The Annals of Applied Statistics
2019, Vol. 13, No. 3, 1430–1463
https://doi.org/10.1214/19-AOAS1242

IDENTIFYING MULTIPLE CHANGES FOR A FUNCTIONAL DATA SEQUENCE WITH APPLICATION TO FREEWAY TRAFFIC SEGMENTATION

BY JENG-MIN CHIOU[*,1], YU-TING CHEN[†] AND TAILEN HSING[‡]

Academia Sinica[*], National Cheng Chi University[†] and University of Michigan[‡]

Motivated by the study of road segmentation partitioned by shifts in traffic conditions along a freeway, we introduce a two-stage procedure, Dynamic Segmentation and Backward Elimination (DSBE), for identifying multiple changes in the mean functions for a sequence of functional data. The Dynamic Segmentation procedure searches for all possible changepoints using the derived global optimality criterion coupled with the local strategy of at-most-one-changepoint by dividing the entire sequence into individual subsequences that are recursively adjusted until convergence. Then, the Backward Elimination procedure verifies these changepoints by iteratively testing the unlikely changes to ensure their significance until no more changepoints can be removed. By combining the local strategy with the global optimal changepoint criterion, the DSBE algorithm is conceptually simple and easy to implement and performs better than the binary segmentation-based approach at detecting small multiple changes. The consistency property of the changepoint estimators and the convergence of the algorithm are proved. We apply DSBE to detect changes in traffic streams through real freeway traffic data. The practical performance of DSBE is also investigated through intensive simulation studies for various scenarios.

1. Introduction. This study is motivated by the application of detecting multiple changes in a long sequence of traffic streams concerning traffic speeds. Homogeneous traffic conditions within a freeway segment are essential for determining the level-of-service, a widely used performance measure to assess the freeway operations [Transportation Research Board (2010)], and are useful to traffic sensor deployment [Hu and Wang (2008)]. It is a common practice to monitor traffic streams through a series of dual loop vehicle detectors located along freeways. The continuously recorded traffic data records allow us to develop a systematic approach to identifying the locations of changes in traffic streams. We regard the daily traffic speed trajectory collected at each detector as a realization sampled from a stochastic process and place it in the context of functional data. Thus, the

Received May 2018; revised October 2018.

[1]Supported by Grants AS-IA-105-M01 from Academia Sinica and MOST-104-2118-M-001-015-MY3 and MOST-107-2118-M-001-001-MY3 from the Ministry of Science and Technology, Taiwan, R.O.C.

Key words and phrases. Changepoint analysis, covariance operator, functional principal component, projection, segmentation.

traffic-speed trajectories collected by a series of vehicle detectors along a freeway form a long sequence of functional data that are likely to contain multiple changepoints rather than just a single one. As such, we formulate the segmentation problem as a multiple changepoint analysis for a sequence of functional observations.

Changepoint analysis has broad applications in various disciplines, such as biology, climatology, economics and engineering. It is an active research area and has been discussed extensively in the univariate and multivariate data settings. While studies with a single changepoint had a long history, problems with multiple changepoints are much more challenging and have received more attention in the last decade. Niu, Hao and Zhang (2016) provided an excellent selective overview on both classical and new multiple changepoint detection strategies, including exhaustive search, stepwise selection, penalization, screening and ranking, and multiscale estimators, such as Killick, Fearnhead and Eckley (2012), Fryzlewicz (2014), Harchaoui and Lévy-Leduc (2010), Niu and Zhang (2012), Frick, Munk and Sieling (2014), to name a few.

Despite the rapidly growing field of functional data, changepoint analysis for functional data has been studied only in recent years. Functional data analysis (FDA) studies theory and analysis of data sampling from infinite-dimensional random objects in the forms of functions, curves or images, or other general objects. We refer to Ramsay and Silverman (2005), Ferraty and Vieu (2006), Horváth and Kokoszka (2012), Zhang (2014), Hsing and Eubank (2015) for general overview and monographs of FDA and a recent review article Wang, Chiou and Müller (2016), among others.

In the literature, most changepoint analyses for functional data use the assumption of at most one changepoint (AMOC). Berkes et al. (2009) proposed a CUSUM-type test for changepoint detection in a sequence of independent functional data, and the asymptotic theory for the changepoint estimator was studied by Aue et al. (2009). Subsequently, the CUSUM-type test was extended further to accommodate functional data under a weakly dependent structure [Hörmann and Kokoszka (2010)]. Additionally, Aston and Kirch (2012) modified the test statistics to be suitable for epidemic changes in which the mean function changes at some time point and then returns to the original one. Very recently, Gromenko, Kokoszka and Reimherr (2017) developed a statistical tool for inferring a change in the pattern of a spatio-temporal process, and the testing statistic reduced to that of Berkes et al. (2009) when there is only a single site.

While these pioneering works mainly assume a single changepoint in a sequence of functional data, this assumption may not be suitable for a long series of data. To relax the assumption in a real-data example, Berkes et al. (2009) extended the method to detecting multiple changes by combining the test of detecting a single change with the binary segmentation algorithm. Although it serves as a remedy for the AMOC assumption, the binary segmentation algorithm requires a condition regarding the minimum segment length and may omit small changes when both large and small segments exist simultaneously [Olshen et al. (2004), Fryzlewicz (2014)];

this inconsistency is due to the violation of the AMOC assumption. A reasonable strategy of detecting multiple changes is to perform a single changepoint detection within a small subsequence that satisfies the AMOC assumption; this is referred as *the local strategy*. Recently, methods adopting the local strategy for multiple changepoint detection have been developed [Niu and Zhang (2012), Fryzlewicz (2014)] but these were designed for scalar rather than functional data.

In addition to hypothesis testing, the detection of multiple changepoints in univariate and multivariate settings is often regarded as a global optimization problem, in which changepoints are characterized as an optimal set, assuming the number of changepoints is known [Pan and Chen (2006)]. For situations in which the number of changepoints is unknown, Yao (1988) and Braun, Braun and Müller (2000) have proposed using the Schwarz criterion to estimate the total number of changepoints; a general criterion can be found in Ciuperca (2011). Although the consistency for these optimization estimators has been established theoretically, the computational burden remains. To reduce the computational complexity, Braun, Braun and Müller (2000) employed dynamic programming and Killick, Fearnhead and Eckley (2012) proposed an algorithm with linear computational cost but under a more restrictive assumption, which may be impractical [Fryzlewicz (2014)].

In this study, we introduce a new two-stage procedure, the Dynamic Segmentation and Backward Elimination (DSBE), for identifying multiple changes in the mean functions for a sequence of functional data. In the first stage, the Dynamic Segmentation (DS) searches for all possible changepoints in each individual subsequences satisfying the AMOC assumption, and these subsequences are recursively adjusted to achieve the objective of global optimality. In the second stage, the Backward Elimination (BE) procedure aims to ensure statistical significance of these changepoints, which iteratively performs hypothesis testing starting from the most unlikely changepoints until no more changepoint candidates can be removed. The concept of DSBE is straightforward and easy to understand. The derived global optimality criterion in DS characterizes the changepoints, and therefore can be quite effective for any general changepoint detection problem for a functional data sequence.

To the best of our knowledge, there is no method specifically designed for identifying multiple changes in a sequence of functional data. The most relevant approaches could be extensions from a single changepoint detection for functional data using CUSUM-based hypothesis testing coupled with algorithms such as the binary segmentation (BS) algorithm and could be further generalized to methods such as wild binary segmentation and circular binary segmentation.

Unlike the BS-based approaches, DSBE accommodates the existence of multiple changepoints and does not impose the AMOC restriction in its global objective function. Most importantly, DSBE achieves its global optimality by incorporating the local strategy of AMOC, which is easy to implement even for a functional data sequence in comparison with those using other global optimization methods for multiple changepoint analysis designed for one-dimensional data.

Besides, since the BS algorithm is a forward selection procedure, the number of tests required for changepoint detection is unknown, and the overall type I error is hard to control. In contrast, the proposed DSBE uses a backward testing algorithm and can handle the multiple testing problems.

The remainder of this article is organized as follows. Section 2 introduces a multiple changepoint model for a functional data sequence and develops a global optimality criterion for detecting multiple changes. Section 3 presents the DSBE algorithm and its theoretical properties. Section 4 illustrates the DSBE method with an application to freeway traffic segmentation. Section 5 investigates the finite sample performance of DSBE by using a simulation study. Discussions and concluding remarks are provided Section 6. Proofs of Lemma 1, Theorems 1 and 2 are given in the Appendix. Additional simulation results are deferred to Supplementary Material (Chiou, Chen and Hsing (2019)).

2. Multiple changepoint model and optimal segmentation estimator.

2.1. *Multiple changepoint model.* Let $\{X_i\}_{i=1}^N$ be a sequence of continuous random functions in $\mathcal{L}^2(\mathcal{T})$, a Hilbert space of square integrable functions on a closed interval $\mathcal{T}$ satisfying $E\int_{\mathcal{T}}|X_i(t)|^2\,dt<\infty$, with the inner product $\langle f,g\rangle=\int_{\mathcal{T}}f(t)g(t)\,dt$ for any f and g in $\mathcal{L}^2(\mathcal{T})$ and the induced norm $\|f\|^2=\langle f,f\rangle$. The functional data X_i are observed at locations within some bounded interval which, without loss of generality, will be taken as $(0,1]$. Assume that X_i is observed at location i/N.

We assume that there are M changes, $M<N$, in the mean functions at the positions $\{\theta_m^*\}_{m=1}^M$, $0<\theta_1^*<\cdots<\theta_M^*<1$, where both M and $\{\theta_m^*\}_{m=1}^M$ are fixed and unknown, and do not depend on N. The boundary points are denoted by $0=\theta_0^*$ and $1=\theta_{M+1}^*$. Furthermore, let μ_m be the mean function of the X_i's, with the associated position lying in the segment $(\theta_{m-1}^*,\theta_m^*]$, for $m=1,\ldots,(M+1)$, and assume that $\mu_m\neq\mu_{m+1}$ for model identifiability.

Under these assumptions, the functional multiple changepoint model for $X_i(t)$, $t\in\mathcal{T}$, is

$$X_i(t)=Y_i(t)+\sum_{m=1}^{M+1}\mu_m(t)\cdot\mathbf{1}_{(\theta_{m-1}^*,\theta_m^*]}(i/N), \tag{1}$$

where $\mathbf{1}_{(\theta_{m-1}^*,\theta_m^*]}(\theta)=1$ if $\theta_{m-1}^*<\theta\le\theta_m^*$ and 0 otherwise, and $\{Y_i\}_{i=1}^N$ is a sequence of identically distributed random functions in $\mathcal{L}^2(\mathcal{T})$ with mean zero and covariance function $c(s,t)=\operatorname{cov}(Y_i(s),Y_i(t))$. The integral operator with the kernel $c(s,t)$ is

$$\mathcal{C}f=\int_{\mathcal{T}}c(s,\cdot)f(s)\,ds,\quad f\in\mathcal{L}^2(\mathcal{T}),$$

which is a bounded linear transformation from $\mathcal{L}^2(\mathcal{T})$ into itself. Since $c(s,t)$ is continuous, $\mathcal{C}$ is a Hilbert–Schmidt operator. In general, the Hilbert–Schmidt norm

of a Hilbert–Schmidt operator $\mathcal{K}$ is defined as $\|\mathcal{K}\|_{\mathrm{HS}} = (\sum_{\ell,d=1}^{\infty}\langle \mathcal{K}e_\ell, e_d\rangle^2)^{1/2}$ with an arbitrary choice of complete orthonormal basis functions $\{e_\ell\}$ of $\mathcal{L}^2(\mathcal{T})$. See Hsing and Eubank (2015) for details of these notions.

Moreover, $\{Y_i\}$ may be dependent because the elements of the functional data sequence can be spatially or temporally correlated. We assume that $\{Y_i\}$ is a $\mathcal{L}^4 - m$-approximable functional data sequence [Hörmann and Kokoszka (2010)]. This assumption is required mainly to achieve estimation consistency as follows. Let

$$\bar{X}_{(\theta^*_{m-1},\theta^*_m]}(t) = \frac{1}{N_k(\boldsymbol{\theta}^*)} \sum_{i=\lfloor N\theta^*_{m-1}\rfloor+1}^{\lfloor N\theta^*_m\rfloor} X_i(t), \tag{2}$$

where $N_k(\boldsymbol{\theta}^*) = \lfloor N\theta^*_m\rfloor - \lfloor N\theta^*_{m-1}\rfloor$, the number of subjects or functional observations with the associated positions falling into $(\theta^*_{m-1}, \theta^*_m]$. Further, let

$$\hat{\mu}_i\big(t \mid \boldsymbol{\theta}^*\big) = \sum_{m=1}^{M+1} \bar{X}_{(\theta^*_{m-1},\theta^*_m]}(t) \cdot \mathbf{1}_{(\theta^*_{m-1},\theta^*_m]}(i/N).$$

Under the $\mathcal{L}^4 - m$-approximable assumption, it holds that given the true changepoints $\boldsymbol{\theta}^*$ the empirical estimator $\hat{\mathcal{C}}_{\boldsymbol{\theta}^*}$ with the kernel

$$\hat{c}\big(s,t \mid \boldsymbol{\theta}^*\big) = \frac{1}{N}\sum_{i=1}^{N}\big\{X_i(s) - \hat{\mu}_i\big(s \mid \boldsymbol{\theta}^*\big)\big\}\big\{X_i(t) - \hat{\mu}_i\big(t \mid \boldsymbol{\theta}^*\big)\big\},$$

converges in probability to the operator $\mathcal{C}$ with the kernel $c(s,t)$ in the Hilbert–Schmidt norm as $N \to \infty$. Nevertheless, other dependency assumption can be considered as long as that $\|\hat{\mathcal{C}}_{\boldsymbol{\theta}^*} - \mathcal{C}\|^2_{\mathrm{HS}} = o_p(1)$ can be satisfied.

In the functional multiple changepoint model (1), the number of changepoints M along with their positions $\{\theta^*_m\}_{m=1}^{M}$ and the segment mean functions $\{\mu_m\}_{m=1}^{M+1}$ are the unknown parameters to be estimated. Before presenting the method of estimating the changepoints, we first characterize the changepoints in a functional data sequence through a certain optimality property, which can be used to derive a criterion for detecting multiple changepoints.

2.2. *Optimal segmentation estimator.* For any $K > 0$, consider the $\Theta = \{\boldsymbol{\theta} = (\theta_1, \ldots, \theta_K) \mid 0 < \theta_1 < \cdots < \theta_K < 1\} \subset (0,1]^K$. An arbitrary $\boldsymbol{\theta} \in \Theta$ forms a K-segmentation on $(0,1]$ with $(K+1)$ nonoverlapping segments. Let $\theta_0 = 0$ and $\theta_{K+1} = 1$ for notational convenience. Let $\boldsymbol{\theta}^* = \{\theta^*_m; m = 1, \ldots, M\}$ be the set of true changepoints, which forms the M-segmentation. In the following, we show that among all possible $\boldsymbol{\theta}$ the true $\boldsymbol{\theta}^*$ possesses an optimality property.

Because $\{Y_i(t)\}$ in (1) are not observable and $\boldsymbol{\theta}^*$ is unknown, the covariance function $c(s,t) = EY_i(s)Y_i(t)$ is estimated through $\{X_i(t)\}$ by the empirical estimator conditional on $\boldsymbol{\theta}$,

$$\hat{c}(s,t|\boldsymbol{\theta}) = \frac{1}{N}\sum_{i=1}^{N}\hat{Y}_i(s|\boldsymbol{\theta})\hat{Y}_i(t|\boldsymbol{\theta}), \tag{3}$$

where

$$\hat{Y}_i(t|\boldsymbol{\theta}) = X_i(t) - \sum_{k=1}^{K+1} \bar{X}_{(\theta_{k-1},\theta_k]}(t) \cdot \mathbf{1}_{(\theta_{k-1},\theta_k]}(i/N) \tag{4}$$

for $i = 1, \ldots, N$ and $\bar{X}_{(\theta_{k-1},\theta_k]}(t)$ is defined as in (2). Similarly to the definition of $\mathcal{C}$, we define the integral operator with the kernel $\hat{c}(s,t|\boldsymbol{\theta})$ as $\hat{\mathcal{C}}_{\boldsymbol{\theta}}$, using the subscript to indicate its dependence on $\boldsymbol{\theta}$. The conditional estimator $\hat{c}(s,t|\boldsymbol{\theta})$ may not be consistent with $c(s,t)$. The discrepancy between $\mathcal{C}$ and $\hat{\mathcal{C}}_{\boldsymbol{\theta}}$ for an arbitrary K-segmentation $\boldsymbol{\theta}$ depends on the relationship between $\boldsymbol{\theta}$ and $\boldsymbol{\theta}^*$, which is explicitly explained as follows.

