50	0.001	1.412	0.003	1.413	0.054	1.412	0.127	1.410
		opt	0.004	0.187	0.007	0.358	0.018	0.656	0.055	0.838
	scenario 4	1	0.012	0.046	0.034	0.044	0.083	0.048	0.123	0.088
		10	0.001	0.473	0.004	0.472	0.026	0.469	0.108	0.460
		20	0.001	0.946	0.002	0.946	0.020	0.944	0.074	0.938
		50	2.499	2.273	2.222	2.272	1.771	2.266	1.452	2.244
		opt	0.003	0.171	0.007	0.343	0.019	0.662	0.049	0.730
MSE	scenario 1	1	0.120	0.054	0.653	0.254	5.764	2.190	31.887	11.756
		10	0.064	0.190	0.330	0.316	2.446	1.332	12.130	5.361
		20	0.062	0.647	0.255	0.760	2.197	1.667	9.499	5.262
		50	0.051	3.869	0.241	3.967	1.591	4.747	8.033	7.861
		opt	0.025	0.101	0.629	2.855	4.553	18.212	113.824	309.688
	scenario 2	1	0.163	0.089	0.821	0.344	7.049	2.569	33.083	12.624
		10	0.153	1.627	0.549	1.822	3.484	3.232	15.844	8.602
		20	0.165	6.304	0.613	6.506	3.296	7.872	13.290	12.663
		50	0.167	39.014	0.669	39.240	4.320	40.528	13.236	44.719
		opt	0.153	0.612	3.775	17.874	0.597	2.387	14.136	36.450
	scenario 3	1	0.212	0.133	1.060	0.425	8.525	2.995	37.348	13.915
		10	0.162	5.548	0.668	5.731	4.068	7.190	20.541	13.045
		20	0.173	22.027	0.649	22.192	4.190	23.503	16.338	28.582
		50	0.184	137.456	0.793	137.554	5.287	138.528	16.810	141.751
		opt	0.151	0.604	3.769	17.614	4.404	14.793	40.825	70.578
	scenario 4	1	0.219	0.238	1.048	0.552	8.253	3.299	42.864	14.546
		10	0.168	15.813	0.672	15.965	4.614	17.425	20.388	23.408
		20	0.171	63.187	0.672	63.251	4.765	64.360	18.580	68.611
		50	503.977	354.523	401.807	354.500	323.822	354.203	268.787	350.345
		opt	0.177	0.707	4.391	22.252	3.002	11.823	39.030	60.301

* opt : optimal λ for each repetition.

and compared in Table 1 and Figure 5.

We can see that, in general, large tuning parameter values lead to large biases for the ℓ_1-trend filtering estimator. For all the four noise levels, $\sigma = 0.1, 0.2, 0.5$ and 1.0, $\hat{\theta}^{\text{L1}}(\lambda)$ show that the absolute value of bias increases in λ. On the contrary, the absolute values of bias of $\hat{\theta}^{\text{BR}}(\lambda)$ do not increase in λ unless the value of λ gets large enough. However, for large values of λ, some true change points are not identified and a large bias can be induced for $\hat{\theta}^{\text{BR}}(\lambda)$ due to the missing true change points.

For all the combinations of the four scenarios and noise levels ($\sigma = 0.1, 0.2, 0.5, 1.0$), the bias-reduced estimator $\hat{\theta}^{\text{BR}}(\lambda)$ shows smaller biases for moderate values of tuning parameter λ. However, for smaller values of λ, $\hat{\theta}^{\text{L1}}(\lambda)$ and $\hat{\theta}^{\text{BR}}(\lambda)$ give similar magnitudes of biases. In addition, for scenario



