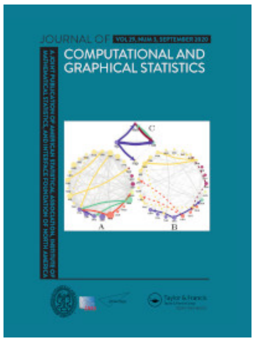

Taylor & Francis
Taylor & Francis Group

Journal of Computational and Graphical Statistics

ISSN: 1061-8600 (Print) 1537-2715 (Online) Journal homepage: https://www.tandfonline.com/loi/ucgs20

Estimating Truncated Functional Linear Models With a Nested Group Bridge Approach

Tianyu Guan, Zhenhua Lin & Jiguo Cao

To cite this article: Tianyu Guan, Zhenhua Lin & Jiguo Cao (2020) Estimating Truncated Functional Linear Models With a Nested Group Bridge Approach, Journal of Computational and Graphical Statistics, 29:3, 620-628, DOI: 10.1080/10618600.2020.1713797

To link to this article: https://doi.org/10.1080/10618600.2020.1713797

View supplementary material

Published online: 21 Feb 2020.

Submit your article to this journal

Article views: 374

View related articles

View Crossmark data

Citing articles: 8 View citing articles

Full Terms & Conditions of access and use can be found at
https://www.tandfonline.com/action/journalInformation?journalCode=ucgs20

JOURNAL OF COMPUTATIONAL AND GRAPHICAL STATISTICS
2020, VOL. 29, NO. 3, 620–628
https://doi.org/10.1080/10618600.2020.1713797

Taylor & Francis
Taylor & Francis Group

Check for updates

Estimating Truncated Functional Linear Models With a Nested Group Bridge Approach

Tianyu Guan[a], Zhenhua Lin[b], and Jiguo Cao[a]

[a]Department of Statistics and Actuarial Science, Simon Fraser University, Burnaby, BC, Canada; [b]Department of Statistics and Applied Probability, National University of Singapore, Singapore, Singapore

ABSTRACT
We study a scalar-on-function truncated linear regression model which assumes that the functional predictor does not influence the response when the time passes a certain cutoff point. We approach this problem from the perspective of locally sparse modeling, where a function is locally sparse if it is zero on a substantial portion of its defining domain. In the truncated linear model, the slope function is exactly a locally sparse function that is zero beyond the cutoff time. A locally sparse estimate then gives rise to an estimate of the cutoff time. We propose a nested group bridge penalty that is able to specifically shrink the tail of a function. Combined with the B-spline basis expansion and penalized least squares, the nested group bridge approach can identify the cutoff time and produce a smooth estimate of the slope function simultaneously. The proposed nested group bridge estimator is shown to be consistent, while its numerical performance is illustrated by simulation studies. The proposed nested group bridge method is demonstrated with an application of determining the effect of the past engine acceleration on the current particulate matter emission. Supplementary materials for this article are available online.

ARTICLE HISTORY
Received September 2018
Revised December 2019

KEYWORDS
B-spline basis functions; Functional data analysis; Functional linear regression; Group bridge approach; Locally sparse; Penalized B-splines

1. Introduction

In this article, we consider a scalar-on-function truncated linear regression model where the functional predictor $X_i(t), i = 1, \dots, n$, is defined on a time interval $[0, T]$ but influences the scalar response Y_i only on $[0, \delta]$ for some unknown cutoff time $\delta \le T$. Specifically, the model is written as

$$Y_i = \mu + \int_0^\delta X_i(t)\beta(t)\,\mathrm{d}t + \varepsilon_i, \qquad (1)$$

where, without loss of generality, $X_i(\cdot)$ is assumed to be centered, that is, $\mathbf{E}X_i(t) \equiv 0$, μ is then the mean of Y_i, $\beta(t)$ is the slope function (or coefficient function), and ε_i represents the noise that is independent of $X_i(\cdot)$.

An example of the scalar-on-function truncated linear regression is to determine the effects of the past engine acceleration on the current particulate matter emission. The response variable is the current particulate matter emission and the explanatory function is the smoothed engine acceleration curve for the past 60 sec. Figure 1(a) displays 108 smoothed engine acceleration curves against the backward time, in which 0 means the current time, while Figure 1(b) shows the slope function estimated by the penalized B-splines method (Cardot, Ferraty, and Sarda 2003). The penalized B-splines method is detailed in the supplementary document. We observe from Figure 1(b) that the acceleration over the past 20–60 sec does not have apparent contribution to predicting the current particulate matter emission. Intuitively, the particulate matter emissions shall depend on the recent acceleration, but not the ancient one. Therefore, if a linear relation between the particulate matter emission and the acceleration curve is assumed, one might naturally use the truncated linear model (1) to analyze such data, where the task includes identifying the cutoff time beyond which the engine acceleration has no influence on the current particulate matter emission.

The degenerate case $\delta = T$ in model (1) corresponds to the classic functional linear regression that has been studied in vast literature. Hastie and Mallows (1993) pioneered the smooth estimation of $\beta(t)$ via penalized least squares and/or smooth basis expansion. Cardot, Ferraty, and Sarda (2003) adopted B-spline basis expansion, while Li and Hsing (2007) used Fourier basis, both with a roughness penalty to control the smoothness of estimated slope functions. Data-driven bases such as eigenfunctions of the covariance function of the predictor process $X_i(t)$ were considered in Cardot, Ferraty, and Sarda (2003), Cai and Hall (2006), and Hall and Horowitz (2007). Yuan and Cai (2010) took a reproducing kernel Hilbert space approach to estimate the slope function. The case of sparsely observed functional data was studied by Yao, Müller, and Wang (2005). These estimation procedures for classic functional linear regression do not apply to the truncated linear model where $\delta \le T$ is often assumed. The functional linear models have many interesting applications. For instance, Ainsworth, Routledge, and Cao (2011) applied a functional linear model to study the impact of river flow to the salmon abundance. Luo et al. (2013) investigated the effect of the time-varying admission intensity on emergency department access block. Liu, Wang, and Cao (2017) proposed a functional linear mixed model when repeated measurements

CONTACT Jiguo Cao jiguo_cao@sfu.ca Department of Statistics and Actuarial Science, Simon Fraser University, 8888 University Drive, Burnaby, BC, Canada, V5A 1S6.
Color versions of one or more of the figures in the article can be found online at www.tandfonline.com/r/JCGS.
Supplementary materials for this article are available online. Please go to www.tandfonline.com/r/JCGS.
© 2020 American Statistical Association, Institute of Mathematical Statistics, and Interface Foundation of North America