LEMMA 1. *Consider an arbitrary K-segmentation $\boldsymbol{\theta}$ of $(0,1]$ such that*

$$\min_{k\in\{1,\ldots,K+1\}} \{\theta_k - \theta_{k-1}\} > \epsilon,$$

for some $\epsilon > 0$. Under model (1) and its assumptions, we have

$$\hat{\mathcal{C}}_{\boldsymbol{\theta}} \xrightarrow{\mathcal{P}} \mathcal{C} + \mathcal{B}(\boldsymbol{\theta}), \tag{5}$$

where $\mathcal{B}(\boldsymbol{\theta})$ is the integral operator with the kernel

$$B_{\boldsymbol{\theta}}(s,t) = \sum_{1\le r<m\le K+1} \alpha_{r,m}(\boldsymbol{\theta})(\mu_r - \mu_m)(s)(\mu_r - \mu_m)(t),$$

and $\alpha_{r,m}(\boldsymbol{\theta}) \ge 0$ are constants depending on $\boldsymbol{\theta}$ as well as $\boldsymbol{\theta}^*$. Moreover, $\alpha_{r,m}(\boldsymbol{\theta}) = 0$ for all r and m if and only if $\boldsymbol{\theta}$ contains $\boldsymbol{\theta}^*$ as a subset; that is, if $\boldsymbol{\theta}^* \subset \boldsymbol{\theta}$.

The condition that the minimal distance of the adjacent changepoints in $\boldsymbol{\theta}$ is bounded away from zero is to assure that $N_k(\boldsymbol{\theta})$ and N are of the same order of magnitude as $N \to \infty$. The proof of Lemma 1 and the specific expression of $\alpha_{r,m}(\boldsymbol{\theta})$ are provided in Appendix A. Lemma 1 indicates that the difference between $\hat{\mathcal{C}}_{\boldsymbol{\theta}}$ and $\mathcal{C}$ hinges on $\alpha_{r,m}(\boldsymbol{\theta})$, the size of which relies on the difference between $\boldsymbol{\theta}$ and $\boldsymbol{\theta}^*$ as shown in the proof. Most importantly, when $\boldsymbol{\theta}^*$ is a subset of $\boldsymbol{\theta}$, $\hat{\mathcal{C}}_{\boldsymbol{\theta}}$ is a consistent estimator of $\mathcal{C}$, and vice versa.

It is easy to see that the kernel $B_{\boldsymbol{\theta}}(s,t)$ in Lemma 1 is nonnegative definite, which implies that $\hat{\mathcal{C}}_{\boldsymbol{\theta}} - \mathcal{C}$ is asymptotically nonnegative definite. Consequently, for any $p = 1, 2, \ldots,$

$$\sum_{\ell=1}^{p} \langle \hat{\mathcal{C}}_{\boldsymbol{\theta}} e_\ell, e_\ell \rangle \xrightarrow{p} \sum_{\ell=1}^{p} \langle \{\mathcal{C} + \mathcal{B}(\boldsymbol{\theta})\} e_\ell, e_\ell \rangle \ge \sum_{\ell=1}^{p} \langle \mathcal{C} e_\ell, e_\ell \rangle \tag{6}$$

and the equality holds if either $\alpha_{r,m}(\boldsymbol{\theta}) = 0$ or $\langle \mu_r - \mu_m, e_\ell \rangle = 0$ for all $1 \le r < m \le K+1$ and $1 \le \ell \le p$. When using $\min_{\boldsymbol{\theta}\in\Theta} \sum_{\ell=1}^{p} \langle \hat{\mathcal{C}}_{\boldsymbol{\theta}} e_\ell, e_\ell \rangle$ as the optimality criterion for $\boldsymbol{\theta}$, we require the following condition on $(\mu_r - \mu_m)(t)$ for the identifiability of $\boldsymbol{\theta}^*$.

DEFINITION. Under model (1), $\boldsymbol{\theta}^*$ is said to be *detectable with respect to* V if $(\mu_{m+1} - \mu_m) \notin V^\perp$ for $m = 1, \ldots, M$ for a subspace $V = \operatorname{span}\{e_1, \ldots, e_p\}$ in $\mathcal{L}^2(\mathcal{T})$.

Under the assumptions of Lemma 1 and assuming that $\boldsymbol{\theta}^*$ is detectable with respect to V, we conclude that the equality in (6) holds if and only if $\boldsymbol{\theta}^* \subset \boldsymbol{\theta}$; that is, $\alpha_{r,m}(\boldsymbol{\theta}) = 0$. Therefore, the true changepoints $\boldsymbol{\theta}^*$ can be characterized as an optimal segmentation that minimizes $\sum_{\ell=1}^{p} \langle \hat{\mathcal{C}}_{\boldsymbol{\theta}} e_\ell, e_\ell \rangle$ with respect to $\boldsymbol{\theta}$.

A data-driven approach to choosing a subspace V uses the eigenspace of the covariance operator corresponding to the empirical estimator $\hat{k}(s,t) = (1/N) \times \sum_{i=1}^{N} \{(X_i - \bar{X})(s)(X_i - \bar{X})(t)\}$, where $\bar{X}(t) = (1/N) \sum_{i=1}^{N} X_i(t)$. Specifically, we consider the spectral decomposition $\hat{k}(s,t) = \sum_{\ell=1}^{\infty} \hat{\lambda}_\ell \hat{\phi}_\ell(s) \hat{\phi}_\ell(t)$, where the eigenvalues $\{\hat{\lambda}_\ell\}$ are in nonascending order and the set of eigenfunctions $\{\hat{\phi}_\ell\}$ forms an orthonormal basis in $\mathcal{L}^2(\mathcal{T})$. Here, $\hat{k}(s,t)$ is the empirical estimator of the covariance kernel function $k(s,t) = \sum_{\ell=1}^{\infty} \lambda_\ell \phi_\ell(s) \phi_\ell(t)$, leading to a subspace $V = \operatorname{span}\{\hat{\phi}_1, \ldots, \hat{\phi}_p\}$, where the dimension p can be determined by the proportion of variance explained by the criterion, such that, given some $0 < \delta < 1$, $p(\delta) = \min\{p : \sum_{\ell=1}^{p} \hat{\lambda}_\ell / \sum_{\ell=1}^{\infty} \hat{\lambda}_\ell > \delta\}$. For instance, δ can be set as 95%.

Choosing V by the data-driven approach using the eigenspace of the covariance operator allows us to efficiently identify the space that maximizes the total variance of the processes with the minimal number of components. Nevertheless, other basis functions such as B-splines and Fourier basis can also be used as the basis functions. Because the estimate of the covariance kernel function $c(s,t)$ depends on the unknown changepoints, it is natural to use its empirical counterpart $\hat{k}(s,t)$. More importantly, it has been shown that working with $\hat{k}(s,t)$ is as effective as $c(s,t)$ in the sense that whenever a changepoint θ_m^* is detectable with respect to the eigenspace of $c(s,t)$, it is also detectable with respect to the eigenspace of $\hat{k}(s,t)$ (Section 4 in Aston and Kirch (2012)). Furthermore, in this study using the eigenspace of the covariance operator in the dynamic segmentation (DS) stage is in line with the tests for the equality of two covariance operators (Fremdt et al. (2013)) that is used in the backward elimination (BS) procedure, which also motivates the use of the eigenspace of the covariance operator in the DS procedure.

Let $T_N(\boldsymbol{\theta}) = \sum_{\ell=1}^{p} \langle \hat{\mathcal{C}}_{\boldsymbol{\theta}} \hat{\phi}_\ell, \hat{\phi}_\ell \rangle$ be the trace of $\hat{\mathcal{C}}_{\boldsymbol{\theta}}$ restricted to the span of $\{\hat{\phi}_\ell; \ell = 1, \ldots, p\}$. By (3), $T_N(\boldsymbol{\theta})$ can be expressed as

$$T_N(\boldsymbol{\theta}) = \frac{1}{N} \sum_{i=1}^{N} \{\hat{\boldsymbol{\xi}}_i^c(\boldsymbol{\theta})\}^\top \{\hat{\boldsymbol{\xi}}_i^c(\boldsymbol{\theta})\}, \tag{7}$$

where $\hat{\boldsymbol{\xi}}_i^c(\boldsymbol{\theta}) = (\hat{\xi}_{i,1}^c(\boldsymbol{\theta}), \ldots, \hat{\xi}_{i,p}^c(\boldsymbol{\theta}))^\top$ and $\hat{\xi}_{i,\ell}^c(\boldsymbol{\theta}) = \langle \hat{Y}_i(\cdot|\boldsymbol{\theta}), \hat{\phi}_\ell \rangle$. We define the optimal segmentation estimator $\tilde{\boldsymbol{\theta}}_N = \{\tilde{\theta}_{N,k}\}_{k=1}^{K}$ to be the changepoint estimator that minimizes $T_N(\boldsymbol{\theta})$ over all K-segmentation $\boldsymbol{\theta}$, $K \geq M$; that is, $\tilde{\boldsymbol{\theta}}_N = \operatorname{argmin}_{\boldsymbol{\theta} \in \Theta} T_N(\boldsymbol{\theta})$. We present the consistency of $\tilde{\boldsymbol{\theta}}_N$ in the following theorem.

THEOREM 1. *Suppose that the assumptions of model* (1) *hold, and* $\boldsymbol{\theta}^*$ *is detectable with respect to* $V = \operatorname{span}\{\phi_1, \ldots, \phi_p\}$. *Assume* $\min_{1 \le l \le p}\{\lambda_l - \lambda_{l+1}\} \ge \eta_0$ *for some* $\eta_0 > 0$. *If* $M \le K < \infty$, *then for each element* θ_m^* *in* $\boldsymbol{\theta}^*$, $1 \le m \le M$, *there exists a corresponding* $\tilde{\theta}_{N,k}$ *for some* $1 \le k \le K$, *such that* $\tilde{\theta}_{N,k} \xrightarrow{p} \theta_m^*$.

The proof of Theorem 1 is provided in Appendix B. The detectable assumption on $\boldsymbol{\theta}^*$ ensures the identifiability of the estimator $\tilde{\theta}_{N,k}$ when using $T_N(\boldsymbol{\theta})$ in (7) as the criterion for detecting multiple changepoints in a functional data sequence. The assumption of minimum spacing of the adjacent eigenvalues is to ensure the consistency of the estimated eigenfunctions $\{\hat{\phi}_\ell(t)\}_{\ell=1}^p$. In the estimation of the functional changepoint model (1), the variation not only comes from the stochastic component $\{Y_i\}$ but also the shifts of the estimated mean functions. In practice, the dimension of V increases with the value of δ in order for the stochastic expansion to explain a certain proportion of variance for the sequence $\{X_i(t)\}$. Therefore, the detectable assumption for changepoints is more likely to be satisfied for higher value of δ. In this study, we set δ as 0.95 with the aim of detecting all the changes.

3. Dynamic segmentation and backward elimination. We propose a two-stage algorithm, Dynamical Segmentation (DS) and Backward Elimination (BE). The DS stage efficiently searches for an optimal K-segmentation with a prespecified K, $K \ge M$ and, in the BE stage, falsely selected changepoint candidates are eliminated through hypothesis testing.

3.1. *Dynamic segmentation.* The basic idea behind DS is to divide the whole sequence of functional observations into subsequences according to their associated positions that cover all possible changepoints and satisfy the AMOC assumption. The boundary points of each subinterval are recursively adjusted to minimize the criterion $T_N(\boldsymbol{\theta})$ in (7).

Given a subinterval $(\theta_l, \theta_r]$ of (0, 1] and any θ in $(\theta_l, \theta_r]$, the sample averages $\bar{X}_{(\theta_l,\theta]}(t)$ and $\bar{X}_{(\theta,\theta_r]}$ are defined analogously as in (2). Similar to the definition of $T_N(\boldsymbol{\theta})$ in (7) for the entire interval (0, 1], we define $S_{(\theta_l,\theta_r]}(\theta)$ for each subinterval $(\theta_l, \theta_r]$ as

$$S_{(\theta_l,\theta_r]}(\theta) = \frac{1}{\lfloor N\theta_r \rfloor - \lfloor N\theta_l \rfloor} \sum_{i=\lfloor N\theta_l \rfloor + 1}^{\lfloor N\theta_r \rfloor} \{\hat{\boldsymbol{\xi}}^c_{i,(\theta_l.\theta_r]}(\theta)\}^\top \{\hat{\boldsymbol{\xi}}^c_{i,(\theta_l.\theta_r]}(\theta)\}, \tag{8}$$

where $\hat{\boldsymbol{\xi}}^c_{i,(\theta_l,\theta_r]}(\theta) = (\hat{\xi}^c_{i1,(\theta_l,\theta_r]}(\theta), \ldots, \hat{\xi}^c_{ip,(\theta_l,\theta_r]}(\theta))^\top$ with the element

$$\hat{\xi}^c_{i\ell,(\theta_l,\theta_r]}(\theta) = \langle X_i - \{\bar{X}_{(\theta_l,\theta]} \cdot \mathbf{1}_{(\theta_l,\theta]}(i/N) + \bar{X}_{(\theta,\theta_r]} \cdot \mathbf{1}_{(\theta,\theta_r]}(i/N)\}, \hat{\phi}_\ell \rangle.$$

The estimate of a single changepoint within $(\theta_l, \theta_r]$ is obtained by minimizing (8),

$$\tilde{\theta}_N = \operatorname*{argmin}_{\theta \in (\theta_l,\theta_r]} S_{(\theta_l,\theta_r]}(\theta).$$

The minimizer $\tilde{\theta}_N$ can be found by implementing a grid search in the elements of the sequence $\{X_i(t); i = (\lfloor N\theta_l \rfloor + 1), \ldots, \lfloor N\theta_r \rfloor\}$ under the AMOC assumption.

Given that the subintervals $(\theta_l, \theta_r]$ are unknown and are estimated recursively, let the index (r) in $\boldsymbol{\theta}^{(r)} = \{\theta_1^{(r)}, \ldots, \theta_K^{(r)}\}$ denote the rth iteration of DS and let $\mathcal{I}_j^{(r)}(h) = (\theta_{j-1}^{(r+1)} + h, \theta_{j+1}^{(r)} - h]$ for some $h > 0$ be the subinterval excluding the parts near the boundaries of $(\theta_{j-1}^{(r+1)}, \theta_{j+1}^{(r)}]$. The proposed procedure is summarized as follows.

Algorithm DS:

(D1) Choose a positive integer K and obtain the initial equally-spaced K-segmentation $\boldsymbol{\theta}^{(0)} = \{d, 2d, \ldots, Kd\}$ with $d = 1/(K+1)$.

(D2) Set $\theta_0^{(r)} = 0$ and $\theta_{K+1}^{(r)} = 1$ for all $r \geq 0$.

(D3) Given $\boldsymbol{\theta}^{(r-1)} = \{\theta_1^{(r-1)}, \ldots, \theta_K^{(r-1)}\} \in \Theta$ at the beginning of the rth iteration, update $\theta_j^{(r-1)}$ sequentially for $j = 1, \ldots, K$ by

$$\theta_j^{(r)} = \inf\Big[\operatorname*{argmin}_{\theta \in \mathcal{I}_j^{(r-1)}(h)} \{S_{(\theta_{j-1}^{(r)}, \theta_{j+1}^{(r-1)}]}(\theta)\}\Big].$$

(D4) Iterate the previous updating step until $\boldsymbol{\theta}^{(r)}$ satisfies the convergence criterion $\max_{1 \leq j \leq K} |\theta_j^{(r+1)} - \theta_j^{(r)}| < 1/N$.

(D5) Obtain the K-segmentation after convergence $\hat{\boldsymbol{\theta}}_N = \{\hat{\theta}_{N,1}, \ldots, \hat{\theta}_{N,K}\}$.

In (D3), the introduction of h in $\mathcal{I}_j^{(r)}(h)$ is necessary in practice. It serves as the possible minimum segment length of $\hat{\boldsymbol{\theta}}_N$ in the recursive updating procedure. Theoretically, $h = h(N)$ tends to 0 as $N \to \infty$. Moreover, because $\{T_N(\boldsymbol{\theta}^{(r)})\}$ forms a positive nonincreasing sequence in r, the convergence of DS is ensured. However, to ensure the consistency of $\hat{\boldsymbol{\theta}}_N$, an additional assumption on the minimal distance between two adjacent changepoints in relation to the algorithm is required. Let $\Delta = \min_{0 \leq m \leq M} |\theta_{m+1}^* - \theta_m^*|$.

THEOREM 2. *The DS algorithm converges*; *that is*, *the K-segmentation $\hat{\boldsymbol{\theta}}_N$ in* (*D*5) *always exists such that* $\lim_{r \to \infty} \boldsymbol{\theta}^{(r)} = \hat{\boldsymbol{\theta}}_N$. *Furthermore*, *under the assumptions of Theorem* 1, *if $d \leq \Delta$*, *there exists a positive value $h = h(N) \to 0$ such that*

$$\lim_{N \to \infty} P\Big(\bigcap_m \{|\theta_m^* - \hat{\theta}_{N,k}| < h \text{ for some } k = 1, \ldots, K\}\Big) = 1.$$

The proof of Theorem 2 is provided in Appendix C. The theorem requires that the distance of the initial equally-spaced K-segmentation d be smaller than the minimal adjacent distance between the true changepoints Δ; thus, the AMOC assumption is satisfied. Because Δ is unknown in practice, we make d small by

choosing a large K. Theorems 1 and 2 ensure that the dynamic segmentation procedure is able to detect all M true changepoints; however, the remaining $(K - M)$ changepoint candidates are falsely selected. Therefore, we require a procedure to identify the M true changepoints among the K changepoint candidates.

