Accounting for Time Series Errors in Partially Linear Model With Single- or Multiple-Runs

Chunming ZHANG, Yu HAN, and Shengji JIA

This article concerns statistical estimation of the partially linear model (PLM) for time course measurements, which are temporally correlated and allow multiple-runs for repeated measurements to enhance experimental accuracy without extending the number of time points within each trial. Such features arise naturally from biomedical data, for example, in brain fMRI, and call for special treatment beyond classical methods in either a purely nonparametric regression model or a PLM with independent errors. We develop a stepwise procedure for estimating the parametric and nonparametric components of the multiple-run PLM and making inference for parameters of interest, adaptive to either single- or multiple-run, in the presence of error temporal dependence. Simulation study and real fMRI data applications illustrate the computational simplicity and effectiveness of the proposed methods. Supplementary material for this article is available online.

Key Words: Autocorrelation matrix; Difference-based method; fMRI; Matrix inverse; Multiple testing; Semiparametric model.

1. INTRODUCTION

Semiparametric models, such as the partially linear model (PLM), play an important role in statistics, biostatistics, economics, and engineering studies (Andrews 1994; Yatchew 1997; Robinson 1988; Speckman 1988). The conventional PLM when applied to time-course responses $Y(t_i)$, observed at time points $t_i = i/n$, $i = 1, \dots, n$, and covariates $\boldsymbol{X}_i = (X_{i1}, \dots, X_{id})^T \in \mathbb{R}^d$, describes the model structure according to

$$Y(t_i) = \boldsymbol{X}_i^T \boldsymbol{\beta}_o + \eta_o(t_i) + \epsilon(t_i), \qquad E\{\epsilon(t_i) \mid \boldsymbol{X}_i\} = 0, \tag{1.1}$$

where $\boldsymbol{\beta}_o = (\beta_{1;o}, \dots, \beta_{d;o})^T$ is a vector of unknown parameters, $\eta_o(\cdot)$ is an unknown smooth function, and $\{\epsilon(t_i)\}$ are error terms. Denoting $\mathbf{y} = (Y(t_1), \dots, Y(t_n))^T$, $\mathbf{X} = (\boldsymbol{X}_1, \dots, \boldsymbol{X}_n)^T$, $\boldsymbol{\eta}_o = (\eta_o(t_1), \dots, \eta_o(t_n))^T$, and $\boldsymbol{\epsilon} = (\epsilon(t_1), \dots, \epsilon(t_n))^T$, model (1.1) is

Chunming Zhang, School of Mathematical Sciences, Nankai University, Tianjin 300071, China; Department of Statistics, University of Wisconsin, Madison, WI 53706 (E-mail: *cmzhang@stat.wisc.edu*). Yu Han, School of Science, Northeast Dianli University, Jilin, Jilin 132013, P.R. China (E-mail: *hanyu@mail.nedu.edu.cn*). Shengji Jia, Department of Statistics, University of Wisconsin, Madison, WI 53706 (E-mail: *shengji@stat.wisc.edu*).

Journal of Computational and Graphical Statistics, Volume 25, Number 1, Pages 123–143
DOI: 10.1080/10618600.2014.966107
Color versions of one or more of the figures in the article can be found online at *www.tandfonline.com/r/jcgs*.

rewritten as

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta}_{\mathrm{o}} + \boldsymbol{\eta}_{\mathrm{o}} + \boldsymbol{\epsilon}. \tag{1.2}$$

A two-step procedure (see, e.g., Fan and Huang 2005) is commonly used for estimating $\boldsymbol{\beta}_{\mathrm{o}}$ and $\boldsymbol{\eta}_{\mathrm{o}}$ by

$$\begin{aligned} \widehat{\boldsymbol{\beta}} &= \{\mathbf{X}^T(\mathbf{I}_n - S_b^T)(\mathbf{I}_n - S_b)\mathbf{X}\}^{-1}\{\mathbf{X}^T(\mathbf{I}_n - S_b^T)(\mathbf{I}_n - S_b)\mathbf{y}\}, \\ \widehat{\boldsymbol{\eta}} &= S_b(\mathbf{y} - \mathbf{X}\widehat{\boldsymbol{\beta}}), \end{aligned} \tag{1.3}$$