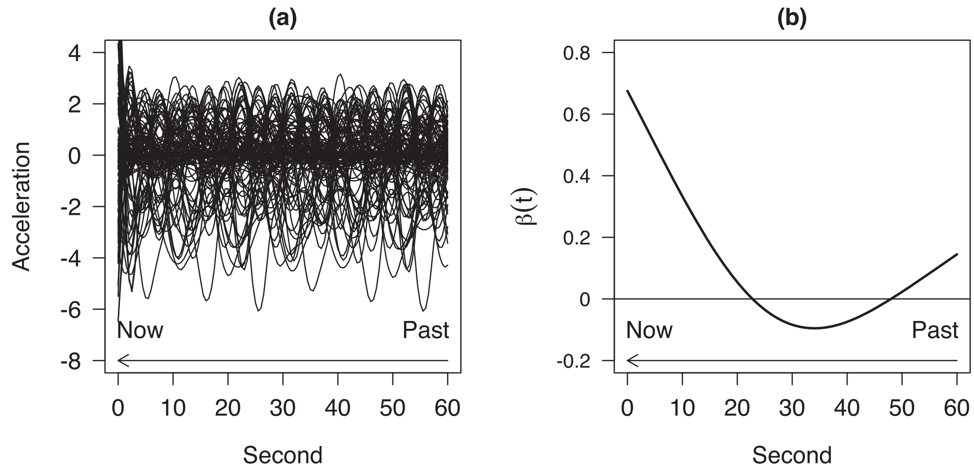

Figure 1. (a) 108 smoothed engine acceleration curves. (b) Estimated slope function using the penalized B-splines approach (Cardot, Ferraty, and Sarda 2003). The arrows indicate the direction of time.

are available on multiple subjects. For models beyond linear regression and a comprehensive introduction to functional data analysis, readers are referred to the monographs by Ramsay and Silverman (2005), Ferraty and Vieu (2006), Hsing and Eubank (2015), and Kokoszka and Reimherr (2017), as well as the review articles by Morris (2015) and Wang, Chiou, and Müller (2016) and references therein.

Model (1) has been investigated by Hall and Hooker (2016) who proposed to estimate $\beta(t)$ and δ by penalized least squares with a penalty on δ^2. The resulting estimates for $\beta(t)$ are discontinuous at $t = \hat{\delta}$ where $\hat{\delta}$ stands for the estimator of δ. This feature might not be desirable when $\beta(t)$ is a priori assumed to be continuous. For example, it is more reasonable to assume the acceleration function influences the particulate matter emission in a continuous and smooth manner. Alternatively, we observe that model (1) is equivalent to a classic functional linear model with $\beta(t) = 0$ for all $t \in [\delta, T]$. Such a slope function $\beta(t)$ is a special case of locally sparse functions which by definition are functions being zero in a substantial portion of their defining domains. Locally sparse slope functions have been studied in Lin et al. (2017), as well as pioneering works of James, Wang, and Zhu (2009) and Zhou, Wang, and Wang (2013). For example, in Lin et al. (2017), a general functional shrinkage regularization technique, called fSCAD, was proposed and demonstrated to be able to encourage the local sparseness. Although these endeavors are able to produce a smooth and locally sparse estimate, they do not specifically focus on the tail region $[\delta, T]$. Therefore, the estimated slope functions produced by such methods might not be zero in the region that is very close to the endpoint T, in particular when the boundary effect is not negligible.

In this article, we propose a new nested group bridge approach to estimate the slope function $\beta(t)$ and the cutoff time δ. Compared to the existing methods, the proposed nested group bridge approach has two features. First, it is based on the B-spline basis expansion and penalized least squares with a roughness penalty. Therefore, the resulting estimator of $\beta(t)$ is continuous and smooth over the entire domain $[0, T]$, contrasting the discontinuous estimator of Hall and Hooker (2016). Second, it employs a new nested group bridge shrinkage method proposed in Section 2 to specifically shrink the estimated function on the tail region $[\delta, T]$. Group bridge was proposed in Huang et al. (2009) for variable selection, and used by Wang and Kai (2015) for locally sparse estimation in the setting of nonparametric regression. In our approach, we creatively organize the coefficients of B-spline basis functions into a sequence of nested groups and apply the group bridge penalty to the groups. With the aid from B-spline basis expansion, such nested structure enables us to shrink the tail of the estimated slope function. This fixes the problem of the aforementioned generic locally sparse estimation procedures. An R package `ngr` has been developed for implementing the proposed method.

We structure the rest of the article as follows. In Section 2, we present the proposed nested group bridge estimation method for the slope function and the cutoff time, and also provide computational details. In Section 3, we investigate the asymptotic properties of the derived estimators. Simulation studies are discussed in Section 4, and an application to the particulate matter emissions data is given in Section 5. Conclusion and discussion are given in Section 6. In the supplementary document, we provide proofs and additional discussion.

2. Methodology

2.1. Nested Group Bridge Approach

Our estimation method uses B-spline basis functions that are detailed in de Boor (2001). Let $\boldsymbol{B}(t) = (B_1(t), \ldots, B_{M+d}(t))^{\mathrm{T}}$ be a vector that contains $M + d$ B-spline basis functions defined on $[0, T]$ with degree d and $M + 1$ equally spaced knots $0 = t_0 < t_1 < \cdots < t_M = T$. For $m \geq 0$, let $\boldsymbol{B}^{(m)}(t) = (B_1^{(m)}(t), \ldots, B_{M+d}^{(m)}(t))^{\mathrm{T}}$ denote the vector of the mth derivatives of the B-spline basis functions. Each of these basis functions is a piecewise polynomial of degree d. B-spline basis functions are well known for their compact support property, that is, each basis function is positive over at most $d+1$ adjacent subintervals. Due to this compact support property, if we approximate $\beta(t)$ by a linear combination of B-spline basis functions, then such approximation is locally sparse if the coefficients are sparse in groups.

We shall further introduce some notations. Let $I_j = (t_{j-1}, t_M)$, and $A_j = \{j, j+1, \ldots, M+d\}$ for $j = 1, \ldots, M$.

Intuitively, each group A_j represents the indices of B-spline basis functions that are nonzero on I_j. For a vector $\boldsymbol{b} = (b_1, \dots, b_{M+d})^{\mathrm{T}}$ of scalars, we denote by $b_{A_j} = \{b_k : k \in A_j\}$ the subvector of elements whose indices are in the jth group A_j. We shall use $\|\boldsymbol{a}\|_1 = |a_1| + \cdots + |a_q|$ to denote the L_1 norm of a generic q-dimensional vector $\boldsymbol{a}$, and use $\|x\|_2$ to denote the L_2 norm of a generic function $x(t)$. As our focus is on the estimation of $\beta(t)$ and δ, without loss of generality, we assume that $\mu = 0$ in model (1) in the sequel.

For a fixed $0 < \gamma < 1$, the historically sparse (zero on the tail region) and smooth estimators for β and δ are defined as

$$\hat{\beta}_n(t) = \hat{\boldsymbol{b}}_n^{\mathrm{T}} \boldsymbol{B}(t), \quad \hat{\delta}_n = t_{J_0 - 1}, \tag{2}$$

where $J_0 = \min\{M+1, \min\{l : \hat{b}_{nk} = 0, \text{for all } k \geq l\}\}$ and $\hat{\boldsymbol{b}}_n = (\hat{b}_{n1}, \dots, \hat{b}_{nM+d})^{\mathrm{T}}$ minimizes the penalized least squares

$$\frac{1}{n}\sum_{i=1}^{n}\left(Y_i - \sum_{k=1}^{M+d} b_k \int_0^T X_i(t)B_k(t)\,\mathrm{d}\,t\right)^2 + \kappa \left\|\boldsymbol{b}^{\mathrm{T}}\boldsymbol{B}^{(m)}\right\|_2^2 + \lambda \sum_{j=1}^{M} c_j \left\|b_{A_j}\right\|_1^{\gamma}, \tag{3}$$

with known weights c_j and nonnegative tuning parameters κ and λ. In the above criterion, the first term is the ordinary least squares error that encourages the fidelity of model fitting, while the second term is a roughness penalty that aims to enforce smoothness of the estimate $\hat{\beta}_n(t)$. In practice, $m = 2$ is a common choice, which corresponds to measuring the roughness of a function by its integrated curvature.

