AIMS Mathematics, 10(3): 4662–4708.
DOI: 10.3934/math.2025216
Received: 19 December 2024
Revised: 19 February 2025
Accepted: 24 February 2025
Published: 04 March 2025

https://www.aimspress.com/journal/Math

Research article

Bayesian multiple changing-points detection

Sang Gil Kang[1], **Woo Dong Lee**[2] **and Yongku Kim**[3,4,*]

[1] Department of Data Science, Sangji University, Wonju, Korea
[2] Department of Self-Design Convergence, Daegu Haany University, Gyeongsan, Korea
[3] Department of Statistics, Kyungpook National University, Daegu, Korea
[4] KNU G-LAMP Research Center, Institute of Basic Sciences, Kyungpook National University, Daegu, Korea

* **Correspondence:** Email: kim.1252@knu.ac.kr; Tel: +82539505365; Fax: +82539505363.

Abstract: This study investigated the application of Bayesian multiple change-point detection techniques in the context of piecewise polynomial signals. Given the limited number of existing methodologies for identifying change-points in such signals, we proposed an objective Bayesian change-point detection approach that accommodated heterogeneous error distributions. Our methodology was grounded in a piecewise polynomial regression framework and employed binary segmentation. Initially, we identified change-points across various signals using Bayesian binary segmentation. Subsequently, we applied Bayesian model selection to ascertain the most suitable polynomial order for the identified segments. This approach facilitated a change-point detection method that minimized reliance on subjective inputs. We incorporated intrinsic priors that allowed for the formulation of Bayes factors and model selection probabilities. To evaluate the efficacy of the proposed change-point detection techniques, we conducted a simulation study alongside two empirical case studies: one involving the Goddard Institute for space studies surface temperature analysis and the other concerning the daily closing stock prices of Samsung Electronics Co.

Keywords: binary segmentation; change-points detection; model selection; piecewise polynomial signals
Mathematics Subject Classification: 62F10, 62N01, 62N02

Abbreviations:

$\mathbf{y} = (y_1, \cdots, y_n)$: observation vector; K: number of change-points; $\boldsymbol{\tau} = (\tau_1, \cdots, \tau_K)$: change-points location vector; $I_i = (\tau_{i-1}, \tau_i]$: change-points interval; $\mathbf{y}_i = (y_{\tau_{i-1}+1}, \cdots, y_{\tau_i})$: observation vector in change-points interval I_i; θ_i: parameter vector of model M_i; $\pi^N(\theta_i)$: noninformative prior for θ_i of

model M_i; $\pi^I(\theta_j|\theta_i)$: the conditional intrinsic prior of θ_j for each point θ_i; $\pi^I(\theta_j)$: the unconditional intrinsic prior of θ_j; $m_i(\mathbf{y})$: marginal distribution in model M_i; $B^I_{ij}(\mathbf{y})$: Bayes factor with the intrinsic priors $\{\pi^N(\theta_i), \pi^I(\theta_j)\}$ for models M_i and M_j; $P(M_i)$: prior probability of model M_i; $P(M_i|\mathbf{y})$: posterior probability of model M_i; λ: threshold criterion; MSE: mean squared error; MSR: mean squared residual; $\hat{K}$: estimated number of change-points; $\hat{\boldsymbol{\tau}} = (\hat{\tau}_1, \cdots, \hat{\tau}_{\hat{K}})$: estimated change-points location vector; d_H: Hausdorff distance

1. Introduction

The issue of change-point detection has been the subject of investigation by scholars across multiple disciplines, including economics, climatology, oceanography, biosciences, and technology, for several decades (Bai and Perron [1]; Killick et al. [2]; Futschik et al. [3]; Galceran et al. [4]). The primary goal of change-point detection is to identify the location and number of structural changes in a given sequence of observations by time or location, as well as to provide an assessment of accuracy.

The most fundamental model for this problem is represented by

$$y_i = f_i + \epsilon_i, \quad i = 1, \cdots, n, \tag{1.1}$$

where the observations y_i, for $i = 1, \cdots, n$, are composed of a deterministic and unknown signal f_i with structural changes at certain points, plus a set of independent and identically distributed Gaussian random variables $\epsilon_1, \cdots, \epsilon_n$ with mean zero and finite variance σ^2.

The canonical multiple change point problem, in which the signal f_i is represented as a piecewise constant function, has been extensively examined in the academic literature, leading to the development of numerous methodologies. Frequently employed techniques involve the optimization of a fitness or cost function, such as log-likelihood or the sum of squared errors, across all potential segmentations. Additionally, model selection strategies are utilized to mitigate the risk of overfitting. In this domain, Yao and Au [5, 6] employed the Bayesian information criterion (BIC), while Zhang and Siegmund [7, 8] utilized a modified variant of this criterion. Bai and Perron [9] introduced a method for estimating structural breaks concurrently by minimizing the residual sum of squares. Furthermore, Braun et al. [10] expanded the BIC methodology to encompass a broader range of distributions in which the variance is proportional to a function of the mean. Additionally, Boysen et al. [11] applied the least squares criterion, incorporating a linear penalty for the number of change-points. The fused Lasso technique, as described by Tibshirani and Wang [12], incorporates an L_1 penalty into the least squares component, which serves to penalize the variations between consecutive segment means. Harchaoui and Lévy-Leduc [13] utilized the least absolute shrinkage and selection operator algorithm to determine the positions of multiple change-points within a one-dimensional piecewise constant signal. Additionally, Jeng et al. [14] introduced a likelihood ratio selection method aimed at identifying signal segments, demonstrating that this approach yields consistent estimations. Frick et al. [15] estimated the unknown step function by minimizing the number of change points in the receptive region of a multi-scale test. Killick et al. [16] and Maidstone et al. [17] proposed methods to optimize the penalized cost function. To address the computational complexity of change-point configurations, strategies such as binary segmentation (Olshen et al. [18]; Venkatraman and Olshen [19]), dynamic programming algorithms (Jackson et al. [20]; Killick et al. [16]; Haynes et al. [21]; Maidstone et al. [17]), and screening and ranking algorithms (Hao

et al. [22]) have been suggested. For the Bayesian inference of change-points, several Markov chain Monte Carlo (MCMC) and direct simulation methods have been proposed. Several MCMC procedures were developed to approximate the posterior distribution of the change point locations (Barry and Hartigan [23]; Chib [24]; Lavielle and Lebarbier [25]; Fearnhead and Clifford [26]; Koop and Potter [27]; Fearnhead and Liu [28]; Giordani and Kohn [29]; Martinez and Mena [30]). Alternatively, Fearnhead [31] and Ruggieri [32] avoided the MCMC approaches by directly sampling from the posterior distribution through the use of dynamic programming-like recursions.

Beyond the canonical change point problem, signals in which f_i is modeled as a piecewise polynomial of order $k \geq 1$ have received less attention in the literature despite many applications, such as monitoring patient health (Aminikhanghahi and Cook [33]), finance (Bianchi et al. [34]) and climate change (Robbins et al. [35]). Bai and Perron [9] proposed a method based on Wald-type sequential tests, while Maidstone et al. [17] devised a dynamic programming applied to an l_0-penalized least square procedure. Kim et al. [36] developed a methodology called l_1-trend filtering in continuous piecewise linear models. Subsequently, Baranowski et al. [37] put forward the narrowest over threshold (NOT) method, and Anastasiou and Fryzlewicz [38] developed an isolate-detect (ID) approach, which both provide asymptotically consistent estimators of the number and locations of change points. Maeng and Fryzlewicz [39] then proposed the TrendSegment methodology for detecting multiple change-points corresponding to linear trend changes or point anomalies. In particular, Mehrizi and Chenouri [40] proposed a change point detection method, pattern recovery using trend filtering, based on trend filtering for identifying change points in piecewise polynomial signals. Gavioli-Akilagun and Fryzlewicz [41] proposed a change-points detection procedure based on performing local tests at a number of scales and locations on a sparse grid, which adapts to the choice of grid in the sense that by choosing a sparser grid one explicitly pays a lower price for multiple testing in piecewise polynomial model. On the other hand, Faulkner and Minin [42] investigated a more aggressive horseshoe prior, which demonstrated superior local adaptivity to abrupt changes or jumps. Kowal et al. [43] proposed dynamic shrinkage processes for Bayesian trend filtering with even stronger localized adaptivity to irregular features through modeling dependence between the local scale parameters. Zhao et al. [44] investigated two accelerated primal-dual mirror dynamical approaches for smooth and nonsmooth convex optimization problems. Shirilord and Dehghan [45] proposed an enhancement to the convergence rate of Landweber's method by incorporating the concept of momentum acceleration in least squares problems.

A crucial condition in most of the studies cited above, with the exception of Baranowski et al. [37], is the assumption of a constant variance σ^2 in (1.1). However, this assumption is often violated in many applications, such as the analysis of array of CGH (comparative genomic hybridization) data (Arlot and Celisse [46]; Muggeo and Adelfio [47]), or economic applications, where the real interest rate is modeled by Bai and Perron [1] as a piecewise linear regression with covariates and heterogeneous noise (see Pein et al. [48]). In this paper, we aim to develop a unified method for objective Bayesian change-point detection in a general setting, where there can be multiple types of features in a signal and the features can have different variances from one another. Our method is capable of detecting change-points in piecewise polynomial signals of order k (where $k = 0, 1, 2, \cdots$) with and without continuity constraints at the locations of change points, as well as with heterogeneous variances. Most of the existing methods have been developed under the assumption of homogeneous variance, and therefore may not perform well under heterogeneous variances.

The problem of detecting change-points is equivalent to a model selection problem, so we propose to develop a method using a Bayesian model selection procedure in the presence of heterogeneous Gaussian error noise and without requiring subjective input. To this end, we construct intrinsic priors which allow us to define Bayes factors (BFs) and model selection probabilities. The intrinsic priors are based on the work of Berger and Pericchi [49] and Moreno et al. [50] and are motivated and justified for model selection by Berger and Pericchi [51] and Moreno [52]. The Bayesian change-point detection methodology we propose comprises two primary components. The first component involves identifying change-points within the underlying signals through the application of a piecewise polynomial model utilizing Bayesian binary segmentation. The second component entails the selection of the most appropriate polynomial model order for the identified segments, achieved through Bayesian model selection. To facilitate this process, we incorporate intrinsic priors that ensure the BFs and model selection probabilities are rigorously defined. Furthermore, our approach demonstrates consistency when the sample size is sufficiently large, as it relies on BFs informed by these intrinsic priors. This methodology is versatile and can be applied to a diverse array of signal types, including those that may or may not adhere to continuity constraints at the change-point locations, signals characterized by heterogeneous variances, and mixed signals that exhibit both linear and nonlinear behaviors—challenges that existing methods fail to address. Additionally, our approach yields the most plausible change trends for each interval surrounding the detected change-points.

2. Bayesian multiple change-points detection methods

For a given sequence of data, assume that there are K change-points located at the vector

$$\boldsymbol{\tau} = (\tau_1, \cdots, \tau_K),$$

where $\tau_i < \tau_j$ if $i < j$, and

$$\tau_0 = 0 \quad \text{and} \quad \tau_{K+1} = n.$$

Thus, the K change-points will divide the data into $K+1$ segments, with the ith segment spanning from time $\tau_{i-1} + 1$ to time τ_i. Each segment follows a k-order polynomial structure, with optional continuity constraints at the change-points.

2.1. Binary segmentation

Binary segmentation, initially introduced by Scott and Knott [53], is the predominant method employed for the identification of multiple change-points. This technique utilizes a singular change-point detection approach and entails a recursive examination of the entire dataset to identify a change-point. Upon the identification of a change-point, the dataset is divided into two subsegments delineated by the change-point. The change-point detection method is subsequently applied to these two subsegments, and this iterative process continues until no additional change-points are identified.

The binary segmentation procedure has been found to be consistent under certain regularity conditions for the number and locations of change-points (Vostrikova [54]). Its advantages include low computational complexity $O(n \log n)$, simplicity in its structure, and ease of coding, even for more sophisticated models (Fryzlewicz [55]). As stated by Killick et al. [16], it can be considered as one of the most widely used change-point search methods.

One issue with binary segmentation is that its power to detect short segments in the data sequence can be low. To address this, a number of adaptations have been proposed: circular binary segmentation (Olshen et al. [18]), wild binary segmentation (Fryzlewicz [55]), NOT (Baranowski et al. [37]) and ID (Anastasiou and Fryzlewicz [38]).

2.2. Bayesian change-points detection method

We propose an objective Bayesian detection method based on binary segmentation. Our process consists of three steps. In the first step, binary segmentation is applied to the entire data sequence to obtain a location vector

$$\boldsymbol{\tau} = (\tau_1, \cdots, \tau_K)$$

of the true change-points. In the second step, we must verify whether the change-point intervals detected by the binary segmentation are truly distinct. This is necessary because signals with frequent change-points and various features may lead to false positives in the binary segmentation. If the signals of the adjacent change-point intervals are the same, then false change-points among the detected change-points can be eliminated. Finally, in each of the change-point intervals found, the order of the polynomial model that has the most appropriate trend for the interval is determined. Thus, the Bayesian detection method can be summarized as follows:

Step 1 (Detection of change-points): First, we apply binary segmentation. Then, we formulate the single change-point detection problem as a hypothesis testing problem. For the given point j, we test the null hypothesis H_0 that the observed data y_i, $i = 1, \cdots, n$, follows a distribution f_0, against the alternative hypothesis H_1 that the data follows two separate distributions f_1 and f_2 for the intervals

$$i = 1, \cdots, j \quad \text{and} \quad i = j+1, \cdots, n,$$

respectively. We calculate the BFs $B_{10}(j)$ for all j, and determine

$$B_M = \max_{1<j<n} B_{10}(j).$$

If B_M exceeds the threshold λ, then we reject the null hypothesis of no change. We then split the sequence into two subsequences according to the detected change-point and test them for additional single change-points. We repeat this process until no further subsequences have change-points.

Step 2 (Merge): Second, for the detected change-point intervals in Step 1, we need to check whether the adjacent two change-point intervals are really different or not. Let

$$\boldsymbol{\tau} = (\tau_1, \cdots, \tau_K)$$

be the detected change-point locations. That is, for the given change-point locations τ_{i-1}, τ_i and τ_{i+1}, we consider the null hypothesis

$$H_0 : y_j \sim f_0, \quad j = \tau_{i-1}+1, \cdots, \tau_{i+1},$$

against the alternative

$$H_1 : y_j \sim \begin{cases} f_1, & j = \tau_{i-1}+1, \cdots, \tau_i, \\ f_2, & j = \tau_i+1, \cdots, \tau_{i+1}, \end{cases}$$

where $i = 1, \cdots, K$. Then, we compute the BF $B_{10}(i)$ to assess the evidence of the alternative hypothesis. If $B_{10}(i)$ does not exceed the threshold λ, then the null hypothesis is accepted and the two change-point intervals combine into one change-point interval. This process is repeated until no further change-point intervals combine into one change-point interval. This step is used to remove the false positives in the detected change-points by binary segmentation.

Step 3 (Determiantion of order of polynomial model): Finally, for each change-point interval identified in Step 2, we determine the optimal polynomial model that best captures the trend of the signal in that interval.

2.3. The decision criteria for the detection of change-points

In general, Bayesian testing uses a threshold of $\lambda = 1$ to divide a sequence of data into two subsequences when the BF $B_M > 1$. However, due to the recursive nature of binary segmentation, multiple change-points may be divided into many small and noisy segments, which can lead to false positives. This can cause difficulty in change-point detection, as each false positive leads to further subdivision of the data, creating more chances for false positives, whereas each false negative terminates the investigation of a particular segment (Jensen [56]).

The traditional interpretations of BFs that are in the ranges of

$$1 < B_M < 3, \;\; 3 < B_M < 10, \;\; 10 < B_M < 30, \;\; \text{and} \;\; B_M > 100$$

are considered to be of no more than a bare mention, substantial, strong, and decisive, respectively. The selection of a decision criterion is dependent upon the main goal of the analysis; for primarily descriptive models, a criterion as low as 3 may be acceptable, whereas for a highly rigorous theoretical test, a criterion as high as 100 may be required. For further guidance on the selection of a decision criterion, we refer to the work of Kass and Raftery [57].

We propose using the decision criterion $\bar{B}$, which is the mean of the computed BFs in a given sequence for detecting a change-point, as an alternative to the decision criterion $B_M > \lambda$. The result of averaging of BFs is largely unaffected by the situation in which the sequence has the segments with some noisy observations. This can prevent the problem of overfitting the true number of change-points in the sequence with false positives. Therefore, we use both decision criteria, $\bar{B} > \lambda$ and $B_M > \lambda$, in our work.

We investigate the choice of the threshold λ for Bayesian change-point detection in terms of the mean square error (MSE), mean square residual (MSR), and accuracy of the estimated change-points through a simulation study in the next section. We consider the decision criterion for λ that minimizes MSE and MSR, and yields the exact number and locations of the change-points.

3. Bayesian method for detection of change-points

We consider the observations $y_1, \cdots, y_n$ from the model

$$y_i = f_i(x_i) + \epsilon_i, \quad i = 1, \cdots, n. \tag{3.1}$$

where $y_i \in R$ are response, f_i is a deterministic and piecewise-polynomial signal containing K change-points, $x_i \in R$ are fixed input points, and the error ϵ_i are independent errors having normal distribution

with mean 0 and variance σ_i^2. It can be assumed that the domain of input points is the compact unit interval [0, 1], that is,

$$x_i = i/n$$

for $i = 1, \cdots, n$.

Our Bayesian change-point detection methodology involves identifying change-points through the application of piecewise polynomial regressions and ascertaining the order of each polynomial regression within the identified change-point intervals. The initial phase of the Bayesian change-point detection process entails the identification of change-points, which delineate $K + 1$ polynomial regressions defined across $K + 1$ intervals, denoted as $I_1, \cdots, I_{K+1}$. Let

$$\boldsymbol{\tau} = (\tau_1, \cdots, \tau_K)$$

represent the locations of change-points, and define the intervals of these change-points as

$$I_r = (\tau_{r-1}, \tau_r]$$

for $r = 1, \cdots, K$, with the conditions

$$\tau_0 = 0 < \tau_1 < \tau_2 < \cdots < \tau_K < \tau_{K+1} = n.$$

This framework establishes a partition of the observations into $K + 1$ segments, denoted by lengths $n_1, \cdots, n_{K+1}$. Consequently, we can articulate our piecewise polynomial regression model as follows: for each $y_j \in \mathbf{y}_i$, $j = 1, \cdots, n$,

$$\mathbf{y}_i = \mathbf{X}_i\boldsymbol{\beta}_i + \boldsymbol{\epsilon}_i, i = 1, \cdots, K + 1, \tag{3.2}$$

where

$$\mathbf{y}_i = (y_{\tau_{i-1}+1}, \cdots, y_{\tau_i}),$$

$\mathbf{X}_i$ is the $n_i \times k$ design matrix and is given by

$$\mathbf{X}_i = \begin{pmatrix} 1 & x_{\tau_{i-1}+1} & x^2_{\tau_{i-1}+1} & \cdots & x^{k-1}_{\tau_{i-1}+1} \\ 1 & x_{\tau_{i-1}+2} & x^2_{\tau_{i-1}+2} & \cdots & x^{k-1}_{\tau_{i-1}+2} \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ 1 & x_{\tau_i} & x^2_{\tau_i} & \cdots & x^{k-1}_{\tau_i} \end{pmatrix}, \tag{3.3}$$

$\boldsymbol{\beta}_i$ is $k \times 1$ vector of the regression coefficients, and $\boldsymbol{\epsilon}_i$ is an error vector distributed as

$$\boldsymbol{\epsilon} \sim N_{n_i}(\mathbf{0}, \sigma_i^2\mathbf{I}_{n_i}).$$

The second part is finding the most reasonable model by determining the order k of the polynomial regression model in the given intervals $I_i, i = 1, \cdots, K + 1$.

3.1. Bayesian testing for the detection of a change-point

Bayesian change-point detection initiates with the identification of change-point locations, denoted as

$$\boldsymbol{\tau} = (\tau_1, \cdots, \tau_K),$$

where

$$1 < \tau_1 < \cdots < \tau_K.$$

These change-points delineate $K+1$ segments, each characterized by distinct signals. We postulate that the random variable Y_i follows a normal distribution with a mean of $\mathbf{X}_i\boldsymbol{\beta}_i$ and a variance of $\sigma_i^2\mathbf{I}_i$ within the intervals $(\tau_{i-1}, \tau_i]$ for $i = 1, \cdots, K+1$. Our approach to detecting intervals with varying signals is grounded in Bayesian hypothesis testing. Specifically, we establish intrinsic priors that ensure the BFs and model selection probabilities are well-defined. The detection methodology primarily employs binary segmentation, which serves to identify the locations of intervals exhibiting different signals.