3.2. *Backward elimination.* To remove the $(K - M)$ falsely selected changepoints from $\hat{\boldsymbol{\theta}}_N$, we propose the BE procedure to ensure that the respective changes are statistically significant. We start by selecting the most unlikely changepoint candidate in the set of $\hat{\boldsymbol{\theta}}_N$ and then perform hypothesis testing to possibly remove the nonstatistically significant changepoint candidate. This process is implemented iteratively until no more candidates can be removed. Because the hypothesis testing procedures are performed at most K times, we control the overall level of type I errors through the Bonferroni method.

Let $\hat{\boldsymbol{\theta}}_{-k} = \hat{\boldsymbol{\theta}}_N \setminus \{\hat{\theta}_{N,k}\}$ denote the subset of $\hat{\boldsymbol{\theta}}_N$ excluding the element $\hat{\theta}_{N,k}$. If $\hat{\theta}_{N,k}$ is a falsely selected changepoint candidate, then by Thereom 1 we expect that the values of $T_N(\hat{\boldsymbol{\theta}}_{-k})$ and $T_N(\hat{\boldsymbol{\theta}}_N)$ are relatively close. Thus, the $\hat{\theta}_{N,k}$ corresponding to the minimum value of $\{|T_N(\hat{\boldsymbol{\theta}}_N) - T_N(\hat{\boldsymbol{\theta}}_{-l})| : 1 \le l \le K\}$ is the most unlikely changepoint candidate. Moreover, by Lemma 1 only the deletion of a true changepoint can cause a change in the corresponding covariance operators. Thus, we test the difference between the covariance operators before and after deleting the changepoint candidate.

Consider the case of deleting $\hat{\theta}_{N,k}$ within the subinterval $(\hat{\theta}_{N,k-1}, \hat{\theta}_{N,k+1}]$. Let $\hat{\mathcal{C}}_{k,b}$ be the integral operator with the empirical estimator $\hat{c}_{k,b}(s,t)$ before deleting $\hat{\theta}_{N,k}$ as its kernel, which is given by

$$\hat{c}_{k,b}(s,t) = \frac{1}{N_k(\hat{\boldsymbol{\theta}}_N) + N_{k+1}(\hat{\boldsymbol{\theta}}_N)} \sum_{i=\lfloor N\hat{\theta}_{N,k-1}\rfloor+1}^{\lfloor N\hat{\theta}_{N,k+1}\rfloor} \{\hat{Y}_i(s|\hat{\boldsymbol{\theta}}_N)\hat{Y}_i(t|\hat{\boldsymbol{\theta}}_N)\},$$

where $\hat{Y}_i(t|\hat{\boldsymbol{\theta}}_N)$ is defined as in (4) and $N_k(\hat{\boldsymbol{\theta}}_N) = (\lfloor N\hat{\theta}_{N,k}\rfloor - \lfloor N\hat{\theta}_{N,k-1}\rfloor)$; $\hat{\mathcal{C}}_{k,a}$ is defined analogously by replacing $\hat{\boldsymbol{\theta}}_N$ in $\hat{Y}_i(t|\hat{\boldsymbol{\theta}}_N)$ with $\hat{\boldsymbol{\theta}}_{-k}$. Moreover, we denote $\mathcal{C}_{k,b} = \lim_{N\to\infty} \hat{\mathcal{C}}_{k,b}$ and $\mathcal{C}_{k,a} = \lim_{N\to\infty} \hat{\mathcal{C}}_{k,a}$. The null hypothesis

$$H_{0k} : \mathcal{C}_{k,b} = \mathcal{C}_{k,a}$$

is to test if $\hat{\theta}_{N,k}$ can be rejected as a true changepoint. By Lemma 1, $\mathcal{C}_{k,b} = \mathcal{C}_{k,a} = \mathcal{C}$ if $\hat{\theta}_{N,k}$ is not a changepoint, and $\mathcal{C}_{k,a} = \mathcal{C}_{k,b} + \mathcal{B}$ if $\hat{\theta}_{N,k} = \theta_m^*$ for some $1 \le m \le M$, where the kernel of $\mathcal{B}$ is $\alpha(1-\alpha)(\mu_{m+1} - \mu_m)(s)(\mu_{m+1} - \mu_m)(t)$ with $\alpha = N_k(\hat{\boldsymbol{\theta}}_N)/\{N_k(\hat{\boldsymbol{\theta}}_N) + N_{k+1}(\hat{\boldsymbol{\theta}}_N)\}$.

Intrinsically, the hypothesis H_0 is a two sample test problem for functional data. Benko, Härdle and Kneip (2009) proposed a bootstrap procedure for testing the equality of spectra such as the eigenvalues and eigenfunctions of covariance operators. Panaretos, Kraus and Maddocks (2010) developed a test based on the Hilbert–Schmidt norm of the difference between covariance operators under the Gaussian

assumption, and Fremdt et al. (2013) extended the method to the non-Gaussian case.

To test H_0, we adopt the method proposed by Fremdt et al. (2013) which uses the weighted Hilbert–Schmidt norm of $(\hat{\mathcal{C}}_{k,b} - \hat{\mathcal{C}}_{k,a})$ as the test statistic. Let A be a $p \times p$ symmetric matrix with entries $A(\ell, d) = \langle (\hat{\mathcal{C}}_{k,b} - \hat{\mathcal{C}}_{k,a})\hat{\phi}_\ell, \hat{\phi}_d \rangle$ and $\boldsymbol{\zeta} = \operatorname{vech}(A)$, where $\operatorname{vech}(A)$ is the half-vectorization of A. If H_0 holds, then $(N_k/2)\boldsymbol{\zeta}^\top \hat{L}^{-1}\boldsymbol{\zeta} \xrightarrow{d} \chi^2_{p(p+1)/2}$. Here, $\hat{L}$ is an estimate of the asymptotic covariance matrix for $(N_k/2)^{1/2}\boldsymbol{\zeta}$, which consists of the elements of $\hat{\boldsymbol{\xi}}^c_{i,b}$ and $\hat{\boldsymbol{\xi}}^c_{i,a}$, where $\hat{\boldsymbol{\xi}}^c_{i,b} = (\hat{\xi}^c_{i1,b}, \ldots, \hat{\xi}^c_{ip,b})^\top$ with $\hat{\xi}^c_{i\ell,b} = \langle \hat{Y}_i(\cdot|\hat{\boldsymbol{\theta}}_N), \hat{\phi}_\ell \rangle$; $\hat{\boldsymbol{\xi}}^c_{i,a}$ is defined analogously.

Let $\boldsymbol{\theta}^{[k]}$ denote the changepoint candidates at the kth BE iteration. Set $\boldsymbol{\theta}^{[0]} = \hat{\boldsymbol{\theta}}_N$. We use square brackets in the superscript rather than parentheses to distinguish the iteration indices of the DS and BE stages. The procedure is as follows.

Algorithm BE:

(B1) *Selecting the most unlikely changepoint candidate.* Given $\boldsymbol{\theta}^{[k-1]}$ at the beginning of the kth iteration, obtain

$$\theta^{[k]}_{\mathrm{del}} = \underset{\theta \in \boldsymbol{\theta}^{[k-1]}}{\operatorname{argmin}} \left| T_N\left(\boldsymbol{\theta}^{[k-1]} \setminus \{\theta\}\right) - T_N\left(\boldsymbol{\theta}^{[k-1]}\right) \right|,$$

and then update $\boldsymbol{\theta}^{[k]} = \boldsymbol{\theta}^{[k-1]} \setminus \{\theta^{[k]}_{\mathrm{del}}\}$.

(B2) *Testing the most unlikely changepoint candidate.* Consider the null hypothesis H_{0k}. If H_{0k} is rejected, then the BE procedure stops and $\boldsymbol{\theta}^{[k-1]} = \hat{\boldsymbol{\theta}}^*_N$ is the estimate of $\boldsymbol{\theta}^*$. Otherwise, the procedure continues to the $(k+1)$th iteration.

If the procedure stops at the k^*th iteration, we conclude that there are $\widehat{M} = (K - k^* + 1)$ changepoints with positions $\boldsymbol{\theta}^{[k^*-1]}$. If none of the K tests are significant, then the DSBE procedure concludes the algorithm there is no changepoint in the sequence.

4. Application to freeway traffic segmentation. We apply the DSBE algorithm to freeway traffic data to identify locations of changes in mean vehicle speeds. The speeds were recorded in 2014 by dual loop vehicle detectors in a 5-minute interval at a sequence of locations along Freeway No. 5 in Taiwan. Let the random function $X_i(t)$ be the vehicle speed (km/hr) recorded by the ith detector at time t on a particular day of the week, such as Sunday or Saturday. Thus, $\{X_i(t)\}_{i=1}^N$ forms a sequence of random functions. The dataset includes a total of $N = 85$ vehicle detectors (VDs). We consider the profiles of daily vehicle speed as realizations of the random functions. For each date j, we take $\{X_{ij}(t)\}_{i=1}^N$ as realizations sampled from the random function $\{X_i(t)\}_{i=1}^N$. In practice, some of the data records may be missing due to various reasons such as occasional road maintenance or package loss during transmission, and such missing values occur

TABLE 1
Estimated number of changepoints $\hat{M}$ along with their locations

	Saturday		*Sunday*	
K	$\hat{M}$	*VD ID*	$\hat{M}$	*VD ID*
6	5	[13 19 50 65 76]	5	[13 27 50 65 76]
7	7	[11 18 28 48 54 65 76]	7	[11 16 27 48 54 65 76]
8	8	[9 13 19 28 48 54 65 76]	7	[9 13 27 50 54 65 76]
9	8	[9 13 19 28 48 54 65 76]	7	[9 13 27 50 54 65 76]
10	8	[9 13 19 28 48 54 65 76]	7	[9 13 27 50 54 65 76]
11	8	[9 13 19 28 48 54 65 76]	7	[9 13 27 50 54 65 76]

at different dates and locations. To utilize all the recorded data, we consider the sequence of functional data, $\{X_{11}(t), \ldots, X_{1r_1}(t), \ldots, X_{N1}(t), \ldots, X_{Nr_N}(t)\}$, where $\{X_{ij}(t),\ j = 1, \ldots, r_i\}$ are replicates of $X_i(t)$.

Here, $\{X_{ij}(t),\ j = 1, \ldots, r_i;\ i = 1 \ldots, N\}$ follow model (1), and the changes in mean functions can occur only at locations where the index i changes. For an arbitrary $\boldsymbol{\theta}$, when there are replications at the same locations, we modify $\bar{X}_{(\theta_{k-1},\theta_k]}(t)$ in (2) and $\hat{c}(s,t|\boldsymbol{\theta})$ in (3) as follows: $\bar{X}_{(\theta_{k-1},\theta_k]}(t) = \sum_{i=\lfloor N\theta_{k-1}\rfloor+1}^{\lfloor N\theta_k\rfloor} \sum_{j=1}^{r_i} X_{ij}(t)/\sum_{i=\lfloor N\theta_{k-1}\rfloor+1}^{\lfloor N\theta_k\rfloor} r_i$, for $k = 1, \ldots, K$, and $\hat{c}(s,t|\boldsymbol{\theta}) = \sum_{i=1}^{N} \sum_{j=1}^{r_i} \hat{Y}_{ij}(s|\boldsymbol{\theta})\hat{Y}_{ij}(t|\boldsymbol{\theta})/\sum_{i=1}^{N} r_i$, where $\hat{Y}_{ij}(t|\boldsymbol{\theta}) = X_{ij}(t) - \sum_{k=1}^{K} \bar{X}_{(\theta_{k-1},\theta_k]}(t) \cdot \mathbf{1}_{(\theta_{k-1},\theta_k]}(i/N)$ for $i = 1, \ldots, N$.

To apply the DSBE algorithm, we set different values of K for the initial K-segmentation and a small value for the parameter $h = 3/84$. The results on the detected changepoints are displayed in Table 1, and their locations are illustrated in Figure 1. Table 1 shows that the DSBE algorithm is quite stable when using different initial values of K. As K increases, the segments of the initial K-segmentation are more likely to satisfy the AMOC assumption and more of the changepoints can be detected. For $K \geq 8$, the number and locations of the detected changepoints are the same for Saturdays and Sundays, indicating the capability of DS to collect candidates containing subset with the true changepoints and the testing power of BE. The changepoint locations detected using $K = 6$ are also detected when we let $K \geq 8$. For $K = 7$, the locations of the detected changepoints are slightly different from those detected for $K \geq 8$, but the differences are still within the acceptable margins given by $h = 3/84$, as supported in Theorem 2.

Figure 1 indicates that the detected changepoint locations on Saturdays and Sundays are very close, although an additional changepoint is detected on Saturdays. This reveals that every changepoint corresponds to either a freeway-interchange entrance or a location inside a tunnel. The first and last changes correspond to the entrances to Yilan and Siding interchanges. After the entrance to the Yilan interchange, the vehicle flow rate increases and the vehicle speed drops. The vehicle

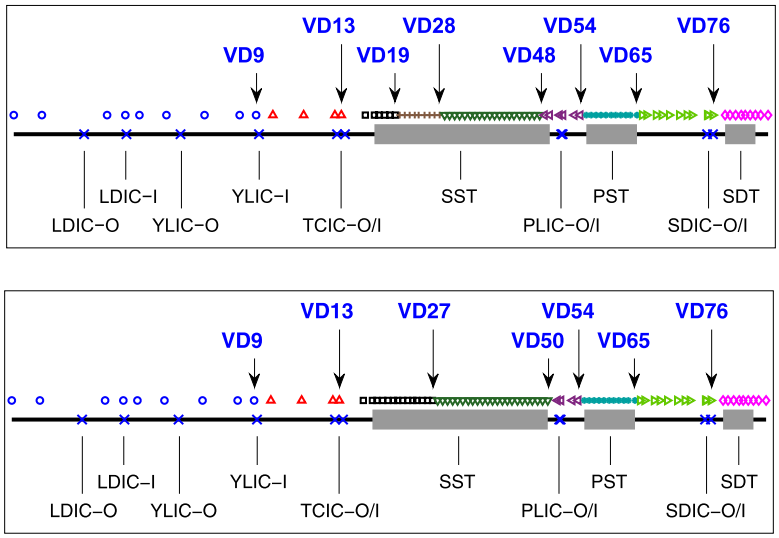

FIG. 1. *Freeway segments on northbound lanes of Freeway No.* 5 *in Taiwan for Saturdays* (*top*) *and Sundays* (*bottom*), *corresponding to the changepoints indicated by vehicle detector IDs and the locations*: *SST, PST and SDT are three tunnels, and IC-O and IC-I stand for the exit and entry of the interchange. For TCIC, PLIC and SDIC, the exit and entry are denoted together by IC-O/I because they are very close.*

speed continues to drop as vehicles approach the interchange before entering the Shueh–Shan Tunnel (SST), which is the longest tunnel in Taiwan at approximately 13 km in length. Drivers in a tunnel of this length may slow down unconsciously. Although a broadcasting system in the tunnel reminds drivers not to lower their speed, it still decreases gradually until the exit of the tunnel (detector No. 48 on Saturdays and No. 50 on Sundays). Finally, the other obvious rise in vehicle speed corresponds to detector No. 65, near the exit of the Peng–Shan Tunnel (PST).

Figures 2 and 3 display the sample mean speed profiles for Saturdays and Sundays, respectively, at each detector superimposed on the segment mean function of vehicle speed. The mean speed trajectories between the adjacent segments differ obviously in the magnitude as well as the pattern of daily traffic speed profiles. Figure 4 illustrates the segment mean functions for Saturdays and Sundays, which demonstrates the multiple changes in the mean functions of the freeway segments.