The last term in the objective function (3) is designed to shrink the estimated slope function toward zero specifically on the tail region. It originates from the group bridge penalty that was introduced by Huang et al. (2009) for simultaneous selection of variables at both the group and within-group individual levels. In (3), the groups have a special structure: $A_1 \supset \cdots \supset A_M$. In other words, the groups are nested as a sequence and hence we call the last term in (3) *nested group bridge*. Due to such nested nature, if $k > j$, then one can observe in (3) that (i) the coefficient b_k appears in all groups where the coefficient b_j also appears, and (ii) b_k appears in more groups than b_j. As a consequence, b_k is always penalized more heavily than b_j. These two features suggest that the nested group bridge penalty spends more effort on shrinking those coefficients of B-spline basis functions whose support is in a closer proximity to T. As B-spline basis functions enjoy the aforementioned compact support property and our estimate is represented by a linear combination of such basis functions as in (2), the progressive shrinkage of nested group bridge encourages the estimate of $\beta(t)$ to be locally sparse specifically on the tail part of the time domain. Such estimate is exactly what we are after in the scalar-on-function truncated linear model (1). The weights c_j are introduced to adjust the number of elements in the set A_j. A simple choice for c_j is $c_j \propto |A_j|^{1-\gamma}$, where $|A_j|$ denotes the cardinality of A_j (Huang et al. 2009). Borrowing the idea of the adaptive lasso (Zou 2006), we practically choose $c_j = |A_j|^{1-\gamma}/\|b_{A_j}^{(0)}\|_2^{\gamma}$, where $\boldsymbol{b}^{(0)}$ can be obtained by the penalized B-splines method (Cardot, Ferraty, and Sarda 2003). As Huang et al. (2009) pointed out, when $\gamma = 1$, the group bridge penalty is the lasso penalty and can only do individual variable selection. When $0 < \gamma < 1$, the group bridge penalty can be used for variable selection at the group and with-in group individual levels simultaneously. We also conduct a simulation study to compare the lasso and the nested group bridge penalty; see the supplementary document for details.

2.2. Computational Method

The objective function (3) is not convex and thus difficult to optimize. Huang et al. (2009) suggested the following formulation that was easier to work with. Based on Proposition 1 of Huang et al. (2009), for $0 < \gamma < 1$, if $\lambda = \tau^{1-\gamma}\gamma^{-\gamma}(1-\gamma)^{\gamma-1}$, then $\hat{\boldsymbol{b}}_n$ minimizes (3) if and only if $(\hat{\boldsymbol{b}}_n, \hat{\boldsymbol{\theta}})$ minimizes

$$\frac{1}{n}\sum_{i=1}^{n}\left(Y_i - \sum_{k=1}^{M+d} b_k \int_0^T X_i(t)B_k(t)\,\mathrm{d}\,t\right)^2 + \kappa \left\|\boldsymbol{b}^{\mathrm{T}}\boldsymbol{B}^{(m)}\right\|_2^2 + \sum_{j=1}^{M} \theta_j^{1-1/\gamma} c_j^{1/\gamma} \|b_{A_j}\|_1 + \tau \sum_{j=1}^{M}\theta_j, \tag{4}$$

subject to $\theta_j \geq 0$ $(j = 1, \dots, M)$, where $\boldsymbol{\theta} = (\theta_1, \dots, \theta_M)^{\mathrm{T}}$ and $\hat{\boldsymbol{\theta}} = (\hat{\theta}_1, \dots, \hat{\theta}_M)^{\mathrm{T}}$. Below we develop an algorithm following this idea.

Let $\boldsymbol{U}$ denote the $n \times (M+d)$ matrix with elements $u_{ij} = \int_0^T X_i(t)B_j(t)\,\mathrm{d}\,t$, and let $\boldsymbol{V}$ denote the $(M+d) \times (M+d)$ matrix with elements $v_{ij} = \int_0^T B_i^{(m)}(t)B_j^{(m)}(t)\,\mathrm{d}\,t$. Let $\boldsymbol{Y} = (Y_1, \dots, Y_n)^{\mathrm{T}}$, then the first term of (4) can be expressed as $1/n\,(\boldsymbol{Y} - \boldsymbol{U}\boldsymbol{b})^{\mathrm{T}}(\boldsymbol{Y} - \boldsymbol{U}\boldsymbol{b})$ and the second term of (4) yields $\kappa \boldsymbol{b}^{\mathrm{T}}\boldsymbol{V}\boldsymbol{b}$. Since $\boldsymbol{V}$ is a positive semidefinite matrix, we write $\boldsymbol{V} = \boldsymbol{W}\boldsymbol{W}$, where $\boldsymbol{W}$ is symmetric. Define

$$\boldsymbol{U}_* = \begin{pmatrix} \boldsymbol{U} \\ \sqrt{n\kappa}\boldsymbol{W} \end{pmatrix} \quad \text{and} \quad \tilde{\boldsymbol{Y}} = \begin{pmatrix} \boldsymbol{Y} \\ \boldsymbol{0} \end{pmatrix},$$

where $\boldsymbol{0}$ is the zero vector of length $M+d$. If we write $g_k = \sum_{j=1}^{\min\{k,M\}} \theta_j^{1-1/\gamma} c_j^{1/\gamma}$ for $k = 1, \dots, M+d$, then (4) can be written in the form

$$\frac{1}{n}\left(\tilde{\boldsymbol{Y}} - \boldsymbol{U}_*\boldsymbol{b}\right)^{\mathrm{T}}\left(\tilde{\boldsymbol{Y}} - \boldsymbol{U}_*\boldsymbol{b}\right) + \sum_{k=1}^{M+d} g_k|b_k| + \tau\sum_{j=1}^{M}\theta_j. \tag{5}$$

Let $\boldsymbol{G}$ be the $(M+d) \times (M+d)$ diagonal matrix with the ith diagonal element $(ng_i)^{-1}$. With notation $\tilde{\boldsymbol{U}} = \boldsymbol{U}_*\boldsymbol{G}$ and $\tilde{\boldsymbol{b}} = \boldsymbol{G}^{-1}\boldsymbol{b}$, (5) can be expressed in a form of lasso problem (Tibshirani 1996),

$$\frac{1}{n}\left\{\left(\tilde{\boldsymbol{Y}} - \tilde{\boldsymbol{U}}\tilde{\boldsymbol{b}}\right)^{\mathrm{T}}\left(\tilde{\boldsymbol{Y}} - \tilde{\boldsymbol{U}}\tilde{\boldsymbol{b}}\right) + \sum_{k=1}^{M+d}|\tilde{b}_k|\right\} + \tau\sum_{j=1}^{M}\theta_j,$$

where $\tilde{b}_k$ denote the kth element of vector $\tilde{\boldsymbol{b}}$. Now, we take the following iterative approach to compute $\hat{\boldsymbol{b}}_n$.

Step 1. Obtain an initial estimate $\boldsymbol{b}^{(0)}$.