Let τ denote a location characterized by the presence of distinct signals. The vector

$$\mathbf{y}_1 = (y_1, \cdots, y_\tau)^T$$

represents independent random samples drawn from the normal distribution $N(\mathbf{X}_1\boldsymbol{\beta}_1, \sigma_1^2\mathbf{I}_1)$, while

$$\mathbf{y}_2 = (y_{\tau+1}, \cdots, y_n)^T$$

constitutes independent random samples from the normal distribution $N(\mathbf{X}_2\boldsymbol{\beta}_2, \sigma_2^2\mathbf{I}_2)$. Consequently, the problem of location detection can be reformulated as a hypothesis testing problem involving the null hypothesis

$$H_0 : \boldsymbol{\beta}_1 = \boldsymbol{\beta}_2$$

and the alternative hypothesis

$$H_2 : \boldsymbol{\beta}_1 \neq \boldsymbol{\beta}_2.$$

The two statistical models under comparison, as specified by the hypotheses, are

$$M_1 : N_n(\mathbf{y}|\mathbf{X}\boldsymbol{\beta}, \sigma^2\mathbf{I}_n), \pi^N(\theta_1) = \frac{c_1}{\sigma} \tag{3.4}$$

and

$$M_2 : N_\tau(\mathbf{y}_1|\mathbf{X}_1\boldsymbol{\beta}_1, \sigma_1^2\mathbf{I}_\tau)N_{n-\tau}(\mathbf{y}_2|\mathbf{X}_2\boldsymbol{\beta}_2, \sigma_2^2\mathbf{I}_{n-\tau}), \pi^N(\theta_2) = \frac{c_2}{\sigma_1\sigma_2}, \tag{3.5}$$

where $\mathbf{X}_1$ is defined as the $\tau \times k$ design matrix, and $\mathbf{X}_2$ is the $(n-\tau) \times k$ design matrix within the overall design matrix $\mathbf{X}$. The vectors $\boldsymbol{\beta}_i$ (for $i = 1, 2$) represent the $k \times 1$ regression coefficient vectors, with

$$\theta_1 = (\boldsymbol{\beta}, \sigma)$$

and

$$\theta_2 = (\boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \sigma_1, \sigma_2)$$

denoting the respective parameters associated with these models.

3.2. Intrinsic priors

Model (3.4) is a subset of (3.5). It is evident that a theoretically minimal training sample consists of the random vectors $\mathbf{z}_1$ and $\mathbf{z}_2$, each of dimension $k+1$. In the context of Model M_1, $\mathbf{z}_1$ follows a normal distribution $N_{k+1}(\mathbf{Z}_1\boldsymbol{\beta}, \sigma^2\mathbf{I}_{k+1})$, while $\mathbf{z}_2$ adheres to a normal distribution $N_{k+1}(\mathbf{Z}_2\boldsymbol{\beta}, \sigma^2\mathbf{I}_{k+1})$. Conversely, under Model M_2, $\mathbf{z}_1$ is characterized by a normal distribution $N_{k+1}(\mathbf{Z}_1\boldsymbol{\beta}_1, \sigma_1^2\mathbf{I}_{k+1})$ and $\mathbf{z}_2$ by a normal distribution $N_{k+1}(\mathbf{Z}_2\boldsymbol{\beta}_2, \sigma_2^2\mathbf{I}_{k+1})$. It is important to note that $\mathbf{Z}_1$ and $\mathbf{Z}_2$ denote the unknown design matrices of dimensions $(k+1)\times k$ associated with $\mathbf{z}_1$ and $\mathbf{z}_2$ in each respective model (refer to Casella and Moreno [58]). The intrinsic prior, as established in the following theorem, is derived from the works of Berger and Pericchi [49] and Moreno et al. [50].

Theorem 1. For each point θ_1, the intrinsic prior of θ_2 under the minimal training sample $\mathbf{z}$ is

$$\pi^I(\theta_2|\theta_1) = \prod_{i=1}^{2} \frac{2}{\pi}\frac{1}{\sigma}\left(1+\frac{\sigma_i^2}{\sigma^2}\right)^{-1} N_k(\boldsymbol{\beta}_i|\boldsymbol{\beta}, (\sigma_i^2+\sigma^2)\mathbf{W}_i^{-1}), \tag{3.6}$$

where

$$\mathbf{W}_i = \mathbf{Z}_i^T\mathbf{Z}_i, \quad i = 1, 2.$$

Proof. See Appendix A. □

It is observed that two closely related forms have been suggested for the evaluation of the matrix $\mathbf{W}_i^{-1}$ (Casella and Moreno [58]; Girón et al. [59]). Although both essentially give a similar posterior answer, the computationally simpler form is given in Girón et al. [59] as

$$\mathbf{W}_i^{-1} = \frac{n_i}{k+1}(\mathbf{X}_i^T\mathbf{X}_i)^{-1}, \quad i = 1, 2, \tag{3.7}$$

where

$$n_1 = \tau \quad \text{and} \quad n_2 = n - \tau.$$

Notice that the mean of the conditional intrinsic prior for $\boldsymbol{\beta}_i$ depends on the model M_1, but the matrix $\mathbf{W}_i^{-1}$ depends on the design matrix of the $\mathbf{X}_1$ and $\mathbf{X}_2$ of model M_2. The model M_1 plays the role of the null hypothesis and the intrinsic prior for the parameters of the model M_2 is centered at the null, which seems a natural requirement for a sharp null hypothesis (Casella and Moreno [58]). The marginal distribution of $\boldsymbol{\beta}_i$ is a proper prior that does not have moments. This implies that the intrinsic prior for $\boldsymbol{\beta}_i$, conditional on the null $(\boldsymbol{\beta}, \sigma)$, has a heavy tail as expected for a default prior.

The unconditional intrinsic prior for $(\boldsymbol{\beta}_i, \sigma_i)$ obtained by integrating out $\boldsymbol{\beta}$ and σ is

$$\pi^I(\boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \sigma_1, \sigma_2) = \int\int \frac{c_1}{\sigma}\prod_{i=1}^{2}\frac{2}{\pi}\frac{\sigma}{\sigma_i^2+\sigma^2} N_k(\boldsymbol{\beta}_i|\boldsymbol{\beta}, (\sigma_i^2+\sigma^2)\mathbf{W}_i^{-1}) d\boldsymbol{\beta} d\sigma. \tag{3.8}$$

The pair $\{\pi^N(\theta_1), \pi^I(\theta_2)\}$ is called the intrinsic priors for comparing models M_1 and M_2 and, although they are improper, they are well calibrated since both depend on the same arbitrary constant c_1. Further, they are a well established limit of proper priors (Moreno et al. [50]).

3.3. The BF

We aim to test whether a location τ is the point of divergence for various signals by utilizing the BF. Consequently, we calculate the BFs employing intrinsic priors. For a given data $(\mathbf{y}, \mathbf{X})$, the BF with the intrinsic priors $\{\pi^N(\theta_1), \pi^I(\theta_2)\}$ is

$$B_{12}^I(\mathbf{y}) = \frac{\int N_n(\mathbf{y}|\mathbf{X}\boldsymbol{\beta}, \sigma^2\mathbf{I}_n)\pi^N(\theta_1)d\theta_1}{\int N_\tau(\mathbf{y}_1|\mathbf{X}_1\boldsymbol{\beta}_1, \sigma_1^2\mathbf{I}_\tau)N_{n-\tau}(\mathbf{y}_2|\mathbf{X}_2\boldsymbol{\beta}_2, \sigma_2^2\mathbf{I}_{n-\tau})\pi^I(\theta_2)d\theta_2}, \tag{3.9}$$

where

$$\theta_1 = (\boldsymbol{\beta}, \sigma)$$

and

$$\theta_2 = (\boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \sigma_1, \sigma_2).$$

The resulting BF is given in Theorem 2.

Theorem 2. The BF with the intrinsic priors is

$$\begin{aligned} B_{21}^I(\mathbf{y}) &= \pi^{-2}|\mathbf{X}^T\mathbf{X}^T|^{\frac{1}{2}}[\mathbf{y}^T(\mathbf{I}_n - \mathbf{H}_1)\mathbf{y}]^{\frac{n-k}{2}} \\ &\quad \times \int_0^\infty \int_0^\infty \frac{\omega_1^{-\frac{1}{2}}\omega_2^{-\frac{1}{2}}}{(1+\omega_1)(1+\omega_2)}|\Sigma_1|^{-\frac{1}{2}}|\Sigma_2|^{-\frac{1}{2}}|\Sigma|^{-\frac{1}{2}}\mathbf{H}_2^{-\frac{n-k}{2}}d\omega_1 d\omega_2, \end{aligned} \tag{3.10}$$

where $n_1 = \tau$, $n_2 = n - \tau$,

$$\begin{aligned} \mathbf{H}_1 &= \mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T, \Sigma = \mathbf{X}_1^T\Sigma_1^{-1}\mathbf{X}_1 + \mathbf{X}_2^T\Sigma_2^{-1}\mathbf{X}_2, \\ \mathbf{H}_2 &= \sum_{i=1}^2 \mathbf{y}_i^T\Sigma_i^{-1}\mathbf{y}_i - \left(\sum_{i=1}^2 \mathbf{X}_i^T\Sigma_i^{-1}\mathbf{y}_i\right)^T \Sigma^{-1}\left(\sum_{i=1}^2 \mathbf{X}_i^T\Sigma_i^{-1}\mathbf{y}_i\right), \\ |\Sigma_i| &= \omega_i^{n_i}\left[1 + \frac{n_i}{k+1}\frac{1+\omega_i}{\omega_i}\right]^k, \quad i = 1, 2, \\ \Sigma_i^{-1} &= \frac{1}{\omega_i}\left[\mathbf{I}_{n_i} - \frac{n_i(1+\omega_i)}{(k+1)\omega_i}\left(1 + \frac{n_i(1+\omega_i)}{(k+1)\omega_i}\right)^{-1}\mathbf{X}_i(\mathbf{X}_i^T\mathbf{X}_i)^{-1}\mathbf{X}_i^T\right], \quad i = 1, 2, \\ \mathbf{X}_i^T\Sigma_i^{-1}\mathbf{X}_i &= \frac{1}{\omega_i}\left[1 - \frac{n_i(1+\omega_i)}{(k+1)\omega_i}\left(1 + \frac{n_i(1+\omega_i)}{(k+1)\omega_i}\right)^{-1}\right]\mathbf{X}_i^T\mathbf{X}_i, \quad i = 1, 2, \\ \mathbf{y}_i^T\Sigma_i^{-1}\mathbf{y}_i &= \frac{1}{\omega_i}\left[\mathbf{y}_i^T\mathbf{y}_i - \frac{n_i(1+\omega_i)}{(k+1)\omega_i}\left(1 + \frac{n_i(1+\omega_i)}{(k+1)\omega_i}\right)^{-1}\mathbf{y}_i^T\mathbf{X}_i(\mathbf{X}_i^T\mathbf{X}_i)^{-1}\mathbf{X}_i^T\mathbf{y}_i\right], \quad i = 1, 2 \end{aligned}$$

and

$$\mathbf{X}_i^T\Sigma_i^{-1}\mathbf{y}_i = \frac{1}{\omega_i}\left[1 - \frac{n_i(1+\omega_i)}{(k+1)\omega_i}\left(1 + \frac{n_i(1+\omega_i)}{(k+1)\omega_i}\right)^{-1}\right]\mathbf{X}_i^T\mathbf{y}_i, \quad i = 1, 2.$$

Proof. See Appendix B. □

It is important to note that the intrinsic BF cannot be derived analytically; rather, it necessitates two-dimensional integration. The identification of change points is initiated using the binary segmentation method, which is based on the aforementioned BF.

4. Model selection in the change-point interval

We now aim to identify the most appropriate polynomial regression model within the specified intervals I_j for $j = 1, \cdots, K+1$ as discussed in the previous section. We will proceed under the assumption that the polynomial regression model for a particular interval I can be articulated in the following manner:

$$\mathbf{y} = \mathbf{X}_k \boldsymbol{\beta}_k + \boldsymbol{\epsilon}, \tag{4.1}$$

where $\mathbf{X}_k$ is the $n_j \times k$ design matrix and is given by

$$\mathbf{X}_k = \begin{pmatrix} 1 & x_1 & x_1^2 & \cdots & x_1^{k-1} \\ 1 & x_2 & x_2^2 & \cdots & x_2^{k-1} \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ 1 & x_{n_j} & x_{n_j}^2 & \cdots & x_{n_j}^{k-1} \end{pmatrix}, \tag{4.2}$$

$\boldsymbol{\beta}_k$ is the $k \times 1$ vector of the regression coefficients and $\boldsymbol{\epsilon}$ is an error vector distributed as

$$\boldsymbol{\epsilon} \sim N_{n_j}(\mathbf{0}, \sigma_k^2 \mathbf{I}_{n_j}).$$

Then we find the reasonable order k for the given polynomial regression model. Therefore the models being compared are

$$M_i : N_{n_j}(\mathbf{y}|\mathbf{X}_i\boldsymbol{\alpha}_i, \sigma_i^2\mathbf{I}_{n_j}), \pi_i^N(\theta_i) = \frac{c_i}{\sigma_i}, \quad i = 1, \cdots, k-1 \tag{4.3}$$

and

$$M_k : N_{n_j}(\mathbf{y}|\mathbf{X}_k\boldsymbol{\beta}_k, \sigma_k^2\mathbf{I}_{n_j}), \pi_k^N(\theta_k) = \frac{c_k}{\sigma_k}, \tag{4.4}$$

where the $n_j \times i$ design matrix $\mathbf{X}_i, i = 1, \cdots, k$ are given by

$$\mathbf{X}_1 = \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix}, \quad \mathbf{X}_2 = \begin{pmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_{n_j} \end{pmatrix}, \cdots, \quad \mathbf{X}_k = \begin{pmatrix} 1 & x_1 & x_1^2 & \cdots & x_1^{k-1} \\ 1 & x_2 & x_2^2 & \cdots & x_2^{k-1} \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ 1 & x_{n_j} & x_{n_j}^2 & \cdots & x_{n_j}^{k-1} \end{pmatrix}, \tag{4.5}$$

$\boldsymbol{\alpha}_i, i = 1, \cdots, k-1$, are the $i \times 1$ regression coefficient vectors, $\boldsymbol{\beta}_k$ is the $k \times 1$ regression coefficient vector,

$$\theta_i = (\boldsymbol{\alpha}_i, \sigma_i), \quad i = 1, \cdots, k-1$$

and

$$\theta_k = (\boldsymbol{\beta}_k, \sigma_k).$$

4.1. Intrinsic priors

For the models (4.3) and (4.4), the Bayes factor $B_{ik}^N(\mathbf{y}, \mathbf{X})$ compares the model $\{N_{n_j}(\mathbf{y}|\ \mathbf{X}_i\boldsymbol{\alpha}_i, \sigma_i^2\mathbf{I}_{n_j}), \pi_i^N(\theta_i)\}$ with the full model $\{N_{n_j}(\mathbf{y}|\mathbf{X}_k\boldsymbol{\beta}_k, \sigma_k^2\mathbf{I}_{n_j}), \pi_k^N(\theta_k)\}$. Since the sampling model $N_{n_j}(\mathbf{y}|\mathbf{X}_i\boldsymbol{\alpha}_i, \sigma_i^2\mathbf{I}_{n_j})$ is nested in $N_{n_j}(\mathbf{y}|\mathbf{X}_k\boldsymbol{\beta}_k, \sigma_k^2\mathbf{I}_{n_j})$, we can apply the intrinsic method to derive intrinsic priors for comparing model M_i and M_k, for any i.

It is easily seen that a theoretical minimal training sample is a random vector $\mathbf{z}$ of dimension $k+1$. Under model M_i, $\mathbf{z}$ has a density $N_{k+1}(\mathbf{Z}_i\boldsymbol{\alpha}_i, \sigma_i^2\mathbf{I}_{k+1})$, and under model M_k, $\mathbf{z}$ has $N_{k+1}(\mathbf{Z}_k\boldsymbol{\beta}_k, \sigma_k^2\mathbf{I}_{k+1})$. We note that $\mathbf{Z}_i$ and $\mathbf{Z}_k$ represent $(k+1)\times i$ and $(k+1)\times k$ unknown design matrics associated with $\mathbf{z}$ in each model. Then, the intrinsic prior is given in Theorem 3.

Theorem 3. For each point $(\boldsymbol{\alpha}_i, \sigma_i)$, the intrinsic prior of $(\boldsymbol{\beta}_k, \sigma_k)$ under the minimal training sample $\mathbf{z}$ is

$$\pi_k^I(\boldsymbol{\beta}_k, \sigma_k|\boldsymbol{\alpha}_i, \sigma_i) = \frac{2}{\pi}\frac{\sigma_i}{\sigma_k^2+\sigma_i^2}N_k(\boldsymbol{\beta}_k|\tilde{\boldsymbol{\alpha}}_i, (\sigma_k^2+\sigma_i^2)\mathbf{W}_k^{-1}), \tag{4.6}$$

where

$$\mathbf{W}_k^{-1} = \frac{n_j}{k+1}(\mathbf{X}_k^T\mathbf{X}_k)^{-1}$$

and

$$\tilde{\boldsymbol{\alpha}}_i = (\boldsymbol{\alpha}_i^T, \mathbf{0}^T)$$

is $k\times 1$ vector with $\mathbf{0}$ being the null vector of $k-i$ components.

Proof. See Appendix C. □

The unconditional intrinsic prior for $(\boldsymbol{\beta}_k, \sigma_k)$ obtained by integrating out $\boldsymbol{\alpha}_i$ and σ_i is

$$\pi_k^I(\boldsymbol{\beta}_k, \sigma_k) = \frac{2c_i}{\pi}\int\frac{1}{\sigma_k^2+\sigma_i^2}N_k(\boldsymbol{\beta}_k|\tilde{\boldsymbol{\alpha}}_i, (\sigma_k^2+\sigma_i^2)\mathbf{W}_k^{-1})d\boldsymbol{\alpha}_i d\sigma_i. \tag{4.7}$$

The pair $\{\pi_i^N(\theta_i), \pi_k^I(\theta_k)\}$ is referred to as the intrinsic priors utilized for the comparison of models M_i and M_k. Despite being classified as improper priors, they exhibit strong calibration properties, as both are contingent upon the same arbitrary constant c_i. Moreover, these intrinsic priors represent a well-recognized limit of proper priors, as noted by Moreno et al. [50].

4.2. The BF and the posterior model probability

We identify the most appropriate model by evaluating the posterior model probabilities. Consequently, we calculate the BFs utilizing intrinsic priors. For a given data $(\mathbf{y}, \mathbf{X})$, the BF with the intrinsic priors $\{\pi_i^N(\theta_i), \pi_k^I(\theta_k)\}$ is

$$B_{ik}^I(\mathbf{y}) = \frac{\int N_{n_j}(\mathbf{y}|\mathbf{X}_i\boldsymbol{\alpha}_i, \sigma_i^2\mathbf{I}_{n_j})\pi_i^N(\theta_i)d\theta_i}{\int N_{n_j}(\mathbf{y}|\mathbf{X}\boldsymbol{\beta}_k, \sigma_k^2\mathbf{I}_{n_j})\pi_k^I(\theta_k)d\theta_k}, \tag{4.8}$$

where

$$\theta_i = (\boldsymbol{\alpha}_i, \sigma_i) \quad \text{and} \quad \theta_k = (\boldsymbol{\beta}_k, \sigma_k).$$

The resulting BF is given in Theorem 4.

Theorem 4. The BF with the intrinsic priors is

$$B_{ik}^{I}(\mathbf{y}) = \frac{1}{\int_0^\infty \pi^{-1} z^{-\frac{1}{2}} (1+z)^{-1} \left[1 + \frac{n_j(z+1)}{k+1}\right]^{-\frac{k-i}{2}} \left\{1 + c(z)^{-1} \frac{\mathbf{y}^T(\mathbf{H}_i - \mathbf{H}_k)\mathbf{y}}{\mathbf{y}^T(\mathbf{I}_{n_j} - \mathbf{H}_i)\mathbf{y}}\right\}^{-\frac{n_j - i}{2}} dz}, \tag{4.9}$$

where

$$c(z) = 1 + \frac{k+1}{n_j(1+z)}, \quad \mathbf{H}_i = \mathbf{X}_i(\mathbf{X}_i^T\mathbf{X}_i)^{-1}\mathbf{X}_i^T$$

and

$$\mathbf{H}_k = \mathbf{X}_k(\mathbf{X}_k^T\mathbf{X}_k)^{-1}\mathbf{X}_k^T.$$

Proof. See Appendix D. □

In general, for a dataset denoted as $(\mathbf{y}, \mathbf{X})$, the BF for a specific model M_i, in relation to the model M_k, is defined as the ratio of their respective marginal distributions

$$B_{ik}(\mathbf{y}) = \frac{m_i(\mathbf{y})}{m_k(\mathbf{y})}. \tag{4.10}$$

For

$$P(M_i) = P(M_k) = 1/2,$$

the posterior probability of M_i is

$$Pr(M_i|\mathbf{y}) = \frac{B_{ik}(\mathbf{y})}{1 + B_{ik}(\mathbf{y})}. \tag{4.11}$$

It is observed that the probability $Pr(M_i|\mathbf{y})$ is a monotonically increasing function of the BF $B_{ik}(\mathbf{y})$. Consequently, the ranking of models based on their posterior probabilities

$$\{Pr(M_i|\mathbf{y}), M_i \in \mathcal{M}\}$$

is equivalent to ranking them according to their respective BFs

$$\{B_{ik}(\mathbf{y}), M_i \in \mathcal{M}\},$$

where $\mathcal{M}$ denotes the comprehensive set of models. Moreover, this ordering remains unchanged when the BF is normalized across all models, resulting in a corresponding set of probabilities

$$P(M_i|\mathbf{y}) = \frac{B_{ik}(\mathbf{y})}{1 + \sum_{j=1}^{k-1} B_{jk}(\mathbf{y}, \mathbf{X})}. \tag{4.12}$$

This approach was examined by Casella and Moreno [58]. Subsequently, the posterior probabilities, as indicated in Eq (4.12), are derived from the BF presented in Eq (4.9), thereby enabling the identification of the polynomial regression order. Once the model is established, the Bayes estimator for the regression coefficient vector, utilizing the reference prior, converges to the ordinary least squares estimator for the specified model.

