Statistica Sinica **34** (2024), 1903-1921
doi:https://doi.org/10.5705/ss.202022.0196

LOCALLY SPARSE ESTIMATOR OF GENERALIZED VARYING COEFFICIENT MODEL FOR ASYNCHRONOUS LONGITUDINAL DATA

Rou Zhong[1], Chunming Zhang[2] and Jingxiao Zhang*[1]

[1]*Renmin University of China and* [2]*University of Wisconsin-Madison*

Abstract: In longitudinal studies, it is common that the response and the covariate are not measured at the same time, which complicates the subsequent analysis. In this study, we consider the estimation of a generalized varying coefficient model with such asynchronous observations. We construct a penalized kernel-weighted estimating equation using the kernel technique in a functional data analysis framework. Moreover, we consider local sparsity in the estimating equation to improve the interpretability of the estimate. We extend the iteratively reweighted least squares algorithm in our computation, and establish the theoretical properties of the proposed method, including the consistency, sparsistency, and asymptotic distribution. Lastly, we use simulation studies to verify the performance of our method, and demonstrate the method by applying it to data from a study on women's health.

Key words and phrases: Asynchronous observation, functional data analysis, generalized varying coefficient model, kernel technique, local sparsity.

1. Introduction

A generalized varying coefficient model (Hastie and Tibshirani (1993); Cai, Fan and Li. (2000)) allows the coefficients to vary over time, significantly widening the application of regression models. Specifically, the model can be expressed as

$$E\{Y(t)|X(t)\} = g\{\beta_0(t) + \beta_1(t)X(t)\}, \quad t \in \mathcal{T}, \tag{1.1}$$

where $Y(t)$ is the response, $X(t)$ is the covariate, $g(\cdot)$ is a known strictly increasing and continuously twice-differentiable link function, $\beta_0(t)$ is the intercept function, $\beta_1(t)$ is the varying coefficient function, and $\mathcal{T}$ is a bounded and closed interval. Here, we propose a new estimating method for a generalized varying coefficient model with longitudinal measurements, from the perspective of functional data.

In practice, it often happens that the covariate and the response are not measured at the same time for each subject in longitudinal observations. Such asynchronous observations make the subsequent analysis more complicated. Two main types of approaches have been proposed to solve this problem. The first

*Corresponding author.

comprises two steps, and is based on synchronizing the measurements of the covariate and the response. For example, Xiong and Dubin (2010) propose a binning method to align the measurement times in order to use traditional longitudinal modeling, and Şentürk et al. (2013) use a functional principal component analysis (FPCA) method to synchronize the data. However, because the data used for modeling is obtained from estimations, errors from each step accumulate. The second approach imposes a kernel weight based on the time difference between the observations of the covariate and the response. These methods are more appealing, because they use all available data. Cao, Zeng and Fine (2015) construct a kernel-weighted estimating equation for a generalized linear model and a generalized varying coefficient model. Cao, Li and Fine (2016) develop a weighted last observation carried forward (LOCF) method, and Chen and Cao (2017) apply the kernel weighting technique to partially linear models. Li et al. (2022) consider models with longitudinal functional covariates, and Sun, Zhao and Sun (2021) examine cases in which the observation times are informative. Most of the above kernel methods work only with models with time invariant coefficients, and only Cao, Zeng and Fine (2015) consider a generalized varying coefficient model. However, their varying coefficients are estimated point by point, which can be time consuming and lacks integrity. Therefore, a new estimating method is required.

Interpreting the varying coefficient function $\beta_1(t)$ is a vital part of a regression analysis. These interpretability can be improved by introducing local sparsity, which means the curve can be strictly equal to zero in some subintervals. Some prior works have achieved local sparsity by imposing a sparseness penalty for various models. For example, James, Wang and Zhu (2009), Zhou, Wang and Wang (2013), and Lin et al. (2017) develop locally sparse estimators for a scalar-on-function regression model, and Tu, Park and Wang (2020) use a group bridge approach to obtain locally sparse estimates for a varying coefficient model. Fang et al. (2020) generalize the method of Lin et al. (2017) to cases in which the response is multivariate, and a function-on-function regression model and a function-on-scalar regression model are considered by Centofanti et al. (2020) and Wang et al. (2020), respectively. However, to the best of our knowledge, local sparsity has not been considered for generalized varying coefficient models.

We use a functional data analysis (FDA) approach, because longitudinal data can be viewed as functional data in a sparse design, and an FDA is more effective than using pointwise methods. Our goal is to propose a novel method that can be applied to asynchronous data, and that can produce estimates that are more interpretable. Specifically, we construct a new kernel-weighted estimating equation with penalties on both the roughness and the sparseness. To solve the estimating equation, we extend the iteratively reweighted least squares (IRLS) method, and design an innovative algorithm for the computation. We also consider the selection of the tuning parameters. We generalize the extended

Bayesian information criterion (EBIC) in Chen and Chen (2008, 2012), to adapt it to asynchronous data, such that the roughness parameter and the sparseness parameter can be chosen accordingly. Moreover, we select the number of basis functions using cross-validation (CV). The proposed method for a generalized varying coefficient model is called LocKer, because we can use it to obtain a locally sparse estimator of $\beta_1(t)$, and we use the kernel technique in the procedure. We also explore the theoretical properties of the proposed approach.

Our work contributes to the literature in three ways. First, we study generalized varying coefficient models in an FDA framework, considering both asynchronous data and local sparsity. Solving this problem will improve the accuracy, utility, and interpretability of the results. Second, the proposed algorithm can be implemented using the R package `LocKer`, available at `https://CRAN.R-project.org/package=LocKer`. Third, we explore the consistency, sparsistency, and asymptotic distribution of our proposed method.

The remainder of the paper proceeds as follows. In Section 2, we construct the penalized kernel-weighted estimating equation, and develop a computation algorithm for the proposed LocKer method. We discuss the theoretical properties of the proposed method in Section 3. In Section 4, we use simulation studies to explore the accuracy of the proposed method and its ability to identify zero-valued subintervals. We apply our method to data from a study on women's health in Section 5, and conclude the paper in Section 6.

2. Methodology

2.1. Estimating equation

Suppose there are n independent subjects in the study. For the ith subject, let $Y_i(t)$ and $X_i(t)$ be realizations of the response process $Y(t)$ and the covariate process $X(t)$, respectively. However, only longitudinal measurements are obtained. Specifically, for $i = 1, \ldots, n$, we observe

$$Y_i(T_{ij}), j = 1, \ldots, L_i, \quad X_i(S_{ik}), k = 1, \ldots, M_i,$$

where T_{ij} is the jth observation time of the response, S_{ik} is the kth observation time of the covariate, L_i is the observation size of the response, and M_i is the observation size of the covariate. Following Cao, Zeng and Fine (2015), the observation times can be viewed as being generated from a bivariate counting process

$$N_i(t, s) = \sum_{j=1}^{L_i} \sum_{k=1}^{M_i} I(T_{ij} \le t, S_{ik} \le s),$$

where $I(\cdot)$ is the indicator function.

To estimate $\beta_0(t)$ and $\beta_1(t)$ in (1.1), we employ the following basis approximation:

$$\beta_0(t) \approx \sum_{l=1}^{L} B_l(t)\gamma_l^{(0)} = \mathbf{B}(t)^\top \boldsymbol{\gamma}^{(0)}, \quad \beta_1(t) \approx \sum_{l=1}^{L} B_l(t)\gamma_l^{(1)} = \mathbf{B}(t)^\top \boldsymbol{\gamma}^{(1)},$$

where $\{B_l(t), l = 1, \ldots, L\}$ are B-spline basis functions with degree d and M interior knots, $\gamma_l^{(0)}$ and $\gamma_l^{(1)}$ are the corresponding coefficients of $\beta_0(t)$ and $\beta_1(t)$, $\mathbf{B}(t) = (B_1(t), \ldots, B_L(t))^\top$, $\boldsymbol{\gamma}^{(0)} = (\gamma_1^{(0)}, \ldots, \gamma_L^{(0)})^\top$, $\boldsymbol{\gamma}^{(1)} = (\gamma_1^{(1)}, \ldots, \gamma_L^{(1)})^\top$, and $L = M + d + 1$ is the number of basis functions. Here, we apply B-spline basis functions; Zhong et al. (2021) explain the reasons for the wide use of B-spline basis functions in local sparse estimation. Let $\boldsymbol{\gamma} = (\boldsymbol{\gamma}^{(0)\top}, \boldsymbol{\gamma}^{(1)\top})^\top$, $\widetilde{X}_l(t) = X(t)B_l(t)$, and $\widetilde{\mathbf{X}}(t) = (\widetilde{X}_1(t), \ldots, \widetilde{X}_L(t))^\top$. Then, the generalized varying coefficient model (1.1) can be approximated by