As illustrated in Figure 4 along with the the changepoint locations in Figure 1, it is interesting to see the changes detected at VD 28 and VD 54 for Saturdays, corresponding to changes from Segments 4 to 5 and from 6 to 7, and also the changes detected at VD 27 and VD 54 for Sundays, corresponding to changes from Segments 3 to 4 and from 5 to 6, respectively. Both VD 27 and VD 28 are located in the long SST, where the road curvatures in the tunnel naturally cause drivers to slow down the speed gradually. Thus the speed in Segment 5 is generally slower than that in Segment 4 on Saturdays, and similarly for the reduction of

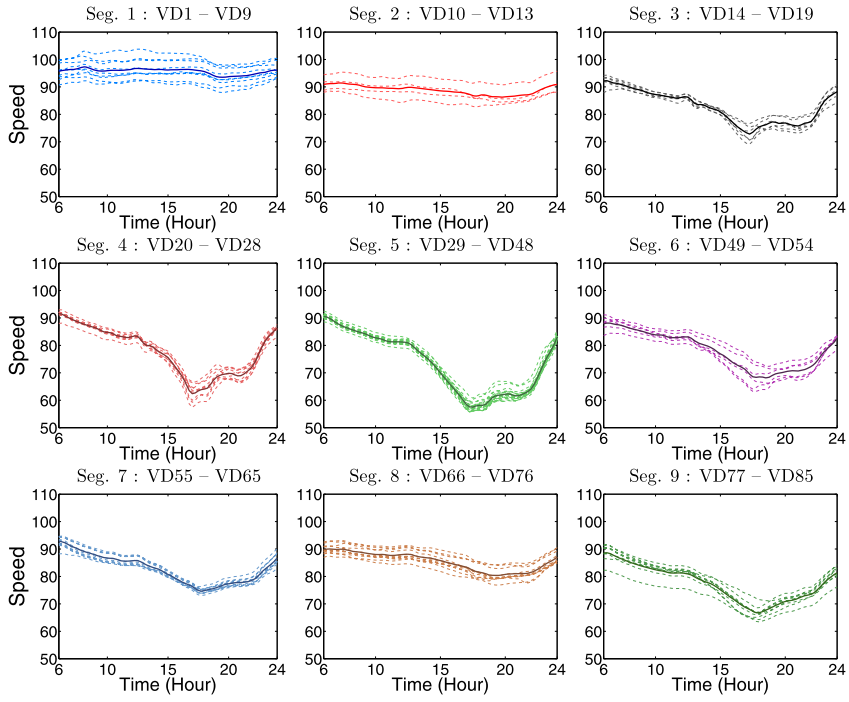

FIG. 2. *The mean functions at each VD along with the corresponding segment mean functions of speed* (*km/hr*) *for Saturday.*

speed from Segment 3 to Segment 4 on Sundays. On the other hand, the speed changes detected at VD 54 from Segment 6 to Segment 7 on Saturdays and from Segment 5 to Segment 6 on Sundays are likely due to the interference of traffic

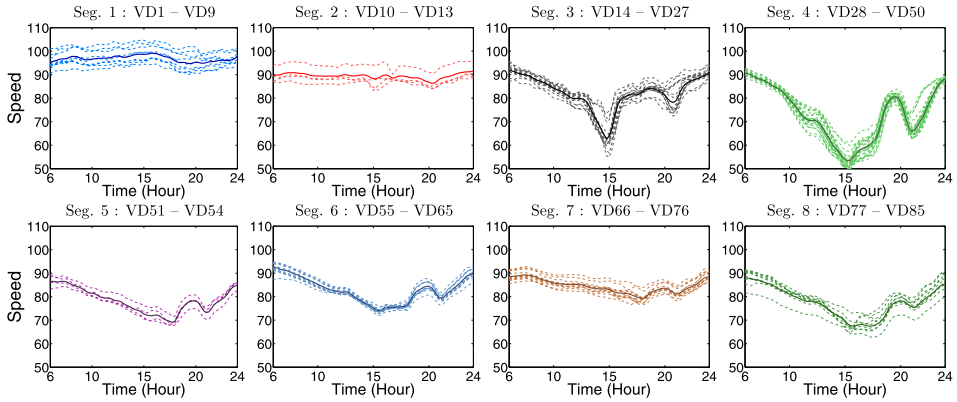

FIG. 3. *The mean functions at each VD along with the corresponding segment mean functions of speed* (*km/hr*) *for Sunday.*

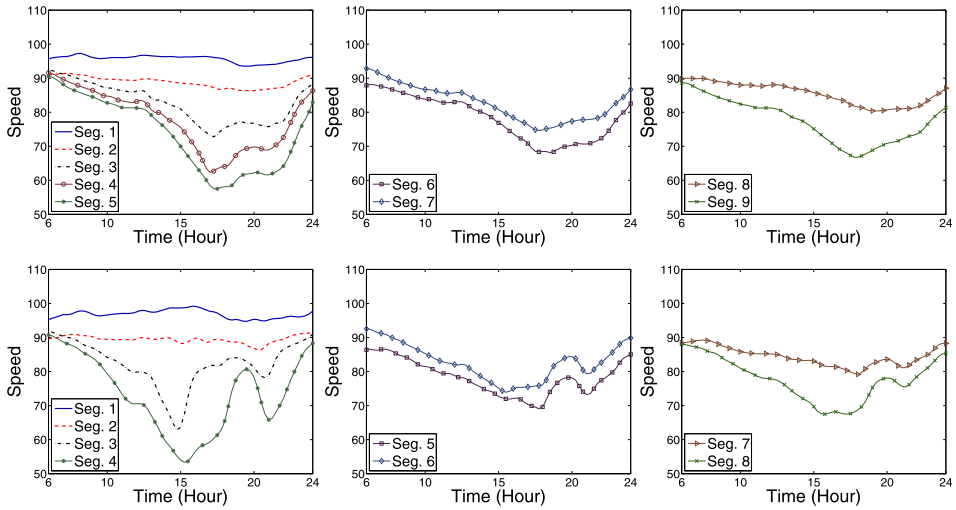

FIG. 4. *The segment mean functions of speed* (*km/hr*) *for Saturdays* (*top panels*) *and Sundays* (*bottom panels*).

flow at Pinling Interchange. Segment 6 on Saturdays and Segment 5 on Sundays contain the interchange, where the in and out traffic flow rates reduce the traffic speed, while the following segments merely cover PST without the interference of traffic flow due to the entrance and the exit of the freeway interchange.

5. Simulation study. In this section, we investigate the finite sample performance of the DSBE algorithm for detecting multiple changepoints. The simulated data are generated analogously to the real traffic data in Section 4 such that $\{X_i(t), 0 \le t \le 1\}_{i=1}^{N}$ is a sequence of random functions following model (1) with r_i replications of $X_i(t)$. We set $r_i = 20$ for all i in the simulation and apply the DSBE algorithm to the sequence of functional data.

5.1. *Simulation settings.* We consider different numbers of changepoints ($M = 0, 1, 2, 3, 4$) and different scales of changes between the mean functions. The segment mean functions $\{\mu_m(t)\}_{m=1}^{M+1}$ are chosen from the following functions: $\psi_1(t) = 5t^2 - \exp(1-20t)$; $\psi_2(t) = 0.5 - 100(t-0.1)(t-0.3)(t-0.5)(t-0.9)$; $\psi_3(t) = \psi_2(t) + 0.8\sin(1+10\pi t)$; and $\psi_4(t) = 1 + 3t^2 - 5t^3 + 0.6\sin(1+10\pi t)$; $\psi_5(t) = 1 + 3t^2 - 5t^3$. The difference between two functions is measured using the L^2 norm; the change from ψ_1 to ψ_2 is the largest, that from ψ_3 to ψ_4 is rather moderate, and those from ψ_2 to ψ_3 and ψ_4 to ψ_5 are relatively small. On the basis of these functions, we design four mean change scenarios with the number of changes M ranging from 1 to 4:

- $M = 1$ (ψ_3,ψ_4): moderate change;

TABLE 2
Summary of simulation settings

Scenario	*Segment mean*	*Changepoint positions*
Null	$\mu_1 = \psi_1$	N/A
A1	$\mu_1 = \psi_3, \mu_2 = \psi_4$	[0.15]
B1		[0.50]
C1		[0.80]
A2	$\mu_1 = \psi_2, \mu_2 = \psi_4, \mu_3 = \psi_5$	[0.15, 0.40]
B2		[0.30, 0.70]
C2		[0.60, 0.75]
A3	$\mu_1 = \psi_1, \mu_2 = \psi_2, \mu_3 = \psi_3, \mu_4 = \psi_4$	[0.10, 0.25, 0.40]
B3		[0.20, 0.70, 0.80]
C3		[0.20, 0.50, 0.75]
A4	$\mu_1 = \psi_1, \mu_2 = \psi_2, \mu_3 = \psi_3, \mu_4 = \psi_4, \mu_5 = \psi_5$	[0.15, 0.25, 0.40, 0.50]
B4		[0.15, 0.60, 0.75, 0.80]
C4		[0.15, 0.25, 0.75, 0.80]

- $M = 2$ (ψ_2,ψ_4,ψ_5): moderate and small changes;
- $M = 3$ (ψ_1,ψ_2,ψ_3,ψ_4): large, small and moderate changes;
- $M = 4$ (ψ_1,ψ_2,ψ_3,ψ_4,ψ_5): large, small, moderate and small changes.

In addition, we establish three different scenarios for the positions of changepoints in each simulation design. Table 2 summarizes the simulation designs with a total of 13 scenarios, and Figure 5 illustrates the segment mean functions along with the locations of the changepoints in each simulation design.

Realizations of the random function $Y_i(t)$ are generated by the basis expansion $Y_i(t) = \sum_{\ell=0}^{L} \sqrt{\lambda_\ell}\tau_{i,\ell}\phi_\ell(t)$, $i = 1, \ldots, N$, where $(\lambda_\ell, \phi_\ell(\cdot))$ are fixed eigenvalue–eigenfunction pairs, and $\{\tau_{i,\ell} : i = 1, \ldots, N\}$, $\ell = 0, \ldots, L$, are random coefficients. The sequence $\{Y_i(t)\}$ depends on the sequence of the random coefficients. For each ℓ, $\{\tau_{i,\ell} : i = 1, \ldots, N\}$ is generated by the AR(1) model such that $\tau_{i,\ell} = \rho\tau_{i-1,\ell} + \varepsilon_{i,\ell}$, where $\varepsilon_{i,\ell}$ is the standard normal random variate. The autocorrelation parameter ρ is set as 0, 0.2, and 0.5 for the independent, low dependence and moderate dependence cases, respectively. We set $L = 150$ and $\lambda_\ell = 0.7 \times 2^{-\ell}$, and set $\{\phi_\ell(t); \ell = 1, \ldots, L\}$ to be the Fourier basis $\sqrt{2}\sin(2\pi kt - \pi)$ for $\ell = 2k - 1$ and $\sqrt{2}\cos(2\pi kt - \pi)$ for $\ell = 2k$, $k = 1, \ldots, L/2$, and $\phi_0(t) = 1$ a constant such that $\int_{\mathcal{T}} \phi_0^2(t)\,dt = 1$.

5.2. *Simulation results.* For each scenario, we generate 500 simulation replicates of functional data sequence with lengths $N = 100$ and 200. We choose $K = 9$ for the initial segmentation. We set $h = 3/99$ for $N = 100$ and $5/199$ for $N = 200$. The rationale and a practical guideline of choosing h are deferred until Section 5.3.

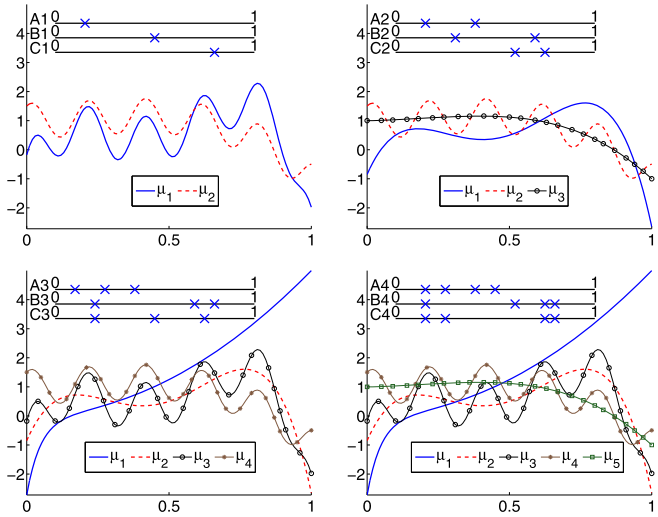

FIG. 5. *Illustration of the simulation designs with the mean functions for* $M = 1$ (*top left*), $M = 2$ (*top right*), $M = 3$ (*bottom left*) *and* $M = 4$ (*bottom right*), *along with the changepoint locations* (*indicated by* $\times$).

For the BE step, we set the overall level of significance at 0.05. To examine the performance, we consider the estimates falling at the exact and neighboring positions of the true changepoints.

Table 3 presents the frequencies of correctly estimated multiple changepoints including the total number and the exact and neighboring positions (in parentheses) in the neighborhood of a true changepoint θ_m^* such that $|\hat{\theta}_{N,k} - \theta_m^*| < 0.02$. The correct detection rates decrease as the strength of autocorrelation increases, especially when more than one changepoint is present. However, as the sample size increases, the accuracy rate increases. When the sequence length is $N = 200$, for cases of zero and low dependence ($\rho = 0$ and 0.2), the correct rate is higher than 85% in all scenarios for the neighboring positions. Even under moderate dependence, the neighboring correctness is over 80%, except for scenario B3 and C4, in which the neighboring correctness is still over 70%. The designed segment lengths in scenario B3 are quite unbalanced (0.20 vs. 0.50 and 0.50 vs. 0.10), which makes it difficult to detect the changepoints. In scenario A4, the change between the mean functions μ_4 and μ_5 is rather small, and the corresponding segment lengths 0.10 and 0.50 are quite unbalanced. Moreover, both scenarios B4 and C4 contain a relatively short segment of length 0.05, which makes the changepoint detection difficult. Therefore, it is not surprising to see the detection accuracies of these difficult scenarios (B3, A4, B4 and C4) are lower than other scenarios. However, even for the difficult scenarios as designed, the DSBE procedure still performs accurately as the sample size increases from $N = 100$ to $N = 200$.

TABLE 3
Frequencies of the correctly estimated changepoints using the DSBE procedure with 500 *simulation replicates*

	$N = 100$			$N = 200$		
$\boldsymbol{\theta}$	$\rho = 0$	$\rho = 0.2$	$\rho = 0.5$	$\rho = 0$	$\rho = 0.2$	$\rho = 0.5$
none	500	500	484	500	500	498
A1	491 (491)	487 (487)	461 (461)	500 (500)	500 (500)	494 (496)
B1	500 (500)	498 (498)	456 (461)	499 (500)	500 (500)	494 (496)
C1	500 (500)	495 (495)	438 (444)	500 (500)	497 (499)	483 (485)
A2	488 (491)	458 (467)	326 (342)	497 (498)	494 (496)	400 (420)
B2	494 (498)	481 (486)	360 (376)	498 (500)	493 (500)	444 (468)
C2	473 (477)	476 (482)	349 (363)	498 (499)	495 (497)	454 (466)
A3	479 (490)	477 (492)	436 (445)	493 (499)	492 (499)	477 (485)
B3	321 (336)	324 (342)	223 (255)	465 (497)	437 (488)	272 (352)
C3	498 (500)	492 (494)	407 (415)	498 (500)	496 (500)	456 (473)
A4	339 (354)	352 (358)	240 (256)	478 (499)	460 (481)	375 (417)
B4	224 (225)	249 (252)	203 (208)	427 (428)	439 (443)	390 (404)
C4	213 (218)	225 (237)	175 (193)	412 (425)	413 (435)	348 (387)

5.3. *Guideline of choosing h in DS algorithm.* Recall the interval $\mathcal{I}_j^{(r)}(h) = (\theta_{j-1}^{(r+1)} + h, \theta_{j+1}^{(r)} - h]$ in (D3) of Algorithm DS. Introducing the parameter h is to prevent the iterative procedure from selecting the updated changepoint ultimately close to the boundaries of $\mathcal{I}_j^{(r)}(h)$, the current changepoints $\theta_{j-1}^{(r+1)}$ and $\theta_{j+1}^{(r)}$, at the current iteration. Given the equally spaced K-segmentation as the initial setting, we require $1/(N-1) \le h < 1/(K+1)$ for $K \ll N$. It is because $\mathcal{I}_1^{(0)}(h) = \varnothing$ if $h \ge 1/(K+1)$, and h serves as the possible minimum segment for DSBE. In practice, we suggest the choice of $h = b/(N-1)$ with b restricted to $b_0/2 \le b \le b_0$, where $b_0 = \max\{b \in \mathbb{N} \mid 2b/(N-1) < 1/(K+1)\}$ for a given value of K.

We investigate the practical performance of selecting h using the guideline suggested above. Under the simulation setting with $N = 200$ and $K = 9$, we obtain $b_0 = 9$ so that b can be chosen as an integer ranging from 5 to 9. When $N = 100$ and $K = 9$, we obtain $b_0 = 4$ and b can be chosen ranging from 2 to 4. For comparisons, we present the the results with $b = 1, 3, 5, 7, 9$, indicated in DSBE(b), for $N = 200$ and $b = 1, 2, 3, 4$ for $N = 100$. The simulation results in Table 4 for $N = 200$ and Table 5 for $N = 100$ indicate that DSBE using the suggested criterion to select h works very well. For smaller values of b such as 1 and 3 for $N = 200$ and 1 for $N = 100$, DSBE performs only slightly less well compared with those using the suggested values. Therefore, DSBE is not sensitive to the value of b within the range selected by the guideline, even for smaller values of b.

In our simulation studies, the settings $h = 3/99 \approx 0.03$ for $N = 100$ and $K = 9$ and $h = 5/199 \approx 0.025$ for $N = 200$ and $K = 9$ indeed meet the above guideline.