4.3. Consistency of Bayesian procedure for model selection

We prove the consistency of the Bayesian model selection procedure encompassed by the model M_k in Section 4.2. Using the idea of proof in [58, Theorem 3], we obtain the following Lemma 1.

Lemma 1. When the model M_i is nested in M_k, for large values of n_j, the Bayes factor given in (4.9) can be approximated by

$$B^I_{ik}(\mathbf{y}) \approx \pi(k+1)^{\frac{i-k}{2}} I(\mathcal{B}^{n_j}_{ik})^{-1} \exp\left\{\frac{k-i}{2}\log n_j + \frac{n_j - i}{2}\log \mathcal{B}^{n_j}_{ik}\right\}, \tag{4.13}$$

where

$$\mathcal{B}^{n_j}_{ik} = \frac{\mathbf{y}^T(\mathbf{I}_{n_j} - \mathbf{H}_F)\mathbf{y}}{\mathbf{y}^T(\mathbf{I}_{n_j} - \mathbf{H}_i)\mathbf{y}}, \quad \mathbf{H}_i = \mathbf{X}_i(\mathbf{X}_i^T\mathbf{X}_i)^{-1}\mathbf{X}_i^T, \quad \mathbf{H}_k = \mathbf{X}_k(\mathbf{X}_k^T\mathbf{X}_k)^{-1}\mathbf{X}_k^T$$

and

$$I(\mathcal{B}^{n_j}_{ik}) = \int_0^\infty z^{-1/2}(1+z)^{\frac{i-k-2}{2}} \exp\left\{\frac{k+1}{2(1+z)}\left(1 - \frac{1}{\mathcal{B}^{n_j}_{ik}}\right)\right\} dz.$$

Proof. We put

$$n = n_j.$$

Then, we can express the integrand in BF (4.9) of Theorem 4 as

$$\begin{aligned}
&\pi^{-1}z^{-\frac{1}{2}}(1+z)^{-1}\left[1 + \frac{n_j(z+1)}{k+1}\right]^{-\frac{k-i}{2}}\left[1 + c(z)^{-1}\mathcal{B}^{n_j}_{ik}\right]^{-\frac{n-i}{2}} \\
&= \pi^{-1}(1+k)^{\frac{k-i}{2}}z^{-\frac{1}{2}}(1+z)^{-1-\frac{n-i}{2}}[1+k+n(z+1)]^{\frac{n-k}{2}}\left[\frac{1+k}{1+z} + n\mathcal{B}^n_{ik}\right]^{-\frac{n-i}{2}}.
\end{aligned} \tag{4.14}$$

Now we have the following equations:

$$\log[k+1+n(z+1)]^{\frac{n-k}{2}} = \frac{n-k}{2}\left[\log n + \log(z+1) + \log\left(1 + \frac{k+1}{n(z+1)}\right)\right] \tag{4.15}$$

and

$$\log\left(\frac{1+k}{1+z} + n\mathcal{B}^n_{ik}\right)^{-\frac{n-i}{2}} = -\frac{n-i}{2}\left[\log n + \log \mathcal{B}^n_{ik} + \log\left(1 + \frac{k+1}{z+1}\frac{1}{n\mathcal{B}^n_{ik}}\right)\right]. \tag{4.16}$$

Using the Eqs (4.15) and (4.16), the integrand (4.14) is expressed as

$$\begin{aligned}
&\pi^{-1}(1+k)^{\frac{k-i}{2}}z^{-\frac{1}{2}}(1+z)^{-1-\frac{n-i}{2}} \\
&\times \exp\left\{\frac{n-k}{2}\log n + \frac{n-k}{2}\log(z+1) + \frac{n-k}{2}\log\left(1 + \frac{k+1}{n(z+1)}\right)\right\} \\
&\times \exp\left\{-\frac{n-i}{2}\log n - \frac{n-i}{2}\log \mathcal{B}^n_{ik} - \frac{n-i}{2}\log\left(1 + \frac{k+1}{z+1}\frac{1}{n\mathcal{B}^n_{ik}}\right)\right\}.
\end{aligned} \tag{4.17}$$

Note that for large n, we have the following approximations:

$$\left(1+\frac{k+1}{n(z+1)}\right)^{\frac{n-k}{2}} \approx \exp\left\{\frac{(k+1)}{2(z+1)}\right\} \tag{4.18}$$

and

$$\left(1+\frac{k+1}{z+1}\frac{1}{n\mathcal{B}_{ik}^{n}}\right)^{\frac{n-i}{2}} \approx \exp\left\{\frac{k+1}{2(1+z)\mathcal{B}_{ik}^{n}}\right\}. \tag{4.19}$$

Then, the integrand (4.17) can be approximated by

$$\pi^{-1}(1+k)^{\frac{k-i}{2}}z^{-\frac{1}{2}}(1+z)^{\frac{i-k-2}{2}}\exp\left\{\frac{k+1}{2(z+1)}\left(1-\frac{1}{\mathcal{B}_{ik}^{n}}\right)\right\}\times\exp\left\{\frac{i-k}{2}\log n+\frac{i-n}{2}\log\mathcal{B}_{ik}^{n}\right\}. \tag{4.20}$$

Therefore, the BF (4.9) can be approximated by

$$B_{ik}^{I}(\mathbf{y}) \approx \pi(k+1)^{\frac{i-k}{2}}I(\mathcal{B}_{ik}^{n})^{-1}\exp\left\{\frac{k-i}{2}\log n+\frac{n-i}{2}\log\mathcal{B}_{ik}^{n}\right\}, \tag{4.21}$$

where

$$I(\mathcal{B}_{ik}^{n}) = \int_{0}^{\infty} z^{-1/2}(1+z)^{\frac{i-k-2}{2}}\exp\left\{\frac{k+1}{2(1+z)}\left(1-\frac{1}{\mathcal{B}_{ik}^{n}}\right)\right\}dz.$$

Hence, the Lemma 1 is proved. □

Ignoring the positive terms that do not depend on n_j from the posterior probability (4.12) and the asymptotic approximation of (4.13) in Lemma 1, we obtain

$$\begin{aligned}P(M_i|\mathbf{y}) &= \frac{B_{ik}^{I}(\mathbf{y})}{1+\sum_{l=1}^{k-1}B_{lk}^{I}(\mathbf{y})}\\ &= \frac{c_{ik}I(\mathcal{B}_{ik}^{n_j})^{-1}\exp\left\{\frac{k-i}{2}\log n_j+\frac{n_j}{2}\log\mathcal{B}_{ik}^{n_j}\right\}}{1+\sum_{l=1}^{k-1}c_{lk}I(\mathcal{B}_{lk}^{n_j})^{-1}\exp\left\{\frac{k-j}{2}\log n_j+\frac{n_j}{2}\log\mathcal{B}_{lk}^{n_j}\right\}},\end{aligned}$$

and similarly, for the true model M_T we have

$$P(M_T|\mathbf{y}) = \frac{c_{Tk}I(\mathcal{B}_{Tk}^{n_j})^{-1}\exp\left\{\frac{k-T}{2}\log n_j+\frac{n_j}{2}\log\mathcal{B}_{Tk}^{n_j}\right\}}{1+\sum_{l=1}^{k-1}c_{lk}I(\mathcal{B}_{lk}^{n_j})^{-1}\exp\left\{\frac{k-j}{2}\log n_j+\frac{n_j}{2}\log\mathcal{B}_{lk}^{n_j}\right\}},$$

where c_{ik}, c_{lk}, and c_{Tk} do not depend on n_j, and $I(\mathcal{B}_{ik}^{n_j})^{-1}$ and $I(\mathcal{B}_{Tk}^{n_j})^{-1}$ are finite for all n_j. Thus we can obtain the ratio of these two probabilities, which is given by

$$\frac{P(M_i|\mathbf{y})}{P(M_T|\mathbf{y})} \approx \exp\left\{\frac{T-i}{2}\log n_j+\frac{n_j}{2}\log\frac{\mathcal{B}_{ik}^{n_j}}{\mathcal{B}_{Tk}^{n_j}}\right\}.$$

From Theorem 4, the intrinsic model selection procedure encompassed by the model M_k in Section 4.2 is consistent. That is, with the sampling from M_T, we have that

$$\frac{P(M_i|\mathbf{y})}{P(M_T|\mathbf{y})} \to 0, [P_t],$$

whenever the model

$$M_i \neq M_T.$$

5. Numerical study

5.1. Simulation study

We compare the performance of Bayesian change-point detection methods, utilizing the BFs with intrinsic priors, to the existing frequentist change-point detection methods: bottom-up (BUP, Keogh et al. [60]), trend filtering (TF, Kim et al. [36]), NOT (Baranowski et al. [37]), continuous-piecewise-linear pruned optimal partitioning (CPOP, Maidstone et al. [17]), ID (Anastasiou and Fryzlewicz [38]), and TrendSegment (TS, Maeng and Fryzlewicz [39]).

Baranowski et al. [37], Anastasiou and Fryzlewicz [38], and Maeng and Fryzlewicz [39] demonstrated that NOT, ID and Trend Segment outperform other competitors when estimating the number of change-points in simulation studies. CPOP was omitted from our comparison due to its performance being nearly identical to ID, but slightly inferior (Anastasiou and Fryzlewicz [38]). TS, on the other hand, demonstrated much better results than CPOP (Maeng and Fryzlewicz). BUP and TF, in comparison, showed no viable results compared to ID, NOT, TS, or CPOP, and therefore were excluded from our comparison.

We aim to compare the performance of the proposed Bayesian detection methods with their frequentist counterparts, namely NOT detection (Baranowski et al. [37]) implemented in the R package NOT from CRAN, ID (Anastasiou and Fryzlewicz [38]) available in the R package IDetect, and TS (Maeng and Fryzlewicz [39]) implemented in the R package trendsegmentR from CRAN.

We constructed underlying signals with a variety of characteristics following approaches similar to Baranowski et al. [37] and Maeng and Fryzlewicz [39]. We simulated data from model (1) using 8 signals: (M1) wave 1, (M2) wave 2, (M3) mix 1, (M4) mix 2, (M5) teeth-linear, (M6) linear, (M7) wave 3, and (M8) quad, shown in Figure 1. (M1) and (M2) are linear signals at change-points, with (M2) having discontinuities. (M3) and (M4) are a mix of continuous and discontinuous change-points with both constant and linear segments, with (M4) having particularly short segments. (M5) is piecewise linear with a teeth shape. (M6) is a linear signal. (M7) is a mix of continuous and discontinuous linear signals with different slopes and segment sizes. Lastly, (M8) is composed of both constant, linear and quadratic signals with a mix of continuous and discontinuous change-points. The underlying signals are as follows:

(M1) wave 1: length 1500; change-points at 150, 300, 450, 600, 750, 900, 1050, 1200, and 1350; values of the slopes between change-points 3/150, -3/150, 3/150, -3/150, 3/150, -3/150, 3/150,- 3/150, 3/150, -3/150, and values of the intercepts between change-points -1, 5, -7, 11, -13, 17, -19, 23, -25, 29. We use two types of standard deviations in the noises for the heterogeneous case. That is, we use values of standard deviations between change-points 1, 0.5, 1, 0.5, 1, 0.5, 1, 0.5, 1, 0.5 (type 1) and 2, 1, 0.5, 0.25, 0.5, 1, 2, 1, 0.5, 0.25 (type 2). In homogeneous case, we put values of standard deviations to 1.

(M2) wave 2: length 1260; change-points at 60, 120, 180, 240, 300, 360, 420, 480, 540, 600, 660, 720, 780, 840, 900, 960, 1020, 1080, 1140, and 1200; values of the slopes between change-points 0.0625, -0.0625, 0.0625, -0.0625, 0.0625, -0.0625, 0.0625, -0.0625, 0.0625, -0.0625, 0.0625, -0.0625, 0.0625, -0.0625, 0.0625, -0.0625, 0.0625, -0.0625, 0.0625, -0.0625, 0.0625, and values of the intercepts between change-points -1, 7.5625, -8.5, 15.0625, -16, 22.5625, -23.5, 30.0625, -31, 37.5625, -38.5, 45.0625, -46, 52.5625, -53.5, 60.0625, -61, 67.5625, -68.5, 75.0625, -76. We use two

types of standard deviations in the noises in the heterogeneous case. That is, we use values of standard deviations between change-points 1, 0.5, 1, 0.5, 1, 0.5, 1, 0.5, 1, 0.5, 1, 0.5, 1, 0.5, 1, 0.5, 1, 0.5, 1, 0.5, 1 (type 1) and 2, 1, 0.5, 0.25, 0.125, 0.25, 0.5, 1, 2, 1, 0.5, 0.25, 0.125, 0.25, 0.5, 1, 2, 1, 0.5, 0.25, 0.125 (type 2). In homogeneous case, we put values of standard deviations to 1.

(M3) mix 1: length 2048; change-points at 256, 512, 768, 1024, 1280, 1536, and 1792; values of the slopes between change-points 0, 4/256, 0, -4/256, 0, 4/256, -4/256, 4/256, and values of the intercepts between change-points 0, -4, 3, 15, -2, -21, 28, -28. We use two types of standard deviations in the noises for the heterogeneous case. That is, we use values of standard deviations between change-points 1, 0.5, 1, 0.5, 1, 0.5, 1, 0.5 (type 1) and 2, 1, 0.5, 0.25, 0.5, 1, 2, 1 (type 2). In homogeneous case, we put values of standard deviations to 1.

(M4) mix 2: length 2048; change-points at 256, 512, 542, 768, 1024, 1280, 1310, 1536, and 1792; values of the slopes between change-points 0, 0, 8/256, -4/256, 4/256, -4/256, 0, 0, 4/256, -4/256, and values of the intercepts between change-points 2, -2, -12.03125, 9.390625, -15.625, 17.390625, -7.609375, -3.609375, -25.625, 30.375. We use two types of standard deviations in the noises for the heterogeneous case. That is, we use values of standard deviations between change-points 1, 0.5, 1, 0.5, 1, 0.5, 1, 0.5 (type 1) and 2, 1, 0.5, 0.25, 0.5, 1, 2, 1 (type 2). In homogeneous case, we put values of standard deviations to 1.

(M5) teeth-linear: length 800; change-points at 100, 200, 300, 400, 500, 600, and 700; values of the slopes between change-points 5/1000, 5/1000, 5/1000, 5/1000, 5/1000, 5/1000, 5/1000, 5/1000, and values of the intercepts between change-points 1, -1, 0, -2, -1, -3, -2, -4. We use two types of standard deviations in the noises for the heterogeneous case. That is, we use values of standard deviations between change-points 1, 0.5, 1, 0.5, 1, 0.5, 1, 0.5 (type 1) and 2, 1, 0.5, 0.25, 0.5, 1, 2, 1 (type 2). In homogeneous case, we put values of standard deviations to 0.5.

(M6) linear: length 1500; values of slope and intercept are 2/1500 and -1, respectively. In homogeneous case, we put values of standard deviations to 1, and thus this signal is a linear signal without change-points. We use two types of standard deviations in the noises for the heterogeneous case. That is, we use values of standard deviations between change-points 1, 0.5, 1, 0.5 (type 1) and 2, 1, 0.5, 0.25 (type 2). The change-points of changes in variances are located at 300, 700, 1200 and thus this signal is a linear signal with changes of variances.

(M7) wave 3: length 1408; change-points at 256, 512, 768, 1024, 1152, 1280, and 1344; values of the slopes between change-points 2^{-8}, -2^{-6}, 2×2^{-6}, -3×2^{-6}, 4×2^{-6}, -5×2^{-6}, 6×2^{-6}, -7×2^{-6} and values of the intercepts between change-points 1, 4, -16, 40, -70, 92, -126, 147. We use two types of standard deviations in the noises for the heterogeneous case. That is, we use values of standard deviations between change-points 1, 0.5, 0.25, 0.5, 1, 0.5, 0.25, 0.5 (type 1) and 2, 1, 0.5, 0.25, 0.5, 1, 2, 1 (type 2). In homogeneous case, we put values of standard deviations to 1.

(M8) quad: length 1408; change-points at 100, 250, and 500; values of the slopes between change-points 0, 0, 1/100, and values of the intercepts between change-points 0, 2, 2.5 and in the quadratic coefficient 10^{-5}, 0, -5. We use two types of standard deviations in the noises for the heterogeneous case. That is, we use values of standard deviations between change-points 1, 0.5, 1, 0.5 (type 1) and 2, 1, 0.5, 0.25 (type 2). In homogeneous case, we put values of standard deviations to 1.

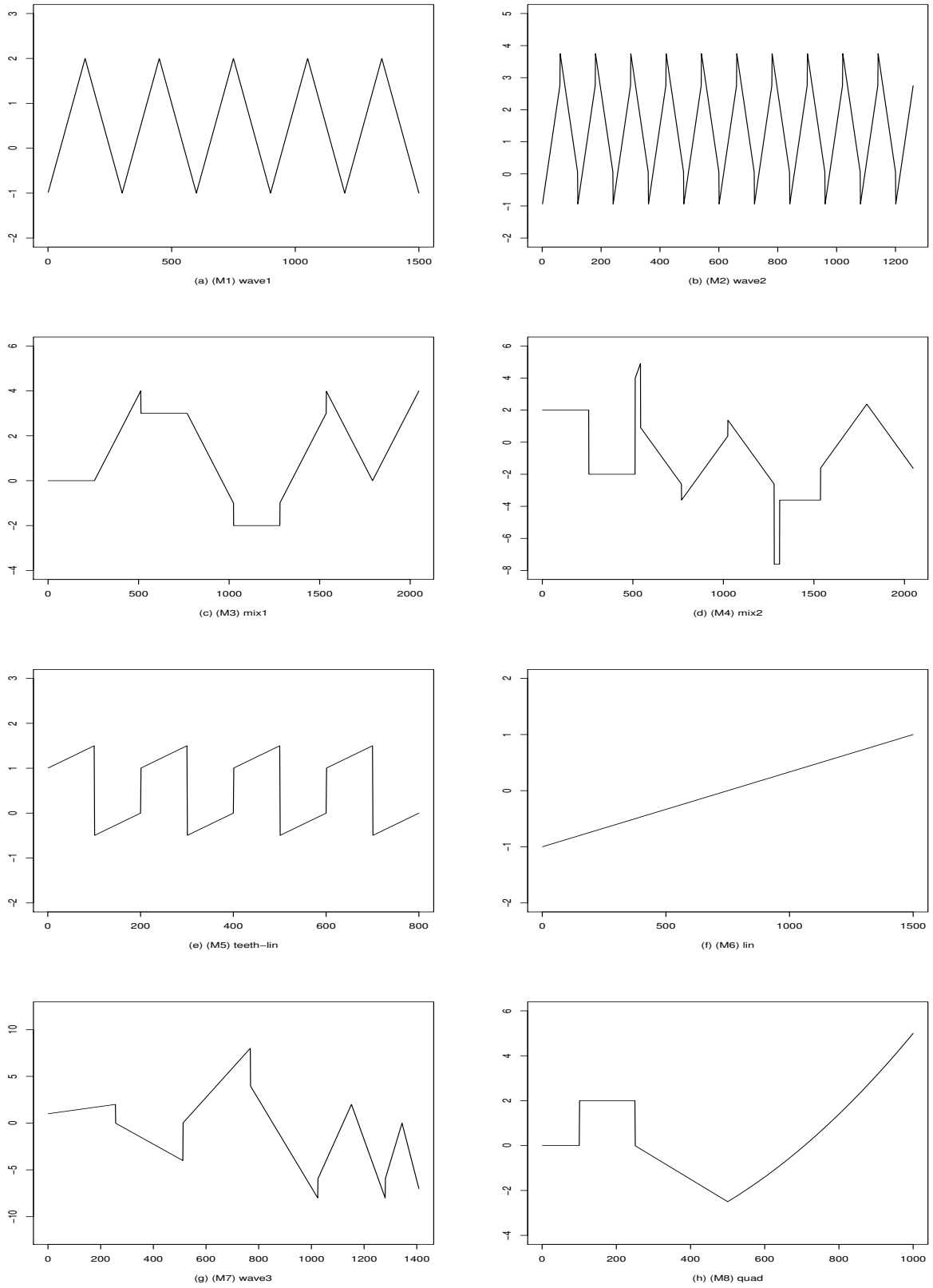

Figure 1. The underlying signals.