$$E\{Y(t)|X(t)\} = g\left\{ \sum_{l=1}^{L} B_l(t)\gamma_l^{(0)} + \sum_{l=1}^{L} \widetilde{X}_l(t)\gamma_l^{(1)} \right\} = g\left\{\widetilde{\mathbf{X}}^\star(t)^\top \boldsymbol{\gamma}\right\},$$

where $\widetilde{\mathbf{X}}^\star(t) = (\mathbf{B}(t)^\top, \widetilde{\mathbf{X}}(t)^\top)^\top$. Following previous works, such as Lin et al. (2017) and Li et al. (2022), we use an equal sign above to denote the approximation. We can obtain estimates of $\beta_0(t)$ and $\beta_1(t)$ using the estimation of $\boldsymbol{\gamma}$. To this end, we construct the following penalized kernel-weighted estimating equation:

$$\begin{aligned} U_n(\boldsymbol{\gamma}) = \frac{1}{N_0} \sum_{i=1}^{n} \sum_{j=1}^{L_i} \sum_{k=1}^{M_i} K_h(T_{ij} - S_{ik}) \widetilde{\mathbf{X}}_i^\star(S_{ik}) \left[Y_i(T_{ij}) - g\left\{\widetilde{\mathbf{X}}_i^\star(S_{ik})^\top \boldsymbol{\gamma}\right\}\right] \\ - \mathbf{V}_{\rho_0,\rho_1}\boldsymbol{\gamma} - \frac{\partial \mathrm{PEN}_\lambda(\boldsymbol{\gamma})}{\partial \boldsymbol{\gamma}} = \mathbf{0}, \end{aligned} \tag{2.1}$$

where $N_0 = \sum_{i=1}^{n} L_i M_i$, $\mathbf{V}_{\rho_0,\rho_1} = \mathrm{diag}(\rho_0 \mathbf{V}, \rho_1 \mathbf{V})$, $\mathbf{V} = \int_{\mathcal{T}} \mathbf{B}^{(2)}(t)\mathbf{B}(t)^{(2)\top} dt$, $\mathbf{B}^{(2)}(t)$ is the second derivative of $\mathbf{B}(t)$, ρ_0 and ρ_1 are the roughness parameters for $\beta_0(t)$ and $\beta_1(t)$, respectively, $K_h(t) = K(t/h)/h$, $K(t)$ is a symmetric kernel function, h is the bandwidth, $\mathrm{PEN}_\lambda(\boldsymbol{\gamma})$ is the sparseness penalty for $\beta_1(t)$, λ is the sparseness parameter, and $\mathbf{0}$ is a zero-valued vector with length $2L$. Here, we use $h = \max(\tau_{0.95}, 0.01)$ as the bandwidth, where $\tau_{0.95}$ is the 0.95-quantile of $\min_{j,k} |T_{ij} - S_{ik}|$. For the first term in (2.1), define the kernel-weighted log-likelihood function as

$$\sum_{i=1}^{n} \sum_{j=1}^{L_i} \sum_{k=1}^{M_i} \left\{ \frac{Y_i(T_{ij})\theta_{ik} - b(\theta_{ik})}{a(\phi)} + c(Y_i(T_{ij}), \phi) \right\} K_h(T_{ij} - S_{ik}),$$

where $\theta_{ik} = \widetilde{\mathbf{X}}_i^\star(S_{ik})^\top \boldsymbol{\gamma}$, $b'(\cdot) = g(\cdot)$, $a(\phi)$ and $c(Y_i(T_{ij}), \phi)$ are both constants. Then, the first term can be viewed as the derivative of the kernel-weighted log-

likelihood function by neglecting a constant multiplier. Here, we consider all possible pairs of response and covariate measurements, using the kernel weights to control the effects of various pairs, such that measurements with close observation times are emphasized. The second term is the derivative of the roughness penalty, which is defined as

$$\begin{aligned}&\frac{\rho_0}{2}\int_{\mathcal{T}}\{\beta_0^{(2)}(t)\}^2dt+\frac{\rho_1}{2}\int_{\mathcal{T}}\{\beta_1^{(2)}(t)\}^2dt\\&=\frac{\rho_0}{2}\boldsymbol{\gamma}^{(0)\top}\mathbf{V}\boldsymbol{\gamma}^{(0)}+\frac{\rho_1}{2}\boldsymbol{\gamma}^{(1)\top}\mathbf{V}\boldsymbol{\gamma}^{(1)}=\frac{1}{2}\boldsymbol{\gamma}^{\top}\mathbf{V}_{\rho_0,\rho_1}\boldsymbol{\gamma},\end{aligned}$$

where $\beta_0^{(2)}(t)$ and $\beta_1^{(2)}(t)$ are the second derivatives of $\beta_0(t)$ and $\beta_1(t)$, respectively. The third term is the derivative of the sparseness penalty $\mathrm{PEN}_\lambda(\boldsymbol{\gamma})$, the expression of which is provided in Section 2.2. Note that the roughness penalty and the sparseness penalty are imposed on the estimating equation by their derivatives. Using (2.1), we can obtain a locally sparse estimator for model (1.1) with asynchronous observations. Though we consider a generalized varying coefficient model with one covariate here, this can be extended easily to cases with more covariates.

2.2. Sparseness penalty

In this section, we introduce the sparseness penalty used in (2.1). We generalize the functional SCAD penalty in (Lin et al. (2017)) to achieve local sparsity of $\beta_1(t)$. Specifically, the sparseness penalty imposed on $\beta_1(t)$ is defined as

$$\mathcal{L}(\beta_1)=\frac{M+1}{2T}\int_{\mathcal{T}}p_\lambda(|\beta_1(t)|)dt\approx\frac{1}{2}\sum_{m=1}^{M+1}p_\lambda\left(\sqrt{\frac{M+1}{T}\int_{\tau_{m-1}}^{\tau_m}\beta_1^2(t)dt}\right),\qquad(2.2)$$

where T is the length of $\mathcal{T}$, τ_m is the knot of the used B-spline basis, and $p_\lambda(\cdot)$ is the SCAD function suggested in (Fan and Li (2001)). We then transform (2.2) to the penalty of $\boldsymbol{\gamma}$ for the sake of computation. Let $\|\beta_{1[m]}\|_2^2=\int_{\tau_{m-1}}^{\tau_m}\beta_1^2(t)dt$. By the local quadratic approximation $p_\lambda(|v|)\approx p_\lambda(|v_0|)+(1/2)\{p_\lambda^{'}(|v_0|)/|v_0|\}(v^2-v_0^2)$ in (Fan and Li (2001)), we have

$$\begin{aligned}&\sum_{m=1}^{M+1}p_\lambda\left(\sqrt{\frac{M+1}{T}}\|\beta_{1[m]}\|_2\right)\\&\approx\sum_{m=1}^{M+1}\left\{p_\lambda\left(\sqrt{\frac{M+1}{T}}\|\beta_{1[m]}^{(0)}\|_2\right)\right.\\&\quad\left.+\frac{1}{2}\frac{p_\lambda'\left(\sqrt{(M+1)/T}\|\beta_{1[m]}^{(0)}\|_2\right)}{\sqrt{(M+1)/T}\|\beta_{1[m]}^{(0)}\|_2}\left(\frac{M+1}{T}\|\beta_{1[m]}\|_2^2-\frac{M+1}{T}\|\beta_{1[m]}^{(0)}\|_2^2\right)\right\}\end{aligned}$$

$$
\begin{aligned}
&= \frac{1}{2}\sum_{m=1}^{M+1}\sqrt{\frac{M+1}{T}}p'_\lambda\left(\sqrt{\frac{M+1}{T}}\|\beta^{(0)}_{1[m]}\|_2\right)\frac{\|\beta_{1[m]}\|^2}{\|\beta^{(0)}_{1[m]}\|^2}+C\\
&= \sum_{m=1}^{M+1}\boldsymbol{\gamma}^{(0)\top}\mathbf{U}_m\boldsymbol{\gamma}^{(0)}+C\\
&= \boldsymbol{\gamma}^\top\mathbf{U}\boldsymbol{\gamma}+C,
\end{aligned}
$$