Step 2. At iteration s, $s = 1, 2, \ldots$, compute

$$\theta_j^{(s)} = c_j \left(\frac{1-\gamma}{\tau\gamma} \right)^{\gamma} \|b_{A_j}^{(s-1)}\|_1^{\gamma}, \quad j = 1, \ldots, M,$$

$$g_k^{(s)} = \sum_{j=1}^{\min\{k,M\}} (\theta_j^{(s)})^{1-1/\gamma} c_j^{1/\gamma}, \quad k = 1, \ldots, M+d,$$

$$\boldsymbol{G}^{(s)} = n^{-1}\mathrm{diag}\left(1/g_1^{(s)}, \ldots, 1/g_{M+d}^{(s)}\right), \quad \tilde{\boldsymbol{U}}^{(s)} = \boldsymbol{U}_* \boldsymbol{G}^{(s)}.$$

Step 3. At iteration s, compute

$$\boldsymbol{b}^{(s)} = \boldsymbol{G}^{(s)} \arg\min_{\tilde{\boldsymbol{b}}} \left(\tilde{\boldsymbol{Y}} - \tilde{\boldsymbol{U}}^{(s)}\tilde{\boldsymbol{b}}\right)^{\mathrm{T}} \left(\tilde{\boldsymbol{Y}} - \tilde{\boldsymbol{U}}^{(s)}\tilde{\boldsymbol{b}}\right) + \sum_{k=1}^{M+d} |\tilde{b}_k|. \tag{6}$$

Step 4. Repeat Step 2 and Step 3 until convergence is reached.

A choice for the initial estimate is $\boldsymbol{b}^{(0)} = (\boldsymbol{U}^{\mathrm{T}}\boldsymbol{U} + n\kappa\boldsymbol{V})^{-1}\boldsymbol{U}^{\mathrm{T}}\boldsymbol{Y}$, which is obtained by the penalized B-splines method (Cardot, Ferraty, and Sarda 2003). Once $\hat{\boldsymbol{b}}_n$ is produced, the estimates for β and δ are given in (2). As the nested group bridge penalty is not convex, the above algorithm converges to a local minimizer. It is worth emphasizing that (6) is a lasso problem, which can be efficiently solved by the least angle regression algorithm (Efron et al. 2004).

In our fitting procedure, there are a few tuning parameters including the smoothing parameter κ, the shrinkage parameter λ, and the parameters for constructing the B-spline basis functions such as the degree d of the B-spline basis and the number of knots $M + 1$. Following the schemes of Marx and Eilers (1999), Cardot, Ferraty, and Sarda (2003), and Lin et al. (2017), we choose M to be relatively large to capture the local features of $\beta(t)$. In addition, δ is estimated by the knot t_{J_0-1}, therefore, a small M may lead to a large bias of the estimator $\hat{\delta}_n$. The effect of potential overfitting caused by a large number of knots can be offset by the roughness penalty. Compared to M, the degree d is of less importance, and therefore, we fix it to a reasonable value, that is, $d = 3$.

Once the number of B-spline basis functions is fixed, we can proceed to select the shrinkage parameter λ, as well as the smoothing parameter κ. In Hall and Hooker (2016) where the idea of penalized least squares is also employed, the shrinkage parameter is selected to minimize the mean-squared error of a parametric surrogate estimator of $\beta(t)$. In our case, for a given finite sample, the estimator in (2), which is represented by a finite number of B-spline basis functions, serves as such a surrogate. Therefore, we can adopt the same strategy to select λ. Instead of the mean-squared error, we employ the Bayesian information criterion (BIC) to encourage model sparsity, as follows.

Let $\hat{\boldsymbol{b}}_n = \hat{\boldsymbol{b}}_n(\kappa, \lambda)$ be the estimate based on a chosen pair of κ and λ. Let $\boldsymbol{U}_{\kappa,\lambda}$ denote the submatrix of $\boldsymbol{U}$ with columns corresponding to the nonzero $\hat{\boldsymbol{b}}_n(\kappa, \lambda)$, and $\boldsymbol{V}_{\kappa,\lambda}$ denote the submatrix of $\boldsymbol{V}$ with rows and columns corresponding to the nonzero $\hat{\boldsymbol{b}}_n(\kappa, \lambda)$. The approximated degree of freedom for κ and λ is

$$\mathrm{df}(\kappa, \lambda) = \mathrm{trace}\left(\boldsymbol{U}_{\kappa,\lambda}(\boldsymbol{U}_{\kappa,\lambda}^{\mathrm{T}}\boldsymbol{U}_{\kappa,\lambda} + n\kappa\boldsymbol{V}_{\kappa,\lambda})^{-1}\boldsymbol{U}_{\kappa,\lambda}^{\mathrm{T}}\right).$$

Then, BIC can be approximated by

$$\mathrm{BIC}(\kappa, \lambda) = n\log\left(\|\boldsymbol{Y} - \boldsymbol{U}\hat{\boldsymbol{b}}_n(\kappa, \lambda)\|_2^2/n\right) + \log(n)\mathrm{df}(\kappa, \lambda).$$

The optimal κ and λ are selected to minimize $\mathrm{BIC}(\kappa, \lambda)$.

3. Asymptotic Properties

Let δ_0 and $\beta_0(t)$ be the true values of the cutoff time δ and the slope function $\beta(t)$, respectively. We assume that realizations $X_1, \ldots, X_n$ are fully observed, while notice that the analysis can be extended to sufficiently densely observed data. Without loss of generality, we assume $T = 1$. If $\delta_0 = 0$, set $J_1 = 0$, and if $\delta_0 = 1$, let $J_1 = M$. Otherwise, let J_1 be an integer such that $\delta_0 \in [t_{J_1-1}, t_{J_1})$. According to Theorem XII(6) of de Boor (2001), there exists some $\beta_s(t) = \sum_{j=1}^{M+d} b_{sj}B_j(t) = \boldsymbol{B}^{\mathrm{T}}\boldsymbol{b}_s$ with $\boldsymbol{b}_s = (b_{s1}, \ldots, b_{sM+d})^{\mathrm{T}}$ with $\inf_j |b_{sj}| \geq C_0' M^{-p_0}$, such that $\|\beta_s - \beta_0\|_\infty \leq C_0 M^{-p_0}$ for some positive constants C_0', C_0 and p_0. More specifically, if $\beta_0(t)$ satisfies condition C.2, then $p_0 = k + \nu$. Define $b_{0j} = b_{sj}I_{(j \leq J_1)}, j = 1, \ldots, M + d$. Define Γ as the covariance operator of the random process X, and Γ_n as the empirical version of Γ, which is defined by

$$(\Gamma_n x)(v) = \frac{1}{n}\sum_{i=1}^{n}\int_0^1 X_i(v)X_i(u)x(u)\,\mathrm{d}u.$$

For tow functions g and f defined on $[0, 1]$, we define the inner product in the Hilbert space L^2 as $\langle g, f\rangle = \int_0^1 g(t)f(t)\,\mathrm{d}t$. Let $\boldsymbol{H}$ be the $(M + d) \times (M + d)$ matrix with elements $h_{i,j} = \langle \Gamma_n B_i, B_j\rangle$. To establish our asymptotic properties, we assume that the following conditions are satisfied.

C.1 $E\|X\|_2^2 < \infty$.