TABLE 4
Frequencies of the correctly estimated changepoints for exact and neighboring (in the parentheses) position ($N = 200$)

Method	$\boldsymbol{\theta}$	$\rho = 0$	$\rho = 0.2$	$\rho = 0.5$	$\boldsymbol{\theta}$	$\rho = 0$	$\rho = 0.2$	$\rho = 0.5$
DSBE (1)	A1	500 (500)	497 (500)	492 (496)	A2	481 (497)	475 (488)	377 (402)
DSBE (3)		500 (500)	499 (500)	493 (496)		497 (499)	487 (490)	381 (403)
DSBE (5)		500 (500)	500 (500)	494 (496)		497 (498)	494 (496)	400 (420)
DSBE (7)		500 (500)	500 (500)	496 (497)		499 (499)	494 (495)	420 (438)
DSBE (9)		500 (500)	500 (500)	495 (496)		500 (500)	498 (499)	440 (456)
DSBE (1)	B1	498 (500)	500 (500)	492 (495)	B2	494 (500)	490 (499)	439 (465)
DSBE (3)		499 (500)	500 (500)	491 (495)		497 (500)	491 (499)	446 (469)
DSBE (5)		499 (500)	500 (500)	494 (496)		498 (500)	493 (500)	444 (468)
DSBE (7)		499 (500)	500 (500)	495 (497)		498 (500)	497 (500)	451 (473)
DSBE (9)		499 (500)	500 (500)	496 (497)		499 (500)	498 (500)	459 (477)
DSBE (1)	C1	495 (500)	495 (499)	476 (480)	C2	495 (499)	489 (495)	446 (462)
DSBE (3)		500 (500)	498 (499)	481 (482)		497 (499)	493 (497)	449 (462)
DSBE (5)		500 (500)	497 (499)	483 (485)		498 (499)	495 (497)	454 (466)
DSBE (7)		500 (500)	499 (500)	485 (487)		497 (498)	495 (497)	460 (470)
DSBE (9)		500 (500)	499 (500)	488 (489)		499 (499)	497 (497)	465 (471)
DSBE (1)	A3	490 (499)	489 (498)	464 (479)	A4	471 (492)	452 (476)	355 (389)
DSBE (3)		493 (499)	492 (498)	471 (479)		477 (498)	461 (483)	369 (407)
DSBE (5)		493 (499)	492 (499)	477 (485)		478 (499)	460 (481)	375 (417)
DSBE (7)		493 (499)	491 (499)	478 (487)		479 (499)	465 (485)	388 (426)
DSBE (9)		492 (499)	492 (499)	479 (488)		479 (499)	467 (486)	393 (430)
DSBE (1)	B3	456 (491)	437 (483)	268 (336)	B4	425 (425)	438 (442)	378 (398)
DSBE (3)		459 (495)	439 (491)	271 (345)		425 (427)	436 (442)	386 (402)
DSBE (5)		465 (497)	437 (488)	272 (352)		427 (428)	439 (443)	390 (404)
DSBE (7)		468 (497)	442 (490)	271 (350)		427 (428)	438 (442)	393 (409)
DSBE (9)		470 (497)	448 (493)	286 (368)		427 (428)	439 (443)	399 (406)
DSBE (1)	C3	497 (500)	496 (500)	458 (473)	C4	403 (415)	405 (427)	333 (370)
DSBE (3)		498 (500)	497 (500)	456 (473)		406 (422)	410 (433)	339 (377)
DSBE (5)		498 (500)	496 (500)	456 (473)		412 (425)	413 (435)	348 (387)
DSBE (7)		497 (499)	496 (500)	462 (473)		413 (425)	414 (436)	353 (392)
DSBE (9)		498 (500)	497 (500)	464 (475)		414 (420)	417 (431)	362 (380)

In our data application, we demonstrate the results along with several values of K ranging from 6 to 11, the selection of $h = 3/84$ also meets the guideline for all these values of K.

5.4. *Comparisons*. As mentioned in Section 1, the CUSUM-type test (CUST) for single changepoint detection in a sequence of functional data can also be applied to multiple changepoint detection through the binary segmentation algorithm (BS$_{\text{CUST}}$). We compare the performance of BS$_{\text{CUST}}$ with the DSBE procedure. Let $\hat{\boldsymbol{\xi}}_i = (\hat{\xi}_{i,1}, \ldots, \hat{\xi}_{i,p})^\top$ with $\hat{\xi}_{i,\ell} = \langle X_i, \hat{\phi}_\ell \rangle$ $\ell = 1, \ldots, p$, where the dimension p and

TABLE 5
Frequencies of the correctly estimated changepoints for exact and neighboring (in the parentheses) position ($N = 100$)

Method	$\boldsymbol{\theta}$	$\rho = 0$	$\rho = 0.2$	$\rho = 0.5$	$\boldsymbol{\theta}$	$\rho = 0$	$\rho = 0.2$	$\rho = 0.5$
DSBE (1)	A1	490 (490)	487 (487)	456 (456)	A2	459 (464)	433 (442)	302 (315)
DSBE (2)		491 (491)	487 (487)	457 (457)		477 (481)	447 (456)	306 (322)
DSBE (3)		491 (491)	487 (487)	461 (461)		488 (491)	458 (467)	326 (342)
DSBE (4)		491 (491)	487 (487)	466 (466)		491 (494)	473 (477)	362 (372)
DSBE (1)	B1	499 (500)	497 (497)	453 (458)	B2	494 (495)	469 (477)	351 (367)
DSBE (2)		500 (500)	498 (498)	456 (461)		494 (498)	477 (483)	356 (372)
DSBE (3)		500 (500)	498 (498)	456 (461)		494 (498)	481 (486)	360 (376)
DSBE (4)		500 (500)	498 (498)	457 (460)		497 (499)	481 (487)	370 (386)
DSBE (1)	C1	495 (496)	487 (487)	426 (427)	C2	469 (476)	470 (478)	341 (360)
DSBE (2)		500 (500)	492 (492)	431 (433)		473 (477)	469 (480)	347 (363)
DSBE (3)		500 (500)	495 (495)	438 (444)		473 (477)	476 (482)	349 (363)
DSBE (4)		500 (500)	494 (495)	442 (447)		475 (477)	475 (482)	367 (381)
DSBE (1)	A3	477 (485)	474 (484)	425 (436)	A4	326 (341)	335 (343)	221 (235)
DSBE (2)		481 (490)	478 (489)	429 (439)		330 (347)	345 (353)	232 (246)
DSBE (3)		479 (490)	477 (492)	436 (445)		339 (354)	352 (358)	240 (256)
DSBE (4)		480 (490)	478 (493)	443 (452)		339 (354)	362 (370)	254 (276)
DSBE (1)	B3	316 (330)	317 (336)	224 (249)	B4	223 (226)	252 (254)	204 (208)
DSBE (2)		321 (335)	327 (348)	226 (257)		225 (225)	249 (252)	204 (209)
DSBE (3)		321 (336)	324 (342)	223 (255)		224 (225)	249 (252)	203 (208)
DSBE (4)		326 (338)	319 (339)	233 (265)		225 (225)	250 (254)	202 (207)
DSBE (1)	C3	496 (500)	491 (492)	404 (409)	C4	202 (208)	218 (231)	164 (182)
DSBE (2)		498 (500)	491 (492)	403 (413)		209 (214)	222 (235)	171 (187)
DSBE (3)		498 (500)	492 (494)	407 (415)		213 (218)	225 (237)	175 (193)
DSBE (4)		498 (500)	493 (494)	409 (419)		217 (219)	228 (240)	173 (187)

$\hat{k}(s,t) = \sum_{\ell=1}^{\infty} \hat{\lambda}_\ell \hat{\phi}_\ell(s)\hat{\phi}_\ell(t)$ are defined as in Section 2.2. The BS$_{\text{CUST}}$ procedure is described as follows.

Algorithm BS$_{\text{CUST}}$:

1. Begin with the initial segment $(s, e] = (0, 1]$ and set the iteration index $k = 1$.

2. At the kth iteration, given a segment $(s, e]$ find the changepoint candidate $\hat{\theta}_k = \hat{h}/N$ by

$$\hat{h} = \underset{\lfloor Ns \rfloor < h \le \lfloor Ne \rfloor}{\operatorname{argmax}} T_{\text{CUS}}(h),$$

where

$$T_{\text{CUS}}(h) = \frac{U(h)^\top \hat{\boldsymbol{\Sigma}}^{-1} U(h)}{\lfloor Ne \rfloor - \lfloor Ns \rfloor}$$

with

$$U(h) = \sum_{i=\lfloor Ns_r \rfloor + 1}^{h} \left\{ \hat{\boldsymbol{\xi}}_i - \frac{\sum_{i=\lfloor Ns_r \rfloor + 1}^{\lfloor Ne_r \rfloor} \hat{\boldsymbol{\xi}}_i}{\lfloor Ne_r \rfloor - \lfloor Ns_r \rfloor} \right\}$$

and the sample long-run covariance $\hat{\boldsymbol{\Sigma}}$ (Berkes et al. (2009); Hörmann and Kokoszka (2010); Aston and Kirch (2012)) for $\{\hat{\boldsymbol{\xi}}_i\}$.

3. Test the null hypothesis H_0: $\hat{\theta}_k$ is NOT a changepoint. If H_0 cannot be rejected, stop; there is no additional changepoint in the interval $(s, e]$. Otherwise, reject H_0; conclude that $\hat{\theta}_k$ is a changepoint and iterate Steps 2–3 for the updated subintervals $(s, e] = (s, \hat{\theta}_k]$ and $(s, e] = (\hat{\theta}_k, e]$, and set $k = k + 1$.

When the sequence $\{X_i(t)\}_{i=1}^N$ consists of independent random functions, $\hat{\boldsymbol{\Sigma}}$ can be reduced to the diagonal matrix given by $\operatorname{diag}(\hat{\lambda}_1, \ldots, \hat{\lambda}_p)$. The calculation of $\hat{\boldsymbol{\Sigma}}$ involves a bandwidth q and its optimal choice remains an unresolved problem. Thus, we use several values of q in BS$_{\text{CUST}}$ for comparison.

Besides the standard binary segmentation BS$_{\text{CUST}}$ for a sequence of functional data, we also compare DSBE with a variant extended from the binary segmentation, the circular binary segmentation (CBS) (Olshen et al., 2004) coupled with an epidemic change model (CBS$_{\text{EC}}$). The basic idea of CBS differs from the conventional binary segmentation. Given an interval $(s, e]$, the conventional binary segmentation searches for a changepoint $\hat{\theta}$ such that the mean functions of the subintervals $(s, \hat{\theta}]$ and $(\hat{\theta}, e]$ differentiate most. In contrast, CBS detects an epidemic change in $(s, e]$ by searching a subinterval $(\hat{\theta}_k, \hat{\theta}_{k'}] \subset (s, e]$ such that the mean function of the subinterval differs most from that of the complement $(s, e] \setminus (\hat{\theta}_k, \hat{\theta}_{k'}]$. In particular, when $\hat{\theta}_k = s$ or $\hat{\theta}_{k'} = e$, CBS reduces to the conventional binary segmentation. Here, CBS$_{\text{EC}}$ applies the searching scheme of CBS coupled with the test statistic proposed by Aston and Kirch (2012) for detecting epidemic changes in a sequence of functional data, which can be viewed as a generalization of BS$_{\text{CUST}}$. The CBS$_{\text{EC}}$ procedure is described as follows.

Algorithm CBS$_{\text{EC}}$:

1. Begin with the initial segment $(s, e] = (0, 1]$ and set the iteration index $k = 1$.

2. At the kth iteration, given a segment $(s, e]$ find the epidemic changepoint candidates $(\hat{\theta}_k, \hat{\theta}_{k'}) = (\hat{h}/N, \hat{h}'/N)$ by

$$(\hat{h}, \hat{h}') = \operatorname*{argmax}_{\lfloor Ns \rfloor < h < h' \le \lfloor Ne \rfloor} T_{\text{EC}}(h, h'),$$

where

$$T_{\text{EC}}(h, h') = \frac{\{U(h') - U(h)\}^\top \hat{\boldsymbol{\Sigma}}^{-1} \{U(h') - U(h)\}}{\lfloor Ne \rfloor - \lfloor Ns \rfloor},$$

where $U(h)$ and $\hat{\boldsymbol{\Sigma}}$ are as defined in BS$_{\text{CUST}}$.

TABLE 6
Frequencies of correctly estimated total numbers and averages (in parentheses) of estimated total numbers of changepoints ($N = 200$)

	$\boldsymbol{\theta} = \text{none}$		
Method	$\rho = 0$	$\rho = 0.2$	$\rho = 0.5$
DSBE	500 (0.00)	500 (0.00)	498 (0.00)
BSCUST (0)	476 (0.05)	364 (0.45)	67 (7.08)
BSCUST (1)	478 (0.05)	423 (0.18)	237 (1.26)
BSCUST (3)	486 (0.03)	468 (0.07)	398 (0.25)
BSCUST (5)	488 (0.02)	484 (0.03)	443 (0.12)
BSCUST (7)	492 (0.02)	488 (0.02)	471 (0.06)
CBSEC (0)	476 (0.10)	294 (1.52)	10 (9.42)
CBSEC (1)	485 (0.05)	417 (0.39)	124 (3.76)
CBSEC (3)	490 (0.03)	463 (0.14)	356 (0.68)
CBSEC (5)	495 (0.02)	480 (0.07)	437 (0.24)
CBSEC (7)	498 (0.01)	491 (0.04)	466 (0.13)

3. Test the null hypothesis H_0: $(\hat{\theta}_k, \hat{\theta}_{k'})$ is NOT a pair of epidemic changepoints. If H_0 cannot be rejected, stop; there is no additional changepoint in the interval $(s, e]$. Otherwise, reject H_0; conclude that both $\hat{\theta}_k$ and $\hat{\theta}_{k'}$ are changepoints and iterate Steps 2–3 for the updated subintervals $(s, e] = (s, \hat{\theta}_k]$, $(s, e] = (\hat{\theta}_k, \hat{\theta}_{k'}]$ and $(s, e] = (\hat{\theta}_{k'}, e]$, and set $k = k + 1$.

The results are shown in Tables 6–8, where the values of q are indicated in BSCUST (q) and CBSEC (q). We present the results with $q = 0, 1, 3, 5, 7$ and $N = 200$. Additional simulation results with the sample size $N = 100$ are available in Supplemental Material (Chiou, Chen and Hsing (2019)).

The performance of BSCUST and CBSEC is influenced by the choice of the bandwidth while we have shown DSBE is stable with respect to the choice of K; however, no optimal choice of q can be consistently identified in the simulation. Second, the correct rates using BSCUST and CBSEC also decrease as the strength of autocorrelation increases, especially for CBSEC. Since CBSEC searches subinterval rather than splitting the interval, the finite sample performance is influenced more severely by the length of segments than BSCUST. When the changepoint locations are evenly distributed, such as in scenarios B2 and C3, BSCUST and CBSEC perform equally well, even comparable with DSBE. However, the correct detection rate of BSCUST and CBSEC drops when the segment lengths are uneven. By contrast, the performance of DSBE is rather stable in these scenarios, with correct detection rates over 80% in most cases (except for scenarios B3 and C4 under moderate dependence). In particular, when there is dependence ($\rho = 0.2$ or 0.5) DSBE outperforms BSCUST and CBSEC as the number of changepoints increases to $M = 4$ regardless of which bandwidth is used. Overall, we conclude that under zero to

TABLE 7
Frequencies of the correctly estimated changepoints for exact and neighboring (in the parentheses) position ($N = 200$)

Method	$\boldsymbol{\theta}$	$\rho = 0$	$\rho = 0.2$	$\rho = 0.5$	$\boldsymbol{\theta}$	$\rho = 0$	$\rho = 0.2$	$\rho = 0.5$
DSBE	A1	500 (500)	500 (500)	494 (496)	A2	497 (498)	494 (496)	400 (420)
BSCUST (0)		432 (445)	281 (295)	25 (27)		420 (420)	198 (198)	6 (6)
BSCUST (1)		453 (466)	383 (406)	177 (194)		459 (459)	376 (376)	113 (113)
BSCUST (3)		450 (474)	423 (446)	337 (365)		487 (487)	453 (453)	372 (372)
BSCUST (5)		423 (459)	410 (441)	336 (368)		308 (315)	339 (342)	363 (366)
BSCUST (7)		366 (414)	355 (383)	265 (295)		0 (0)	0 (0)	2 (2)
CBSEC (0)		451 (465)	253 (268)	10 (11)		455 (455)	205 (205)	0 (0)
CBSEC (1)		463 (476)	384 (409)	111 (121)		477 (477)	412 (412)	86 (86)
CBSEC (3)		442 (471)	420 (445)	276 (300)		353 (354)	401 (401)	374 (374)
CBSEC (5)		282 (316)	261 (288)	199 (218)		16 (20)	38 (41)	91 (97)
CBSEC (7)		20 (26)	29 (34)	25 (27)		0 (0)	0 (0)	0 (0)
DSBE	B1	499 (500)	500 (500)	494 (496)	B2	498 (500)	493 (500)	444 (468)
BSCUST (0)		434 (434)	275 (275)	33 (33)		409 (409)	209 (209)	9 (9)
BSCUST (1)		452 (452)	375 (375)	177 (177)		446 (446)	335 (335)	98 (98)
BSCUST (3)		474 (474)	455 (455)	348 (348)		484 (484)	461 (461)	352 (352)
BSCUST (5)		488 (488)	480 (480)	443 (443)		499 (499)	492 (492)	481 (482)
BSCUST (7)		498 (498)	491 (491)	480 (480)		489 (499)	495 (500)	490 (492)
CBSEC (0)		457 (457)	243 (243)	3 (3)		437 (437)	162 (162)	0 (0)
CBSEC (1)		469 (469)	368 (368)	84 (84)		473 (473)	366 (366)	46 (46)
CBSEC (3)		488 (488)	463 (463)	330 (330)		496 (496)	482 (482)	383 (383)
CBSEC (5)		495 (495)	488 (488)	444 (444)		499 (499)	500 (500)	494 (494)
CBSEC (7)		499 (499)	496 (496)	487 (487)		497 (500)	500 (500)	500 (500)
DSBE	C1	500 (500)	497 (499)	483 (485)	C2	498 (499)	495 (497)	454 (466)
BSCUST (0)		436 (440)	288 (289)	28 (28)		417 (417)	208 (208)	6 (6)
BSCUST (1)		452 (457)	389 (393)	170 (174)		450 (450)	366 (366)	112 (112)
BSCUST (3)		473 (480)	462 (471)	373 (386)		488 (488)	473 (475)	405 (406)
BSCUST (5)		476 (485)	459 (473)	426 (443)		397 (418)	419 (435)	391 (396)
BSCUST (7)		456 (471)	450 (470)	418 (439)		27 (28)	22 (27)	16 (18)
CBSEC (0)		437 (441)	233 (235)	7 (8)		441 (441)	185 (185)	0 (0)
CBSEC (1)		459 (464)	381 (385)	92 (95)		481 (481)	388 (388)	65 (65)
CBSEC (3)		477 (484)	456 (466)	340 (353)		427 (432)	438 (439)	376 (377)
CBSEC (5)		467 (483)	441 (462)	380 (399)		62 (66)	89 (100)	133 (138)
CBSEC (7)		246 (263)	268 (284)	245 (260)		1 (1)	3 (4)	3 (5)

low dependence, all these three methods are comparable with each other when there is at most one changepoint, and DSBE outperforms BSCUST and CBSEC in the detection of multiple changepoints.