where

$$
\begin{aligned}
\mathbf{U}_m &= \sqrt{\frac{M+1}{T}}\,\frac{p'_\lambda\Big(\sqrt{(M+1)/T}\|\beta^{(0)}_{1[m]}\|_2\Big)}{2\|\beta^{(0)}_{1[m]}\|_2}\mathbf{T}_m,\\
\mathbf{T}_m &= \int_{\tau_{m-1}}^{\tau_m}\mathbf{B}(t)\mathbf{B}(t)^\top dt, \quad \mathbf{U} = \operatorname{diag}\Big(\mathbf{O}, \sum_{m=1}^{M+1}\mathbf{U}_m\Big), \qquad (2.3)\\
C &= \sum_{m=1}^{M+1}p_\lambda\left(\sqrt{\frac{M+1}{T}}\|\beta^{(0)}_{1[m]}\|_2\right)\\
&\quad -\frac{1}{2}\sum_{m=1}^{M+1}\sqrt{\frac{M+1}{T}}p'_\lambda\left(\sqrt{\frac{M+1}{T}}\|\beta^{(0)}_{1[m]}\|_2\right)\|\beta^{(0)}_{1[m]}\|_2,
\end{aligned}
$$

and $\mathbf{O}$ is an $L\times L$ matrix with all elements being zero. Here, $\|\beta^{(0)}_{1[m]}\|_2$ is obtained from the initial value or the estimate in the previous iteration. Then, the sparseness penalty in (2.1) can be expressed as

$$\mathrm{PEN}_\lambda(\boldsymbol{\gamma}) = \frac{1}{2}\boldsymbol{\gamma}^\top\mathbf{U}\boldsymbol{\gamma}.$$

Here, the value of $\mathbf{U}$ depends on the value of $\|\beta^{(0)}_{1[m]}\|_2$, so it varies in the iteration process introduced in Section 2.3.

2.3. Algorithm

We generalize the IRLS algorithm to solve our estimating equation proposed in Section 2.1. To this end, we first rewrite (2.1) in matrix form, and introduce some additional notation. Let $\widetilde{\mathbf{X}}_i^\star = (\widetilde{\mathbf{X}}_i^\star(S_{i1}),\ldots,\widetilde{\mathbf{X}}_i^\star(S_{iM_i}))^\top$, $\widetilde{\mathbf{X}}^\star = (\mathbf{1}_{L_1}^\top\otimes\widetilde{\mathbf{X}}_1^{\star\top},\ldots,\mathbf{1}_{L_n}^\top\otimes\widetilde{\mathbf{X}}_n^{\star\top})^\top$, $\mathbf{Y}_i = (Y_i(T_{i1}),\ldots,Y_i(T_{iL_i}))^\top$, $\mathbf{Y} = (\mathbf{Y}_1^\top\otimes\mathbf{1}_{M_1}^\top,\ldots,\mathbf{Y}_n^\top\otimes\mathbf{1}_{M_n}^\top)^\top$, $\boldsymbol{\eta} = \widetilde{\mathbf{X}}^\star\boldsymbol{\gamma}$, $\mathbf{Z} = \boldsymbol{\eta} + \{\mathbf{Y}-g(\boldsymbol{\eta})\}\cdot f'\{g(\boldsymbol{\eta})\}$, $\mathbf{W} = \operatorname{diag}\{K_h(T_{11}-S_{11}),\ldots,K_h(T_{11}-S_{1M_1}),K_h(T_{12}-S_{11}),\ldots,K_h(T_{nL_n}-S_{nM_n})\}$, and $\mathbf{H} = \operatorname{diag}[1/f'\{g(\boldsymbol{\eta})\}]$, where $\otimes$ is the Kronecker product, $\mathbf{1}_{L_i}$ and $\mathbf{1}_{M_i}$ are vectors of length L_i and M_i, respectively, with all elements being one, and $f'(\cdot)$ is the first derivative of $f(\cdot)$, which is the inverse function of $g(\cdot)$. Then, the penalized kernel-weighted

estimating equation (2.1) becomes

$$U_n(\boldsymbol{\gamma}) = \frac{1}{N_0}\widetilde{\mathbf{X}}^{\star\top}\mathbf{WH}(\mathbf{Z}-\boldsymbol{\eta}) - \mathbf{V}_{\rho_0,\rho_1}\boldsymbol{\gamma} - \mathbf{U}\boldsymbol{\gamma} = \mathbf{0}, \tag{2.4}$$

where $\mathbf{H}$, $\mathbf{Z}$, $\boldsymbol{\eta}$, and $\mathbf{U}$ are computed using the initial value of $\boldsymbol{\gamma}$ or its estimate in the previous iteration. From (2.4), we obtain the new estimate as

$$\widehat{\boldsymbol{\gamma}} = (\widetilde{\mathbf{X}}^{\star\top}\mathbf{WH}\widetilde{\mathbf{X}}^{\star} + N_0\mathbf{V}_{\rho_1,\rho_2} + N_0\mathbf{U})^{-1}\widetilde{\mathbf{X}}^{\star\top}\mathbf{WHZ}. \tag{2.5}$$

Moreover, following (Lin et al. (2017)) and (Zhong et al. (2021)), the small elements of $\widehat{\boldsymbol{\gamma}}$ are shrunk to zero in the iteration such that $\widetilde{\mathbf{X}}^{\star\top}\mathbf{WH}\widetilde{\mathbf{X}}^{\star} + N_0\mathbf{V}_{\rho_1,\rho_2} + N_0\mathbf{U}$ is not singular. Then, the estimates of $\beta_0(t)$ and $\beta_1(t)$ are given by

$$\widehat{\beta}_0(t) = \mathbf{B}(t)^\top\widehat{\boldsymbol{\gamma}}^{(0)} \quad \text{and} \quad \widehat{\beta}_1(t) = \mathbf{B}(t)^\top\widehat{\boldsymbol{\gamma}}^{(1)}, \tag{2.6}$$

respectively, where $\widehat{\boldsymbol{\gamma}}^{(0)}$ and $\widehat{\boldsymbol{\gamma}}^{(1)}$ are obtained from the final estimate of $\boldsymbol{\gamma}$ using the definition $\boldsymbol{\gamma} = (\boldsymbol{\gamma}^{(0)\top}, \boldsymbol{\gamma}^{(1)\top})^\top$.

The whole algorithm is summarized as follows:

Step 1: Give the initial value of $\boldsymbol{\gamma}$, which we denote as $\boldsymbol{\gamma}^{[0]}$. Here, we use a least squares estimate with a kernel weight, and consider the roughness penalty in the initial estimate, that is, $\boldsymbol{\gamma}^{[0]} = (\widetilde{\mathbf{X}}^{\star\top}\mathbf{W}\widetilde{\mathbf{X}}^{\star} + N_0\mathbf{V}_{\rho_1,\rho_2})^{-1}\widetilde{\mathbf{X}}^{\star\top}\mathbf{WY}$.

Step 2: Start with $q = 1$. For the qth iteration,

(1) $\boldsymbol{\eta}^{[q]} = \widetilde{\mathbf{X}}^{\star}\boldsymbol{\gamma}^{[q-1]}$.

(2) $\mathbf{Z}^{[q]} = \boldsymbol{\eta}^{[q]} + \{\mathbf{Y} - g(\boldsymbol{\eta}^{[q]})\} \cdot f'\{g(\boldsymbol{\eta}^{[q]})\}$ and $\mathbf{H}^{[q]} = \mathrm{diag}[1/f'\{g(\boldsymbol{\eta}^{[q]})\}]$.

(3) Compute $\mathbf{U}^{[q]}$ from (2.3).

(4) $\boldsymbol{\gamma}^{[q]} = (\widetilde{\mathbf{X}}^{\star\top}\mathbf{WH}^{[q]}\widetilde{\mathbf{X}}^{\star} + N_0\mathbf{V}_{\rho_1,\rho_2} + N_0\mathbf{U}^{[q]})^{-1}\widetilde{\mathbf{X}}^{\star\top}\mathbf{WH}^{[q]}\mathbf{Z}^{[q]}$ from (2.5).

(5) Repeat Step 2(1)–(4) until convergence.

Step 3: Let $\widehat{\boldsymbol{\gamma}} = \boldsymbol{\gamma}^{[q]}$. Then, compute $\widehat{\beta}_0(t)$ and $\widehat{\beta}_1(t)$ using (2.6).