For the given true signals, we detect the change-points using the NOT, ID, and TS methods in the corresponding R packages. We use the recommended contrast functions for NOT and ID methods, and the default values for the other tuning parameters.

In Bayesian change-point detection methods, we consider the threshold criteria

$$\lambda = 1, 3, 10, 20, \text{ and } 100$$

following the guidelines of Jeffreys and Kass and Raftery as described in Section 2.3, and use the appropriate order of polynomials for the respective true signals. We denote detection by B_M as BF, and detection by both B_M and $\bar{B}$ as minimum Bayes factor (MBF).

We generated 1000 datasets for each true signal in all the simulations. For each dataset, we calculated the MSE and MSR values for each true signal to evaluate the performance of each method. The MSE and MSR are defined as follows:

$$\begin{aligned} MSE &= \frac{1}{n}\sum_{i=1}^{n}(\hat{f}_i(x_i) - f_i(x_i))^2, \\ MSR &= \frac{1}{n}\sum_{i=1}^{n}(\hat{f}_i(x_i) - y_i)^2, \end{aligned} \tag{5.1}$$

where $\hat{f}_i(x_i)$ is the estimated signal and $f_i(x_i)$ is the true signal between two successive change-points, and y_i is observation. The empirical distribution of the difference between the estimated number of change-points ($\hat{K}$) and the true number of change-points (K) is reported. To gauge the accuracy of the estimated change-point locations ($\hat{\tau}_i$), we present estimates of the scaled Hausdorff distance as

$$d_H = \frac{1}{n}\max\left\{\max_i \min_j |\tau_i - \hat{\tau}_j|, \max_j \min_i |\hat{\tau}_j - \tau_i|\right\}, \tag{5.2}$$

where

$$i = 0, \cdots, K+1$$

and

$$j = 0, \cdots, \hat{K}+1$$

with the convention

$$\tau_0 = \hat{\tau}_0 = 0, \quad \tau_{K+1} = \hat{\tau}_{\hat{K}+1} = n$$

and $\hat{\tau}$ and τ denote estimated and true locations of the change-points. The smaller the Hausdorff distance, the more accurate the estimation of the change-point locations. The summary of the results for all models and methods can be found in Tables 1–6.

Table 1. Type 1 heterogeneous variances: distribution of $\hat{K} - K$ for the various competing methods, the average MSE, the average MSR, and the average Hausdorff distance d_H for models (M1)–(M4).

		$\hat{K} - K$									
Model	Method	≤ -3	-2	-1	0	1	2	≥ 3	MSR	MSE	d_H
(M1)	ID	0	0	0	82	17	1	0	0.6528	0.0234	0.0167
$K = 9$	NOT	0	0	0	100	0	0	0	0.6145	0.0103	0.0070
	TS	0	0	0	58	27	8	7	0.6101	0.0297	0.0236
	BF: $\lambda = 1$	0	0	0	0	0	0	100	0.5910	0.0343	0.0376
	BF: $\lambda = 3$	0	0	0	11	13	21	55	0.6043	0.0214	0.0260
	BF: $\lambda = 10$	0	0	0	44	28	19	9	0.6094	0.0161	0.0183
	BF: $\lambda = 20$	0	0	0	65	22	10	3	0.6108	0.0156	0.0170
	BF: $\lambda = 100$	0	0	0	88	11	1	0	0.6122	0.0160	0.0162
	MBF: $\lambda = 1$	0	0	0	45	21	19	15	0.6091	0.0172	0.0201
	MBF: $\lambda = 3$	0	0	0	70	24	6	0	0.6114	0.0163	0.0173
	MBF: $\lambda = 10$	0	0	0	83	17	0	0	0.6125	0.0175	0.0180
	MBF: $\lambda = 20$	0	0	0	84	16	0	0	0.6128	0.0184	0.0185
	MBF: $\lambda = 100$	0	0	0	92	8	0	0	0.6140	0.0204	0.0194
(M2)	ID	0	0	0	69	21	9	1	0.6702	0.0753	0.0130
$K = 20$	NOT	0	0	0	98	2	0	0	0.6012	0.0621	0.0143
	TS	0	0	0	71	17	9	3	0.6203	0.0722	0.0155
	BF: $\lambda = 1$	0	0	0	0	0	0	100	0.5670	0.0943	0.0220
	BF: $\lambda = 3$	0	0	0	1	4	8	87	0.5988	0.0701	0.0191
	BF: $\lambda = 10$	0	0	0	27	27	21	25	0.6146	0.0657	0.0157
	BF: $\lambda = 20$	0	0	0	48	32	13	7	0.6194	0.0670	0.0155
	BF: $\lambda = 100$	0	0	0	75	22	3	0	0.6279	0.0763	0.0162
	MBF: $\lambda = 1$	0	0	0	4	14	16	66	0.6106	0.0722	0.0181
	MBF: $\lambda = 3$	0	0	0	30	33	24	13	0.6221	0.0735	0.0173
	MBF: $\lambda = 10$	0	0	0	68	19	11	2	0.6302	0.0811	0.0169
	MBF: $\lambda = 20$	0	0	0	76	19	5	0	0.6361	0.0872	0.0173
	MBF: $\lambda = 100$	0	0	0	92	8	0	0	0.6464	0.0995	0.0184
(M3)	ID	0	0	0	31	42	15	12	0.6628	0.0512	0.0206
$K = 7$	NOT	0	0	0	97	3	0	0	0.6148	0.0141	0.0199
	TS	0	0	0	49	15	27	9	0.6183	0.0239	0.0299
	BF: $\lambda = 1$	0	0	0	0	5	3	92	0.6101	0.0164	0.0399
	BF: $\lambda = 3$	0	0	0	20	13	25	42	0.6162	0.0107	0.0280
	BF: $\lambda = 10$	0	0	0	66	16	13	5	0.6196	0.0077	0.0129
	BF: $\lambda = 20$	0	0	0	80	10	7	3	0.6199	0.0074	0.0096
	BF: $\lambda = 100$	0	0	0	96	3	1	0	0.6204	0.0069	0.0059
	MBF: $\lambda = 1$	0	0	0	87	7	5	1	0.6201	0.0072	0.0087
	MBF: $\lambda = 3$	0	0	0	93	5	1	1	0.6203	0.0070	0.0070
	MBF: $\lambda = 10$	0	0	0	97	2	1	0	0.6205	0.0070	0.0060
	MBF: $\lambda = 20$	0	0	0	97	2	1	0	0.6205	0.0070	0.0060
	MBF: $\lambda = 100$	0	0	0	100	0	0	0	0.6206	0.0069	0.0055
(M4)	ID	0	0	0	0	0	0	100	0.6351	0.1037	0.0279
$K = 9$	NOT	0	0	0	94	5	1	0	0.5450	0.0119	0.0105
	TS	0	0	0	30	21	21	28	0.5419	0.0217	0.0233
	BF: $\lambda = 1$	0	0	0	0	0	3	97	0.5362	0.0223	0.0305
	BF: $\lambda = 3$	0	0	0	9	26	24	41	0.5442	0.0145	0.0160
	BF: $\lambda = 10$	0	0	0	39	38	16	7	0.5475	0.0115	0.0078
	BF: $\lambda = 20$	0	0	0	57	33	8	2	0.5482	0.0114	0.0060
	BF: $\lambda = 100$	0	0	0	87	13	0	0	0.5498	0.0110	0.0030
	MBF: $\lambda = 1$	0	0	0	68	23	7	2	0.5483	0.0111	0.0039
	MBF: $\lambda = 3$	0	0	0	78	20	2	0	0.5490	0.0109	0.0033
	MBF: $\lambda = 10$	0	0	0	90	10	0	0	0.5500	0.0116	0.0031
	MBF: $\lambda = 20$	0	0	0	94	6	0	0	0.5501	0.0118	0.0031
	MBF: $\lambda = 100$	0	0	0	96	4	0	0	0.5502	0.0120	0.0031

Table 2. Type 1 heterogeneous variances: distribution of $\hat{K} - K$ for the various competing methods, the average MSE, the average MSR, and the average Hausdorff distance d_H for models (M5)–(M8).

		$\hat{K} - K$									
Model	Method	≤ -3	-2	-1	0	1	2	≥ 3	MSR	MSE	d_H
(M5)	ID	0	0	0	7	49	12	32	0.6728	0.0949	0.1139
$K = 7$	NOT	35	32	19	11	3	0	0	0.6295	0.0574	0.1166
	TS	4	19	32	21	18	3	3	0.6223	0.0621	0.1045
	BF: $\lambda = 1$	0	0	0	0	5	15	80	0.5907	0.0462	0.0292
	BF: $\lambda = 3$	0	0	0	24	35	24	17	0.60283	0.0353	0.0150
	BF: $\lambda = 10$	0	0	0	65	30	4	1	0.60805	0.0318	0.0078
	BF: $\lambda = 20$	0	0	0	79	19	2	0	0.60884	0.0311	0.0069
	BF: $\lambda = 100$	0	0	0	96	4	0	0	0.61071	0.0303	0.0046
	MBF: $\lambda = 1$	0	0	0	87	13	0	0	0.60983	0.0308	0.0061
	MBF: $\lambda = 3$	0	0	0	96	4	0	0	0.61057	0.0303	0.0045
	MBF: $\lambda = 10$	0	0	0	98	2	0	0	0.61061	0.0303	0.0043
	MBF: $\lambda = 20$	0	0	0	99	1	0	0	0.61080	0.0301	0.0043
	MBF: $\lambda = 100$	0	0	0	100	0	0	0	0.61097	0.0300	0.0042
(M6)	ID	98	2	0	0	0	0	0	0.6494	0.0010	0.4640
$K = 3$	NOT	100	0	0	0	0	0	0	0.6495	0.0008	0.4667
	TS	72	1	24	0	3	0	0	0.6460	0.0044	0.4105
	BF: $\lambda = 1$	0	0	0	17	21	27	35	0.6416	0.0118	0.0200
	BF: $\lambda = 3$	0	0	0	54	28	15	3	0.6450	0.0081	0.0147
	BF: $\lambda = 10$	0	0	0	83	16	1	0	0.6467	0.0061	0.0107
	BF: $\lambda = 20$	0	0	0	92	8	0	0	0.6472	0.0057	0.0088
	BF: $\lambda = 100$	0	0	0	100	0	0	0	0.6477	0.0053	0.0079
	MBF: $\lambda = 1$	0	0	0	98	2	0	0	0.6476	0.0054	0.0091
	MBF: $\lambda = 3$	0	0	0	98	2	0	0	0.6476	0.0053	0.0097
	MBF: $\lambda = 10$	0	0	0	99	1	0	0	0.6477	0.0053	0.0093
	MBF: $\lambda = 20$	0	0	0	99	1	0	0	0.6477	0.0053	0.0093
	MBF: $\lambda = 100$	0	3	0	97	0	0	0	0.6476	0.0052	0.0187
(M7)	ID	0	0	0	0	0	4	96	0.6630	0.2729	0.0262
$K = 7$	NOT	0	0	0	81	8	9	2	0.4037	0.0105	0.0104
	TS	0	0	0	1	2	1	96	0.3654	0.0524	0.0555
	BF: $\lambda = 1$	0	0	0	0	5	14	81	0.4008	0.0130	0.0191
	BF: $\lambda = 3$	0	0	0	21	34	27	18	0.4046	0.0094	0.0096
	BF: $\lambda = 10$	0	0	0	66	26	7	1	0.4060	0.0081	0.0056
	BF: $\lambda = 20$	0	0	0	76	21	2	1	0.4063	0.0080	0.0049
	BF: $\lambda = 100$	0	0	0	93	7	0	0	0.4072	0.0079	0.0037
	MBF: $\lambda = 1$	0	0	0	71	25	4	0	0.4064	0.0080	0.0059
	MBF: $\lambda = 3$	0	0	0	88	12	0	0	0.4070	0.0079	0.0043
	MBF: $\lambda = 10$	0	0	0	94	6	0	0	0.4074	0.0080	0.0039
	MBF: $\lambda = 20$	0	0	0	94	6	0	0	0.4076	0.0081	0.0040
	MBF: $\lambda = 100$	0	0	0	100	0	0	0	0.4089	0.0093	0.0044
(M8)	ID	0	0	0	0	3	10	87	0.5177	0.0285	0.1679
$K = 3$	NOT	0	0	0	83	15	1	1	0.4940	0.0164	0.0292
	TS	0	0	0	3	42	12	43	0.4935	0.0297	0.1184
	BF: $\lambda = 1$	0	0	0	30	27	23	20	0.4945	0.0150	0.0170
	BF: $\lambda = 3$	0	0	0	63	26	11	0	0.4996	0.0102	0.0096
	BF: $\lambda = 10$	0	0	0	90	9	1	0	0.5018	0.0082	0.0072
	BF: $\lambda = 20$	0	0	0	95	4	1	0	0.5022	0.0081	0.0071
	BF: $\lambda = 100$	0	0	0	99	1	0	0	0.5025	0.0081	0.0077
	MBF: $\lambda = 1$	0	0	0	95	3	2	0	0.5022	0.0082	0.0088
	MBF: $\lambda = 3$	0	0	0	99	1	0	0	0.5025	0.0080	0.0087
	MBF: $\lambda = 10$	0	0	0	100	0	0	0	0.5025	0.0081	0.0086
	MBF: $\lambda = 20$	0	0	0	100	0	0	0	0.5025	0.0081	0.0086
	MBF: $\lambda = 100$	0	0	0	100	0	0	0	0.5026	0.0081	0.0090

Table 3. Type 2 heterogeneous variances: Distribution of $\hat{K} - K$ for the various competing methods, the average MSE, the average MSR, and the average Hausdorff distance d_H for models (M1)–(M4).

		$\hat{K} - K$									
Model	Method	≤ -3	-2	-1	0	1	2	≥ 3	MSR	MSE	d_H
(M1)	ID	0	0	0	38	11	15	36	1.1569	0.0584	0.0272
$K = 9$	NOT	0	0	0	90	7	2	1	1.1616	0.0227	0.0129
	TS	0	0	0	0	0	0	100	0.6916	0.4961	0.0451
	BF: $\lambda = 1$	0	0	0	1	1	1	97	1.1293	0.0564	0.0374
	BF: $\lambda = 3$	0	0	0	6	18	20	56	1.1463	0.0378	0.0263
	BF: $\lambda = 10$	0	0	0	44	28	16	12	1.1557	0.0291	0.0186
	BF: $\lambda = 20$	0	0	0	65	22	12	1	1.1586	0.0264	0.0162
	BF: $\lambda = 100$	0	0	0	89	9	2	0	1.1598	0.0266	0.0163
	MBF: $\lambda = 1$	0	0	0	50	29	11	10	1.1561	0.0286	0.0201
	MBF: $\lambda = 3$	0	0	0	74	23	3	0	1.1583	0.0274	0.0178
	MBF: $\lambda = 10$	0	0	0	85	14	1	0	1.1595	0.0287	0.0187
	MBF: $\lambda = 20$	0	0	0	90	10	0	0	1.1601	0.0288	0.0179
	MBF: $\lambda = 100$	0	0	0	92	8	0	0	1.1632	0.0305	0.0212
(M2)	ID	0	0	0	21	6	11	62	0.8789	0.1352	0.0181
$K = 20$	NOT	0	0	3	50	18	9	20	0.7994	0.1210	0.0194
	TS	0	0	0	0	0	0	100	0.3338	0.5677	0.0222
	BF: $\lambda = 1$	0	0	0	0	0	0	100	0.7870	0.1216	0.0213
	BF: $\lambda = 3$	0	0	0	0	2	8	90	0.8268	0.0843	0.0190
	BF: $\lambda = 10$	0	0	0	37	26	19	18	0.8419	0.0773	0.0160
	BF: $\lambda = 20$	0	0	0	62	24	11	3	0.8472	0.0757	0.0153
	BF: $\lambda = 100$	0	0	0	84	15	1	0	0.8514	0.0816	0.0171
	MBF: $\lambda = 1$	0	0	0	18	30	18	34	0.8419	0.0797	0.0171
	MBF: $\lambda = 3$	0	0	0	53	32	10	5	0.8483	0.0798	0.0168
	MBF: $\lambda = 10$	0	0	0	82	15	2	1	0.8534	0.0833	0.0175
	MBF: $\lambda = 20$	0	0	1	87	10	1	1	0.8565	0.0867	0.0182
	MBF: $\lambda = 100$	0	0	1	94	5	0	0	0.8615	0.0939	0.0190
(M3)	ID	0	0	0	52	12	11	25	1.4718	0.0707	0.0303
$K = 7$	NOT	0	0	0	64	20	7	9	1.4067	0.0521	0.0314
	TS	0	0	0	0	0	0	100	1.0654	0.3940	0.0559
	BF: $\lambda = 1$	0	0	0	1	2	4	93	1.4074	0.0417	0.0373
	BF: $\lambda = 3$	0	0	0	17	18	28	37	1.4238	0.0249	0.0207
	BF: $\lambda = 10$	0	0	0	63	22	13	2	1.4321	0.0164	0.0102
	BF: $\lambda = 20$	0	0	0	82	14	4	0	1.4335	0.0151	0.0071
	BF: $\lambda = 100$	0	0	0	97	3	0	0	1.4350	0.0140	0.0057
	MBF: $\lambda = 1$	0	0	0	91	5	4	0	1.4342	0.0148	0.0066
	MBF: $\lambda = 3$	0	0	0	95	3	2	0	1.4347	0.0145	0.0060
	MBF: $\lambda = 10$	0	0	0	97	3	0	0	1.4351	0.0141	0.0061
	MBF: $\lambda = 20$	0	0	0	98	2	0	0	1.4351	0.0143	0.0062
	MBF: $\lambda = 100$	0	0	0	98	2	0	0	1.4351	0.0143	0.0063
(M4)	ID	0	0	0	0	0	0	100	1.0555	0.1444	0.0344
$K = 9$	NOT	0	0	0	41	11	17	31	0.9614	0.0536	0.0247
	TS	0	0	0	0	0	0	100	0.6253	0.3887	0.0564
	BF: $\lambda = 1$	0	0	0	0	1	1	98	0.9752	0.0339	0.0302
	BF: $\lambda = 3$	0	0	0	10	18	23	49	0.9859	0.0251	0.0189
	BF: $\lambda = 10$	0	0	0	47	25	22	6	0.9913	0.0225	0.0079
	BF: $\lambda = 20$	0	0	0	58	31	10	1	0.9934	0.0221	0.0057
	BF: $\lambda = 100$	0	0	0	87	12	1	0	0.9962	0.0207	0.0035
	MBF: $\lambda = 1$	0	0	0	65	18	8	9	0.9942	0.0212	0.0057
	MBF: $\lambda = 3$	0	0	0	80	15	4	1	0.9958	0.0210	0.0041
	MBF: $\lambda = 10$	0	0	0	89	11	0	0	0.9961	0.0208	0.0035
	MBF: $\lambda = 20$	0	0	0	89	11	0	0	0.9961	0.0208	0.0037
	MBF: $\lambda = 100$	0	0	0	92	8	0	0	0.9968	0.0217	0.0035

Table 4. Type 2 heterogeneous variances: distribution of $\hat{K} - K$ for the various competing methods, the average MSE, the average MSR, and the average Hausdorff distance d_H for models (M5)–(M8).