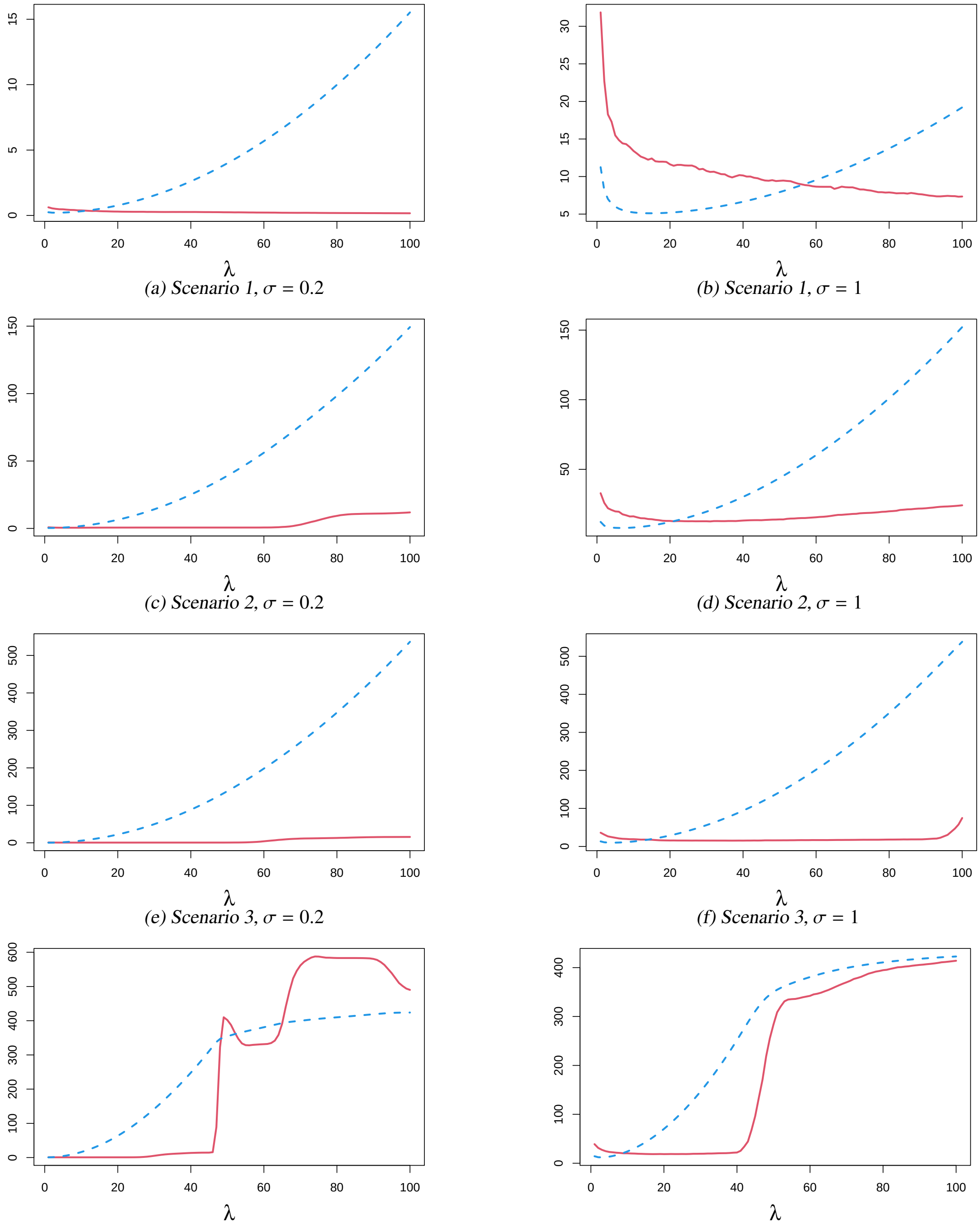
(a) Scenario 1, $\sigma = 0.2$ (b) Scenario 1, $\sigma = 1$

(c) Scenario 2, $\sigma = 0.2$ (d) Scenario 2, $\sigma = 1$

(e) Scenario 3, $\sigma = 0.2$ (f) Scenario 3, $\sigma = 1$

(g) Scenario 4, $\sigma = 0.2$ (h) Scenario 4, $\sigma = 1$

Figure 4: *Empirical biases as a function of λ for $\epsilon_i \sim N(0, \sigma^2)$ where $\sigma = 0.2$ and 1 (red solid line: bias-reduced estimator, blue dashed line: ℓ_1-trend filtered estimates).*



(a) *Scenario 1*, $\sigma = 0.2$ (b) *Scenario 1*, $\sigma = 1$

(c) *Scenario 2*, $\sigma = 0.2$ (d) *Scenario 2*, $\sigma = 1$

(e) *Scenario 3*, $\sigma = 0.2$ (f) *Scenario 3*, $\sigma = 1$

(g) *Scenario 4*, $\sigma = 0.2$ (h) *Scenario 4*, $\sigma = 1$

Figure 5: *Empirical MSE as a function of* λ *for* $\epsilon_i \sim N(0, \sigma^2)$ *where* $\sigma = 0.2$ *and* 1 *(red solid line: bias-reduced estimator, blue dashed line:* ℓ_1*-trend filtered estimates).*

4 which contains more change points than the other scenarios, $\hat{\theta}^{BR}(\lambda)$ gives comparable bias levels to $\hat{\theta}^{L1}(\lambda)$ for large tuning parameter values. For the optimal values of the tuning parameter λ, average absolute biases of the bias-reduced estimator are smaller than those of the ℓ_1-trend filtered estimator.