2.4. Selection of tuning parameters

Recall that the bandwidth is chosen as $h = \max(\tau_{0.95}, 0.01)$, where $\tau_{0.95}$ is the 0.95-quantile of $\min_{j,k}|T_{ij} - S_{ik}|$. In this section, we discuss selecting the other tuning parameters in the computation, including the roughness parameters, sparseness parameter, and number of B-spline basis functions, with the bandwidth determined already. For clarity, let $\rho_0 = \rho_1 \triangleq \widetilde{\rho}$, which means $\beta_0(t)$ and $\beta_1(t)$ share the same roughness parameter. However, our selection criterion can be extended easily to the case $\rho_0 \neq \rho_1$.

The roughness parameter $\widetilde{\rho}$ and the sparseness parameter λ are considered jointly. We generalize the EBIC in (Chen and Chen (2008, 2012)) to adapt it to the asynchronous observations. More specifically, define

$$\mathrm{EBIC}(\widetilde{\rho},\lambda)=\log(\mathrm{Dev})+df\cdot\frac{\log(n_0)}{n_0}+\nu\cdot df\cdot\frac{\log(2L)}{n_0}, \tag{2.7}$$

where Dev represents the deviance of the estimate, df is the degrees of freedom, $n_0=\#\{K_h(T_{ij}-S_{ik})\neq 0, i=1,\ldots,n; j=1,\ldots L_i; k=1,\ldots,M_i\}$, and $0\leq\nu\leq 1$. We use $\nu=0.5$, as suggested by Huang, Horowitz and Wei (2010). Moreover, Dev is given by

$$\mathrm{Dev}=-2\sum_{i=1}^{n}\sum_{j=1}^{L_i}\sum_{k=1}^{M_i}\{Y_i(T_{ij})\widehat{\theta}_{ik}-b(\widehat{\theta}_{ik})\}K_h(T_{ij}-S_{ik}),$$

where $\widehat{\theta}_{ik}=g(\widehat{Y}_i(S_{ik}))$. Then by ignoring some constant, we have

$$\mathrm{Dev}=\sum_{i=1}^{n}\sum_{j=1}^{L_i}\sum_{k=1}^{M_i}\{Y_i(T_{ij})-\widehat{Y}_i(S_{ik})\}^2K_h(T_{ij}-S_{ik})$$

for a Gaussian response,

$$\mathrm{Dev}=2\sum_{i=1}^{n}\sum_{j=1}^{L_i}\sum_{k=1}^{M_i}\left[Y_i(T_{ij})\log\frac{Y_i(T_{ij})}{\widehat{Y}_i(S_{ik})}+\{1-Y_i(T_{ij})\}\log\frac{1-Y_i(T_{ij})}{1-\widehat{Y}_i(S_{ik})}\right]K_h(T_{ij}-S_{ik})$$

for a Bernoulli response, and,

$$\mathrm{Dev}=2\sum_{i=1}^{n}\sum_{j=1}^{L_i}\sum_{k=1}^{M_i}[\widehat{Y}_i(S_{ik})-Y_i(T_{ij})\log\{\widehat{Y}_i(S_{ik})\}]K_h(T_{ij}-S_{ik})$$

for a Poisson response. Furthermore, df is computed by

$$df=\mathrm{tr}\{\widetilde{\mathbf{X}}^{\star}_{\mathcal{A}}(\widetilde{\mathbf{X}}^{\star\top}_{\mathcal{A}}\mathbf{W}_{\mathcal{A}}\widetilde{\mathbf{X}}^{\star}_{\mathcal{A}}+N_0\mathbf{V}_{\rho_1,\rho_2\mathcal{A}})^{-1}\widetilde{\mathbf{X}}^{\star\top}_{\mathcal{A}}\mathbf{W}_{\mathcal{A}}\},$$

where $\mathcal{A}$ is a set indexing the nonzero elements in $\widehat{\boldsymbol{\gamma}}$. For the third term in (2.7), $2L$ is the length of $\boldsymbol{\gamma}$, and if more covariates are considered, it should be varied accordingly.

We choose the number of B-spline basis functions using CV. For a given L, we first select the best $\widetilde{\rho}$ and λ using the EBIC, and then calculate the CV score using the same method as Dev when facing responses with various distributions. The effect of L is discussed in our simulation study in Section 4.2.

3. Theoretical results

We study the asymptotic properties of our method in this section. Let $\eta(t,\boldsymbol{\beta}) = \beta_0(t) + X(t)\beta_1(t)$, where $\boldsymbol{\beta}(t) = (\beta_0(t), \beta_1(t))^\top$. Let $\boldsymbol{\beta}_0(t)$ be the true value of $\boldsymbol{\beta}(t)$. Define $\mathbf{X}^\star(t) = (1, X(t))^\top$. Let $\text{var}\{Y(t)|X(t)\} = \sigma\{t, X(t)\}^2$ and $\text{cov}\{Y(s), Y(t)|X(s), X(t)\} = r\{s, t, X(s), X(t)\}$. Moreover, denote $\text{NULL}(f) = \{t \in \mathcal{T} : f(t) = 0\}$ and $\text{SUPP}(f) = \{t \in \mathcal{T} : f(t) \neq 0\}$, for any function $f(t)$. Denote $\rho = \max(\rho_0, \rho_1)$. The needed assumptions are listed as follows:

Assumption 1. *There exists some constant $c > 0$ such that $|\beta_0^{(p')}(t_1) - \beta_0^{(p')}(t_2)| \leq c|t_1 - t_2|^\nu$ and $|\beta_1^{(p')}(t_1) - \beta_1^{(p')}(t_2)| \leq c|t_1 - t_2|^\nu$, for $\nu \in [0, 1]$. Let $r = p' + \nu$, and assume that $3/2 < r \leq d$, where d is the degree of the B-spline basis.*

Assumption 2. *The counting process $N_i(t, s)$ is independent of (Y_i, X_i) and $E\{dN_i(t, s)\} = \lambda(t, s)dtds$, where $\lambda(t, s)$ is a bounded twice-continuous differentiable function for any $t, s \in \mathcal{T}$. The Borel measure for $\mathcal{G} = \{\lambda(t, t) > 0, t \in \mathcal{T}\}$ is strictly positive. Moreover, $P\{dN(t_1, t_2) = 1|N(s_1, s_2) - N(s_1-, s_2-) = 1\} = f(t_1, t_2, s_1, s_2)dt_1dt_2$, for $t_1 \neq s_1$ and $t_2 \neq s_2$, where $f(t_1, t_2, s_1, s_2)$ is continuous and $f(t_1\pm, t_2\pm, s_1\pm, s_2\pm)$ exists.*

Assumption 3. *The tuning parameter $\lambda \to 0$ as $n \to \infty$. Assume that $\sqrt{\int_{SUPP_{(\beta_1)}} p'_\lambda(|\beta_1(t)|)^2 dt} = O(n^{-1/2}M^{-3/2})$, $\sqrt{\int_{SUPP_{(\beta_1)}} p''_\lambda(|\beta_1(t)|)^2 dt} = o(M^{-3/2})$.*

Assumption 4. *For any $\boldsymbol{\beta}$ in a neighborhood of $\boldsymbol{\beta}_0$, we assume that $E[\boldsymbol{X}^\star(s)\, g\{\eta(t,\boldsymbol{\beta})\}]$ and $E[\boldsymbol{X}^\star(s)g'\{\eta(t,\boldsymbol{\beta})\}X^b(t)]$ are twice-continuous differentiable for any $(t, s) \in \mathcal{T}^2$, where $b = 0, 1$. Moreover, we assume that $E[\boldsymbol{X}^\star(s_1)\boldsymbol{X}^\star(s_2)^\top\, g\{\eta(t_1,\boldsymbol{\beta})\}g\{\eta(t_2,\boldsymbol{\beta})\}]$ and $E[\boldsymbol{X}^\star(s_1)\boldsymbol{X}^\star(s_2)^\top r\{t_1, t_2, X(t_1), X(t_2)\}]$ are twice-continuous differentiable for any $(t_1, t_2, s_1, s_2) \in \mathcal{T}^4$.*

Assumption 5. *For any $\boldsymbol{\beta}$ in a neighborhood of $\boldsymbol{\beta}_0$, we assume that $E[\boldsymbol{X}_2^\star(s)\, \boldsymbol{X}_2^\star(s)^\top g'\{\eta(s,\boldsymbol{\beta})\}^2]$ and $E[\boldsymbol{X}^\star(s)\sigma\{s, X(s)\}^2]$ are uniformly bounded in s, where $\boldsymbol{X}_2^\star(s) = (1, X^2(t))^\top$.*

Assumption 6. *If ψ_0 and ψ_1 satisfy $\psi_0(s) + \psi_1(s)X(s) = 0$, for $\forall s \in \mathcal{G}$, with probability one, then $\psi_0 = 0$ and $\psi_1 = 0$.*

Assumption 7. *The kernel function $K(\cdot)$ is a symmetric density function. Assume that $\int z^2K(z)dz < \infty$ and $\int K(z)^2dz < \infty$.*

Assumption 1 is similar to (C2) in (Lin et al. (2017)), and is used to justify the B-spline approximation. The requirement for the counting process is presented in Assumption 2, and is the same as Condition 1 in (Cao, Zeng and Fine (2015)) and Assumption 3 in (Li et al. (2022)). Assumption 3 is analogous to (C3) in (Lin et al. (2017)), and Assumptions 4–6 are parallel to assumptions in

(Li et al. (2022)). Furthermore, Assumption 7 is a common assumption for a kernel function.