		$\hat{K} - K$									
Model	Method	≤ -3	-2	-1	0	1	2	≥ 3	MSR	MSE	d_H
(M5)	ID	0	0	0	12	14	11	63	1.4279	0.1828	0.1194
$K = 7$	NOT	49	17	17	6	6	3	2	1.3909	0.1490	0.1250
	TS	0	0	0	0	1	0	99	1.0149	0.4906	0.1161
	BF: $\lambda = 1$	0	0	0	0	5	11	84	1.3494	0.0980	0.0307
	BF: $\lambda = 3$	0	0	0	19	40	23	18	1.3841	0.0678	0.0169
	BF: $\lambda = 10$	0	0	0	62	31	7	0	1.3972	0.0564	0.0097
	BF: $\lambda = 20$	0	0	0	77	21	2	0	1.4002	0.0536	0.0085
	BF: $\lambda = 100$	0	0	0	95	5	0	0	1.4031	0.0525	0.0070
	MBF: $\lambda = 1$	0	0	0	81	18	1	0	1.3992	0.0551	0.0091
	MBF: $\lambda = 3$	0	0	0	91	9	0	0	1.4019	0.0527	0.0078
	MBF: $\lambda = 10$	0	0	0	97	3	0	0	1.4026	0.0547	0.0083
	MBF: $\lambda = 20$	0	0	0	100	0	0	0	1.4040	0.0536	0.0082
	MBF: $\lambda = 100$	0	0	1	99	0	0	0	1.4038	0.0545	0.0091
(M6)	ID	71	9	9	8	2	0	1	1.1458	0.1110	0.4438
$K = 3$	NOT	96	0	3	0	0	1	0	1.1521	0.0087	0.4613
	TS	0	0	0	0	0	0	100	0.6391	0.5178	0.2372
	BF: $\lambda = 1$	0	0	0	21	28	22	29	1.1433	0.0202	0.0185
	BF: $\lambda = 3$	0	0	0	57	28	13	2	1.1489	0.0141	0.0127
	BF: $\lambda = 10$	0	0	0	87	12	1	0	1.1526	0.0104	0.0075
	BF: $\lambda = 20$	0	0	0	95	4	1	0	1.1537	0.0094	0.0056
	BF: $\lambda = 100$	0	0	0	100	0	0	0	1.1539	0.0091	0.0056
	MBF: $\lambda = 1$	0	0	0	100	0	0	0	1.1539	0.0091	0.0056
	MBF: $\lambda = 3$	0	0	0	100	0	0	0	1.1539	0.0091	0.0056
	MBF: $\lambda = 10$	0	0	0	100	0	0	0	1.1539	0.0091	0.0056
	MBF: $\lambda = 20$	0	0	0	100	0	0	0	1.1539	0.0091	0.0056
	MBF: $\lambda = 100$	0	0	0	100	0	0	0	1.1539	0.0091	0.0056
(M7)	ID	0	0	0	0	1	7	92	1.5276	0.3525	0.0489
$K = 7$	NOT	0	0	0	58	12	15	15	1.2558	0.0743	0.0238
	TS	0	0	0	0	0	0	100	0.7915	0.5345	0.0814
	BF: $\lambda = 1$	0	0	0	0	5	14	81	1.2678	0.0587	0.0168
	BF: $\lambda = 3$	0	0	0	17	34	22	27	1.2822	0.0463	0.0091
	BF: $\lambda = 10$	0	0	0	58	30	11	1	1.2890	0.0424	0.0064
	BF: $\lambda = 20$	0	0	0	70	22	7	1	1.2900	0.0420	0.0058
	BF: $\lambda = 100$	0	0	0	88	12	0	0	1.2936	0.0399	0.0053
	MBF: $\lambda = 1$	0	0	0	69	23	6	2	1.2908	0.0421	0.0063
	MBF: $\lambda = 3$	0	0	0	85	15	0	0	1.2931	0.0413	0.0057
	MBF: $\lambda = 10$	0	0	0	96	4	0	0	1.2939	0.0422	0.0053
	MBF: $\lambda = 20$	0	0	0	98	2	0	0	1.2942	0.0430	0.0053
	MBF: $\lambda = 100$	0	0	0	99	1	0	0	1.2948	0.0437	0.0056
(M8)	ID	0	0	0	0	0	2	98	0.5922	0.0802	0.0948
$K = 3$	NOT	0	0	0	3	5	8	84	0.5032	0.1358	0.0383
	TS	0	0	0	0	0	0	100	0.2211	0.4181	0.0893
	BF: $\lambda = 1$	0	0	0	30	19	25	26	0.6090	0.0274	0.0170
	BF: $\lambda = 3$	0	0	0	58	27	8	7	0.6182	0.0190	0.0097
	BF: $\lambda = 10$	0	0	0	90	7	2	1	0.6219	0.0159	0.0084
	BF: $\lambda = 20$	0	0	0	92	6	2	0	0.6219	0.0159	0.0083
	BF: $\lambda = 100$	0	0	0	96	4	0	0	0.6228	0.0152	0.0088
	MBF: $\lambda = 1$	0	0	0	93	5	2	0	0.6220	0.0159	0.0092
	MBF: $\lambda = 3$	0	0	0	97	3	0	0	0.6231	0.0147	0.0090
	MBF: $\lambda = 10$	0	0	0	97	3	0	0	0.6232	0.0148	0.0104
	MBF: $\lambda = 20$	0	0	0	97	3	0	0	0.6232	0.0148	0.0104
	MBF: $\lambda = 100$	0	0	0	97	3	0	0	0.6232	0.0148	0.0110

Table 5. Homogeneous variances: distribution of $\hat{K} - K$ for the various competing methods, the average MSE, the average MSR, and the average Hausdorff distance d_H for models (M1)–(M4).

		$\hat{K} - K$									
Model	Method	≤ -3	-2	-1	0	1	2	≥ 3	MSR	MSE	d_H
(M1)	ID	0	0	0	97	3	0	0	0.9985	0.0305	0.0154
$K = 9$	NOT	0	0	0	100	0	0	0	0.9862	0.0159	0.0090
	TS	0	0	0	99	1	0	0	0.9937	0.0444	0.0290
	BF: $\lambda = 1$	0	0	0	0	0	2	98	0.9477	0.0839	0.0440
	BF: $\lambda = 3$	0	0	0	11	19	18	52	0.9717	0.0728	0.0392
	BF: $\lambda = 10$	0	0	0	51	26	13	10	0.9833	0.0746	0.0366
	BF: $\lambda = 20$	0	0	0	64	24	7	5	0.9869	0.0775	0.0371
	BF: $\lambda = 100$	0	0	1	77	20	2	0	0.9950	0.0869	0.0388
	BFM: $\lambda = 1$	0	0	0	39	42	11	8	0.9835	0.0761	0.0386
	BFM: $\lambda = 3$	0	0	0	61	34	5	0	0.9914	0.0836	0.0384
	BFM: $\lambda = 10$	0	0	1	76	23	0	0	0.9978	0.0905	0.0398
	BFM: $\lambda = 20$	0	0	4	79	17	0	0	1.0005	0.0938	0.0402
	BFM: $\lambda = 100$	0	0	7	85	8	0	0	1.0083	0.1040	0.0434
(M2)	ID	0	0	0	98	2	0	0	1.0204	0.0902	0.0142
$K = 20$	NOT	0	0	0	100	0	0	0	0.9374	0.0888	0.0155
	TS	0	0	2	98	0	0	0	0.9818	0.1087	0.0189
	BF: $\lambda = 1$	0	0	0	0	0	0	100	0.8967	0.1773	0.0226
	BF: $\lambda = 3$	0	0	0	2	5	11	82	0.9467	0.1613	0.0217
	BF: $\lambda = 10$	0	0	0	25	22	30	23	0.9713	0.1687	0.0216
	BF: $\lambda = 20$	0	0	0	49	33	15	3	0.9810	0.1728	0.0218
	BF: $\lambda = 100$	0	0	4	80	15	1	0	0.9969	0.1865	0.0228
	BFM: $\lambda = 1$	0	0	0	19	23	27	31	0.9714	0.1700	0.0220
	BFM: $\lambda = 3$	0	0	0	43	40	14	3	0.9714	0.1784	0.0223
	BFM: $\lambda = 10$	0	0	2	73	22	3	0	0.9984	0.1880	0.0230
	BFM: $\lambda = 20$	0	0	10	76	14	0	0	1.0065	0.1968	0.0238
	BFM: $\lambda = 100$	3	2	30	63	2	0	0	1.0440	0.2346	0.0329
(M3)	ID	0	0	0	89	10	1	0	1.0314	0.0475	0.0264
$K = 7$	NOT	0	0	0	100	0	0	0	0.9847	0.0191	0.0215
	TS	0	0	0	99	1	0	0	1.0034	0.0307	0.0312
	BF: $\lambda = 1$	0	0	0	0	3	6	91	0.9728	0.0413	0.0490
	BF: $\lambda = 1$	0	0	0	0	3	6	91	0.9728	0.0413	0.0490
	BF: $\lambda = 3$	0	0	0	15	20	25	40	0.9834	0.0338	0.0421
	BF: $\lambda = 10$	0	0	0	59	22	16	3	0.9895	0.0312	0.0392
	BF: $\lambda = 20$	0	0	0	73	21	5	1	0.9912	0.0318	0.0392
	BF: $\lambda = 100$	0	0	0	95	4	1	0	0.9943	0.0340	0.0428
	BFM: $\lambda = 1$	0	0	0	86	11	3	0	0.9930	0.0334	0.0423
	BFM: $\lambda = 3$	0	0	0	89	9	2	0	0.9937	0.0343	0.0436
	BFM: $\lambda = 10$	0	0	0	95	5	0	0	0.9960	0.0363	0.0448
	BFM: $\lambda = 20$	0	0	1	95	4	0	0	0.9970	0.0378	0.0457
	BFM: $\lambda = 100$	0	0	1	98	1	0	0	0.9989	0.0400	0.0471
(M4)	ID	0	0	0	0	2	8	90	1.2104	0.2413	0.0730
$K = 9$	NOT	0	0	0	98	2	0	0	0.9867	0.0171	0.0097
	TS	0	0	0	50	40	9	1	1.0010	0.0349	0.0158
	BF: $\lambda = 1$	0	0	0	0	0	1	99	0.9713	0.0393	0.0301
	BF: $\lambda = 3$	0	0	0	8	19	28	45	0.9832	0.0296	0.0201
	BF: $\lambda = 10$	0	0	0	41	32	19	8	0.9884	0.0268	0.0153
	BF: $\lambda = 20$	0	0	0	54	34	11	1	0.9898	0.0261	0.0140
	BF: $\lambda = 100$	0	0	0	82	18	0	0	0.9921	0.0261	0.0133
	BFM: $\lambda = 1$	0	0	0	54	34	7	5	0.9898	0.0258	0.0148
	BFM: $\lambda = 3$	0	0	0	75	21	3	1	0.9917	0.0264	0.0138
	BFM: $\lambda = 10$	0	0	0	85	14	1	0	0.9928	0.0271	0.0135
	BFM: $\lambda = 20$	0	0	0	88	11	1	0	0.9936	0.0283	0.0136
	BFM: $\lambda = 100$	0	0	0	93	7	0	0	0.9956	0.0305	0.0148

Table 6. Homogeneous variances: distribution of $\hat{K} - K$ for the various competing methods, the average MSE, the average MSR, and the average Hausdorff distance d_H for models (M5)–(M8).

		$\hat{K} - K$									
Model	Method	≤ -3	-2	-1	0	1	2	≥ 3	MSR	MSE	d_H
(M5)	ID	0	0	0	0	25	0	75	0.2857	0.0555	0.1142
$K = 7$	NOT	0	0	8	68	24	0	0	0.2432	0.0126	0.0222
	TS	1	0	17	47	25	9	1	0.2523	0.0216	0.0413
	BF: $\lambda = 1$	0	0	0	1	5	11	83	0.2370	0.0207	0.0295
	BF: $\lambda = 3$	0	0	0	20	37	23	20	0.2421	0.0162	0.0144
	BF: $\lambda = 10$	0	0	0	61	35	4	0	0.2442	0.0150	0.0072
	BF: $\lambda = 20$	0	0	3	73	22	2	0	0.2446	0.0151	0.0095
	BF: $\lambda = 100$	0	0	5	86	9	0	0	0.2456	0.0155	0.0100
	MBF: $\lambda = 1$	0	0	4	77	17	1	1	0.2449	0.0152	0.0105
	MBF: $\lambda = 3$	0	0	6	84	9	1	0	0.2456	0.0156	0.0117
	MBF: $\lambda = 10$	0	0	10	85	5	0	0	0.2461	0.0160	0.0156
	MBF: $\lambda = 20$	0	0	13	84	3	0	0	0.2465	0.0164	0.0191
	MBF: $\lambda = 100$	0	0	22	77	1	0	0	0.2476	0.0178	0.0299
(M6)	ID	0	0	0	100	0	0	0	0.9988	0.0014	0.0000
$K = 0$	NOT	0	0	0	100	0	0	0	0.9987	0.0016	0.0000
	TS	0	0	0	100	0	0	0	0.9988	0.0014	0.0000
	BF: $\lambda = 1$	0	0	0	85	11	2	2	0.9977	0.0025	0.0019
	BF: $\lambda = 3$	0	0	0	91	9	0	0	0.9985	0.0018	0.0007
	BF: $\lambda = 10$	0	0	0	98	2	0	0	0.9988	0.0015	0.0001
	BF: $\lambda = 20$	0	0	0	99	1	0	0	0.9988	0.0015	0.0001
	BF: $\lambda = 100$	0	0	0	100	0	0	0	0.9988	0.0015	0.0000
	MBF: $\lambda = 1$	0	0	0	100	0	0	0	0.9988	0.0014	0.0000
	MBF: $\lambda = 3$	0	0	0	100	0	0	0	0.9988	0.0014	0.0000
	MBF: $\lambda = 10$	0	0	0	100	0	0	0	0.9988	0.0014	0.0000
	MBF: $\lambda = 20$	0	0	0	100	0	0	0	0.9988	0.0014	0.0000
	MBF: $\lambda = 100$	0	0	0	100	0	0	0	0.9988	0.0014	0.0000
(M7)	ID	0	0	0	0	2	65	33	1.1164	0.1495	0.0120
$K = 7$	NOT	0	0	0	100	0	0	0	0.9845	0.0223	0.0052
	TS	0	0	0	71	27	2	0	1.0030	0.0424	0.0092
	BF: $\lambda = 1$	0	0	0	0	2	10	88	0.9691	0.0444	0.0196
	BF: $\lambda = 3$	0	0	0	19	25	27	29	0.9807	0.0375	0.0122
	BF: $\lambda = 10$	0	0	0	55	33	9	3	0.9873	0.0365	0.0096
	BF: $\lambda = 20$	0	0	0	69	25	6	0	0.9891	0.0363	0.0088
	BF: $\lambda = 100$	0	0	0	91	9	0	0	0.9923	0.0382	0.0078
	MBF: $\lambda = 1$	0	0	0	59	32	7	2	0.9893	0.0384	0.0097
	MBF: $\lambda = 3$	0	0	0	80	17	3	0	0.9915	0.0390	0.0088
	MBF: $\lambda = 10$	0	0	0	90	10	0	0	0.9940	0.0412	0.0082
	MBF: $\lambda = 20$	0	0	0	91	9	0	0	0.9948	0.0421	0.0083
	MBF: $\lambda = 100$	0	0	0	100	0	0	0	0.9981	0.0451	0.0085
(M8)	ID	0	0	1	37	40	15	7	1.1052	0.1365	0.0933
$K = 3$	NOT	0	0	0	99	1	0	0	0.9792	0.0222	0.0288
	TS	0	0	3	59	29	9	0	1.0112	0.0470	0.0580
	BF: $\lambda = 1$	0	0	0	34	25	21	20	0.9726	0.0304	0.0363
	BF: $\lambda = 3$	0	0	0	69	18	11	2	0.9796	0.0239	0.0331
	BF: $\lambda = 10$	0	0	0	89	9	2	0	0.9826	0.0212	0.0313
	BF: $\lambda = 20$	0	0	0	93	7	0	0	0.9830	0.0207	0.0313
	BF: $\lambda = 100$	0	0	0	97	3	0	0	0.9834	0.0204	0.0312
	MBF: $\lambda = 1$	0	0	0	96	4	0	0	0.9832	0.0206	0.0315
	MBF: $\lambda = 3$	0	0	0	99	1	0	0	0.9836	0.0202	0.0312
	MBF: $\lambda = 10$	0	0	0	99	1	0	0	0.9836	0.0202	0.0312
	MBF: $\lambda = 20$	0	0	1	98	1	0	0	0.9840	0.0206	0.0319
	MBF: $\lambda = 100$	0	0	3	96	1	0	0	0.9850	0.0221	0.0338

5.1.1. Cases of type 1 heterogenous variances

We start with type 1 of standard deviations in the noises for the heterogeneous case. A good change-points detection method is one that has a small Hausdorff distance d_H, a small MSE, and can accurately estimate the number of change-points.

(M1) wave 1: NOT is the most competitive method in terms of the estimation of the number of change-points, MSE, and Hausdorff distance. MBF with $\lambda = 100$ and BF with $\lambda = 100$ show comparable performance to NOT, although they are less attractive when it comes to MSE and the estimated locations of change-points. ID shows good performance, with comparable results to MBFs with

$$\lambda = 10 \quad \text{and} \quad \lambda = 20.$$

On the other hand, TS tends to overestimate the number of change-points. In this model, NOT, MBF with $\lambda = 100$ and BF with $\lambda = 100$ are the optimal methods in terms of the number of change-points, MSE, and Hausdorff distance. However, the result of NOT is obtained by selecting a contrast function that matches the signal characteristics of model M1.

(M2) wave 2: Among the most competitive methods in terms of the estimation of the number of change-points, MSE, and Hausdorff distance, MBF with $\lambda = 100$ and NOT perform the best, followed by MBF with $\lambda = 20$ and BF with $\lambda = 100$. ID, TS, BF with $\lambda = 100$, and MBF with

$$\lambda = 10 \quad \text{and} \quad \lambda = 20$$

also show relatively similar performance. In this model, NOT and MBF with $\lambda = 100$ methods show the best performance in terms of the number of change points, MSE, and Hausdorff distance. Note that the result of NOT is obtained by selecting a contrast function that matches the signal characteristics of model M2.

(M3) mix 1: MBFs with λ values of 3, 10, 20, and 100, as well as the BF with $\lambda = 100$, perform best in terms of the estimation of the number of change-points, MSE, and Hausdorff distance. NOT has similar performance in terms of the estimation of the number of change-points, though it is less desirable in terms of MSE and the estimated locations of change-points. Furthermore, ID and TS tend to overestimate the number of change-points. In this model, MBFs with

$$\lambda = 10, 20, 100$$

are the most competitive methods in terms of the number of change-points, MSE, and Hausdorff distance.

(M4) mix2: MBFs with

$$\lambda = 10, 20, 100$$

show some of the best performance in terms of estimating the number of change-points, MSE, and Hausdorff distance. NOT is comparable to MBF with $\lambda = 20$ in terms of the estimated number of change-points, however, it does not perform as well with respect to the estimated locations of change-points. In contrast, ID fails to provide any useful results, while TS tends to overestimate the number of change-points. In this model, MBFs with

$$\lambda = 20, 100$$

are the best methods in terms of the number of change points, MSE, and Hausdorff distance.

(M5) teeth-linear: MBFs with

$$\lambda = 3, 10, 20, 100$$

and BF with $\lambda = 100$ demonstrate the best performance across all methods for the estimation of the number of change-points, MSE and Hausdorff distance. MBF with $\lambda = 1$ and BF with $\lambda = 20$ also show good performance. In contrast, ID tends to overestimate the number of change-points significantly, while both NOT and TS underestimate the number of change-points significantly. In this model, MBFs with

$$\lambda = 10, 20, 100$$

are the optimal methods in terms of the number of change-points, MSE, and Hausdorff distance.

(M6) linear: All MBF and BF methods with

$$\lambda = 20, 100$$

demonstrate the highest performance across all metrics, including the estimation of the number of change-points, MSE, and Hausdorff distance. BF with $\lambda = 10$ also performs well. In contrast, ID, NOT, and TS fail to accurately capture the number of changes. Notably, ID and NOT are unable to detect changes in variance in linear signals with changes in variance. In this model, BF with $\lambda = 100$ and all MBF methods show the best performance in terms of the number of change points, MSE, and Hausdorff distance.

(M7) wave 3: MBF with

$$\lambda = 10, 20, 100$$

and BF with $\lambda = 100$ are among the best methods in terms of estimating the number of change-points, MSE, and Hausdorff distance. Additionally, MBF with $\lambda = 3$ performs well. NOT is comparable to MBF with $\lambda = 3$ for estimating the number of change-points, but is less effective for MSE and the estimation of change-point locations. In contrast, both ID and TS fail to accurately estimate the number of change-points. In this model, MBFs with

$$\lambda = 10, 20, 100$$

and BF with $\lambda = 100$ are the most competitive methods in terms of the number of change-points, MSE, and Hausdorff distance.

(M8) quad: All MBFs and BFs with

$$\lambda = 20, 100$$

demonstrate superior performance across all methods in terms of the estimation of the number of change-points, MSE, and Hausdorff distance. BF with $\lambda = 10$ also shows good performance. Although NOT exhibits comparable performance to BF with $\lambda = 10$ in terms of the number of change-points detected, it lags behind in terms of MSE and estimated locations of change-points. Meanwhile, ID fails to work, and TS tends to overestimate the number of change-points in signals with a quadratic trend. In this model, MBFs with

$$\lambda = 3, 10, 20, 100$$

and BF with $\lambda = 100$ methods show the best performance in terms of the number of change points, MSE, and Hausdorff distance.