C.2 The kth derivative $\beta^{(k)}(t)$ exists and satisfies the Hölder condition with exponent ν, that is $|\beta^{(k)}(t') - \beta^{(k)}(t)| \leq c|t' - t|^{\nu}$, for some constant $c > 0$, $\nu \in (0, 1]$. Define $p = k + \nu$. Assume $3/2 < p \leq d$.

C.3 $M = o(n^{1/2})$, $M = \omega(n^{\frac{1}{2p}})$ and $\kappa = o(n^{-1/2}M^{1/2-2m})$.

C.4 There are constants $C_{\max} > C_{\min} > 0$ such that

$$C_{\min}M^{-1} \leq \rho_{\min}(\boldsymbol{H}) \leq \rho_{\max}(\boldsymbol{H}) \leq C_{\max}M^{-1}$$

with probability tending to one as n goes to infinity, where $\rho_{\min}$ and $\rho_{\max}$ denote the smallest and largest eigenvalues of a matrix, respectively.

C.5 $\lambda = O(n^{-1/2}M^{-1/2}\eta^{-1})$, where $\eta = \left(\sum_{j=1}^{J_1} c_j^2\|b_{0A_j}\|_1^{2\gamma-2}|A_j|\right)^{1/2}$ with $c_j \propto |A_j|^{1-\gamma}$.

C.6 $\dfrac{\lambda}{M^{1-\gamma}n^{\gamma/2-1}} \to \infty$.

The condition C.1 assures the existence of the covariance function of X. The second condition concerns the smoothness of the slope function β, which has been used by Cardot, Ferraty, and Sarda (2003) and Lin et al. (2017). In condition C.3, we set the growth rate for the smoothing tuning parameter κ. Our analysis applies to $m = 0$, which is equivalent to Tikhonov regularization in Hall and Horowitz (2007) and simplifies our analysis. A similar result can be derived for $m > 0$. The last two conditions together pose certain constraints on the decay rate of λ. Similar conditions appear in Wang and Kai (2015). Here, η is a sequence of constants varying with M and determined by β_0 and γ. It can be shown that, when $\beta_0(t) \neq 0$ for some t, $C_1 M^{1/2} \leq \eta \leq C_2 M^{(2-\gamma)+(1-\gamma)p}$ for constants $C_1, C_2 > 0$, and otherwise $\eta \equiv 0$. These conditions can be realized, for example, by $\lambda \asymp n^{-1/2} M^{\gamma-(1-\gamma)p-5/2}$ and $M \asymp n^{(1-\gamma)/(8-4\gamma+2p-2p\gamma)}$.

Below we state the main results, and relegate their proofs to the supplementary document. Our first result provides the convergence rate of the estimator $\hat{\beta}_n$ defined in (2).

Theorem 1 (Convergence rate). Suppose that Conditions C.1–C.6 hold. Then, $\|\hat{\beta}_n - \beta_0\|_2 = O_p(Mn^{-1/2} + M^{-p})$.

The convergence rate consists of two competing components, the variance term $Mn^{-1/2}$ and the bias term M^{-p}. With an increase of M, the approximation to $\beta(t)$ by B-spline basis functions is improved, however, at the cost of increased variance.

In addition, we observe that the smoothing parameter κ has negligible impact on the rate of the proposed estimator when its asymptotic rate is bounded by the threshold stated in the condition C.3. This is aligned with the classic results for penalized spline estimator (e.g., Claeskens, Krivobokova, and Opsomer 2009, Theorem 1). Moreover, as the nested group bridge penalty has the effect of shrinkage, it also penalizes the roughness of the estimator. This partially explains why the κ shall be chosen smaller than the one in Claeskens, Krivobokova, and Opsomer (2009). On the other hand, in practice, as the sample size is often limited, κ plays an important role in regulating the roughness/variability of the estimator, in particular when a large number of B-spline basis functions are required to reduce estimation bias. The next result shows that the null tail of $\beta(t)$, as well as the cutoff time δ, can be consistently estimated.

Theorem 2 (Consistency). Suppose that Conditions C.1–C.6 hold.

(i) For any $\zeta \in (0, 1 - \delta_0)$, $\hat{\beta}_n(t) = 0$ for all $t \in [\delta_0 + \zeta, 1]$ with probability tending to 1.

(ii) $\hat{\delta}_n$ converges to δ_0 in probability.

4. Simulation Studies

We conduct simulation studies to evaluate the numerical performance of the proposed nested group bridge method, and compare the results with the penalized B-splines approach (Cardot, Ferraty, and Sarda 2003), the two truncation methods (Hall and Hooker 2016), and two locally sparse modeling methods, the FLiRTI method (James, Wang, and Zhu 2009) and the SLoS method (Lin et al. 2017). The truncation methods first expand the slope function with a sequence of principal component functions and then penalize δ by adding a penalty on δ^2 to the least squares. Two estimation procedures were suggested by Hall and Hooker (2016). The first one (called Method A) estimates δ and $\beta(t)$ simultaneously, while the second one (called Method B) estimates them in an iterative fashion. The FLiRTI method proposed by James, Wang, and Zhu (2009) achieves local sparseness by applying variable selection to various derivatives at some discrete grid points. The SLoS method is based on fSCAD, a functional regularization technique.

In our studies, for the purpose of fair comparison, we consider the same scenarios for $\beta(t)$ in Hall and Hooker (2016), namely,

Scenario I. $\beta(t) = I_{(0 \leq t < 0.5)}$,

Scenario II. $\beta(t) = \sin(2\pi t) I_{(0 \leq t < 0.5)}$,

Scenario III. $\beta(t) = (\cos(2\pi t) + 1)\, I_{(0 \leq t < 0.5)}$,

where $I_{(\cdot)}$ denotes the indicator function. For all cases the slope function $\beta(t) > 0$ on $(0, 0.5)$ and $\beta(t) = 0$ on $[0.5, 1]$. As illustrated in Figure 2, the slope function is discontinuous for scenario I, and the first and second derivatives of the slope functions are discontinuous for scenarios II and III, respectively. The predictor functions $X_i(t)$ are generated by $X_i(t) = \sum a_{ij} B_j(t)$, where $B_j(t)$ are cubic B-spline basis functions defined on 64 (the number 64 is randomly selected between 50 and 100) equally spaced knots over $[0, 1]$, and the coefficients a_{ij} are generated independently from the standard normal distribution. The errors ε are normally distributed and sampled so that the signal-to-noise ratio equals to 2. We consider sample sizes $n = 100$ and $n = 500$. For each of the three scenarios and for each sample size, we replicate the simulation independently for 200 times. We also consider smooth functional covariates, which are generated in the same set up, except that the signal-to-noise ratio is 5 and $X_i(t)$ are generated as a linear combinations of 25 Fourier basis functions $1, \sin(2\pi t), \cos(2\pi t), \ldots, \sin(2^{12}\pi t), \cos(2^{12}\pi t)$ defined on $[0, 1]$, with the coefficients corresponding to the jth Fourier basis function generated independently from the normal distribution with mean 0 and variance $1/j^{1.2}$, $j = 1, \ldots, 25$. The results regarding the smooth functional covariates are provided in the supplementary document.