6. Discussions and concluding remarks. In the application of freeway traffic segmentation, our interest is to identify the locations with significant changes in the aggregated traffic speed from the viewpoint of macroscopic traffic flow. A com-

TABLE 8
Frequencies of the correctly estimated changepoints for exact and neighboring (in the parentheses) position ($N = 200$)

Method	$\boldsymbol{\theta}$	$\rho = 0$	$\rho = 0.2$	$\rho = 0.5$	$\boldsymbol{\theta}$	$\rho = 0$	$\rho = 0.2$	$\rho = 0.5$
DSBE	A3	493 (499)	492 (499)	477 (485)	A4	478 (499)	460 (481)	375 (417)
BSCUST (0)		401 (407)	215 (216)	9 (9)		350 (366)	162 (171)	3 (3)
BSCUST (1)		449 (453)	376 (377)	165 (165)		415 (430)	316 (329)	124 (133)
BSCUST (3)		466 (473)	457 (465)	412 (414)		453 (473)	422 (444)	368 (384)
BSCUST (5)		4 (5)	6 (6)	3 (3)		314 (337)	288 (311)	95 (109)
BSCUST (7)		0 (0)	0 (0)	0 (0)		0 (0)	5 (5)	1 (1)
CBSEC (0)		448 (453)	293 (295)	8 (8)		428 (449)	260 (276)	8 (9)
CBSEC (1)		467 (475)	416 (420)	152 (152)		445 (462)	385 (403)	151 (153)
CBSEC (3)		478 (488)	458 (470)	405 (408)		240 (294)	259 (302)	156 (170)
CBSEC (5)		437 (467)	439 (462)	432 (442)		2 (4)	6 (7)	1 (1)
CBSEC (7)		161 (207)	175 (214)	235 (255)		0 (0)	0 (0)	0 (0)
DSBE	B3	465 (497)	437 (488)	272 (352)	B4	427 (428)	439 (443)	390 (404)
BSCUST (0)		357 (371)	183 (196)	17 (17)		365 (381)	182 (190)	8 (8)
BSCUST (1)		394 (415)	309 (333)	96 (115)		407 (436)	341 (359)	94 (100)
BSCUST (3)		402 (450)	388 (437)	255 (314)		124 (144)	111 (126)	52 (64)
BSCUST (5)		355 (424)	346 (420)	270 (337)		1 (1)	0 (0)	0 (0)
BSCUST (7)		14 (30)	31 (42)	37 (48)		0 (0)	0 (0)	0 (0)
CBSEC (0)		359 (415)	152 (196)	4 (6)		442 (447)	219 (220)	4 (4)
CBSEC (1)		389 (448)	270 (339)	51 (81)		434 (443)	364 (370)	87 (90)
CBSEC (3)		335 (403)	300 (372)	179 (236)		4 (7)	8 (9)	5 (10)
CBSEC (5)		86 (114)	73 (100)	69 (108)		0 (0)	0 (0)	0 (0)
CBSEC (7)		18 (21)	18 (21)	16 (23)		0 (0)	0 (0)	0 (0)
DSBE	C3	498 (500)	496 (500)	456 (473)	C4	412 (425)	413 (435)	348 (387)
BSCUST (0)		380 (380)	199 (199)	12 (12)		370 (385)	176 (183)	5 (5)
BSCUST (1)		416 (416)	324 (327)	109 (109)		390 (424)	321 (347)	112 (120)
BSCUST (3)		464 (464)	430 (433)	357 (358)		266 (312)	263 (308)	152 (175)
BSCUST (5)		491 (493)	482 (486)	471 (474)		2 (2)	0 (2)	4 (5)
BSCUST (7)		493 (500)	490 (499)	489 (494)		0 (0)	0 (0)	0 (0)
CBSEC (0)		429 (429)	185 (185)	3 (3)		346 (395)	183 (213)	6 (7)
CBSEC (1)		459 (460)	358 (358)	67 (67)		288 (371)	243 (295)	71 (91)
CBSEC (3)		485 (493)	470 (474)	382 (383)		23 (28)	23 (35)	16 (20)
CBSEC (5)		461 (499)	473 (494)	426 (435)		0 (0)	0 (0)	0 (0)
CBSEC (7)		356 (425)	297 (339)	215 (226)		0 (0)	0 (0)	0 (0)

bination of various factors may induce such changes. For instance, it could be the geometric designs of roads that affect driving behaviors, such as lane width, the number of lanes, slope gradient, road curve radius, curvature change rate, sight distance, etc., or it could also be the interferences or changes of traffic flow that are present at the entrance or the exit ramp of an interchange.

While traffic shockwaves [May (1990)] exist whenever traffic conditions change, propagation speed induced by traffic shockwaves might also play a role

in speed change. For instance, traffic congestions cause shockwaves from the upstream to the downstream roads may result in a delaying effect or propagation of speeds. However, these considerations are complex and may involve other factors including those mentioned in the previous paragraph. Moreover, how these delaying effects interact with the detection of changepoints is also unclear. For instance, the plots in Figure 4 shows little evidence that the shift of mean speed due to propogation of shockwaves plays a significant role on changepoint detection. In general, measuring and modeling propagation speed under a variety of traffic conditions is an important and challenging topic. In particular, the influence of propagation speed on detection of changepoints requires further study and is beyond the scope of the present paper. A better understanding of these issues would potentially enhance the efficacy of online travel time prediction using traffic speed data along with the tools introduced in this paper.

Unlike most previous approaches on changepoint detection for functional data, which assume the AMOC assumption holds, our approach accommodates the existence of multiple changepoints in a functional data sequence and designs an optimality criterion accordingly. The proposed DSBE procedure adopts the local AMOC strategy to achieve the optimal global criterion for multiple changepoint detection, which is conceptually straightforward, easy to implement, and computationally feasible and leads to consistent estimators for multiple changepoint detection. The numerical results show that the proposed DSBE procedure performs well, and outperforms other binary segmentation-based approaches for multiple changepoint detection in many scenarios.

While our objective is to deal with the multiple changepoint problem from the viewpoint of global optimization and introduce the proposed optimality criterion, we use the empirical estimator in estimation for ease of presentation, assuming that the random functions can be observed completely. The DSBE approach certainly is valid for sparsely collected functional data that could be contaminated with measurement errors. Even though this study was motivated by the interest to study freeway traffic segmentation, the DSBE algorithm can be broadly applied to applications with a functional data sequence. Besides the vehicle speed, other traffic parameters can also be considered for freeway traffic segmentation, in which cases the tools for multivariate functional data analysis [Chiou, Chen and Yang (2014), Happ and Greven (2018)] can also be used with DSBE. We conclude that the proposed DSBE algorithm is a simple and useful approach for the general problem of multiple changepoint detection.

APPENDIX A: PROOF OF LEMMA 1

For an arbitrary K-segmentation $\boldsymbol{\theta} \in \Theta$, let

$$Z_{i,\boldsymbol{\theta}}(t) = \sum_{k=1}^{K+1} \{Y_i(t) - \bar{Y}_{k,\boldsymbol{\theta}}(t)\} \cdot \mathbf{1}_{(\theta_{k-1},\theta_k]}(i/N)$$

and

$$v_{i,\boldsymbol{\theta}}(t) = \sum_{k=1}^{K+1} \{\mu_i(t) - \bar{\mu}_{k,\boldsymbol{\theta}}(t)\} \cdot \mathbf{1}_{(\theta_{k-1},\theta_k]}(i/N),$$

where

$$\bar{Y}_{k,\boldsymbol{\theta}}(t) = \tfrac{1}{N_k(\boldsymbol{\theta})} \textstyle\sum_{i=1}^{N} Y_i(t) \cdot \mathbf{1}_{(\theta_{k-1},\theta_k]}(i/N);$$
$$\mu_i(t) = \mathrm{E}X_i(t) = \textstyle\sum_{m=1}^{M+1} \mu_m(t) \cdot \mathbf{1}_{(\theta^*_{m-1},\theta^*_m]}(i/N);$$
$$\bar{\mu}_{k,\boldsymbol{\theta}}(t) = \tfrac{1}{N_k(\boldsymbol{\theta})} \textstyle\sum_{i=1}^{N} \mu_i(t) \cdot \mathbf{1}_{(\theta_{k-1},\theta_k]}(i/N).$$

Given the K-segmentation $\boldsymbol{\theta}$, define the associated empirical segment mean function to be $\hat{\mu}_{i,\boldsymbol{\theta}}(t) = \sum_{k=1}^{K+1} \bar{X}_{(\theta_{k-1},\theta_k]}(t) \cdot \mathbf{1}_{(\theta_{k-1},\theta_k]}(i/N)$, where $\bar{X}_{(\theta_{k-1},\theta_k]}(t)$ is as defined in (2). Then, the empirical covariance function estimator conditional on $\boldsymbol{\theta}$ in (3) can be expressed as

$$\begin{aligned}
\hat{c}(s,t \mid \boldsymbol{\theta}) &= \frac{1}{N}\sum_{i=1}^{N}\{X_i(s) - \hat{\mu}_{i,\boldsymbol{\theta}}(s)\}\{X_i(t) - \hat{\mu}_{i,\boldsymbol{\theta}}(t)\} \\
&= \frac{1}{N}\sum_{i=1}^{N}\{Z_{i,\boldsymbol{\theta}}(s) + v_{i,\boldsymbol{\theta}}(s)\}\{Z_{i,\boldsymbol{\theta}}(t) + v_{i,\boldsymbol{\theta}}(t)\} \\
&= \frac{1}{N}\sum_{i=1}^{N} Z_{i,\boldsymbol{\theta}}(s)Z_{i,\boldsymbol{\theta}}(t) + \frac{1}{N}\sum_{i=1}^{N} v_{i,\boldsymbol{\theta}}(s)v_{i,\boldsymbol{\theta}}(t) \\
&\quad + \frac{1}{N}\sum_{i=1}^{N} Z_{i,\boldsymbol{\theta}}(s)v_{i,\boldsymbol{\theta}}(t) + \frac{1}{N}\sum_{i=1}^{N} v_{i,\boldsymbol{\theta}}(s)Z_{i,\boldsymbol{\theta}}(t) \\
&=: A_1(s,t) + A_2(s,t) + A_3(s,t) + A_4(s,t).
\end{aligned}$$

For notational convenience, denote the kth segment of $\boldsymbol{\theta}$ by $\mathcal{I}_k(\boldsymbol{\theta}) = (\theta_{k-1}, \theta_k]$. The following observations are used in the proof:

(I) If $i/N, j/N \in \mathcal{I}_m(\boldsymbol{\theta}^*)$ for some m, then $\mu_i(t) = \mu_j(t) = \mu_m(t)$.

(II) The $\bar{\mu}_{k,\boldsymbol{\theta}}(t)$ is a linear combination of $\{\mu_m(t), m = 1, \ldots, M+1\}$. Specifically, we can write $\bar{\mu}_{k,\boldsymbol{\theta}}(t) = \sum_{m=1}^{M+1} \tilde{\beta}_{\mathrm{km}}(\boldsymbol{\theta})\mu_m(t)$, where $0 \le \tilde{\beta}_{\mathrm{km}}(\boldsymbol{\theta}) \le 1$ and $\sum_{m=1}^{M+1} \tilde{\beta}_{\mathrm{km}}(\boldsymbol{\theta}) = 1$;

(III) When $\mathcal{I}_k(\boldsymbol{\theta})$ does not contain any true changepoint; that is, $\theta^*_m \notin \mathcal{I}_k(\boldsymbol{\theta})$ for all m, it holds that $\mathcal{I}_k(\boldsymbol{\theta}) \subseteq \mathcal{I}_{m_0}(\boldsymbol{\theta}^*)$ for some $1 \le m_0 \le M+1$. In this case, $\mu_i(t) = \mu_{m_0}(t)$ for all $i/N \in \mathcal{I}_k(\boldsymbol{\theta})$ according to (I) and, therefore, $\bar{\mu}_{k,\boldsymbol{\theta}}(t) = \mu_i(t) = \mu_{m_0}(t)$ for all $i/N \in \mathcal{I}_k(\boldsymbol{\theta})$;

(IV) If $i/N, j/N \in \mathcal{I}_l(\boldsymbol{\theta} \cup \boldsymbol{\theta}^*)$ for some l, then $v_{i,\boldsymbol{\theta}}(t) = v_{j,\boldsymbol{\theta}}(t)$.

The coefficients $\tilde{\beta}_{\mathrm{km}}(\boldsymbol{\theta})$ in (II) rely on the relative positions between $\boldsymbol{\theta}$ and $\boldsymbol{\theta}^*$. For example, if $\mathcal{I}_k(\boldsymbol{\theta})$ contains no true changpoints θ^*_m as in (III), then $\tilde{\beta}_{\mathrm{km}}(\boldsymbol{\theta}) = 0$

if $m \neq m_0$ and $\tilde{\beta}_{\mathrm{km}_0}(\boldsymbol{\theta}) = 1$. If there exists exactly one changepoint $\theta^*_{m_0}$ in $\mathcal{I}_k(\boldsymbol{\theta})$, then $\bar{\mu}_{k,\boldsymbol{\theta}}(t) = \frac{\lfloor N\theta^*_{m_0}\rfloor - \lfloor N\theta_{k-1}\rfloor}{N_k(\boldsymbol{\theta})}\mu_{m_0}(t) + \frac{\lfloor N\theta_k\rfloor - \lfloor N\theta^*_{m_0}\rfloor}{N_k(\boldsymbol{\theta})}\mu_{m_0+1}(t)$. For other scenarios, the value of $\tilde{\beta}_{\mathrm{km}}(\boldsymbol{\theta})$ can also be obtained through similar arguments.

Now,

$$\begin{aligned} A_1(s,t) &= \frac{1}{N}\sum_{i=1}^{N} Z_{i,\boldsymbol{\theta}}(s)Z_{i,\boldsymbol{\theta}}(t) \\ &= \sum_{k=1}^{K+1} \frac{N_k(\boldsymbol{\theta})}{N} \cdot \frac{1}{N_k(\boldsymbol{\theta})} \sum_{i/N\in\mathcal{I}_k(\boldsymbol{\theta})} Z_{i,\boldsymbol{\theta}}(s)Z_{i,\boldsymbol{\theta}}(t). \end{aligned}$$

Under the assumptions that $\{Y_i\}$ is $\mathcal{L}^4 - m$-approximable and $N_k(\boldsymbol{\theta})$ and N are the same order of magnitude as $N \to \infty$, the fact that $N_k(\boldsymbol{\theta})/N \to \gamma_k := \theta_k - \theta_{k-1}$ and $\sum_k \gamma_k = 1$ implies

$$\mathcal{A}_1 \xrightarrow{\mathcal{P}} \sum_{k=1}^{K+1} \gamma_k \mathcal{C} = \mathcal{C},$$

where $\mathcal{A}_1$ is the integral operator with kernel $A_1(s,t)$ and $\mathcal{C}$ is the integral operator with kernel $c(s,t) = \mathrm{cov}(Y_i(s), Y_i(t))$.