In summary, the Bayesian method is the best overall when it comes to accurately estimating the number of change-points, MSE, and d_H for the signals considered. For both continuous piecewise-linear (M1) and discontinuous piecewise-linear (M2) signals, the Bayesian method is always in the top 10% of the best methods for any aspect of the estimation of the number of change-points, MSE, and d_H. The performance of ID and TS is not good in most of the signals considered, while NOT shows relatively good performance. It should be noted that the results of NOT are obtained by selecting the most suitable contrast function for the characteristics of the given signals. However, it is not known which contrast function is suitable for signals in practical applications. In addition, the NOT method cannot determine the contrast function for mixed signals with various change trends. In particular, no estimation of the number of change-points is possible with ID in signals M4, M6, and M7, and NOT cannot be used in signal M6. Additionally, TS does not work in signals M6 and M7.

5.1.2. Cases of type 2 heterogenous variances

Now we consider the type 2 of standard deviations in the noises for the heterogeneous case.

(M1) wave 1: MBFs with

$$\lambda = 20 \quad \text{and} \quad \lambda = 100,$$

BF with $\lambda = 100$, and NOT are among the most competitive methods in terms of the estimation of the number of change-points, MSE, and Hausdorff distance. MBF with $\lambda = 10$ shows comparable performance to BF with $\lambda = 100$. On the contrary, ID tends to overestimate the number of change-points, while the TS method does not seem to be effective in this regard. In this model, MBFs with $\lambda = 20, 100$ and NOT are the optimal methods in terms of the number of change-points, MSE, and Hausdorff distance. Note that the result of NOT is obtained by selecting a contrast function that matches the signal characteristics of model M1.

(M2) wave 2: MBF with $\lambda = 100$ and MBF with $\lambda = 10$ both perform as well as MBF with $\lambda = 20$ in terms of estimating the number of change-points, MSE, and Hausdorff distance. Conversely, ID and NOT tend to overestimate the number of change-points, while TS fails to provide reliable estimations at all. In this model, the MBF with $\lambda = 100$ method shows the best performance in terms of the number of change points, MSE, and Hausdorff distance.

(M3) mix 1: All MBFs and BF with $\lambda = 100$ demonstrate the highest performance across all methods in terms of change-point estimation, MSE, and Hausdorff distance. BF with $\lambda = 20$ also shows relatively good results. On the contrary, ID and NOT tend to overestimate the number of change-points. This indicates that the TS method does not effectively estimate the number of change-points. In this model, BF with $\lambda = 100$ and MBFs with

$$\lambda = 10, 20, 100$$

are the most competitive methods in terms of the number of change-points, MSE, and Hausdorff distance.

(M4) mix 2: MBFs with

$$\lambda = 10, 20, 100$$

and a BF with $\lambda = 100$ are among the most competitive methods in terms of estimating the number of change-points, MSE, and Hausdorff distance. MBF with $\lambda = 3$ also shows relatively good performance. On the other hand, NOT tends to overestimate the number of change-points, while ID and TS methods fail to accurately estimate the number of change-points. In this model, BF with $\lambda = 100$ and MBFs with

$$\lambda = 10, 20, 100$$

methods show the best performance in terms of the number of change points, MSE, and Hausdorff distance.

(M5) teeth-linear: MBFs with λ values of 10, 20, and 100, and BF with a λ value of 100, show the best performance across all methods in terms of the estimation of the number of change-points, MSE and Hausdorff distance. Next, MBF with a λ value of 3 is comparable to the performance of BF with λ=100. Furthermore, BF with a λ value of 20 and MBF with a λ value of 1 provide relatively good performance. In contrast, ID tends to considerably overestimate the number of change-points and NOT tends to considerably underestimate the number of change-points. It is evident that the TS method does not perform well for estimating the number of change-points. In this model, BF with $\lambda = 100$ and MBFs with

$$\lambda = 10, 20, 100$$

are the best methods in terms of the number of change-points, MSE, and Hausdorff distance.

(M6) linear: All MBFs and BFs with

$$\lambda = 20 \quad \text{and} \quad \lambda = 100$$

exhibit the best performance across all methods in terms of estimating the number of change-points, minimizing the MSE and minimizing the Hausdorff distance. Meanwhile, BF with $\lambda = 10$ exhibits good performance. In contrast, both ID, NOT and TS methods are unable to detect any change-points of variances in linear signals where the variances fluctuate. In this model, BF with $\lambda = 100$ and all MBF are the most competitive methods in terms of the number of change points, MSE, and Hausdorff distance.

(M7) wave 3: The most competitive methods for estimating the number of change-points, MSE, and Hausdorff distance are the MBF with λ values of 10, 20, and 100. Additionally, MBF with $\lambda = 3$ and BF with $\lambda = 100$ perform well. However, NOT tends to overestimate the number of change-points, and while it shows relatively similar performance to BF with $\lambda = 10$ in terms of the number of change-points estimated, its MSE and estimated location of change-points are not attractive. ID and TS, on the other hand, struggle to estimate the number of change-points effectively. In this model, MBFs with

$$\lambda = 10, 20, 100$$

methods show the best performance in terms of the number of change-points, MSE, and Hausdorff distance.

(M8) quad: All MBFs and BFs with

$$\lambda = 10, 20, 100$$

achieved the best performance across all methods in terms of the number of change-points, MSE and Hausdorff distance estimation. On the other hand, NOT failed to successfully detect the number of

change-points. Additionally, ID and TS, which are not specifically designed for detecting change-points in signals containing quadratic trends, also failed to provide satisfactory results. In this model, BF with $\lambda = 100$ and MBFs with

$$\lambda = 3, 10, 20, 100$$

are the best methods in terms of the number of change points, MSE, and Hausdorff distance.

In summary, the Bayesian method stands out overall as the best approach for estimating the number of change-points, MSE and d_H in all models. The ID, NOT, and TS methods generally perform poorly, that is, existing methods ID, NOT, and TS are not effective for mixed signals. However, in continuous piecewise-linear (M1) signals, NOT can be one of the top 10% of the best methods when considering accuracy in any aspect of the estimation of the number of change-points, MSE, and d_H, as long as the most suitable contrast function is selected. Additionally, ID is not performed at all in models M4, M7 and M8, while NOT and TS do not work at all in models M6 and M8.

5.1.3. Case of homogenous variance

Our Bayesian change-point detection methods are designed to work in situations where the variances between change-point intervals are heterogeneous. To evaluate the performance of our methods compared to competitors, we will also analyze signals with homogeneous variance.

(M1) wave 1: NOT is the most competitive method in terms of the estimation of the number of change-points, MSE, and Hausdorff distance. ID and TS exhibit comparable performance to NOT in terms of the estimation of the number of change-points, while they are slightly less effective in terms of the estimated locations of change-points and MSE. Among Bayesian methods, MBF with

$$\lambda = 10, 20, 100,$$

and BF with $\lambda = 100$ show relatively good performance. In this model, NOT is the optimal method in terms of the number of change-points, MSE, and Hausdorff distance. Note that the result of NOT is obtained by selecting a contrast function that matches the signal characteristics of model M1.

(M2) wave 2: ID, NOT, and TS are among the most competitive methods when it comes to estimating the number of change-points, as evidenced by their low MSE and Hausdorff distance scores. Among Bayesian methods, BF with $\lambda = 100$ and MBF with

$$\lambda = 10, 20$$

have shown relatively good performance. In this model, NOT, ID and TS methods show the best performance in terms of the number of change points, MSE, and Hausdorff distance. However, the results of NOT and ID are obtained by selecting a contrast function that matches the signal characteristics of model M2.

(M3) mix 1: NOT, TS, MBF with

$$\lambda = 10, 20, 100$$

and BF with $\lambda = 100$ demonstrate the best performance across all methods in terms of the estimation of the number of change-points, MSE and Hausdorff distance. However, in terms of the estimated locations of change-points and MSE, TS, MBF with

$$\lambda = 10, 20, 100$$

and BF with λ= 100 are slightly less effective than NOT. In this model, NOT, TS, and MBF with λ=100 are the most competitive methods in terms of the number of change points, MSE, and Hausdorff distance. However, the result of NOT is obtained by selecting a contrast function that matches the signal characteristics of model M3.

(M4) mix 2: NOT and MBF with $\lambda = 100$ show the best performance across all metrics evaluated, namely, the estimation of the number of change-points, MSE, and Hausdorff distance. MBF with $\lambda = 100$ performs comparably to NOT when estimating the number of change-points, but shows slightly lower performance for estimated locations of change-points and MSE. MBF with

$$\lambda = 10, 20$$

and BF with $\lambda = 100$ also display favorable results. Conversely, TS tends to overestimate the number of change-points, while ID fails to adequately estimate the number of change-points. In this model, NOT and MBF with $\lambda = 100$ are the best methods in terms of the number of change points, MSE, and Hausdorff distance. Note that the result of NOT is obtained by selecting a contrast function that matches the signal characteristics of model M4.

(M5) teeth-linear: MBFs with

$$\lambda = 3, 10, 20$$

and BF with $\lambda = 100$ show the best performance across all methods in terms of the estimation of the number of change-points, MSE, and Hausdorff distance. MBFs with

$$\lambda = 1, 100$$

and BF with $\lambda = 20$ exhibit relatively good performance. NOT performs comparably to BF with $\lambda = 20$ in the estimation of the number of change-points, though it is slightly less accurate in terms of the estimated locations of change-points. TS tends to overestimate the number of change-points, while ID fails to produce reliable estimates of the number of change-points. In this model, BF with $\lambda = 100$ and MBFs with

$$\lambda = 3, 10, 20$$

are the most competitive methods in terms of the number of change-points, MSE, and Hausdorff distance.

(M6) linear: All MBFs and BFs with

$$\lambda = 10, 20, 100$$

show the best performance in terms of the estimation of the number of change-points, MSE, and Hausdorff distance. Additionally, BFs with

$$\lambda = 1, 3$$

also show good performance. In this model, ID, NOT, TS, BFs with

$$\lambda = 10, 20, 100$$

and all MBF are the best methods in terms of the number of change points, MSE, and Hausdorff distance. However, the results of NOT and ID are obtained by selecting a contrast function that matches the signal characteristics of model M6.

(M7) wave 3: MBF with

$$\lambda = 100$$

and NOT are among the most competitive methods when it comes to the estimation of the number of change-points, MSE, and Hausdorff distance. MBF with

$$\lambda = 100$$

shows comparable performance to NOT in terms of the estimation of the number of change-points, while its MSE is slightly lower. MBF with

$$\lambda = 10, 20$$

and BF with

$$\lambda = 100$$

also perform well. TS tends to underestimate and is comparable to BF with

$$\lambda = 20.$$

Unfortunately, ID does not perform well and tends to overestimate the number of change-points considerably. In this model, MBF with

$$\lambda = 100$$

and NOT methods show the best performance in terms of the number of change-points, MSE, and Hausdorff distance.

(M8) quad: NOT, all MBFs and BFs with

$$\lambda = 100$$

demonstrate the best performance across all methods when it comes to the estimation of the number of change-points, MSE and Hausdorff distance. However, BFs with

$$\lambda = 10, 20$$

still show satisfactory results. Conversely, both ID and TS do not perform well and tend to overestimate the number of change-points, even though ID and TS are not suitable for detecting change-points in signals with a quadratic trend. In this model, NOT, BF with

$$\lambda = 100$$

and MBFs with

$$\lambda = 3, 10, 20$$

are the optimal methods in terms of the number of change points, MSE, and Hausdorff distance.

In summary, Bayesian methods consistently perform among the top 10% of the best methods in terms of change-point estimation, MSE, and d_H, being the best overall except for continuous and discontinuous piecewise-linear (M1 and M2) signals. NOT provides good performance overall, especially in model M5, outperforming ID and TS. However, its performance largely depends on the selection of suitable contrast functions for the specific characteristics of the signals, and it also cannot determine the contrast function for mixed signals with various change trends and and does not work well in cases where signals have heterogeneous variances. ID is not even attempted in models M4, M5, and M7 for change-point estimation.

5.2. Real examples

Example 1. The Goddard Institute for space studies (GISS) monitors broad global changes around the world. The GISS Surface Temperature Analysis (GISTEMP) is an estimate of the global surface temperature changes, which are expressed as temperature anomalies. A temperature anomaly is the difference between an observed temperature and the average or baseline temperature, which is normally calculated by taking the mean of thirty or more years of data (from 1951 to 1980 in the current dataset). A positive anomaly indicates that the observed temperature was higher than the baseline, whereas a negative anomaly implies that the observed temperature was lower than the baseline. For further information, please refer to Hansen et al. [61] and Lenssen et al. [62]. The GISTEMP dataset has been widely explored in the literature of change point analysis, such as in Ruggieri [32], James and Matteson [63], Baranowski et al. [37], and Mehrizi and Chenouri [40]. The monthly land-ocean temperature anomalies from January 1880 to December 2020 can be found on the website *https://data.giss.nasa.gov/gistemp*. It can be shown that the presence of a linear trend with varying numbers of change-points is present in the dataset for each method. The estimates of the piecewise linear signals are found using ID, NOT1, NOT2, and TS methods, as well as Bayesian methods. The number of change-points and mean square residuals for each method are given in Table 7.

Table 7. MSR and number of change-points for the GISTEMP dataset.

Methods	ID	NOT1	NOT2	TS	BF($\lambda = 1$)
MSR	0.0111	0.0186	0.0149	0.0093	0.0058
Number of change-points	55	9	13	51	144
Methods	BF($\lambda = 3$)	BF($\lambda = 10$)	BF($\lambda = 20$)	BF($\lambda = 100$)	MBF($\lambda = 1$)
MSR	0.0085	0.0099	0.0102	0.0137	0.0099
Number of change-points	72	48	44	21	50
Methods	MBF($\lambda = 3$)	MBF($\lambda = 10$)	MBF($\lambda = 20$)	MBF($\lambda = 100$)	
MSR	0.0135	0.0142	0.0149	0.0161	
Number of change-points	24	17	15	10	

Table 7 shows that ID, TS, BF with

$$\lambda = 10$$

and MBF with

$$\lambda = 1$$

detect 55, 51, 48, and 50 change-points, respectively, indicating a relatively similar detection rate of change-points. NOT1, NOT2, MBF with

$$\lambda = 20$$

and MBF with

$$\lambda = 100$$

detect 9, 13, 15 and 10 change-points, respectively, indicating a similar level of change-point detection. These results show that ID and TS methods are consistent with the results of simulation studies that overestimate the number of change-points. Although the analysis is conducted under the assumption of linear trend for comparison with existing methods, the proposed method does not require the assumption of a contrast function such as a linear trend. Thus, the proposed method can be applied to various types of signals, such as mixed signals containing constant, linear, and nonlinear components, which cannot be solved by ID, TS, and NOT methods.

The estimated locations of the change-points for each method are provided in Table 8.

Table 8. The locations of change-points for the GISTEMP.

Methods	Locations
ID	97 110 121 144 157 213 219 250 296 316 376 384 394 424 445 467 542 553 588 625 648 719 800 854 860 921 932 1007 1012 1079 1105 1117 1162 1169 1184 1221 1279 1298 1347 1353 1380 1405 1418 1450 1466 1521 1525 1537 1563 1573 1628 1635 1638 1662 1682
NOT1	85 255 291 422 669 771 850 1157 1627
NOT2	84 115 241 297 406 449 719 856 927 1163 1321 1370 1627
TS	84 109 121 154 216 241 302 348 371 387 394 423 443 455 459 539 555 586 625 719 765 854 888 918 940 1003 1013 1080 1104 1117 1130 1164 1194 1235 1253 1306 1322 1344 1355 1369 1382 1400 1424 1449 1467 1536 1571 1629 1633 1638 1647
BF with $\lambda = 10$	14 24 38 50 58 84 109 116 154 158 167 217 241 297 324 371 381 390 406 434 449 680 707 718 745 765 796 856 888 927 1008 1091 1116 1127 1145 1163 1195 1224 1235 1253 1347 1391 1424 1530 1535 1571 1629 1635
BF with $\lambda = 20$	14 38 50 58 84 109 116 154 158 167 217 241 297 324 371 381 390 406 434 449 680 718 745 765 856 888 927 1008 1091 1116 1127 1145 1163 1195 1235 1253 1347 1391 1424 1530 1535 1571 1629 1635
BF with $\lambda = 100$	84 116 241 297 324 371 381 390 406 434 449 680 718 765 856 888 927 1008 1571 1629 1635
MBF with $\lambda = 10$	50 84 116 241 297 324 406 434 449 680 718 796 856 927 1008 1629 1635
MBF with $\lambda = 20$	50 84 116 241 279 324 406 434 449 680 718 796 1008 1629 1635
MBF with $\lambda = 100$	50 279 324 406 434 718 796 1008 1629 1635

Example 2. (Samsung stock prices) We consider the daily closing stock prices of Samsung Electronics Co. from July 2012 to July 2017. Anastasiou and Fryzlewicz [38] studied estimation of the number of change-points for this data, which is available from *https://finance.yahoo.com/quote/005930.KS/history?p=005930.KS*. We examine changes in both continuous piecewise-linear signals and piecewise-linear signals in the NOT method. The results for the ID, NOT1, NOT2, TS and Bayesian methods can be compared. Additionally, Table 9 summarizes the number of change-points and mean square residuals for each method.

Table 9. Mean square residual and number of change-points for the daily closing stock prices of Samsung Electronics Co.

Methods	ID	NOT1	NOT2	TrendSegment	BF($\lambda = 1$)
MSR	164225.5	677431.3	345964.4	105789.9	67969.1
Number of change-points	90	13	25	86	170
Methods	BF($\lambda = 3$)	BF($\lambda = 10$)	BF($\lambda = 20$)	BF($\lambda = 100$)	BFM($\lambda = 1$)
MSR	95825.4	129145.9	141981.5	191847.5	118456.65
Number of change-points	115	78	66	48	102
Methods	BFM($\lambda = 3$)	BFM($\lambda = 10$)	BFM($\lambda = 20$)	BFM($\lambda = 100$)	
MSR	141337.8	182680.45	223614.48	318590.1	
Number of change-points	72	54	45	33	

From the results of Table 9, the ID, NOT1, NOT2, and TS methods detect 90, 13, 25, and 86 change-points, respectively. MBF with

$$\lambda = (10, 20, 100)$$

and BF with

$$\lambda = (10, 20, 100)$$

detect (54, 45, 33) and (78, 66, 48) change-points, respectively. These methods show different estimates for the number of change-points, making it difficult to determine which method gives the most accurate estimate. However, it appears that ID and TS detect too many change-points, while NOT gives a more accurate estimate of the overall changing trends. Meanwhile, the Bayesian method tends to provide an intermediate estimate of the number of change-points relative to these competitors. The locations of change-points for each method are provided in Table 10.

Looking more closely at the time interval (0, 250), ID and TS methods detect 14 and 16 change-points, respectively, whereas NOT1 and NOT2 detect 2 and 5 change-points, respectively.

Table 10. The locations of change-points for the stock prices of Samsung Electronics Co.