Theorem 1. *Under Assumptions* 1–7*, if* $M^{1/2}h^2 \to 0$*,* $n^{-1/2}M^{3/2}h^{-1/2} \to 0$*,* $\rho \to 0$*, and* $M^{-r} \to 0$*, then we have*

$$\sup_{t\in\mathcal{T}} |\widehat{\beta}_0(t) - \beta_0(t)| = O_p(M^{1/2}h^2 + n^{-1/2}M^{1/2}h^{-1/2} + \rho M^{-1/2} + M^{-r}),$$
$$\sup_{t\in\mathcal{T}} |\widehat{\beta}_1(t) - \beta_1(t)| = O_p(M^{1/2}h^2 + n^{-1/2}M^{1/2}h^{-1/2} + \rho M^{-1/2} + M^{-r}).$$

The above theorem states the consistency of both $\beta_0(t)$ and $\beta_1(t)$, and the convergence rates are also given. To achieve the best convergence rate in Theorem 1, we can set $h = O(n^{-1/5})$, $M = O(n^{4/(5(1+2r))})$, and $\rho = O(n^{(-4r+2)/(5(1+2r))})$. Then, we have $\sup_{t\in\mathcal{T}} |\widehat{\beta}_0(t) - \beta_0(t)| = O_p(n^{-4r/(5(1+2r))})$ and $\sup_{t\in\mathcal{T}} |\widehat{\beta}_1(t) - \beta_1(t)| = O_p(n^{-4r/(5(1+2r))})$. We discuss the sparsistency of $\beta_1(t)$ in the following theorem.

Theorem 2. *Suppose that the conditions of Theorem* 1 *are satisfied. If* $nh^5 = O(1)$*,* $nhM^{-2r} = o(1)$*,* $\rho = o(n^{-1/2})$*, and* $\lambda n^{1/2}M^{-1/2}h^{1/2} \to \infty$*, then we have* $NULL(\widehat{\beta}_1) \to NULL(\beta_1)$ *and* $SUPP(\widehat{\beta}_1) \to SUPP(\beta_1)$ *in probability, as* $n \to \infty$*.*

According to Theorem 2, the zero-valued subintervals of our estimate $\widehat{\beta}_1(t)$ are consistent with the true zero-valued subintervals. That means we have $\widehat{\beta}_1(t) = 0$ for any $t \in \text{NULL}(\beta_1)$, and $\widehat{\beta}_1(t) \neq 0$ for any $t \in \text{SUPP}(\beta_1)$ in probability. Next, we discuss the asymptotic distribution of $\widehat{\boldsymbol{\gamma}}$. Let $\boldsymbol{\gamma}_0 = (\boldsymbol{\gamma}_0^{(0)\top}, \boldsymbol{\gamma}_0^{(1)\top})^\top$ be the coefficient vector that satisfies $\|\boldsymbol{\gamma}_0^{(0)\top}\mathbf{B} - \beta_0\|_\infty = O(M^{-r})$ and $\|\boldsymbol{\gamma}_0^{(1)\top}\mathbf{B} - \beta_1\|_\infty = O(M^{-r})$ (de Boor (2001); Zhong et al. (2021)).

Theorem 3. *Suppose that the conditions of Theorem* 1 *are satisfied. If* $nh^5M = o(1)$*,* $nhM^{-2r} = O(1)$*,* $n^{-1}M^2 = o(1)$*, and* $\rho = o(n^{-1/2})$*, then*

$$\frac{nh(\widehat{\boldsymbol{\gamma}} - \boldsymbol{\gamma}_0)^\top \Omega_n^2 (\widehat{\boldsymbol{\gamma}} - \boldsymbol{\gamma}_0) - tr(\Sigma_0)}{\sqrt{2tr(\Sigma_0^2)}} \xrightarrow{d} N(0,1),$$

where

$$\Omega_n = n^{-1}\sum_{i=1}^{n} \int\int K_h(t-s)\widetilde{\boldsymbol{X}}_i^\star(s) g'\{\eta_i(s,\boldsymbol{\beta}_0)\}\widetilde{\boldsymbol{X}}_i^\star(s)^\top dN_i(t,s),$$
$$\Sigma_0 = var\left(h^{1/2}\int\int K_h(t-s)\widetilde{\boldsymbol{X}}^\star(s)[Y(t) - g\{\eta(s,\boldsymbol{\beta}_0)\}]dN(t,s)\right).$$

The asymptotic distribution of $\widehat{\boldsymbol{\gamma}}$ is examined further using simulated data in the Supplementary Material, where we also explore the pointwise asymptotic distributions of $\widehat{\beta}_0(t)$ and $\widehat{\beta}_1(t)$, and provide proofs of all theorems.

4. Simulation Studies

4.1. Numerical performance

In this section, we discuss the performance of the proposed method by simulation studies. The simulated data sets are generated from model (1.1), and Gaussian response, Bernoulli response and Poisson response are all in consideration. Moreover, for each distribution, both nonsparse coefficient function and coefficient function with local sparsity are taken into account. The detailed settings are as follows:

- Gaussian cases: The intercept function is set as $\beta_0(t) = \cos(2\pi t)$, for $t \in [0, 1]$. For the nonsparse setting, the coefficient function $\beta_1(t) = \sin(2\pi t)$, and for the sparse setting, $\beta_1(t) = 2 \cdot \{B_6(t) + B_7(t)\}$, where $B_l(t)$ is the lth B-spline basis on $[0, 1]$, with degree three and nine equally spaced interior knots. We generate the covariate functions in the same way as in Lin et al. (2017), that is, $X_i(t) = \sum_{l=1} a_{il} B_l^X(t)$, where a_{ij} is obtained from the standard normal distribution, and $B_l^X(t)$ is the lth B-spline basis on $[0, 1]$, with degree four and 69 equally spaced interior knots. The sample size is set as $n = 200$. Then, $Y_i(t)$ is generated from Gaussian distribution with mean $\beta_0(t) + \beta_1(t)X_i(t)$ and standard error one. To obtain asynchronous data, we generate the observation sizes of the response and the covariate independently from a Poisson distribution, with one additional observation to avoid cases with no measurement. Here, the response and the covariate share the same intensity rate m, and m is set as 15 and 20. Then, the observation times are uniformly selected on $[0, 1]$.

- Bernoulli cases: The settings are the same as those in the Gaussian cases, except that $Y_i(t)$ is generated from a Bernoulli distribution with mean $\beta_0(t) + \beta_1(t)X_i(t)$.

- Poisson cases: The settings are the same as those in the Gaussian cases, except that $Y_i(t)$ is generated from a Poisson distribution with mean $\beta_0(t) + \beta_1(t)X_i(t)$.

The proposed LocKer method is compared with other four approaches in the simulation. The first is a reconstruction method that synchronizes the response and the covariate using PACE (Yao, Müller and Wang (2005)), as in Şentürk et al. (2013), and then employs the traditional IRLS algorithm. We also consider the moment method of (Şentürk et al. (2013)), the approach in Cao, Zeng and Fine (2015), and the penalized least squares estimating (PLSE) method investigated by Tu, Park and Wang (2020). Note that Tu, Park and Wang (2020) investigated a local sparse estimator for a varying coefficient model with synchronous observations. Therefore, to implement their method for asynchronous cases, we first synchronize the data using smoothing, and then

apply the PLSE to the synchronized data. These four methods are denoted as Recon, Moment, Cao, and PLSE, respectively. Note that the Cao method is available for regression models with Bernoulli and Poisson responses, but is quite slow for these non-Gaussian cases, because it is a pointwise method. Hence, we use an identity link for the Cao method in all considered cases. Moreover, the PLSE is only applicable to a regression model with a Gaussian response; thus, we view the responses as Gaussian for the PLSE in all cases.