For the proposed nested group bridge method, the penalized B-splines approach and the SLoS method, we expand the slope function with cubic B-splines with 101 equally spaced knots. For the FLiRTI method, we use cubic B-splines with the number of knots selected according to the model selection method introduced in James, Wang, and Zhu (2009). For the proposed nested group bridge method, we follow Huang et al. (2009) and set the group bridge parameter $\gamma = 0.5$ in all numerical studies. We discuss the effect of γ in the supplementary document. The tuning parameters of the proposed nested group bridge method are chosen by the procedure reported in Section 2.2. The smoothing parameter of the penalized B-splines approach is chosen by BIC. For the two truncation methods, the number of empirical principal components is chosen from 2 to 15 by BIC. The FLiRTI method is implemented by the Dantzig selector (Candes and Tao 2007). The two truncation methods and the FLiRTI and SLoS estimators are implemented and tuned according to Hall and Hooker (2016), James, Wang, and Zhu (2009), and Lin et al. (2017), respectively.

Table 1 summarizes the Monte Carlo mean and standard deviation of $\hat{\delta}$. The results suggest that the proposed nested

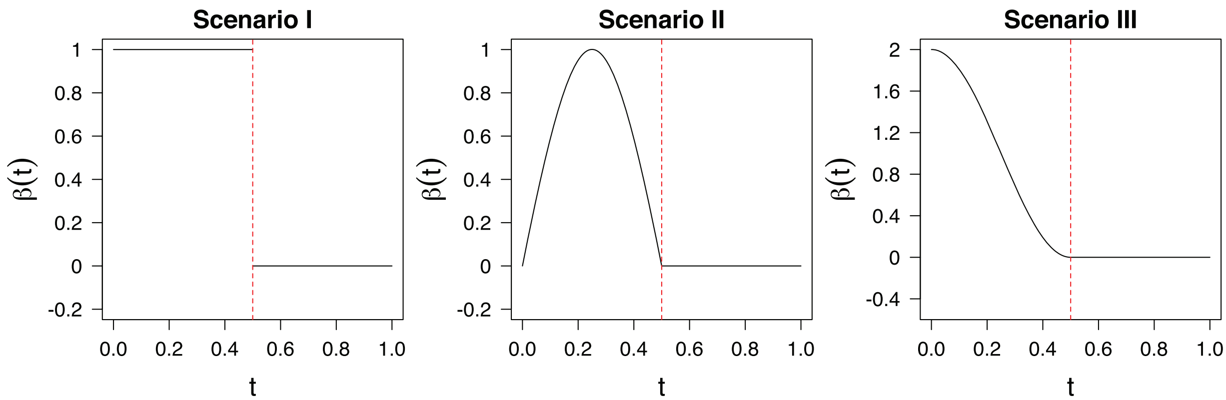

Figure 2. The slope functions in three scenarios. The dashed vertical lines indicate the true values of δ.

Table 1. The mean of estimators for δ based on 200 simulation replications with the corresponding Monte Carlo standard deviation included in parentheses.

	NGR	TR (Method A)	TR (Method B)	FLiRTI	SLoS	True value
Scenario I						
$n = 100$	0.66 (0.06)	0.48 (0.04)	0.35 (0.07)	0.81 (0.18)	0.69 (0.18)	0.50
$n = 500$	0.65 (0.05)	0.50 (0.02)	0.48 (0.05)	0.83 (0.17)	0.60 (0.09)	0.50
Scenario II						
$n = 100$	0.60 (0.07)	0.41 (0.04)	0.38 (0.06)	0.77 (0.21)	0.61 (0.18)	0.50
$n = 500$	0.59 (0.03)	0.45 (0.02)	0.45 (0.03)	0.71 (0.19)	0.55 (0.08)	0.50
Scenario III						
$n = 100$	0.50 (0.09)	0.31 (0.04)	0.30 (0.03)	0.73 (0.25)	0.55 (0.21)	0.50
$n = 500$	0.51 (0.04)	0.34 (0.03)	0.33 (0.04)	0.72 (0.23)	0.49 (0.08)	0.50

NGR, the proposed nested group bridge method; TR (Method A), the truncation method that estimates δ and $\beta(t)$ simultaneously proposed by Hall and Hooker (2016); TR (Method B), the truncation method that estimates δ and $\beta(t)$ iteratively (Hall and Hooker 2016); FLiRTI, the FLiRTI method proposed by James, Wang, and Zhu (2009); SLoS, the SLoS method proposed by Lin et al. (2017).

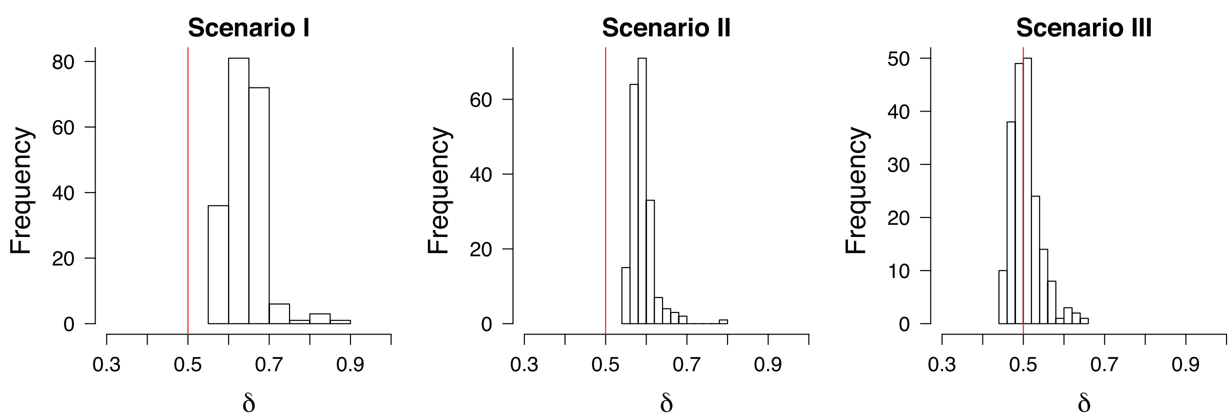

Figure 3. Histograms of the estimated $\hat{\delta}$ in 200 simulation replications in the three scenarios. The results were obtained based on 200 Monte Carlo simulations with $n = 500$. The vertical lines indicate the true values of δ.

group bridge estimator is more accurate than the other methods in scenario III when the second derivative of the slope function is discontinuous. In scenario II when the first derivative of the slope function is discontinuous, the proposed nested group bridge method is comparable to the truncation methods. In scenario I when the slope function is discontinuous, truncation method A is the most accurate. The FLiRTI and SLoS method do not focus on the tail region and therefore exhibit larger variability. The results for the smooth functional covariates reported in the supplementary document are similar. The histograms shown in Figure 3 provide more details of the performance of our method. They indicate that when $\beta(t)$ is not smooth, the proposed nested group bridge estimator is conservative, in the sense that the estimate $\hat{\delta} > \delta_0$.

To examine the quality of the estimation for $\beta(t)$, we report the mean integrated squared errors of the estimated $\hat{\beta}(t)$ in Table 2. It is observed that in general, the proposed nested group bridge estimator outperforms the other methods. The truncation methods do not regularize the roughness of the estimated slope function, which leads to a less favorable performance when the predictor function is relatively rough. The penalized B-splines method, the FLiRTI method and the SLoS method are comparable to the proposed nested group bridge method in terms of the estimation accuracy of $\beta(t)$, but the penalized B-splines method is unable to provide an estimate for δ. The results for the smooth functional covariates are reported in the supplementary document, which shows that the FLiRTI method does not perform as well as the other methods. To display

Table 2. Mean integrated squared errors of estimators for $\beta(t)$ based on 200 simulation replications with the corresponding Monte Carlo standard deviation included in parentheses.