Methods	Locations
ID	8 32 39 69 97 116 134 144 167 179 183 225 228 237 255 257 277 302 305 332 337 351 362 367 374 396 412 426 436 457 465 476 487 502 514 517 573 578 593 603 610 613 628 631 652 668 678 685 703 750 752 779 797 806 808 825 835 862 873 902 921 966 970 985 1000 1019 1023 1036 1038 1052 1067 1070 1080 1096 1116 1120 1128 1131 1141 1151 1166 1169 1184 1188 1197 1222 1231 1237 1243 1249
NOT1	118 228 254 350 390 467 566 683 782 825 886 1023 1079
NOT2	33 69 125 152 232 275 334 372 430 477 516 574 694 781 807 866 911 966 1020 1077 1137 1162 1192 1224 1248
TS	9 19 33 48 69 97 120 126 152 169 181 182 199 232 246 247 256 279 286 302 311 336 350 362 368 378 406 429 435 459 477 485 503 514 517 572 575 594 602 610 619 631 655 673 685 710 720 752 765 775 780 797 807 823 834 848 860 877 895 926 966 979 986 1020 1035 1038 1043 1051 1054 1060 1067 1076 1079 1090 1115 1121 1129 1142 1149 1165 1187 1197 1204 1225 1235 1243
BF with $\lambda = 10$	5 34 39 64 69 100 109 135 142 169 181 199 227 232 246 256 264 279 286 290 293 303 311 323 334 354 372 384 395 398 406 430 434 453 460 481 485 503 516 537 548 574 594 629 655 673 680 694 732 761 785 807 814 822 833 866 871 911 926 953 966 1020 1036 1054 1078 1088 1129 1140 1149 1168 1184 1190 1197 1204 1225 1234 1241 1249
BF with $\lambda = 20$	5 34 39 69 100 135 142 169 181 199 227 232 246 256 279 286 290 303 311 334 354 372 398 406 430 434 460 481 503 516 537 548 574 594 629 655 673 680 694 732 761 785 807 814 822 833 866 871 911 926 966 1020 1036 1054 1078 1088 1129 1140 1168 1184 1190 1197 1204 1225 1241 1249
BF with $\lambda = 100$	5 34 39 69 135 154 169 232 256 286 303 311 334 372 434 460 481 506 516 574 594 629 655 694 732 761 785 807 814 822 866 871 911 926 966 1020 1036 1054 1078 1088 1129 1140 1168 1190 1197 1225 1241 1249
MBF with $\lambda = 10$	5 34 39 69 76 109 135 154 169 232 256 274 279 286 290 303 334 359 372 434 460 481 506 516 574 594 629 655 694 761 785 807 814 822 866 871 911 926 966 1020 1036 1054 1078 1088 1137 1140 1168 1184 1190 1197 1204 1225 1241 1249
MBF with $\lambda = 20$	5 34 51 69 76 135 154 232 256 274 290 334 359 372 434 460 481 506 516 574 594 629 655 694 761 785 807 814 822 866 871 911 926 966 1020 1036 1088 1137 1140 1168 1190 1197 1204 1241 1249
MBF with $\lambda = 100$	5 34 76 135 154 232 256 274 334 372 434 460 481 506 516 574 594 629 655 694 761 785 807 822 866 871 911 926 966 1020 1036 1088 1137

On the other hand, MBF with

$$\lambda = (20, 100)$$

and BF with

$$\lambda = (20, 100)$$

detect 8 and 13 change-points, respectively. For each method, we can see that NOT methods tend to detect the point of change at which the feature of change occurs in the overall stream of the signals. On the other hand, although ID and TS appear to detect detailed changes in the signal, these results are consistent with simulation studies showing that ID and TS methods overestimate the number of change-points. The Bayesian methods appear to find changes by integrating and reflecting the findings of their respective competitors. Therefore, we think that the Bayesian methods give change-points that better reflect the change of signal compared to the other methods. Moreover, we consider that the Bayesian methods can provide a variety of fits that give users the flexibility to choose according to their preferences for the choice of threshold. Although the analysis is conducted under the assumption of linear trend for comparison with existing methods, the proposed method can be applied to various types of signals, such as mixed signals containing constant, linear, and nonlinear components, which cannot be solved by ID, TS, and NOT methods.

6. Conclusions

In this paper, we introduce a nonparametric Bayesian change-point detection method that consists of two components:

(1) the detection of change-points in the underlying signals based on the piecewise polynomial model via Bayesian binary segmentation;

(2) the selection of the most suitable order of the polynomial model in the detected segments via Bayesian model selection.

We provide intrinsic priors to ensure that the BFs and model selection probabilities are well-defined. Moreover, when the sample size is large, our method based on the BFs with the intrinsic priors is consistent. Our method is applicable to different types of signals, such as those with or without continuity constraints at the location of the point of change, signals with heterogeneous variances, and mixed signals with linearity and beyond. From our numerical results, we demonstrate that our proposed method has superior performance compared to its competitors in various scenarios with heterogeneous variances. Our method is also effective in terms of the estimation of the number of change-points, MSE, and the estimated locations of change-points. Notably, our method is the best choice when signals are mixed with different types of changes (e.g., constant, linear, quadratic) with heterogeneous variances, or when signals have discontinuous changes (e.g., signals with jumps). For the problem of detecting change-points of variances in a signal, our method can be used to identify change-points in linear signals with heterogeneous variances. This is not a straightforward problem that can be solved by existing change-point detection methods. The Bayesian change-point detection relies on a decision criterion λ. Using a small λ value will result in a lower MSE or MSR but a greater number of change-points; conversely, a large λ value will incur a higher MSE or MSR but a smaller number of change-points. Empirically, we have found that values of λ around 20 or 100 yield good results in terms of both MSE/MSR and the number of detected change-points; therefore, we

recommend these values for most applications. The developed method requires a lot of computing time because it uses numerical integration. Therefore, future research is needed to find approximate methods to reduce computing time for practical users. In addition, it has been confirmed that ambient noise plays an important role in the dynamics of nonlinear systems in various fields. The perturbations in the natural world are often Gaussian noise. However, many recent experiments have confirmed that the noise source must be non-Gaussian noise. The study of Bayesian methods for finding change points and selecting optimal trends in nonlinear models with non-Gaussian noise is considered an interesting topic in the future and will be the subject of the next research.

Author contributions

Sang Gil Kang: conceptualization, methodology, writing–original draft; Woo Dong Lee: data curation, methodology, writing–original draft; Yongku Kim: conceptualization, formal analysis, writing–original draft, writing–review & editing. All three authors read and approved the final version of the article.

Use of Generative-AI tools declaration

The authors declare they have not used Artificial Intelligence (AI) tools in the creation of this article.

Acknowledgments

This research was supported by Global-Learning & Academic research institution for Master's · PhD students, and Postdocs (G-LAMP) Program of the National Research Foundation of Korea (NRF) grant funded by the Ministry of Education (No. RS-2023-00301914).

Conflict of interest

All authors declare no conflicts of interest in this paper.

References

1. J. Bai, P. Perron, Computation and analysis of multiple structural change models, *J. Appl. Econometrics*, **18** (2003), 1–22. https://doi.org/10.1002/jae.659
2. R. Killick, I. A. Eckley, P. Jonathan, A wavelet-based approach for detecting changes in second order structure within nonstationary time series, *Electron. J. Stat.*, **7** (2013), 1167–1183. https://doi.org/10.1214/13-EJS799
3. A. Futschik, T. Hotz, A. Munk, H. Sieling, Multiscale DNA partitioning: statistical evidence for segments, *Bioinformatics*, **30** (2014), 2255–2262. https://doi.org/10.1093/bioinformatics/btu180
4. E. Galceran, A. G. Cunningham, R. M. Eustice, E. Olson, Multipolicy decisionmaking for autonomous driving via changepoint-based behavior prediction: theory and experiment, *Auton. Robot.*, **41** (2017), 1367–1382. https://doi.org/10.1007/s10514-017-9619-z

5. Y. C. Yao, Estimating the number of change-points via Schwarz criterion, *Stat. Probab. Lett.*, **6** (1988), 181–189. https://doi.org/10.1016/0167-7152(88)90118-6
6. Y. C. Yao, S. T. Au, Least-squares estimation of a step function, *Sankhya A*, **51** (1989), 370–381.
7. N. R. Zhang, D. O. Siegmund, A modified Bayes information criterion with applications to the analysis of comparative genomic hybridization data, *Biometrics*, **63** (2007), 22–32. https://doi.org/10.1111/j.1541-0420.2006.00662.x
8. N. R. Zhang, D. O. Siegmund, Model selection for high-dimensional, multi-sequence change-point problems, *Stat. Sin.*, **22** (2012), 1057–1538. https://doi.org/10.5705/ss.2010.257
9. J. Bai, P. Perron, Estimating and testing linear models with multiple structural changes, *Econometrica*, **66** (1998), 57–78. https://doi.org/10.2307/2998540
10. J. Braun, R. K. Braun, H. G. Müller, Multiple changepoint fitting via quasilikelihood with application to DNA sequence segmentation, *Biometrika*, **87** (2000), 301–304. https://doi.org/10.1093/biomet/87.2.301
11. L. Boysen, A. Kempe, V. Liebsher, A. Munk, O. Wittich, Consistencies and rates of convergence of jump-penalized least squares estimator, *Ann. Stat.*, **37** (2009), 157–183. https://doi.org/10.1214/07-AOS558
12. R. Tibshirani, P. Wang, Spatial smoothing and hot spot detection for CGH data using the fused lasso, *Biostatistics*, **9** (2008), 18–29. https://doi.org/10.1093/biostatistics/kxm013
13. Z. Harchaoui, C. Lévy-Leduc, Multiple change-point estimation with a total variation penalty, *J. Amer. Stat. Assoc.*, **105** (2010), 1480–1493. https://doi.org/10.1198/jasa.2010.tm09181
14. X. J. Jeng, T. T. Cai, H. Li, Optimal sparse segment identification with application in copy number variation analysis, *J. Amer. Stat. Assoc.*, **105** (2010), 1156–1166. https://doi.org/10.1198/jasa.2010.tm10083
15. K. Frick, A. Munk, H. Sieling, Multiscale change-point inference, *J. R. Stat. Soc. Ser. B*, **76** (2014), 495–580. https://doi.org/10.1111/rssb.12047
16. R. Killick, P. Fearnhead, I. A. Eckley, Optimal detection of changepoints with a linear computational cost, *J. Amer. Stat. Assoc.*, **107** (2012), 1590–1598. https://doi.org/10.1080/01621459.2012.737745
17. R. Maidstone, T. Hocking, G. Rigaill, P. Fearnhead, On optimal multiple changepoint algorithms for large data, *Stat. Comput.*, **27** (2017), 519–533. https://doi.org/10.1007/s11222-016-9636-3
18. A. B. Olshen, E. S. Venkatraman, R. Lucito, M. Wigler, Circular binary segmentation for the analysis of array-based DNA copy number data, *Biostatistics*, **5** (2004), 557–572. https://doi.org/10.1093/biostatistics/kxh008
19. E. Venkatraman, A. B. Olshen, A faster circular binary segmentation algorithm for the analysis of array CGH data, *Bioinformatics*, **23** (2007), 657–663. https://doi.org/10.1093/bioinformatics/btl646
20. B. Jackson, J. D. Scargle, D. Barnes, S. Arabhi, P. Gioumousis, E. Gwin, et al., An algorithm for optimal partitioning of data on an interval, *EEE Signal Process. Lett.*, **12** (2005), 105–108. https://doi.org/10.1109/LSP.2001.838216

21. K. Haynes, I. Eckley, P. Fearnhead, Computationally efficient changepoint detection for a range of penalties, *J. Comput. Graph. Stat.*, **26** (2017), 134–143. https://doi.org/10.1080/10618600.2015.1116445

22. N. Hao, Y. S. Niu, H. Zhang, Multiple change-point detection via a screening and ranking algorithm, *Stat. Sin.*, **23** (2013), 1553–1572. https://doi.org/10.5705/ss.2012.018s

23. D. Barry, J. Hartigan, A Bayesian analysis for change-point problems, *J. Amer. Stat. Assoc.*, **88** (1993), 309–319. https://doi.org/10.1080/01621459.1993.10594323

24. S. Chib, Estimation and comparison of multiple change-point models, *J. Econom.*, **86** (1998), 221–241. https://doi.org/10.1016/S0304-4076(97)00115-2

25. M. Lavielle, E. Lebarbier, An application of MCMC methods for the multiple change-points problem, *Digit. Signal Process.*, **81** (2001), 39–53. https://doi.org/10.1016/S0165-1684(00)00189-4

26. P. Fearnhead, P. Clifford, On-line inference for hidden Markov models via particle filters, *J. R. Stat. Soc. Ser. B*, **65** (2003), 887–899. https://doi.org/10.1111/1467-9868.00421

27. G. Koop, S. M. Potter, Estimation and forecasting in models with multiple breaks, *J. Econ. Stud.*, **74** (2007), 763–789. https://doi.org/10.1111/j.1467-937X.2007.00436.x

28. P. Fearnhead, Z. Liu, On-line inference for multiple changepoint problems, *J. R. Stat. Soc. Ser. B*, **69** (2007), 589–605. https://doi.org/10.1111/j.1467-9868.2007.00601.x

29. P. Giordani, R. Kohn, Efficient Bayesian inference for multiple change-point and mixture innovation models, *J. Bus. Econ. Stat.*, **26** (2008), 66–77. https://doi.org/10.1198/073500107000000241

30. A. F. Martinez, R. H. Mena, On a nonparametric change point detection model in Markovian regimes, *Bayesian Anal.*, **9** (2014), 823–858. https://doi.org/ 10.1214/14-BA878

31. P. Fearnhead, Exact and efficient Bayesian inference for multiple changepoint problems, *Stat. Comput.*, **16** (2006), 203–213. https://doi.org/10.1007/s11222-006-8450-8

32. E. Ruggieri, A Bayesian approach to detecting change points in climatic records, *Int. J. Climatol.*, **33** (2013), 520–528. https://doi.org/10.1002/joc.3447

33. S. Aminikhanghahi, D. J. Cook, A survey of methods for time series change point detection, *Knowl. Inf. Syst.*, **51** (2017), 339–367. https://doi.org/10.1007/s10115-016-0987-z

34. M. Bianchi, M. Boyle, D. Hollingsworth, A comparison of methods for trend estimation, *Appl. Econ. Lett.*, **6** (1999), 103–109. https://doi.org/10.1080/135048599353726

35. M. W. Robbins, R. B. Lund, C. M. Gallagher, Q. Lu, Changepoints in the north atlantic tropical cyclone record, *J. Am. Stat. Assoc.*, **106** (2011), 89–99. https://doi.org/10.1198/jasa.2011.ap10023

36. S. J. Kim, K. Koh, S. Boyd, D. Gorinevsky, l_1 trend filtering, *SIAM Rev.*, **51** (2009), 339–360. https://doi.org/10.1137/070690274

37. R. Baranowski, Y. Chen, P. Fryzlewicz, Narrowest-over-threshold detection of multiple change points and change-point-like features, *J. R. Stat. Soc. B*, **81** (2019), 649–672. https://doi.org/10.1111/rssb.12322

38. A. Anastasiou, P. Fryzlewicz, Detecting multiple generalized change-points by isolating single ones, *Metrika*, **85** (2022), 141–174. https://doi.org/10.1007/s00184-021-00821-6

39. H. Maeng, P. Fryzlewicz, Detecting linear trend changes in data sequences, *Stat. Papers*, **65** (2024), 1645–1675. https://doi.org/10.1007/s00362-023-01458-5

40. R. V. Mehrizi, S. Chenouri, Detection of change points in piecewise polynomial signals using trend filtering, *ArXiv*, 2020. https://doi.org/10.48550/arXiv.2009.08573

41. S. Gavioli-Akilagun, P. Fryzlewicz, Fast and optimal inference for change points in piecewise polynomials via differencing, *ArXiv*, 2023. https://doi.org/10.48550/arXiv.2307.03639

42. J. R. Faulkner, V. N. Minin, Locally adaptive smoothing with Markov random fields and shrinkage priors, *Bayesian Anal.*, **13** (2018), 225–252. https://doi.org/10.1214/17-BA1050

43. D. R. Kowal, D. S. Matteson, D. Ruppert, Dynamic shrinkage processes, *J. R. Stat. Soc. Ser. B*, **81** (2019), 781–804. https://doi.org/10.1111/rssb.12325

44. Y. Zhao, X. Liao, X. He, M. Zhou, C. Li, Accelerated primal-dual mirror dynamics for centralized and distributed constrained convex optimization problems, *J. Mach. Learn. Res.*, **24** (2023), 16413–16471. https://doi.org/10.5555/3648699.3649042

45. A. Shirilord, M. Dehghan, Stationary Landweber method with momentum acceleration for solving least squares problems, *Appl. Math. Lett.*, **157** (2024), 109174. https://doi.org/10.1016/j.aml.2024.109174

46. S. Arlot, A. Celisse, Segmentation of the mean of heteroscedastic data via cross-validation, *Stat. Comput.*, **21** (2011), 613–632. https://doi.org/10.1007/s11222-010-9196-x

47. V. M. R. Muggeo, G. Adelfio, Efficient change point detection for genomic sequences of continuous measurements, *Bioinformatics*, **27** (2011), 161–166. https://doi.org/10.1093/bioinformatics/btq647

48. F. Pein, H. Sieling, A. Munk, Heterogeneous change point inference. *J. R. Stat. Soc. Ser. B*, **79** (2017), 1207–1227. https://doi.org/10.1111/rssb.12202

49. J. O. Berger, L. R. Pericchi, The intrinsic Bayes factor for model selection and prediction, *J. Amer. Stat. Assoc.*, **91** (1996), 109–122. https://doi.org/10.2307/2291387

50. E. Moreno, F. Bertolino, W. Racugno, An intrinsic limiting procedure for model selection and hypotheses testing, *J. Amer. Stat. Assoc.*, **93** (1998), 1451–1460. https://doi.org/10.2307/2670059

51. J. O. Berger, L. R. Pericchi, On criticism and comparison of default Bayes factor for model selection and hypothesis testing, *Proceedings of the Workshop on Model Selection*, 1998.

52. E. Moreno, Bayes factor for intrinsic and fractional priors in nested models. Bayesian robustness, *IMS Lecture Notes Monogr. Ser.*, **1997** (1997), 257–270. https://doi.org/10.1214/LNMS/1215454142

53. A. J. Scott, M. Knott, A cluster analysis method for grouping means in the analysis of variance, *Biometrics*, **30** (1974), 507–512. https://doi.org/10.2307/2529204

54. L. J. Vostrikova, Detecting "disorder" in multidimensional random processes, *Sov. Math. Dokl.*, **24** (1981), 55–59.

55. P. Fryziewicz, Wild binary segmentation for multiple change-point detection, *Ann. Stat.*, **42** (2014), 2243–2281. https://doi.org/10.1214/14-AOS1245

56. G. Jensen, Closed-form estimation of multiple change-point models, *PeerJ Preprint*, **3** (2013), e90v3. https://doi.org/10.7287/peerj.preprints.90v3

57. R. E. Kass, A. Raftery, Bayes factors, *J. Amer. Stat. Assoc.*, **90** (1995), 773–795. https://doi.org/10.2307/2291091

58. G. Casella, E. Moreno, Objective Bayesian variable selection, *J. Amer. Stat. Assoc.*, **101** (2009), 157–167. https://doi.org/10.1198/016214505000000646

59. F. J. Girón, L. L. Martínez, E. Moreno, F. Torres, Objective testing procedures in linear models, *Scand. J. Stat.*, **33** (2006), 765–784. https://doi.org/10.1111/j.1467-9469.2006.00514.x

60. E. Keogh, S. Chu, D. Hart, M. Pazzani, Segmenting time series: a survey and novel approach, *Series in Machine Perception and Artificial Intelligence*, 2004. https://doi.org/10.1142/9789812565402_0001

61. J. Hansen, R. Ruedy, M. Sato, K. Lo, Global surface temperature change, *Rev. Geophys.*, **48** (2010), RG4004. https://doi.org/10.1029/2010RG000345

62. N. J. Lenssen, G. A. Schmidt, J. E. Hansen, M. J. Menne, A. Persin, R. Ruedy, et al., Improvements in the GISTEMP uncertainty model, *J. Geophys. Res. Atmos.*, **124** (2019), 6307–6326. https://doi.org/10.1029/2018JD029522

63. N. A. James, D. S. Matteson, Change points via probabilistically pruned objectives, *ArXiv*, 2015. https://doi.org/10.48550/arXiv.1505.04302

Appendix

Appendix A

Proof of Theorem 1. Consider the model M_1,

$$M_1 : N_{k+1}(\mathbf{Z}_1\boldsymbol{\beta}, \sigma^2\mathbf{I}_{k+1})N_{k+1}(\mathbf{Z}_2\boldsymbol{\beta}, \sigma^2\mathbf{I}_{k+1}), \quad \pi^N(\theta_1) = \frac{c_1}{\sigma},$$

and the model M_2,

$$M_2 : N_{k+1}(\mathbf{Z}_1\boldsymbol{\beta}_1, \sigma_1^2\mathbf{I}_{k+1})N_{k+1}(\mathbf{Z}_2\boldsymbol{\beta}_2, \sigma_2^2\mathbf{I}_{k+1}), \quad \pi^N(\theta_2) = \frac{c_2}{\sigma_1\sigma_2},$$

where c_1 and c_2 are the arbitrary positive constants. For the minimal training sample vectors $\mathbf{z}_1$ and $\mathbf{z}_2$, we have

$$B_{12}^N(\mathbf{z}_1, \mathbf{z}_2) = \frac{N_{k+1}(\mathbf{Z}_1\boldsymbol{\beta}, \sigma^2\mathbf{I}_{k+1})N_{k+1}(\mathbf{Z}_2\boldsymbol{\beta}, \sigma^2\mathbf{I}_{k+1})}{c_2\prod_{i=1}^2 \frac{1}{2}|\mathbf{Z}_i^T\mathbf{Z}_i|^{-1/2}\left[\mathbf{z}_i^T(\mathbf{I}-\mathbf{H}_i)\mathbf{z}_i\right]^{-1/2}},$$

where

$$\mathbf{H}_i = \mathbf{Z}_i(\mathbf{Z}_i^T\mathbf{Z}_i)^{-1}\mathbf{Z}_i^T, \quad i = 1, 2.$$