In Table 1, we find that the empirical biases of $\hat{\theta}^{L1}(\lambda)$ are not influenced by the noise level except in the case of small and optimal tuning parameter values. However, for most cases, biases of $\hat{\theta}^{BR}(\lambda)$ tend to increase as the noise level σ gets larger. We can conclude that although the bias-reduced estimator $\hat{\theta}^{BR}(\lambda)$ is less biased, it is not entirely bias-free. One of the possible sources of bias of the estimator $\hat{\theta}^{BR}(\lambda)$ is the restriction that the estimator should be connected at each change point.

MSE shows patterns similar to that of bias. However, contrary to the bias patterns, in some cases, for example under the combination of $\sigma = 0.2$ and $\lambda = 10$ in scenario 1, MSE of the bias-reduced estimator is larger than that of the conventional ℓ_1-trend filter. For optimal tuning parameters, MSEs of the bias-reduced estimators are consistently better than those of the ℓ_1-trend filters for all the scenarios and noise levels.

5. Conclusion and discussion

The ℓ_1-trend filtering can be considered as an extension of the fused lasso (Tibshirani *et al.*, 2005) which is used for the estimating piecewise constant mean models. Although the piecewise linear trend level and sparse change points set can be found at the same time by the ℓ_1-trend filtering, the resulting estimates are biased. In this paper, we showed the biasedness of the ℓ_1-trend filtered estimates and proposed a bias-reduced procedure based on the change points set of the ℓ_1-trend filtered estimates. The proposed procedure can be easily optimized and we found that the proposed estimator has smaller bias than the original ℓ_1-trend filtering estimator.

The proposed bias reduction method is based on the least squares methods while requiring the estimated line segments to be connected at change points. Thus, as long as the ℓ_1-trend filtering is valid for the irregularly-spaced data, the proposed method can also be applied. The original paper on the ℓ_1-trend filtering (Kim *et al.*, 2009) assumed that the explanatory variables are evenly spaced. However, with an adjusted difference operator, the ℓ_1-trend filtering can be extended to arbitrarily spaced data (Ramdas and Tibshirani, 2016). Thus, we can expect that the proposed bias reduction procedure can also be applied to arbitrarily spaced observations.

References

Guntuboyina A, Lieu D, Chatterjee S, and Sen B (2020). Adaptive risk bounds in univariate total variation denoising and trend filtering, *The Annals of Statistics*, **48**, 205–229.

Hodrick RJ and Prescott EC (1997). Postwar US business cycles: An empirical investigation, *Journal of Money, Credit, and Banking*, **29**, 1–16.

Kim SJ, Koh K, Boyd S, and Gorinevsky D (2009). ℓ-1 trend filtering, *SIAM Review*, **51**, 339–360.

Ramdas A and Tibshirani RJ (2016). Fast and flexible ADMM algorithms for trend filtering, *Journal of Computational and Graphical Statistics*, **25**, 839–858.

Rojas CR and Wahlberg B (2015). How to monitor and mitigate stair-casing in l1 trend filtering. In *Proceedings of 2015 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, South Brisbane, QLD, Australia, 3946–3950.

Selvin S, Ajay SG, Gowri BG, Sowmya V, and Soman KP (2016). ℓ_1-trend filter for image denoising, *Procedia Computer Science*, **93**, 495–502.

Tibshirani R, Saunders M, Rosset S, Zhu J, and Knight K (2005). Sparsity and smoothness via the fused lasso, *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, **67**, 91–

108.

Tibshirani RJ and Taylor J (2011). The solution path of the generalized lasso, *Annals of Statistics*, **39**, 1335–1371.

Tibshirani RJ (2014). Adaptive piecewise polynomial estimation via trend filtering, *The Annals of Statistics*, **42**, 285–323.

Yamada H and Jin L (2013). Japan's output gap estimation and ℓ_1-trend filtering, *Empirical Economics*, **45**, 81–88.

Yu D, Lim J, and Son W (2022). An empirical evidence of inconsistency of the ℓ_1-trend filtering in change point detection, *Korean Journal of Applied Statistics*, **35**, 371–384.

Received August 11, 2022; Revised November 27, 2022; Accepted December 26, 2022