	NGR	PS	TR (Method A)	TR (Method B)	FLiRTI	SLoS
Scenario I ($\times 10^{-2}$)						
$n = 100$	2.54 (0.93)	4.57 (1.70)	14.08 (5.13)	28.48 (8.54)	4.97 (2.22)	3.04 (1.21)
$n = 500$	1.42 (0.38)	1.89 (0.50)	3.34 (1.11)	9.65 (5.17)	1.88 (0.53)	1.38 (0.35)
Scenario II ($\times 10^{-2}$)						
$n = 100$	0.64 (0.44)	1.44 (0.70)	5.69 (2.12)	10.33 (4.43)	1.40 (0.93)	0.95 (0.69)
$n = 500$	0.21 (0.11)	0.24 (0.15)	1.17 (0.41)	3.08 (1.33)	0.30 (0.16)	0.14 (0.10)
Scenario III ($\times 10^{-2}$)						
$n = 100$	1.36 (1.05)	2.46 (1.50)	14.55 (6.99)	29.68 (13.91)	4.48 (3.16)	1.97 (1.67)
$n = 500$	0.34 (0.25)	0.64 (0.44)	4.25 (1.44)	11.68 (5.07)	0.87 (0.52)	0.46 (0.33)

NGR, the proposed nested group bridge method; PS, the penalized B-splines method; TR (Method A), the truncation method that estimates δ and $\beta(t)$ simultaneously proposed by Hall and Hooker (2016); TR (Method B), the truncation method that estimates δ and $\beta(t)$ iteratively (Hall and Hooker 2016); FLiRTI, the FLiRTI method proposed by James, Wang, and Zhu (2009); SLoS, the SLoS method proposed by Lin et al. (2017).

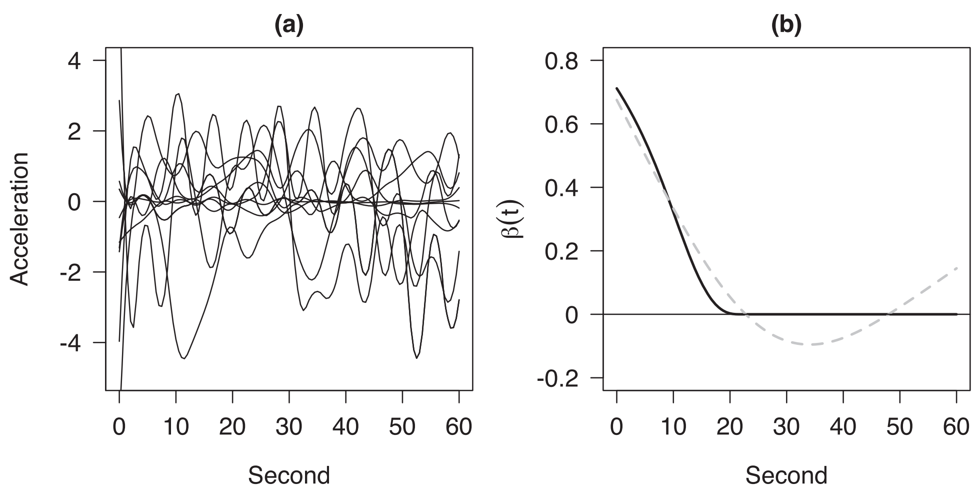

Figure 4. (a) Ten randomly selected smoothed acceleration curves. (b) Estimated $\hat{\beta}(t)$ using the penalized B-splines method (dashed line) and the proposed nested group bridge approach (solid line).

the results more intuitively, we provide in the supplementary document the figures that compare the estimated coefficient functions for various methods.

5. Application: Particulate Matter Emissions Data

In this section, we demonstrate the proposed nested group bridge approach to analyze the particulate matter emissions data which are taken from the Coordinating Research Councils E55/E59 research project (Clark et al. 2007). In this project, trucks were placed on the chassis dynamometer bed to mimic inertia and particulate matter was measured by an emission analyzer on standard test cycles. The engine acceleration of diesel trucks was also recorded. We are interested in determining the effects of the past engine acceleration on the current particulate matter emission, and in particular, identifying the cutoff time in the past that has a predicting power on the current particulate matter emission. The problem was originally addressed by Asencio, Hooker, and Gao (2014) in their case study. As noted in Hall and Hooker (2016), we obtain observation every 10 sec after the first 120 sec to remove dependences in the data. Let Y_i be the logarithm of the particulate matter emission measured at the ith 10 sec after the first 120 sec, and $X_i(t), t \in [0, 60]$, be the corresponding engine acceleration at the past time t. Both Y_i and $X_i(t)$ are centered such that $\mathbf{E}Y_i \equiv 0$ and $\mathbf{E}X_i(t) \equiv 0$. We estimate the functional linear model (1), where $\mu = 0$, the engine acceleration in the past 60 sec $X_i(t)$ is the predictor curve, and $T = 60$. In total, we have 108 such samples. Figure 4(a) displays 10 randomly selected smoothed engine acceleration curves recorded on every second for 60 sec.

Figure 4(b) provides estimates for $\beta(t)$ obtained by the proposed nested group bridge approach with the group bridge parameter $\gamma = 0.5$ and the penalized B-splines method, respectively, both of which use cubic B-spline basis functions with 121 knots. We choose the number of knots to be equal to the number of time points of the observed acceleration, which is 121. With a sample size 108 and number of knots 121, the roughness penalty plays an important role of reducing the variability of the estimates. The proposed nested group bridge estimate $\hat{\beta}(t)$ is zero over [20, 60] and the estimate for δ is 20 sec. It suggests that the engine acceleration influences particulate matter emission for no longer than 20 sec. A similar trend can be observed for the penalized B-splines method which, however, does not give a clear cutoff time of the influence of acceleration on particulate matter emission. Hall and Hooker (2016) suggested that the point estimate for δ is 13 sec using Method A and 15 sec using Method B, both of which are more aggressive than our estimator.

6. Conclusion and Discussion

In this article, we consider to study the relation between a scalar response and a functional predictor in a truncated functional linear model. We propose a nested group bridge approach to achieve the historical sparseness, which reduces the variability and enhances the interpretability. Compared with the truncation methods by Hall and Hooker (2016), the proposed nested group bridge approach is able to provide a smooth and continuous estimate for the coefficient function and performs much better when the coefficient function tends to zero more smoothly. The proposed nested group bridge estimator of the cutoff time enjoys the estimation consistency. We demonstrate in simulation studies and a real data application that the proposed nested group bridge approach performs well for predictor functions that are not very smooth. We also show that even when the signal-to-noise ratio is low, the proposed nested group bridge approach can still accommodate the situation very well.