We evaluate the integrated square error (ISE) of the estimated intercept function and coefficient function for each method. Specifically,

$$\mathrm{ISE}_0 = \int_{\mathcal{T}} \{\widehat{\beta}_0(t) - \beta_0(t)\}^2 dt,$$
$$\mathrm{ISE}_1 = \int_{\mathcal{T}} \{\widehat{\beta}_1(t) - \beta_1(t)\}^2 dt.$$

In the simulation, 100 runs are conducted for each case. The average ISE and the standard deviation are compared between the methods.

Table 1 reports the average ISE_0 and ISE_1 of the Gaussian cases. With various settings for the coefficient function $\beta_1(t)$ and the observation rate m, the simulation results show similar trends. For the estimation of the intercept function $\beta_0(t)$, all five methods give promising results, with minor differences in ISE_0. On the other hand, it is evident that our LocKer method exhibits significant advantages for the estimation of $\beta_1(t)$, regardless of whether or not the true $\beta_1(t)$ is sparse. These results demonstrate that synchronizing and pointwise approaches are not adequate, further indicating the importance of using the observed data directly and taking sufficient account of the smoothness in the estimation. Moreover, the estimating results become more precise for each method as the observation rate increases.

Simulation results for the Bernoulli cases are presented in Table 2. The ISE_0 and ISE_1 are higher than the errors in the Gaussian cases, which implies that a Bernoulli response is more difficult to handle. However, the proposed LocKer method still outperforms the other four methods in terms of estimating $\beta_1(t)$ for both nonsparse and sparse settings, though the Recon and Moment methods are slightly better in terms of estimating $\beta_0(t)$. The reason for the invalid behavior of the Cao and PLSE methods is that they simply treat the Bernoulli response as a Gaussian response here. Table 3 displays the simulation results for the Poisson cases. We find that the proposed LocKer method achieves the most accurate estimates for both $\beta_0(t)$ and $\beta_1(t)$ in each considered setting.

In summary, our LocKer method yields encouraging estimation results for each case compared with those of the other methods. We conjecture that the superiority of our method is because we use an FDA approach and a kernel technique, as well as considering local sparsity.

Table 1. The average ISE_0 and ISE_1 across 100 runs for five methods in Gaussian cases, with the standard deviation in parentheses.

		$n = 200, m = 15$		$n = 200, m = 20$	
		ISE_0	ISE_1	ISE_0	ISE_1
Nonsparse	Recon	0.0050 (0.0022)	0.2768 (0.0505)	0.0044 (0.0019)	0.1889 (0.0455)
	Moment	0.0045 (0.0022)	0.4154 (0.1826)	0.0033 (0.0017)	0.4001 (0.0581)
	Cao	0.0072 (0.0031)	0.3000 (0.0326)	0.0059 (0.0028)	0.2841 (0.0344)
	PLSE	0.0244 (0.0106)	0.3994 (0.0839)	0.0145 (0.0066)	0.2966 (0.0998)
	LocKer	0.0170 (0.0081)	0.0385 (0.0255)	0.0094 (0.0062)	0.0217 (0.0148)
Sparse	Recon	0.0049 (0.0025)	0.2329 (0.0713)	0.0045 (0.0023)	0.1578 (0.0516)
	Moment	0.0052 (0.0059)	0.5350 (0.2588)	0.0033 (0.0016)	0.4972 (0.0648)
	Cao	0.0071 (0.0035)	0.3176 (0.0627)	0.0057 (0.0033)	0.3124 (0.0514)
	PLSE	0.0216 (0.0081)	0.3025 (0.0992)	0.0153 (0.0057)	0.2147 (0.0780)
	LocKer	0.0131 (0.0075)	0.0515 (0.0303)	0.0087 (0.0043)	0.0302 (0.0173)

Table 2. The average ISE_0 and ISE_1 across 100 runs for five methods in Bernoulli cases, with the standard deviation in parentheses.

		$n = 200, m = 15$		$n = 200, m = 20$	
		ISE_0	ISE_1	ISE_0	ISE_1
Nonsparse	Recon	0.0128 (0.0061)	0.3123 (0.0824)	0.0106 (0.0057)	0.2264 (0.0791)
	Moment	0.0171 (0.0085)	0.6108 (0.3848)	0.0131 (0.0064)	0.4744 (0.2760)
	Cao	0.5600 (0.0139)	0.4530 (0.0133)	0.5590 (0.0135)	0.4480 (0.0142)
	PLSE	0.5132 (0.0170)	0.4856 (0.0195)	0.5163 (0.0150)	0.4721 (0.0255)
	LocKer	0.0531 (0.0267)	0.1777 (0.0973)	0.0332 (0.0155)	0.1074 (0.0578)
Sparse	Recon	0.0182 (0.0075)	0.2898 (0.0966)	0.0172 (0.0067)	0.2444 (0.0892)
	Moment	0.0230 (0.0113)	0.6906 (0.3372)	0.0193 (0.0074)	0.5646 (0.1204)
	Cao	0.5751 (0.0150)	0.5259 (0.0148)	0.5753 (0.0113)	0.5239 (0.0140)
	PLSE	0.5272 (0.0175)	0.5490 (0.0301)	0.5311 (0.0119)	0.5381 (0.0331)
	LocKer	0.0426 (0.0235)	0.2600 (0.1094)	0.0291 (0.0147)	0.1773 (0.0805)

4.2. The effect of L

In Section 4.1, we focused on the accuracy of the estimation. In this section, we explore how the number of B-spline basis functions influences the estimation, especially the ability of the model to identify zero-valued subintervals of $\beta_1(t)$. Because local sparsity is also considered for the PLSE method, we include the PLSE in the comparison in this section. The settings are the same as those in Section 4.1, except that the response and the covariate are set to be observed at the same time to make the comparison with the PLSE more meaningful. To quantify the identifying ability, we compute the values of $\beta_1(t)$ and $\widehat{\beta}_1(t)$ at a sequence of dense grids on $[0, 1]$, and calculate the rates of the grids that correctly identified being zero and falsely estimated being zero, which are denoted by TP and FN, respectively. Moreover, the closer TP is to one and the closer FN is to

Table 3. The average ISE_0 and ISE_1 across 100 runs for five methods in Poisson cases, with the standard deviation in parentheses.

		$n = 200, m = 15$		$n = 200, m = 20$	
		ISE_0	ISE_1	ISE_0	ISE_1
Nonsparse	Recon	0.0257 (0.0078)	0.2789 (0.0573)	0.0234 (0.0064)	0.1929 (0.0437)
	Moment	0.0285 (0.0083)	0.3335 (0.1157)	0.0253 (0.0067)	0.3597 (0.0489)
	Cao	1.9949 (0.1056)	0.2645 (0.0378)	1.9772 (0.0794)	0.2496 (0.0371)
	PLSE	1.6426 (0.1044)	0.3555 (0.0909)	1.7170 (0.0942)	0.2408 (0.0773)
	LocKer	0.0163 (0.0103)	0.0345 (0.0186)	0.0096 (0.0069)	0.0192 (0.0128)
Sparse	Recon	0.0660 (0.0166)	0.2462 (0.0940)	0.0660 (0.0146)	0.1670 (0.0903)
	Moment	0.0730 (0.0234)	0.4579 (0.0962)	0.0745 (0.0175)	0.4791 (0.0647)
	Cao	1.8242 (0.0954)	0.4116 (0.0511)	1.8220 (0.0752)	0.3991 (0.0496)
	PLSE	1.4866 (0.0916)	0.4303 (0.1172)	1.5611 (0.0819)	0.3346 (0.1184)
	LocKer	0.0268 (0.0128)	0.0912 (0.0604)	0.0185 (0.0097)	0.0465 (0.0225)

zero, the better the identifying ability is.

Tables 4–5 list the simulation results with different values of L in the Gaussian cases. For the nonsparse settings, ISE_0 and ISE_1 of the proposed LocKer method decrease with an increase in L, and are better than those of the PLSE. Moreover, TP does not exist for nonsparse settings, so only FN is reported. Here, both methods achieve zero-valued FN, which means no grid is falsely identified, indicating that subintervals can be identified effectively for a coefficient function without local sparsity by both methods.