To consider the convergence of $A_2(s,t)$, by (II) we have, for $i/N \in \mathcal{I}_k(\boldsymbol{\theta})$,

$$\begin{aligned} \nu_{i,\boldsymbol{\theta}}(t) &= \mu_i(t) - \bar{\mu}_{k,\boldsymbol{\theta}}(t) \\ &= \sum_{m=1}^{M+1} \tilde{\beta}_{\mathrm{km}}(\boldsymbol{\theta})\mu_i(t) - \sum_{m=1}^{M+1} \tilde{\beta}_{\mathrm{km}}(\boldsymbol{\theta})\mu_m(t) \\ &= \sum_{m=1}^{M+1} \tilde{\beta}_{\mathrm{km}}(\boldsymbol{\theta})\{\mu_i(t) - \mu_m(t)\}. \end{aligned}$$

By further calculations, we have

$$\begin{aligned} &\frac{1}{N_k(\boldsymbol{\theta})} \sum_{i/N\in\mathcal{I}_k(\boldsymbol{\theta})} \nu_{i,\boldsymbol{\theta}}(s)\nu_{i,\boldsymbol{\theta}}(t) \\ &\quad = \sum_{r<m} \tilde{\beta}_{kr}(\boldsymbol{\theta})\tilde{\beta}_{km}(\boldsymbol{\theta})\{\mu_r(s) - \mu_m(s)\}\{\mu_r(t) - \mu_m(t)\}. \end{aligned}$$

Therefore,

$$\begin{aligned} A_2(s,t) &= \frac{1}{N}\sum_{i=1}^{N} \nu_{i,\boldsymbol{\theta}}(s)\nu_{i,\boldsymbol{\theta}}(t) \\ &= \sum_{k=1}^{K+1} \frac{N_k(\boldsymbol{\theta})}{N} \cdot \frac{1}{N_k(\boldsymbol{\theta})} \sum_{i/N\in\mathcal{I}_k(\boldsymbol{\theta})} \nu_{i,\boldsymbol{\theta}}(s)\nu_{i,\boldsymbol{\theta}}(t) \end{aligned}$$

$$\to \sum_{k=1}^{K+1} \gamma_k \sum_{1\le r<m\le K+1} \beta_{\mathrm{kr}}(\boldsymbol{\theta})\beta_{\mathrm{km}}(\boldsymbol{\theta})\{\mu_r(s)-\mu_m(s)\}\{\mu_r(t)-\mu_m(t)\},$$

where $\beta_{\mathrm{km}}(\boldsymbol{\theta}) = \lim_{N\to\infty} \tilde{\beta}_{\mathrm{km}}(\boldsymbol{\theta})$ is determined by the length of $\{\mathcal{I}_l(\boldsymbol{\theta}\cup\boldsymbol{\theta}^*)\}$. Let

$$\alpha_{r,m}(\boldsymbol{\theta}) = \sum_{k=1}^{K+1} \gamma_k \beta_{\mathrm{kr}}(\boldsymbol{\theta})\beta_{\mathrm{km}}(\boldsymbol{\theta}). \tag{9}$$

Then,

$$B_{\boldsymbol{\theta}}(s,t) = \sum_{1\le r<m\le K+1} \alpha_{r,m}(s,t)\{\mu_r(s)-\mu_m(s)\}\{\mu_r(t)-\mu_m(t)\}$$

is the kernel of $\mathcal{B}(\boldsymbol{\theta})$. When $\boldsymbol{\theta}$ contains $\boldsymbol{\theta}^*$ as a subset, by (III) we have $v_{i,\boldsymbol{\theta}}(t) \equiv 0$ for all i and, hence, $A_2(s,t) = 0$ and $B_{\boldsymbol{\theta}}(s,t) = 0$ for all s and t.

By (IV),

$$\begin{aligned} A_3(s,t) &= \frac{1}{N}\sum_{i=1}^{N} Z_{i,\boldsymbol{\theta}}(s)v_{i,\boldsymbol{\theta}}(t) \\ &= \sum_{l=1}^{K'+1} \frac{v_l(t)}{N} \sum_{i/N\in\mathcal{I}_l(\boldsymbol{\theta}\cup\boldsymbol{\theta}^*)} Z_{i,\boldsymbol{\theta}}(s), \end{aligned}$$

where K' is the number of elements in $\boldsymbol{\theta}\cup\boldsymbol{\theta}^*$, $v_l(t) = v_{i,\boldsymbol{\theta}}(t)$ for some $i/N \in \mathcal{I}_l(\boldsymbol{\theta}\cup\boldsymbol{\theta}^*)$. Since we assume $X_i(t)$ is continuous in $\mathcal{T}$, so are $\mu_m(t)$'s. Moreover, because $\mathcal{T}$ is a closed interval, we have $\sup_{t\in\mathcal{T}}|\mu_m(t)| < \infty$ for each m. It then follows that $\sup_{t\in\mathcal{T}}|v_l(t)| < \infty$ for each l and $\frac{v_l(t)}{N}\sum_{i/N\in\mathcal{I}_l(\boldsymbol{\theta}\cup\boldsymbol{\theta}^*)} Z_{i,\boldsymbol{\theta}}(s) \xrightarrow{p} 0$ because $\mathrm{E}Y_i(t) = 0$ and thus $\mathrm{E}Z_{i,\boldsymbol{\theta}}(t) = 0$. Since the number $|\boldsymbol{\theta}\cup\boldsymbol{\theta}^*| \le M+K+1 < \infty$, we have $A_3(s,t) \xrightarrow{p} 0$. Similarly, we can prove $A_4(s,t) \xrightarrow{p} 0$. Let $\mathcal{A}_i$ denote the integral operator with the kernel $A_i(s,t)$, $i = 1,\dots,4$. Then, for an arbitrary $\boldsymbol{\theta}$,

$$\begin{aligned} \|\hat{\mathcal{C}}_{\boldsymbol{\theta}} - \mathcal{C} - \mathcal{B}(\boldsymbol{\theta})\|_{\mathrm{HS}} &\le \|\mathcal{A}_1 - \mathcal{C}\|_{\mathrm{HS}} + \|\mathcal{A}_2 - \mathcal{B}(\boldsymbol{\theta})\|_{\mathrm{HS}} + \|\mathcal{A}_3\|_{\mathrm{HS}} + \|\mathcal{A}_4\|_{\mathrm{HS}} \\ &\xrightarrow{\mathcal{P}} 0, \end{aligned}$$

which completes the proof.

APPENDIX B: PROOF OF THEOREM 1

We denote the $\mathcal{C} + \mathcal{B}(\boldsymbol{\theta})$ in (5) as $\mathcal{C}_{\boldsymbol{\theta}}$ for notational convenience. We first use Lemma 1 to argue that $\sup_{\boldsymbol{\theta}\in\Theta}|T_N(\boldsymbol{\theta}) - T(\boldsymbol{\theta})| \xrightarrow{p} 0$, and then use the argmin continuous mapping theorem to show the corresponding convergence outcome, $\tilde{\theta}_{N,k} \xrightarrow{p} \theta_m^*$.

For a fixed ℓ, it holds that

$$\begin{aligned}\big|\langle \mathcal{C}_{\boldsymbol{\theta}}\phi_\ell,\phi_\ell\rangle - \langle \hat{\mathcal{C}}_{\boldsymbol{\theta}}\hat{\phi}_\ell,\hat{\phi}_\ell\rangle\big| &\le \big|\langle \mathcal{C}_{\boldsymbol{\theta}}\phi_\ell,\phi_\ell\rangle - \langle \mathcal{C}_{\boldsymbol{\theta}}\phi_\ell,\hat{\phi}_\ell\rangle\big| + \big|\langle \mathcal{C}_{\boldsymbol{\theta}}\phi_\ell,\hat{\phi}_\ell\rangle - \langle \mathcal{C}_{\boldsymbol{\theta}}\hat{\phi}_\ell,\hat{\phi}_\ell\rangle\big| \\ &\quad + \big|\langle \mathcal{C}_{\boldsymbol{\theta}}\hat{\phi}_\ell,\hat{\phi}_\ell\rangle - \langle \hat{\mathcal{C}}_{\boldsymbol{\theta}}\hat{\phi}_\ell,\hat{\phi}_\ell\rangle\big| \\ &= \big|\langle \mathcal{C}_{\boldsymbol{\theta}}\phi_\ell,(\phi_\ell-\hat{\phi}_\ell)\rangle\big| + \big|\langle \mathcal{C}_{\boldsymbol{\theta}}(\phi_\ell-\hat{\phi}_\ell),\hat{\phi}_\ell\rangle\big| \\ &\quad + \big|\langle (\mathcal{C}_{\boldsymbol{\theta}}-\hat{\mathcal{C}}_{\boldsymbol{\theta}})\hat{\phi}_\ell,\hat{\phi}_\ell\rangle\big| \\ &\le 2\|\mathcal{C}_{\boldsymbol{\theta}}\|_{\mathrm{HS}}\cdot\|\phi_\ell-\hat{\phi}_\ell\| + \|\mathcal{C}_{\boldsymbol{\theta}}-\hat{\mathcal{C}}_{\boldsymbol{\theta}}\|_{\mathrm{HS}}.\end{aligned}$$

The above inequality implies, for all K-segmentation $\boldsymbol{\theta}$, that

$$\big|\langle \hat{\mathcal{C}}_{\boldsymbol{\theta}}\hat{\phi}_\ell,\hat{\phi}_\ell\rangle - \langle \mathcal{C}_{\boldsymbol{\theta}}\phi_\ell,\phi_\ell\rangle\big| \xrightarrow{p} 0$$

if $\|\hat{\phi}_\ell-\phi_\ell\| = o_p(1)$ and $\|\mathcal{C}_{\boldsymbol{\theta}}-\hat{\mathcal{C}}_{\boldsymbol{\theta}}\|_{\mathrm{HS}} = o_p(1)$. The former holds if $\eta_0>0$ by Lemma 4.3 in Bosq (2000) and the latter follows by Lemma 1. Let $T(\boldsymbol{\theta}) = \sum_{\ell=1}^{p}\langle \mathcal{C}_{\boldsymbol{\theta}}\phi_\ell,\phi_\ell\rangle$. Taking the summation over ℓ on the previous convergence result, we have, for an arbitrary K-segmentation $\boldsymbol{\theta}$,

$$\big|T_N(\boldsymbol{\theta}) - T(\boldsymbol{\theta})\big| \xrightarrow{p} 0. \tag{10}$$

To show that the stochastic equicontinuity of $T_N(\boldsymbol{\theta})$, for two arbitrary K-segmentation $\boldsymbol{\theta}$ and $\boldsymbol{\theta}'$, let $d(\boldsymbol{\theta},\boldsymbol{\theta}') = \sup_k|\theta_k-\theta_k'|$ be the distance between $\boldsymbol{\theta}$ and $\boldsymbol{\theta}'$. By definition, we write

$$\hat{c}(s,t|\boldsymbol{\theta}) = \frac{1}{N}\sum_{i=1}^{N}X_i(s)X_i(t) - \frac{1}{N}\sum_{k=1}^{K+1}N_k(\boldsymbol{\theta})\bar{X}_{(\theta_{k-1},\theta_k]}(s)\bar{X}_{(\theta_{k-1},\theta_k]}(t),$$

where only the second term depends on $\boldsymbol{\theta}$. Let

$$B_k = \max\Big\{\sup_{t\in\mathcal{T}}|\bar{X}_{(\theta_{k-1},\theta_k]}(t)|, \sup_{t\in\mathcal{T}}|\bar{X}_{(\theta'_{k-1},\theta'_k]}(t)|\Big\}.$$

Then, $B_k = O_p(1)$ because $X_i(t)\in\mathcal{L}^2(\mathcal{T})$ and is continuous in t. For N sufficiently large, $N_k(\boldsymbol{\theta})/N = (\lfloor N\theta_k\rfloor - \lfloor N\theta_{k-1}\rfloor)/N \to \theta_k-\theta_{k-1}$, and

$$\begin{aligned}\big|\hat{c}(s,t|\boldsymbol{\theta}) - \hat{c}(s,t|\boldsymbol{\theta}')\big| &\le \sum_{k=1}^{K+1}2B_k^2\frac{|N_k(\boldsymbol{\theta})-N_k(\boldsymbol{\theta}')|}{N} \\ &\le 4\Bigg\{\sum_{k=1}^{K+1}B_k^2\Bigg\}d(\boldsymbol{\theta},\boldsymbol{\theta}'),\end{aligned}$$

when $N\gg 0$. By the definition of $T_N(\boldsymbol{\theta})$, it follows that $|T_N(\boldsymbol{\theta})-T_N(\boldsymbol{\theta}')|\le O_p(1)\,d(\boldsymbol{\theta},\boldsymbol{\theta}')$ and, thus, $T_N(\boldsymbol{\theta})$ is equicontinuous. Because $T(\boldsymbol{\theta})$ is continuous the totally bounded domain Θ, by Corollary 2.2 in Newey (1991) or Theorem 3.1 in Pötscher and Prucha (1994), we conclude that

$$\sup_{\boldsymbol{\theta}\in\Theta}\big|T_N(\boldsymbol{\theta}) - T(\boldsymbol{\theta})\big| \xrightarrow{p} 0. \tag{11}$$

Because $\boldsymbol{\theta}^*$ is a minimizer of $T(\boldsymbol{\theta})$, for every $\varepsilon > 0$,

$$\inf_{\boldsymbol{\theta}:\|\boldsymbol{\theta}-\boldsymbol{\theta}^*\|\geq\varepsilon} T(\boldsymbol{\theta}) > T\big(\boldsymbol{\theta}^*\big). \tag{12}$$

Moreover, since $\tilde{\boldsymbol{\theta}}_N$ is the minimizer of $T_N(\boldsymbol{\theta})$, the sequence of the estimators $\tilde{\boldsymbol{\theta}}_N$ satisfies

$$T_N(\tilde{\boldsymbol{\theta}}_N) \leq T_N\big(\boldsymbol{\theta}^*\big) + o_p(1). \tag{13}$$

Given that T_N is a random function of $\boldsymbol{\theta}$ and T is a fixed function of $\boldsymbol{\theta}$ satisfying the conditions (11) and (12) and for a sequence of the estimator $\tilde{\boldsymbol{\theta}}_N$ satisfying (13), it follows, by the argmin continuous mapping theorem [van der Vaart and Wellner (1996)] and Lemma 1, that there are M out of the K elements in $\tilde{\boldsymbol{\theta}}_N$ that converge to the M elements in $\boldsymbol{\theta}^*$, respectively.

APPENDIX C: PROOF OF THEOREM 2

By restricting the entire interval to a subinterval $(\theta_l, \theta_r]$ and following the argument similarly to the proof in Theorem 1, we have the following lemma:

LEMMA 2. *Under the assumption of Theorem 1, given a subinterval* $(\theta_l, \theta_r] \subset (0, 1]$, *if there are* L $(L \geq 1)$ *changepoints* $\{\theta^*_{m+1}, \theta^*_{m+2}, \ldots, \theta^*_{m+L}\}$ *in* $(\theta_l, \theta_r]$, *then*

$$\operatorname*{argmin}_{\theta\in(\theta_l,\theta_r]}\big\{S_{(\theta_l,\theta_r]}(\theta)\big\} \xrightarrow{p} \theta^*_{m_0},$$

where $\theta^*_{m_0} = \operatorname{argmin}_{1\leq q\leq L}\{\sum_{\ell=1}^{p}\langle\mathcal{B}_{(\theta_l,\theta_r]}(\theta^*_{m+q})\phi_\ell, \phi_\ell\rangle\}$ *and* $\mathcal{B}_{(\theta_l,\theta_r]}(\theta^*_{m+q})$ *is the counterpart of* $\mathcal{B}(\boldsymbol{\theta})$ *when restricting to* $(\theta_l, \theta_r]$.

As an illustrative example, consider that $\theta_l < \theta^*_1 < \theta^*_2 < \theta_r$. Let $\alpha_1 = (\theta^*_1 - \theta_l)/(\theta_r - \theta_l)$, $\alpha_2 = (\theta^*_2 - \theta^*_1)/(\theta_r - \theta_l)$ and $\alpha_3 = (\theta_r - \theta^*_2)/(\theta_r - \theta_l)$. By direct calculation it follows that $\sum_{\ell=1}^{p}\langle\mathcal{B}_{(\theta_l,\theta_r]}(\theta^*_1)\phi_\ell, \phi_\ell\rangle = \frac{\alpha_2\alpha_3}{\alpha_2+\alpha_3}\sum_{\ell=1}^{p}(\langle\mu_2-\mu_3, \phi_\ell\rangle)^2$ and $\sum_{\ell=1}^{p}\langle\mathcal{B}_{(\theta_l,\theta_r]}(\theta^*_1)\phi_\ell, \phi_\ell\rangle = \frac{\alpha_1\alpha_2}{\alpha_1+\alpha_2}\sum_{\ell=1}^{p}(\langle\mu_1 - \mu_2, \phi_\ell\rangle)^2$. Therefore we have $\operatorname{argmin}_{\theta\in(\theta_l,\theta_r]}\{S_{(\theta_l,\theta_r]}(\theta)\} \xrightarrow{p} \theta^*_1$ if

$$\frac{\alpha_2\alpha_3}{\alpha_2+\alpha_3}\sum_{\ell=1}^{p}\big(\langle\mu_2 - \mu_3, \phi_\ell\rangle\big)^2 < \frac{\alpha_1\alpha_2}{\alpha_1+\alpha_2}\sum_{\ell=1}^{p}\big(\langle\mu_1 - \mu_2, \phi_\ell\rangle\big)^2;$$

otherwise, $\operatorname{argmin}_{\theta\in(\theta_l,\theta_r]}\{S_{(\theta_l,\theta_r]}(\theta)\} \xrightarrow{p} \theta^*_2$. Similar arguments can be derived for the cases of more than two changepoints.

In particular, we note that a changepoint near the endpoint θ_l or θ_r will never be chosen through criterion $S_{(\theta_l,\theta_r]}(\theta)$ in the presence of other changepoints within $(\theta_l, \theta_r]$. For instance, in the above example we have $\operatorname{argmin}_{\theta\in(\theta_l,\theta_r]}\{S_{(\theta_l,\theta_r]}(\theta)\} \xrightarrow{p} \theta^*_1$ if $\theta_r - \theta^*_2 \approx 0$, which implies $\alpha_3 \approx 0$.