The question then arises as to in practice whether to use the proposed nested group bridge method or the truncation methods. We believe it depends on how smoothly the coefficient function tends to zero and how smooth the functional covariates are. Based on our simulation studies, we know that when the coefficient function is discontinuous at the cutoff time, the truncation methods perform better than the proposed nested group bridge method in terms of estimating the cutoff time. However, for relatively rough functional covariates, the truncation methods estimate the coefficient function less accurately than the proposed method. When the coefficient function goes to zero more smoothly, the proposed nested group bridge method outperforms the truncation methods in both estimating the cutoff time and the coefficient function. In practice, we can first obtain an estimate of $\beta(t)$ using penalized B-splines method. If the estimated $\hat{\beta}(t)$ does not have a steep slope at the tail region, the proposed nested group bridge method is recommended. When the estimated $\hat{\beta}(t)$ goes steeply to the tail region, for more accurate estimate of the cutoff time, the truncation methods should be applied. However, if the functional covariates are relatively rough, the proposed nested group bridge method provides more accurate estimate for the coefficient function.

Supplementary Materials

R code: We provide the R codes, which can be used to replicate the simulation studies and real data analysis included in the article. Please read file README contained in the zip file for more details. (NGR.zip, zip archive)

R-package: An R package `ngr` has been developed for implementing the proposed method. The R package and a demonstration are provided. (ngr.tar.gz, GNU zipped tar file; example.R)

Supplementary document: A supplementary document is available, which includes the introduction of the penalized B-splines method, additional results in Section 4, and detailed proofs of the theoretical results. (supplementarydoc.pdf)

Acknowledgments

The authors are very grateful to the editor, the associate editor, and two reviewers for their very constructive comments. These comments are extremely helpful for us to improve our work.

Funding

This research was supported by the Discovery grant, the Strategic Partnership grant, and the Canada Research Chair grant from the Natural Sciences and Engineering Research Council of Canada (NSERC) to J. Cao.

References

Ainsworth, L. M., Routledge, R., and Cao, J. (2011), "Functional Data Analysis in Ecosystem Research: The Decline of Oweekeno Lake Sockeye Salmon and Wannock River Flow," *Journal of Agricultural, Biological, and Environmental Statistics*, 16, 282–300. [620]

Asencio, M., Hooker, G., and Gao, H. O. (2014), "Functional Convolution Models," *Statistical Modelling*, 14, 315–335. [626]

Cai, T. T., and Hall, P. (2006), "Prediction in Functional Linear Regression," *The Annals of Statistics*, 34, 2159–2179. [620]

Candes, E., and Tao, T. (2007), "The Dantzig Selector: Statistical Estimation When p Is Much Larger Than n" (with discussion), *The Annals of Statistics*, 35, 2313–2351. [624]

Cardot, H., Ferraty, F., and Sarda, P. (2003), "Spline Estimators for the Functional Linear Model," *Statistica Sinica*, 13, 571–591. [620,621,622,623,624]

Claeskens, G., Krivobokova, T., and Opsomer, J. D. (2009), "Asymptotic Properties of Penalized Spline Estimators," *Biometrika*, 96, 529–544. [624]

Clark, N., Gautam, M., Wayne, W., Lyons, D., Thompson, G., and Zielinska, B. (2007), "Heavy-Duty Vehicle Chassis Dynamometer Testing for Emissions Inventory, Air Quality Modeling, Source Apportionment and Air Toxics Emissions Inventory: E55/59 All Phases," Coordinating Research Council, Alpharetta. [626]

de Boor, C. (2001), *A Practical Guide to Splines*, New York: Springer-Verlag. [621,623]

Efron, B., Hastie, T., Johnstone, I., and Tibshirani, R. (2004), "Least Angle Regression," *The Annals of Statistics*, 32, 407–499. [623]

Ferraty, F., and Vieu, P. (2006), *Nonparametric Functional Data Analysis: Theory and Practice*, New York: Springer-Verlag. [621]

Hall, P., and Hooker, G. (2016), "Truncated Linear Models for Functional Data," *Journal of Royal Statistical Society*, Series B, 78, 637–653. [621,623,624,625,626,627]

Hall, P., and Horowitz, J. L. (2007), "Methodology and Convergence Rates for Functional Linear Regression," *The Annals of Statistics*, 35, 70–91. [620,624]

Hastie, T., and Mallows, C. (1993), "A Statistical View of Some Chemometrics Regression Tools," *Technometrics*, 35, 140–143. [620]

Hsing, T., and Eubank, R. (2015), *Theoretical Foundations of Functional Data Analysis, With an Introduction to Linear Operators*, Chichester: Wiley. [621]

Huang, J., Ma, S., Xie, H., and Zhang, C. H. (2009), "A Group Bridge Approach for Variable Selection," *Biometrika*, 96, 339–355. [621,622,624]

James, G. M., Wang, J., and Zhu, J. (2009), "Functional Linear Regression That's Interpretable," *The Annals of Statistics*, 37, 2083–2108. [621,624,625,626]

Kokoszka, P., and Reimherr, M. (2017), *Introduction to Functional Data Analysis*, Boca Raton, FL: Chapman and Hall/CRC. [621]

Li, Y., and Hsing, T. (2007), "On Rates of Convergence in Functional Linear Regression," *Journal of Multivariate Analysis*, 98, 1782–1804. [620]

Lin, Z., Cao, J., Wang, L., and Wang, H. (2017), "Locally Sparse Estimator for Functional Linear Regression Models," *Journal of Computational and Graphical Statistics*, 26, 306–318. [621,623,624,625,626]

Liu, B., Wang, L., and Cao, J. (2017), "Estimating Functional Linear Mixed-Effects Regression Models," *Computational Statistics & Data Analysis*, 106, 153–164. [620]

Luo, W., Cao, J., Gallagher, M., and Wiles, J. (2013), "Estimating the Intensity of Ward Admission and Its Effect on Emergency Department Access Block," *Statistics in Medicine*, 32, 2681–2694. [620]

Marx, B. D., and Eilers, P. H. C. (1999), "Generalized Linear Regression on Sampled Signals and Curves: A *P*-Spline Approach," *Technometrics*, 41, 1–13. [623]

Morris, J. S. (2015), "Functional Regression," *Annual Review of Statistics and Its Application*, 2, 321–359. [621]

Ramsay, J. O., and Silverman, B. W. (2005), *Functional Data Analysis*, New York: Springer. [621]

Tibshirani, R. (1996), "Regression Shrinkage and Selection via the Lasso," *Journal of the Royal Statistical Society*, Series B, 58, 267–288. [622]

Wang, H., and Kai, B. (2015), "Functional Sparsity: Global Vesus Local," *Statistica Sinica*, 25, 1337–1354. [621,624]

Wang, J.-L., Chiou, J.-M., and Müller, H.-G. (2016), "Review of Functional Data Analysis," *Annual Review of Statistics and Its Application*, 3, 257–295. [621]

Yao, F., Müller, H. G., and Wang, J.-L. (2005), "Functional Linear Regression Analysis for Longitudinal Data," *The Annals of Statistics*, 33, 2873–2903. [620]

Yuan, M., and Cai, T. T. (2010), "A Reproducing Kernel Hilbert Space Approach to Functional Linear Regression," *The Annals of Statistics*, 38, 3412–3444. [620]

Zhou, J., Wang, N.-Y., and Wang, N. (2013), "Functional Linear Model With Zero-Value Coefficient Function at Sub-Regions," *Statistica Sinica*, 23, 25–50. [621]

Zou, H. (2006), "The Adaptive Lasso and Its Oracle Properties," *Journal of the American Statistical Association*, 101, 1418–1429. [622]