For the sparse settings, the estimation of $\beta_0(t)$ becomes better as L increases. However, both methods give the best estimation of $\beta_1(t)$ when $L = 13$. The reason is related to the setting of $\beta_1(t)$. Recall that to ensure local sparsity of $\beta_1(t)$, we use the B-spline basis with degree three and nine equally spaced interior knots in the setup. Therefore, the B-spline basis used in the setup is coincided with the B-spline basis applied in the estimation, yielding good performance of our method for $L = 13$. Except when $L = 13$, a larger value of L can yield a better estimation in terms of both accuracy and identifying ability. Compared with the PLSE, our method produces estimates that are more precise for $\beta_0(t)$, but ISE_1 is slightly higher than that of PLSE. However, for the identifying ability, the proposed LocKer method is much better than the PLSE in terms of both TP and FN, showing the advantage of our method in identifying zero-valued subintervals.

To sum up, although a larger value of L is beneficial for the identification in some general cases, more B-spline basis functions mean more parameters in the estimation, thus increasing the difficulty of the estimation. We discuss the results for the Bernoulli and Poisson cases in the Supplementary Material.

Table 4. The average ISE_0, ISE_1, TP, and FN across 100 runs for PLSE and LocKer using various values of L when $n = 200$ and $m = 15$ in Gaussian cases, with standard deviations in parentheses.

			ISE_0	ISE_1	TP	FN
$L = 10$	Nonsparse	PLSE	0.0120 (0.0054)	0.0196 (0.0064)	–	0 (0)
		LocKer	0.0115 (0.0039)	0.0139 (0.0048)	–	0 (0)
	Sparse	PLSE	0.0209 (0.0079)	0.0159 (0.0062)	0.1740 (0.2254)	0 (0)
		LocKer	0.0099 (0.0036)	0.0169 (0.0060)	0.5564 (0.1486)	0 (0)
$L = 13$	Nonsparse	PLSE	0.0123 (0.0049)	0.0189 (0.0060)	–	0 (0)
		LocKer	0.0077 (0.0029)	0.0115 (0.0054)	–	0 (0)
	Sparse	PLSE	0.0209 (0.0075)	0.0070 (0.0049)	0.6109 (0.3012)	0 (0)
		LocKer	0.0065 (0.0031)	0.0056 (0.0041)	0.9777 (0.0625)	0 (0)
$L = 15$	Nonsparse	PLSE	0.0093 (0.0038)	0.0186 (0.0064)	–	0 (0)
		LocKer	0.0063 (0.0025)	0.0095 (0.0054)	–	0 (0)
	Sparse	PLSE	0.0152 (0.0059)	0.0081 (0.0039)	0.3925 (0.2461)	0.0230 (0.0365)
		LocKer	0.0053 (0.0022)	0.0161 (0.0072)	0.8619 (0.0461)	0.0195 (0.0359)
$L = 20$	Nonsparse	PLSE	0.0093 (0.0039)	0.0204 (0.0062)	–	0 (0)
		LocKer	0.0047 (0.0021)	0.0076 (0.0055)	–	0 (0)
	Sparse	PLSE	0.0179 (0.0065)	0.0098 (0.0049)	0.5022 (0.2323)	0.0786 (0.0619)
		LocKer	0.0043 (0.0018)	0.0135 (0.0092)	0.9086 (0.0631)	0.0042 (0.0167)

Table 5. The average ISE_0, ISE_1, TP, and FN across 100 runs for PLSE and LocKer using various values of L when $n = 200$ and $m = 20$ in Gaussian cases, with standard deviations in parentheses.

			ISE_0	ISE_1	TP	FN
$L = 10$	Nonsparse	PLSE	0.0065 (0.0031)	0.0128 (0.0049)	–	0 (0)
		LocKer	0.0071 (0.0027)	0.0089 (0.0044)	–	0 (0)
	Sparse	PLSE	0.0128 (0.0057)	0.0136 (0.0051)	0.1621 (0.2108)	0 (0)
		LocKer	0.0061 (0.0027)	0.0159 (0.0045)	0.5587 (0.1517)	0 (0)
$L = 13$	Nonsparse	PLSE	0.0066 (0.0029)	0.0131 (0.0045)	–	0 (0)
		LocKer	0.0045 (0.0020)	0.0075 (0.0046)	–	0 (0)
	Sparse	PLSE	0.0143 (0.0046)	0.0050 (0.0033)	0.6009 (0.2522)	0 (0)
		LocKer	0.0038 (0.0016)	0.0049 (0.0034)	0.9838 (0.0542)	0 (0)
$L = 15$	Nonsparse	PLSE	0.0056 (0.0023)	0.0134 (0.0044)	–	0 (0)
		LocKer	0.0041 (0.0018)	0.0075 (0.0043)	–	0 (0)
	Sparse	PLSE	0.0092 (0.0035)	0.0064 (0.0035)	0.3104 (0.2266)	0.0126 (0.0261)
		LocKer	0.0033 (0.0014)	0.0096 (0.0038)	0.8654 (0.0613)	0.0241 (0.0345)
$L = 20$	Nonsparse	PLSE	0.0065 (0.0024)	0.0150 (0.0047)	–	0 (0)
		LocKer	0.0035 (0.0018)	0.0057 (0.0039)	–	0 (0)
	Sparse	PLSE	0.0128 (0.0049)	0.0078 (0.0033)	0.5393 (0.1997)	0.0748 (0.0582)
		LocKer	0.0028 (0.0015)	0.0073 (0.0033)	0.9484 (0.0242)	0.0116 (0.0268)

5. Real Data Analysis

Menopause in women is often accompanied by several physical changes. For example, follicle stimulating hormone (FSH) begins to increase in the

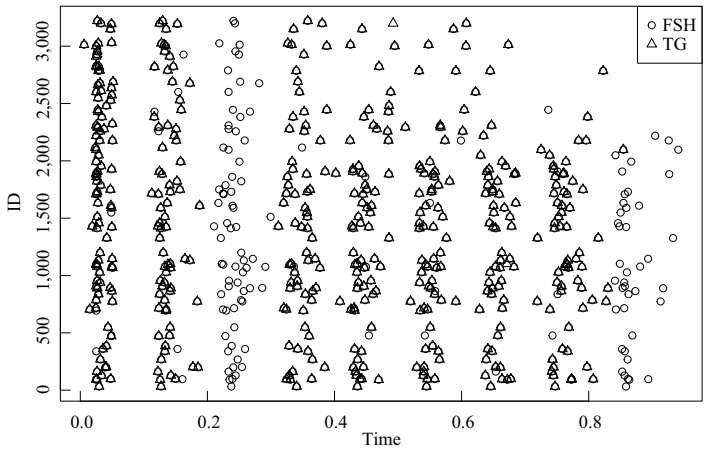

Figure 1. Observation times of FSH and TG for 100 randomly selected women.

perimenopausal stage (Wang et al. (2020)). Some studies showed that FSH has an influence on cardiovascular disease (CVD) risk (El Khoudary et al. (2016); Wang et al. (2017)). Serviente et al. (2019) propose that the association between FSH and CVD risk may be related to the effect of FSH on lipid levels. In this section, we aim to explore the relationship between FSH and triglycerides (TG), one of the lipid variables, using the proposed LocKer method.

The Study of Womens Health Across the Nation (SWAN) focuses on the health of women during their middle years. Between 1996 and 1997, 3302 women enrolled in this study, and 10 visits were conducted from 1997 to 2008. Moreover, both FSH and TG were recorded in this study and the data can be download from `https://www.swanstudy.org/`. Since TG was not measured in the last two visits, only the baseline and the first eight visits are taken into account in our analysis. Furthermore, we exclusively consider women who were early perimenopause or pre-menopausal at the baseline. Then, after removing individuals with no FSH or TG data, we have $n = 3224$ women in the study. Figure 1 displays the observation times of FSH and TG for 100 randomly selected women; note that the observation times are transformed to take values in $[0, 1]$. The figure shows that although some of the observation times for FSH and TG are the same, the asynchronous problem remains, particularly on $[0.2, 0.3]$ and $[0.8, 1]$, owing to the absence of TG records in the second and eighth visits.