Using these preliminaries, we can prove Theorem 2. We first show the convergence of DS; that is, for each fixed N, $\boldsymbol{\theta}^{(r)}$ converges as $r \to \infty$. Then, we show the consistency of the changepoint estimators obtained by DS.

CLAIM 1. *For any fixed N, there exists a K-segmentation $\hat{\boldsymbol{\theta}}_N$ such that* $\lim_{r\to\infty} \boldsymbol{\theta}^{(r)} = \hat{\boldsymbol{\theta}}_N$.

We first show that $\{T_N(\boldsymbol{\theta}^{(r)})\}$ forms a positive nonincreasing sequence in r with the updating step (D3). Let $\boldsymbol{\theta}^{(r,k)} = \{\theta_1^{(r)}, \ldots, \theta_k^{(r)}, \theta_{k+1}^{(r-1)}, \theta_K^{(r-1)}\}$ be the segmentation after k updating steps in the rth iteration. Here, $\boldsymbol{\theta}^{(r,0)} = \boldsymbol{\theta}^{(r-1)}$ and $\boldsymbol{\theta}^{(r,K+1)} = \boldsymbol{\theta}^{(r)}$. By (7),

$$T_N\big(\boldsymbol{\theta}^{(r,k)}\big) - T_N\big(\boldsymbol{\theta}^{(r,k+1)}\big) = \frac{N_k^{(r)}}{N}\big\{S_{(\theta_k^{(r)},\theta_{k+2}^{(r-1)}]}\big(\theta_{k+1}^{(r-1)}\big) - S_{(\theta_k^{(r)},\theta_{k+2}^{(r-1)}]}\big(\theta_{k+1}^{(r)}\big)\big\},$$

where $N_k^{(r)} = \lfloor N\theta_{k+2}^{(r-1)} \rfloor - \lfloor N\theta_k^{(r)} \rfloor$.

By definition of $\theta_k^{(r)}$, $\theta_{k+1}^{(r-1)} - \theta_k^{(r)} > h$, and thus $\theta_{k+1}^{(r-1)} \in \mathcal{I}_{k+1}^{(r-1)}$. In addition, because $\theta_{k+1}^{(r)} = \inf[\operatorname{argmin}_{\theta\in\mathcal{I}_{k+1}^{(r-1)}(h)}\{S_{(\theta_k^{(r)},\theta_{k+2}^{(r-1)}]}(\theta)\}]$, we have $T_N(\boldsymbol{\theta}^{(r,k)}) - T_N(\boldsymbol{\theta}^{(r,k+1)}) \geq 0$ in probability. By applying this argument iteratively for each k, we conclude that $T_N(\boldsymbol{\theta}^{(r-1)}) - T_N(\boldsymbol{\theta}^{(r)}) \geq 0$ in probability; thus, $\{T_N(\boldsymbol{\theta}^{(r)})\}$ forms a positive nonincreasing sequence in r. Therefore, the sequence $\{T_N(\boldsymbol{\theta}^{(r)})\}$ converges in r. As a consequence, with the infimum in the definition of $\boldsymbol{\theta}^{(r)}$ for its uniqueness, the convergence of $\{\boldsymbol{\theta}^{(r)}\}$ follows directly from the convergence of $\{T_N(\boldsymbol{\theta}^{(r)})\}$.

CLAIM 2. $\lim_{N\to\infty} P(\bigcap_m\{|\theta_m^* - \theta_k^{(r)}| < h \text{ for some } k = 1, \ldots, K\}) = 1$ *for any* $r \geq 1$.

To prove the claim, it suffices to show for any $h, \epsilon > 0$,

$$P\left(\bigcup_m\Big\{\min_{1\leq k\leq K}|\theta_m^* - \theta_k^{(r)}| \geq h\Big\}\right) < \epsilon$$

for $r = 1$, given $N > N_0$. By the discussion of Lemma 2, changepoints near the endpoint cannot happen, and there is at most one changepoint within $\mathcal{I}_j^{(1)}(h)$ for each $1 \leq j \leq K$. For a sufficiently small h, we have by Lemma 2, $P(|\theta_m^* - \theta_j^{(2)}| \geq h\}) < \epsilon$ if there exists a $\theta_m^* \in \mathcal{I}_j^{(1)}(h)$. The same argument also applies to the subsequent iterations. For $r = 1$, we show $P(\bigcup_m\{\min_{1\leq k\leq K}|\theta_m^* - \theta_k^{(1)}| \geq h\}) < \epsilon$ through mathematical induction.

Assume that $P(\min_{1\leq k\leq K}|\theta_m^* - \theta_k^{(1)}| \geq h) < \epsilon$ for each $\theta_m^* \leq \theta_{K_0}^{(1)}$. Because $d = \lfloor N/(K+1) \rfloor \leq \Delta = \min_{0\leq m\leq M}|\theta_{m+1}^* - \theta_m^*|$, the AMOC assumption holds in $(\theta_{K_0+1}^{(0)}, \theta_{K_0+2}^{(0)}]$. We now discuss all the possible scenarios in the following.

First, if there is no changepoint between $\theta_{K_0}^{(1)}$ and $\theta_{K_0+1}^{(0)}$, then one of the following must hold:

(i) there is no changepoint between $\theta_{K_0+1}^{(0)}$ and $\theta_{K_0+2}^{(0)}$;
(ii) there exists exactly one changepoint $\theta_{m_0}^*$ such that $\theta_{K_0+2}^{(0)} - \theta_{m_0}^* \le h$;
(iii) there exists exactly one changepoint $\theta_{m_0}^*$ such that $\theta_{K_0+2}^{(0)} - \theta_{m_0}^* > h$.

For cases (i) and (ii), there is no other changepoint in $\mathcal{I}_{K_0+1}^{(0)}(h) = (\theta_{K_0}^{(1)} + h, \theta_{K_0+2}^{(0)} - h]$; since $\theta_{K_0+1}^{(1)}$ is chosen from $\mathcal{I}_{K_0+1}^{(0)}(h)$, there are no more changepoints in $(\theta_{K_0}^{(1)}, \theta_{K_0+1}^{(1)}]$. Therefore, the desired conclusion still holds for $\theta_{K_0+1}^{(1)}$. For case (iii), $\theta_{m_0}^*$ is the only changepoint in $\mathcal{I}_{K_0+1}^{(0)}(h)$ and Lemma 2 directly applies.

Second, assume that there exists some changepoint between $\theta_{K_0}^{(1)}$ and $\theta_{K_0+1}^{(0)}$. Because the assumption of AMOC holds in $(\theta_{K_0}^{(0)}, \theta_{K_0+1}^{(0)}]$, it also holds in $(\theta_{K_0}^{(1)}, \theta_{K_0+1}^{(0)}]$. By a similar argument above, for $\theta_{m_0-1}^* \in (\theta_{K_0}^{(1)}, \theta_{K_0+1}^{(0)}]$, either

(iv) $\theta_{m_0-1}^* - \theta_{K_0}^{(1)} \le h$, or
(v) $\theta_{K_0+1}^{(0)} - \theta_{m_0-1}^* \le h$.

In case (iv), $\theta_{m_0-1}^*$ is detected in the previous iteration, and in case (v) because $\theta_{m_0-1}^*$ is too close to the right boundary $\theta_{K_0+1}^{(0)}$; thus, it cannot be chosen in the previous iteration. These two cases are discussed according to whether there is a changepoint in $(\theta_{K_0+1}^{(0)}, \theta_{K_0+2}^{(0)}]$.

If there is no changepoint in $(\theta_{K_0+1}^{(0)}, \theta_{K_0+2}^{(0)}]$, then for case (iv), there is no more changepoint to be detected in $\mathcal{I}_{K_0+1}^{(0)}(h)$ and for case (v), $\theta_{m_0-1}^*$ is the only changepoint in $\mathcal{I}_{K_0+1}^{(0)}(h)$. Both cases are discussed as above. Conversely, if there is a changepoint $\theta_{m_0}^*$ in $(\theta_{K_0+1}^{(0)}, \theta_{K_0+2}^{(0)}]$, then for case (iv), $\theta_{m_0}^*$ is either being close to $\theta_{K_0+2}^{(0)}$, which results no more changepoint presented in $\mathcal{I}_{K_0+1}^{(0)}(h)$. Or $\theta_{m_0}^*$ is the only changepoint in $\mathcal{I}_{K_0+1}^{(0)}(h)$, which attains the desired conclusion by letting h be sufficiently small and applying Lemma 2. For case (v), because $d \le \Delta$, it must hold that $\theta_{K_0+2}^{(0)} - \theta_{m_0}^* \le h$, and thus $\theta_{m_0-1}^*$ is the only changepoint in $\mathcal{I}_{K_0+1}^{(0)}(h)$, which reduces to the case just discussed. Therefore, in all possible cases, $P(\min_{1\le k\le K} |\theta_m^* - \theta_k^{(r)}| \ge h) < \epsilon$ for each $\theta_m^* \le \theta_{K_0+1}^{(1)}$.

Finally, we check whether the assumption $P(\min_{1\le k\le K} |\theta_m^* - \theta_k^{(1)}| \ge h) < \epsilon$ for each $\theta_m^* \le \theta_{K_0}^{(1)}$ holds for $K_0 = 1$. Because $d \le \Delta$, the AMOC assumption holds in $(\theta_0^{(1)} = 0, \theta_2^{(0)}]$. If there exists θ_1^* in $\mathcal{I}_1^{(0)}(h)$, Lemma 2 applies and we have $P(\{|\theta_1^* - \theta_1^{(1)}| > \delta) < \epsilon$. Otherwise, no changepoints exists in $(0, \theta_1^{(1)}]$ and the same conclusion still holds. By mathematical induction, the claim is proved.

Acknowledgments. The authors are grateful to the Editor, Associate Editor and referees for their constructive comments that have greatly improved the paper.

SUPPLEMENTARY MATERIAL

Additional simulation results (DOI: 10.1214/19-AOAS1242SUPP; .pdf). We provide additional simulation results of Section 5.4 for DSBE, BS$_{\text{CUST}}$ and CBS$_{\text{EC}}$ with the sample size $N = 100$ in comparison with those with $N = 200$.

REFERENCES

ASTON, J. A. D. and KIRCH, C. (2012). Detecting and estimating changes in dependent functional data. *J. Multivariate Anal.* **109** 204–220. MR2922864

AUE, A., GABRYS, R., HORVÁTH, L. and KOKOSZKA, P. (2009). Estimation of a change-point in the mean function of functional data. *J. Multivariate Anal.* **100** 2254–2269. MR2560367

BENKO, M., HÄRDLE, W. and KNEIP, A. (2009). Common functional principal components. *Ann. Statist.* **37** 1–34. MR2488343

BERKES, I., GABRYS, R., HORVÁTH, L. and KOKOSZKA, P. (2009). Detecting changes in the mean of functional observations. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **71** 927–946. MR2750251

BOSQ, D. (2000). *Linear Processes in Function Spaces*: *Theory and Applications*. *Lecture Notes in Statistics* **149**. Springer, New York. MR1783138

BRAUN, J. V., BRAUN, R. K. and MÜLLER, H.-G. (2000). Multiple changepoint fitting via quasilikelihood, with application to DNA sequence segmentation. *Biometrika* **87** 301–314. MR1782480

CHIOU, J.-M., CHEN, Y.-T. and HSING, T. (2019). Supplement to "Identifying multiple changes for a functional data sequence with application to freeway traffic segmentation." DOI:10.1214/19-AOAS1242SUPP.

CHIOU, J.-M., CHEN, Y.-T. and YANG, Y.-F. (2014). Multivariate functional principal component analysis: A normalization approach. *Statist. Sinica* **24** 1571–1596. MR3308652

CIUPERCA, G. (2011). A general criterion to determine the number of change-points. *Statist. Probab. Lett.* **81** 1267–1275. MR2803773

FERRATY, F. and VIEU, P. (2006). *Nonparametric Functional Data Analysis*: *Theory and Practice*. *Springer Series in Statistics*. Springer, New York. MR2229687

FREMDT, S., STEINEBACH, J. G., HORVÁTH, L. and KOKOSZKA, P. (2013). Testing the equality of covariance operators in functional samples. *Scand. J. Stat.* **40** 138–152. MR3024036

FRICK, K., MUNK, A. and SIELING, H. (2014). Multiscale change point inference. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **76** 495–580. MR3210728

FRYZLEWICZ, P. (2014). Wild binary segmentation for multiple change-point detection. *Ann. Statist.* **42** 2243–2281. MR3269979

GROMENKO, O., KOKOSZKA, P. and REIMHERR, M. (2017). Detection of change in the spatiotemporal mean function. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **79** 29–50. MR3597963

HAPP, C. and GREVEN, S. (2018). Multivariate functional principal component analysis for data observed on different (dimensional) domains. *J. Amer. Statist. Assoc.* **113** 649–659. MR3832216

HARCHAOUI, Z. and LÉVY-LEDUC, C. (2010). Multiple change-point estimation with a total variation penalty. *J. Amer. Statist. Assoc.* **105** 1480–1493. MR2796565

HÖRMANN, S. and KOKOSZKA, P. (2010). Weakly dependent functional data. *Ann. Statist.* **38** 1845–1884. MR2662361

HORVÁTH, L. and KOKOSZKA, P. (2012). *Inference for Functional Data with Applications*. *Springer Series in Statistics*. Springer, New York. MR2920735

HSING, T. and EUBANK, R. (2015). *Theoretical Foundations of Functional Data Analysis, with an Introduction to Linear Operators*. *Wiley Series in Probability and Statistics*. Wiley, Chichester. MR3379106

HU, S. R. and WANG, C. M. (2008). Vehicle detector deployment strategies for the estimation of network origin-destination demands using link traffic counts. *IEEE Transactions on Intelligent Transportation Systems* **9** 283–300.

KILLICK, R., FEARNHEAD, P. and ECKLEY, I. A. (2012). Optimal detection of changepoints with a linear computational cost. *J. Amer. Statist. Assoc.* **107** 1590–1598. MR3036418

MAY, A. D. (1990). *Traffic Flow Fundamentals*. Prentice Hall, Englewood Cliffs, NJ.

NEWEY, W. K. (1991). Uniform convergence in probability and stochastic equicontinuity. *Econometrica* **59** 1161–1167. MR1113551

NIU, Y. S., HAO, N. and ZHANG, H. (2016). Multiple change-point detection: A selective overview. *Statist. Sci.* **31** 611–623. MR3598742

NIU, Y. S. and ZHANG, H. (2012). The screening and ranking algorithm to detect DNA copy number variations. *Ann. Appl. Stat.* **6** 1306–1326. MR3012531

OLSHEN, A. B., VENKATRAMAN, E., LUCITO, R. and WIGLER, M. (2004). Circular binary segmentation for the analysis of array-based DNA copy number data. *Biostatistics* **5** 557–572.

PAN, J. and CHEN, J. (2006). Application of modified information criterion to multiple change point problems. *J. Multivariate Anal.* **97** 2221–2241. MR2301636

PANARETOS, V. M., KRAUS, D. and MADDOCKS, J. H. (2010). Second-order comparison of Gaussian random functions and the geometry of DNA minicircles. *J. Amer. Statist. Assoc.* **105** 670–682. MR2724851

PÖTSCHER, B. M. and PRUCHA, I. R. (1994). Generic uniform convergence and equicontinuity concepts for random functions: An exploration of the basic structure. *J. Econometrics* **60** 23–63. MR1247816

RAMSAY, J. O. and SILVERMAN, B. W. (2005). *Functional Data Analysis*, 2nd ed. *Springer Series in Statistics*. Springer, New York. MR2168993

TRANSPORTATION RESEARCH BOARD (2010). *HCM 2010: Highway Capacity Manual*. TRB, Washington, DC.

VAN DER VAART, A. W. and WELLNER, J. A. (1996). *Weak Convergence and Empirical Processes: With Applications to Statistics*. *Springer Series in Statistics*. Springer, New York. MR1385671

WANG, J.-L., CHIOU, J.-M. and MÜLLER, H.-G. (2016). Functional data analysis. *Annual Review of Statistics and Its Application* **3** 257–295.

YAO, Y.-C. (1988). Estimating the number of change-points via Schwarz' criterion. *Statist. Probab. Lett.* **6** 181–189. MR0919373

ZHANG, J.-T. (2014). *Analysis of Variance for Functional Data*. *Monographs on Statistics and Applied Probability* **127**. CRC Press, Boca Raton, FL. MR3185072

J. M. CHIOU
INSTITUTE OF STATISTICAL SCIENCE
ACADEMIA SINICA
128 SEC. 2 ACADEMIA RD.
NANKANG DISTRICT
TAIPEI CITY 11529 TAIWAN R.O.C.
E-MAIL: jmchiou@stat.sinica.edu.tw

Y.-T. CHEN
DEPARTMENT OF STATISTICS
NATIONAL CHENG CHI UNIVERSITY
64 SEC.2 ZHINAN RD., WENSHAN DISTRICT
TAIPEI CITY 11605 TAIWAN
R.O.C.
E-MAIL: bboy0302@gmail.com

T. HSING
DEPARTMENT OF STATISTICS
UNIVERSITY OF MICHIGAN, ANN ARBOR
500 S. STATE STREET
ANN ARBOR, MICHIGAN 48109
USA
E-MAIL: thsing@umich.edu