Since

$$E^{M_2}_{\mathbf{z}_1,\mathbf{z}_2|\theta_2}[B^N_{12}(\mathbf{z}_1,\mathbf{z}_2)] = \frac{1}{c_2}\prod_{i=1}^{2}\frac{2}{\pi}\frac{\sigma_i\sigma}{\sigma_i^2+\sigma^2}N_k(\boldsymbol{\beta}_i|\boldsymbol{\beta},(\sigma^2+\sigma_i^2)(\mathbf{Z}_i^T\mathbf{Z}_i)^{-1}),$$

the conditional intrinsic prior of θ_2 becomes

$$\begin{aligned}\pi^I(\theta_2|\theta_1) &= \pi^N(\theta_2)E^{M_2}_{\mathbf{z}_1,\mathbf{z}_2|\theta_2}[B^N_{12}(\mathbf{z}_1,\mathbf{z}_2)]\\ &= \prod_{i=1}^{2}\frac{2}{\pi}\frac{\sigma}{\sigma_i^2+\sigma^2}N_k(\boldsymbol{\beta}_i|\boldsymbol{\beta},(\sigma_i^2+\sigma^2)(\mathbf{Z}_i^T\mathbf{Z}_i)^{-1}).\end{aligned}$$

Hence, the Theorem 1 is proved. □

Appendix B

Proof of Theorem 2. We first compute the BF for comparing model M_2 versus M_1 with the intrinsic prior $\pi^I(\theta_2)$. Now

$$f(\mathbf{y}|\theta_1)\pi^N(\theta_1) = \frac{c_1}{\sigma_1}N_n(\mathbf{y}|\mathbf{X}\boldsymbol{\beta},\sigma^2\mathbf{I}_n). \tag{B.1}$$

Integrating with respect to $\boldsymbol{\beta}$ and σ in (B.1) yields

$$m_1(\mathbf{y}) = \frac{c_1}{2}\pi^{-\frac{n-k}{2}}\Gamma\left(\frac{n-k}{2}\right)|\mathbf{X}^T\mathbf{X}|^{-\frac{1}{2}}\left[\mathbf{y}^T(\mathbf{I}-\mathbf{H}_1)\mathbf{y}\right]^{-\frac{n-k}{2}}, \tag{B.2}$$

where

$$\mathbf{H}_1 = \mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T.$$

Next we have

$$f(\mathbf{y}|\theta_2)\pi^I(\theta_2|\theta_1)\pi^N(\theta_1) = \frac{4}{\pi^2}\frac{c_1}{\sigma}\prod_{i=1}^{2}\frac{\sigma}{\sigma_i^2+\sigma^2}N_k(\boldsymbol{\beta}_i|\boldsymbol{\beta},(\sigma_i^2+\sigma^2)\mathbf{W}_i^{-1})N_{n_i}(\mathbf{y}_i|\mathbf{X}_i\boldsymbol{\beta}_i,\sigma_i^2\mathbf{I}_{n_i}), \tag{B.3}$$

where

$$n_1 = \tau \quad \text{and} \quad n_2 = n-\tau.$$

Integrating with respect to $\boldsymbol{\beta}_1$, $\boldsymbol{\beta}_2$ and $\boldsymbol{\beta}$ in (B.3), then we get

$$f(\mathbf{y}|\sigma,\sigma_1,\sigma_2) = \frac{4}{\pi^2}\frac{c_1\sigma}{(\sigma_1^2+\sigma^2)(\sigma_2^2+\sigma^2)}\frac{\exp\{-\frac{H_{2*}}{2}\}}{(2\pi)^{\frac{n-k}{2}}|\Sigma_{1*}|^{\frac{1}{2}}|\Sigma_{2*}|^{\frac{1}{2}}|\Sigma_*|^{\frac{1}{2}}}, \tag{B.4}$$

where $n_1 = \tau$, $n_2 = n-\tau$,

$$\begin{aligned}&\mathbf{W}_1^{-1} = \frac{n_1}{k+1}(\mathbf{X}_1^T\mathbf{X}_1)^{-1}, \mathbf{W}_2^{-1} = \frac{n_2}{k+1}(\mathbf{X}_2^T\mathbf{X}_2)^{-1}, \Sigma_{1*} = \sigma_1^2\mathbf{I}_{n_1} + (\sigma_1^2+\sigma^2)\mathbf{X}_1\mathbf{W}_1^{-1}\mathbf{X}_1^T,\\ &\Sigma_{2*} = \sigma_2^2\mathbf{I}_{n_2} + (\sigma_2^2+\sigma^2)\mathbf{X}_2\mathbf{W}_2^{-1}\mathbf{X}_2^T, \Sigma_* = \mathbf{X}_1^T\Sigma_{1*}^{-1}\mathbf{X}_1 + \mathbf{X}_2^T\Sigma_{2*}^{-1}\mathbf{X}_2,\\ &\mathbf{H}_{2*} = \mathbf{y}_1^T\Sigma_{1*}^{-1}\mathbf{y}_1 + \mathbf{y}_2^T\Sigma_{2*}^{-1}\mathbf{y}_2 - (\mathbf{X}_1^T\Sigma_{1*}^{-1}\mathbf{y}_1 + \mathbf{X}_2^T\Sigma_{2*}^{-1}\mathbf{y}_2)^T\Sigma_*^{-1}(\mathbf{X}_1^T\Sigma_{1*}^{-1}\mathbf{y}_1 + \mathbf{X}_2^T\Sigma_{2*}^{-1}\mathbf{y}_2).\end{aligned}$$

Note that

$$\begin{aligned}
\Sigma_{1*} &= \sigma^2\left[\frac{\sigma_1^2}{\sigma^2}\mathbf{I}_{n_1} + \left(\frac{\sigma_1^2}{\sigma^2}+1\right)\mathbf{X}_1\mathbf{W}_1^{-1}\mathbf{X}_1^T\right] = \sigma^2\Sigma_1,\\
\Sigma_{2*} &= \sigma^2\left[\frac{\sigma_2^2}{\sigma^2}\mathbf{I}_{n_2} + \left(\frac{\sigma_2^2}{\sigma^2}+1\right)\mathbf{X}_2\mathbf{W}_2^{-1}\mathbf{X}_2^T\right] = \sigma^2\Sigma_2,\\
\Sigma_* &= \sigma^{-2}\left[\mathbf{X}_1^T\Sigma_1^{-1}\mathbf{X}_1 + \mathbf{X}_2^T\Sigma_2^{-1}\mathbf{X}_2\right] = \sigma^{-2}\Sigma,\\
\mathbf{H}_{2*} &= \sigma^{-2}\Big[\mathbf{y}_1^T\Sigma_1^{-1}\mathbf{y}_1 + \mathbf{y}_2^T\Sigma_2^{-1}\mathbf{y}_2\\
&\quad -(\mathbf{X}_1^T\Sigma_1^{-1}\mathbf{y}_1 + \mathbf{X}_2^T\Sigma_2^{-1}\mathbf{y}_2)^T\Sigma^{-1}(\mathbf{X}_1^T\Sigma_1^{-1}\mathbf{y}_1 + \mathbf{X}_2^T\Sigma_2^{-1}\mathbf{y}_2)\Big] = \sigma^{-2}H_2.
\end{aligned}$$

Let

$$\omega_1 = \sigma_1^2/\sigma^2, \quad \omega_2 = \sigma_2^2/\sigma^2 \quad \text{and} \quad z = \sigma^2.$$

Integrating with respect to z in (B.4), then we obtain

$$\begin{aligned}
m_2(\mathbf{y}) = \int_0^\infty\int_0^\infty \frac{c_1}{2\pi^2}(2\pi)^{-\frac{n-k}{2}}\Gamma\left(\frac{n-k}{2}\right)\omega_1^{-\frac{1}{2}}\omega_2^{-\frac{1}{2}}(1+\omega_1)^{-1}(1+\omega_2)^{-1}\\
\times|\Sigma_1|^{-\frac{1}{2}}|\Sigma_2|^{-\frac{1}{2}}|\Sigma|^{-\frac{1}{2}}\left(\frac{H_2}{2}\right)^{-\frac{n-k}{2}}d\omega_1 d\omega_2,
\end{aligned} \tag{B.5}$$

where

$$\begin{aligned}
|\Sigma_1| &= |\omega_1[\mathbf{I}_{n_1} + \frac{n_1}{k+1}\frac{1+\omega_1}{\omega_1}\mathbf{X}_1(\mathbf{X}_1^T\mathbf{X}_1)^{-1}\mathbf{X}_1^T]|\\
&= \omega_1^{n_1}|\mathbf{I}_k + \frac{n_1}{k+1}\frac{1+\omega_1}{\omega_1}\mathbf{X}_1^T\mathbf{X}_1(\mathbf{X}_1^T\mathbf{X}_1)^{-1}|\\
&= \omega_1^{n_1}\left[1+\frac{n_1}{k+1}\frac{1+\omega_1}{\omega_1}\right]^k,\\
|\Sigma_2| &= \omega_2^{n_2}\left[1+\frac{n_2}{k+1}\frac{1+\omega_2}{\omega_2}\right]^k,\\
\Sigma &= \mathbf{X}_1^T\Sigma_1^{-1}\mathbf{X}_1 + \mathbf{X}_2^T\Sigma_2^{-1}\mathbf{X}_2,\\
\Sigma_1^{-1} &= \left[\omega_1\mathbf{I}_{n_1} + \frac{n_1}{k+1}(1+\omega_1)\mathbf{X}_1(\mathbf{X}_1^T\mathbf{X}_1)^{-1}\mathbf{X}_1^T\right]^{-1}\\
&= \omega_1^{-1}\left[\mathbf{I}_{n_1} - \frac{n_1(1+\omega_1)}{(k+1)\omega_1}\left(1+\frac{n_1(1+\omega_1)}{(k+1)\omega_1}\right)^{-1}\mathbf{X}_1(\mathbf{X}_1^T\mathbf{X}_1)^{-1}\mathbf{X}_1^T\right],\\
\Sigma_2^{-1} &= \omega_2^{-1}\left[\mathbf{I}_{n_2} - \frac{n_2(1+\omega_2)}{(k+1)\omega_2}\left(1+\frac{n_2(1+\omega_2)}{(k+1)\omega_2}\right)^{-1}\mathbf{X}_2(\mathbf{X}_2^T\mathbf{X}_2)^{-1}\mathbf{X}_2^T\right],\\
\mathbf{X}_i^T\Sigma_i^{-1}\mathbf{X}_i &= \omega_i^{-1}\left[1-\frac{n_i(1+\omega_i)}{(k+1)\omega_i}\left(1+\frac{n_i(1+\omega_i)}{(k+1)\omega_i}\right)^{-1}\right]\mathbf{X}_i^T\mathbf{X}_i, \quad i=1,2,\\
\mathbf{y}_i^T\Sigma_i^{-1}\mathbf{y}_i &= \omega_i^{-1}\left[\mathbf{y}_i^T\mathbf{y}_i - \frac{n_i(1+\omega_i)}{(k+1)\omega_i}\left(1+\frac{n_i(1+\omega_i)}{(k+1)\omega_i}\right)^{-1}\mathbf{y}_i^T\mathbf{X}_i(\mathbf{X}_i^T\mathbf{X}_i)^{-1}\mathbf{X}_i^T\mathbf{y}_i\right], \quad i=1,2,
\end{aligned}$$

$$\mathbf{X}_i^T\Sigma_i^{-1}\mathbf{y}_i = \omega_i^{-1}\left[1 - \frac{n_i(1+\omega_i)}{(k+1)\omega_i}\left(1 + \frac{n_i(1+\omega_i)}{(k+1)\omega_i}\right)^{-1}\right]\mathbf{X}_i^T\mathbf{y}_i, \quad i = 1, 2$$

and

$$\mathbf{H}_2 = \sum_{i=1}^{2}\mathbf{y}_i^T\Sigma_i^{-1}\mathbf{y}_i - \left(\sum_{i=1}^{2}\mathbf{X}_i^T\Sigma_i^{-1}\mathbf{y}_i\right)^T\Sigma^{-1}\left(\sum_{i=1}^{2}\mathbf{X}_i^T\Sigma_i^{-1}\mathbf{y}_i\right).$$

Hence, the Theorem 2 is proved. □

Appendix C

Proof of Theorem 3. Consider the model M_i,

$$M_i : N_{k+1}(\mathbf{Z}_i\boldsymbol{\alpha}_i, \sigma_i^2\mathbf{I}_{k+1}), \quad \pi_i^N(\theta_i) = \frac{c_i}{\sigma_i},$$

and the model M_k,

$$M_k : N_{k+1}(\mathbf{Z}_k\boldsymbol{\beta}_k, \sigma_k^2\mathbf{I}_{k+1}), \quad \pi_k^N(\theta_k) = \frac{c_k}{\sigma_k},$$

where c_i and c_k are an arbitrary positive constants. For the minimal training sample vector $\mathbf{z}$, we have

$$B_{ik}^N(\mathbf{z}) = \frac{N_{k+1}(\mathbf{Z}_i\boldsymbol{\alpha}_i, \sigma_i^2\mathbf{I})}{\frac{c_k}{2}|\mathbf{Z}_k^T\mathbf{Z}_k|^{-1/2}\left[\mathbf{z}^T(\mathbf{I} - \mathbf{H}_k)\mathbf{z}\right]^{-1/2}},$$

where

$$\mathbf{H}_k = \mathbf{Z}_k(\mathbf{Z}_k^T\mathbf{Z}_k)^{-1}\mathbf{Z}_k^T.$$

Since

$$E_{\mathbf{z}|\theta_k}^{M_k}[B_{ik}^N(\mathbf{z})] = \frac{1}{c_k}\frac{2}{\pi}\frac{\sigma_i\sigma_k}{\sigma_k^2 + \sigma_i^2}N_k(\boldsymbol{\beta}_k|\tilde{\boldsymbol{\alpha}}_i, (\sigma_k^2 + \sigma_i^2)(\mathbf{Z}_k^T\mathbf{Z}_k)^{-1}),$$

where

$$\tilde{\boldsymbol{\alpha}}_i = (\alpha_1, \cdots, \alpha_i, 0, \cdots, 0)^T$$

is $k \times 1$ vector, the conditional intrinsic prior of θ_2 becomes

$$\begin{aligned}\pi_k^I(\theta_k|\theta_i) &= \pi_k^N(\theta_k)E_{\mathbf{z}|\theta_k}^{M_k}[B_{ik}^N(\mathbf{z})] \\ &= \frac{2}{\pi}\frac{\sigma_i}{\sigma_k^2 + \sigma_i^2}N_k(\boldsymbol{\beta}_k|\tilde{\boldsymbol{\alpha}}_i, (\sigma_k^2 + \sigma_i^2)(\mathbf{Z}_k^T\mathbf{Z}_k)^{-1}).\end{aligned}$$

In addition, $(\mathbf{Z}_k^T\mathbf{Z}_k)^{-1}$ is $\frac{n_j}{k+1}(\mathbf{X}_k^T\mathbf{X}_k)^{-1}$ by Girón et al. [59].

Hence, the Theorem 3 is proved. □

Appendix D

Proof of Theorem 4. We first compute the BF for comparing model M_i versus M_k with the intrinsic prior $\pi_k^I(\theta_k)$. We put $n = n_j$ in our proof. Now

$$f(\mathbf{y}|\theta_i)\pi_i^N(\theta_i) = \frac{c_i}{\sigma_i} N_n(\mathbf{y}|\mathbf{X}_i\boldsymbol{\alpha}_i, \sigma_i^2\mathbf{I}_n). \tag{D.1}$$

Integrating with respect to $\boldsymbol{\alpha}_i$ and σ_i in (D.1) yields

$$m_i(\mathbf{y}) = \frac{c_i}{2}(2\pi)^{-\frac{n-i}{2}}\Gamma\left(\frac{n-i}{2}\right)|\mathbf{X}_i^T\mathbf{X}_i|^{-\frac{1}{2}}\left[\frac{1}{2}\mathbf{y}^T(\mathbf{I}_n - \mathbf{H}_i)\mathbf{y}\right]^{-\frac{n-i}{2}}, \tag{D.2}$$

where

$$\mathbf{H}_i = \mathbf{X}_i(\mathbf{X}_i^T\mathbf{X}_i)^{-1}\mathbf{X}_i^T.$$

Next we have

$$f(\mathbf{x}|\theta_k)\pi_k^I(\theta_k|\theta_i)\pi_k^N(\theta_i) = \frac{2c_i}{\pi}\frac{1}{\sigma_k^2+\sigma_i^2} N_n(\mathbf{y}|\mathbf{X}_k\boldsymbol{\beta}_k, \sigma_k^2\mathbf{I}_n) N_k(\boldsymbol{\beta}_k|\tilde{\boldsymbol{\alpha}}_i, \frac{n}{k+1}(\sigma_k^2+\sigma_i^2)(\mathbf{X}_k^T\mathbf{X}_k)^{-1}). \tag{D.3}$$

Integrating with respect to $\boldsymbol{\beta}_k$ and $\boldsymbol{\alpha}_i$ in (D.3), then we get

$$\frac{2c_i}{\pi}\frac{1}{\sigma_k^2+\sigma_i^2}\frac{\exp\{-\frac{1}{2}\mathbf{y}^T\Sigma_A\mathbf{y}\}}{(2\pi)^{\frac{n-i}{2}}|\Sigma|^{\frac{1}{2}}|\mathbf{X}_i^T\Sigma^{-1}\mathbf{X}_i|^{\frac{1}{2}}}, \tag{D.4}$$

where

$$\Sigma = \sigma_k^2\mathbf{I}_n + \frac{n(\sigma_k^2+\sigma_i^2)}{k+1}\mathbf{X}_k(\mathbf{X}_k^T\mathbf{X}_k)^{-1}\mathbf{X}_k^T$$

and

$$\Sigma_A = \Sigma^{-1} - \Sigma^{-1}\mathbf{X}_i(\mathbf{X}_i^T\Sigma^{-1}\mathbf{X}_i)^{-1}\mathbf{X}_i^T\Sigma^{-1}.$$

Note that

$$\begin{aligned}
|\Sigma| &= |\sigma_k^2[\mathbf{I}_n + \frac{n}{k+1}(1+\sigma_i^2/\sigma_k^2)\mathbf{X}_k(\mathbf{X}_k^T\mathbf{X}_k)^{-1}\mathbf{X}_k^T]| \\
&= \sigma_k^{2n}|\mathbf{I}_k + \frac{n}{k+1}(1+\sigma_i^2/\sigma_k^2)\mathbf{X}_k^T\mathbf{X}_k(\mathbf{X}_k^T\mathbf{X}_k)^{-1}| \\
&= \sigma_k^{2n}\left[1 + \frac{n}{k+1}(1+\sigma_i^2/\sigma_k^2)\right]^k, \\
\Sigma^{-1} &= \sigma_k^{-2}\left[\mathbf{I}_n - \left(1 + \frac{k+1}{n(1+\sigma_i^2/\sigma_k^2)}\right)^{-1}\mathbf{X}_k(\mathbf{X}_k^T\mathbf{X}_k)^{-1}\mathbf{X}_k^T\right], \\
\mathbf{X}_i^T\Sigma^{-1}\mathbf{X}_i &= \sigma_k^{-2}\left[1 - \left(1 + \frac{k+1}{n(1+\sigma_i^2/\sigma_k^2)}\right)^{-1}\right]\mathbf{X}_i^T\mathbf{X}_i
\end{aligned}$$

and

$$\Sigma_A = \sigma_k^{-2}\left\{\mathbf{I}_n - \left(1 + \frac{k+1}{n(1+\sigma_i^2/\sigma_k^2)}\right)^{-1}\mathbf{X}_k(\mathbf{X}_k^T\mathbf{X}_k)^{-1}\mathbf{X}_k^T - \left[1 - \left(1 + \frac{k+1}{n(1+\sigma_i^2/\sigma_F^2)}\right)^{-1}\right]\mathbf{X}_i(\mathbf{X}_i^T\mathbf{X}_i)^{-1}\mathbf{X}_i^T\right\}.$$

Let

$$w = \sigma_k^2 \quad \text{and} \quad z = \sigma_i^2/\sigma_k^2.$$

Integrating with respect to w in (B.4), then we obtain

$$m_k(\mathbf{y}) = \int_0^\infty \frac{c_i}{2\pi}(2\pi)^{-\frac{n-i}{2}}\Gamma\left(\frac{n-i}{2}\right)z^{-\frac{1}{2}}(1+z)^{-1}\left[1+\frac{n}{k+1}(z+1)\right]^{-\frac{k-i}{2}}$$
$$\times|\mathbf{X}_i^T\mathbf{X}_i|^{-\frac{1}{2}}\left\{\frac{1}{2}\mathbf{y}^T\left[\mathbf{I}_n-\mathbf{H}_i+c(z)^{-1}(\mathbf{H}_i-\mathbf{H}_k)\right]\mathbf{y}\right\}^{-\frac{n-i}{2}}dz, \tag{D.5}$$

where

$$c(z) = 1+\frac{k+1}{n(1+z)}, \quad \mathbf{H}_i = \mathbf{X}_i(\mathbf{X}_i^T\mathbf{X}_i)^{-1}\mathbf{X}_i^T$$

and

$$\mathbf{H}_k = \mathbf{X}_k(\mathbf{X}_k^T\mathbf{X}_k)^{-1}\mathbf{X}_k^T.$$

Hence, the Theorem 4 is proved. □

AIMS Press