We apply the LocKer method by treating FSH as the covariate and treating TG as the response. Both FSH and TG are centralized after being log-transformed. The roughness parameter and sparseness parameter are selected as introduced in Section 2.4. Figure 2 shows the estimated coefficient function using LocKer. Our results show a negative association between FSH and TG, which is consistent with the findings of Wang et al. (2020). Furthermore, additional

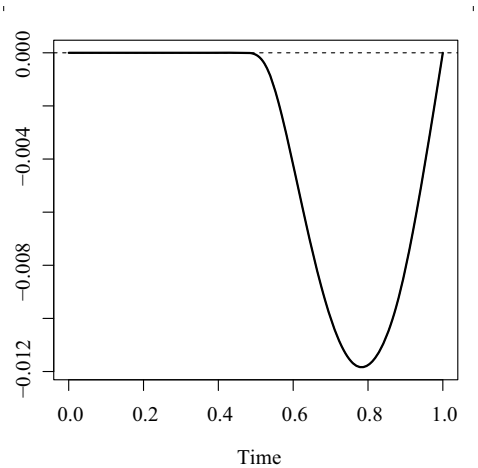

Figure 2. Estimate of the coefficient function obtained using the proposed LocKer method for the relationship between FSH and TG in women enrolled in the SWAN study.

findings can be achieved by local sparsity of our estimate. The estimate is zero-valued in the early stage, indicating that FSH has a minor effect on TG at the start of the menopausal transition. This effect begins to increase at about $t = 0.5$, and reaches a maximum at around $t = 0.8$, which implies a stronger relationship between FSH and TG in the later stage.

6. Conclusion

In this paper, we employ FDA method in the estimation of generalized varying coefficient model. Moreover, we use the kernel technique to solve the asynchronous problem, and impose a sparseness penalty to improve the accuracy and interpretability of the estimates. Our theoretical study verifies both the consistency and the sparsistency of the proposed LocKer method, and provides an asymptotic distribution of the estimator. The results of extensive simulation experiments and a practical application suggest that the LocKer method performs well.

However, we focus on the generalized varying coefficient model, which means only response and covariate values recorded at the same time are relevant. A more general model can be expressed as

$$E\{Y(t)|X(s), s \in \mathcal{T}\} = g\left\{\beta_0(t) + \int_{\mathcal{T}} X(s)\beta_1(s,t)ds\right\}, t \in \mathcal{T}.$$

In the above model, the response is related to the value of the covariate on the whole interval $\mathcal{T}$, rather than at one exact time point, which is more practical in real-world data sets. In future research, we will consider the asynchronous problem and local sparsity in this model in greater detail.

Supplementary Material

The online Supplementary Material contains proofs of Theorems 1–3, and some additional theoretical and simulation results.

Acknowledgments

The authors are grateful to the editor, associate editor, and two reviewers for their careful reading and constructive suggestions. This work was supported by Public Health & Disease Control and Prevention, Major Innovation & Planning Interdisciplinary Platform for the "Double-First Class" Initiative, Renmin University of China. This work was also supported by the Outstanding Innovative Talents Cultivation Funded Programs 2021 of Renmin University of China. C. Zhang's work was partially supported by U.S. National Science Foundation grants DMS-2013486 and DMS-1712418.

References

Cai, Z., Fan, J. and Li, R. (2000). Efficient estimation and inferences for varying-coefficient models. *Journal of the American Statistical Association* **95**, 888–902.

Cao, H., Li, J. and Fine, J. P. (2016). On last observation carried forward and asynchronous longitudinal regression analysis. *Electronic Journal of Statistics* **10**, 1155–1180.

Cao, H., Zeng, D. and Fine, J. P. (2015). Regression analysis of sparse asynchronous longitudinal data. *Journal of the Royal Statistical Society. Series B (Statistical Methodology)* **77**, 755–776.

Centofanti, F., Fontana, M., Lepore, A. and Vantini, S. (2020). Smooth LASSO estimator for the function-on-function linear regression model. *arXiv:2007.00529*.

Chen, J. and Chen, Z. (2008). Extended Bayesian information criteria for model selection with large model spaces. *Biometrika* **95**, 759–771.

Chen, J. and Chen, Z. (2012). Extended BIC for small-n-large-P sparse GLM. *Statistica Sinica* **22**, 555–574.

Chen, L. and Cao, H. (2017). Analysis of asynchronous longitudinal data with partially linear models. *Electronic Journal of Statistics* **11**, 1549–1569.

de Boor, C. (2001). *A Practical Guide to Splines*. Springer-Verlag, New York.

El Khoudary, S. R., Santoro, N., Chen, H.-Y., Tepper, P. G., Brooks, M. M., Thurston, R. C. et al. (2016). Trajectories of estradiol and follicle-stimulating hormone over the menopause transition and early markers of atherosclerosis after menopause. *European Journal of Preventive Cardiology* **23**, 694–703.

Fan, J. and Li, R. (2001). Variable selection via nonconcave penalized likelihood and its oracle properties. *Journal of the American Statistical Association* **96**, 1348–1360.

Fang, K., Zhang, X., Ma, S. and Zhang, Q. (2020). Smooth and locally sparse estimation for multiple-output functional linear regression. *Journal of Statistical Computation and Simulation* **90**, 341–354.

Hastie, T. and Tibshirani, R. (1993). Varying-coefficient models. *Journal of the Royal Statistical Society. Series B (Methodological)* **55**, 757–779.

Huang, J., Horowitz, J. L. and Wei, F. (2010). Variable selection in nonparametric additive models. *The Annals of Statistics* **38**, 2282–2313.

James, G. M., Wang, J. and Zhu, J. (2009). Functional linear regression that's interpretable. *The Annals of Statistics* **37**, 2083–2108.

Li, T., Li, T., Zhu, Z. and Zhu, H. (2022). Regression analysis of asynchronous longitudinal functional and scalar data. *Journal of the American Statistical Association* **117**, 1228–1242.

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.

Şentürk, D., Dalrymple, L. S., Mohammed, S. M., Kaysen, G. A. and Nguyen, D. V. (2013). Modeling time-varying effects with generalized and unsynchronized longitudinal data. *Statistics in Medicine* **32**, 2971–2987.

Serviente, C., Tuomainen, T.-P., Virtanen, J., Witkowski, S., Niskanen, L. and Bertone-Johnson, E. (2019). Follicle stimulating hormone is associated with lipids in postmenopausal women. *Menopause: The Journal of The North American Menopause Society* **26**, 540–545.

Sun, D., Zhao, H. and Sun, J. (2021). Regression analysis of asynchronous longitudinal data with informative observation processes. *Computational Statistics & Data Analysis* **157**, 107161.

Tu, C. Y., Park, J. and Wang, H. (2020). Estimation of functional sparsity in nonparametric varying coefficient models for longitudinal data analysis. *Statistica Sinica* **30**, 439–465.

Wang, N., Shao, H., Chen, Y., Xia, F., Chi, C., Li, Q. et al. (2017). Follicle-stimulating hormone, its association with cardiometabolic risk factors, and 10-year risk of cardiovascular disease in postmenopausal women. *Journal of the American Heart Association* **6**, e005918.

Wang, X., Zhang, H., Chen, Y., Du, Y., Jin, X. and Zhang, Z. (2020). Follicle stimulating hormone, its association with glucose and lipid metabolism during the menopausal transition. *Journal of Obstetrics and Gynaecology Research* **46**, 1419–1424.

Wang, Z., Magnotti, J., Beauchamp, M. S. and Li, M. (2020). Functional group bridge for simultaneous regression and support estimation. *arXiv:2006.10163*.

Xiong, X. and Dubin, J. A. (2010). A binning method for analyzing mixed longitudinal data measured at distinct time points. *Statistics in Medicine* **29**, 1919–1931.

Yao, F., Müller, H.-G. and Wang, J.-L. (2005). Functional data analysis for sparse longitudinal data. *Journal of the American Statistical Association* **100**, 577–590.

Zhong, R., Liu, S., Li, H. and Zhang, J. (2021). Sparse logistic functional principal component analysis for binary data. *arXiv:2109.08009*.

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.

Rou Zhong

Center for Applied Statistics, School of Statistics, Renmin University of China, Beijing 100872, China.

E-mail: zhong_rou@163.com

Chunming Zhang

Department of Statistics, University of Wisconsin-Madison, Madison, WI 53706, USA.

E-mail: cmzhang@stat.wisc.edu

Jingxiao Zhang

Center for Applied Statistics, School of Statistics, Renmin University of China, Beijing 100872, China.

E-mail: zhjxiaoruc@163.com

(Received June 2022; accepted January 2023)