where $\mathbf{I}_n$ denotes an $n \times n$ identity matrix, $b > 0$ is the bandwidth parameter, and S_b is an $n \times n$ smoothing matrix (to be defined in Section 2). The classical PLM and estimation method have some limitations. First, to enhance statistical estimation efficiency of model (1.2), it is preferable to increase the number n of time points. But in some real applications, a smaller number of time points will be more feasible and advantageous to experimental outcomes. Second, the estimator (1.3) ignores the temporal correlation of $\boldsymbol{\epsilon}$ in many applications, that is, assumes $\mathrm{cov}(\boldsymbol{\epsilon}, \boldsymbol{\epsilon} \mid \mathbf{X}) = \sigma^2\mathbf{I}_n$. Third, the bandwidth parameter b is used in estimating both the parametric and nonparametric components, but most of existing work on data-driven selection of b, which is suitable to either a purely nonparametric regression model (i.e., $\mathbf{y} = \boldsymbol{\eta}_{\mathrm{o}} + \boldsymbol{\epsilon}$ in Hart 1991 and Xiao et al. 2003) or a PLM with independent errors, is not directly applicable to the PLM with time series errors. For example, the cross-validation method does not account for temporal autocorrelation directly and, furthermore, is computationally intensive. Some recent work on dimension reduction of the generalized additive PLM includes Lian et al. (2014) and references therein.

1.1 PLM With Time Series Errors Allowing Multiple-Run

This article concerns statistical estimation of the partially linear model (PLM) for time course measurements, which are temporally correlated and allow multiple-runs for repeated measurements to enhance experimental accuracy without extending the number n of time points within each trial. The multiple-run PLM can be described as follows:

$$Y_k(t_i) = \boldsymbol{X}_{k;i}^T\boldsymbol{\beta}_{\mathrm{o}} + \eta_{\mathrm{o};k}(t_i) + \epsilon_k(t_i), \quad i = 1, \ldots, n;\ k = 1, \ldots, \mathrm{Run},$$

that is,

$$\begin{pmatrix} \mathbf{y}_{\mathrm{run}\,1} \\ \vdots \\ \mathbf{y}_{\mathrm{run}\,\mathrm{Run}} \end{pmatrix} = \begin{pmatrix} \mathbf{X}_{\mathrm{run}\,1} \\ \vdots \\ \mathbf{X}_{\mathrm{run}\,\mathrm{Run}} \end{pmatrix}\boldsymbol{\beta}_{\mathrm{o}} + \begin{pmatrix} \boldsymbol{\eta}_{\mathrm{o;run}\,1} \\ \vdots \\ \boldsymbol{\eta}_{\mathrm{o;run}\,\mathrm{Run}} \end{pmatrix} + \begin{pmatrix} \boldsymbol{\epsilon}_{\mathrm{run}\,1} \\ \vdots \\ \boldsymbol{\epsilon}_{\mathrm{run}\,\mathrm{Run}} \end{pmatrix}, \tag{1.4}$$

where Run denotes the total number of runs, and for the kth run, $\mathbf{y}_{\mathrm{run}\,k} = (Y_k(t_1), \ldots, Y_k(t_n))^T$, $\mathbf{X}_{\mathrm{run}\,k} = (\boldsymbol{X}_{k;1}, \ldots, \boldsymbol{X}_{k;n})^T$, $\boldsymbol{\eta}_{\mathrm{o;run}\,k} = (\eta_{\mathrm{o};k}(t_1), \ldots, \eta_{\mathrm{o};k}(t_n))^T$ with an unknown smooth function $\eta_{\mathrm{o};k}(\cdot)$, and $\boldsymbol{\epsilon}_{\mathrm{run}\,k} = (\epsilon_k(t_1), \ldots, \epsilon_k(t_n))^T$. Measurements made across different runs are independent. Model (1.4) can also be rewritten as

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta}_{\mathrm{o}} + \boldsymbol{\eta}_{\mathrm{o}}^* + \boldsymbol{\epsilon}, \tag{1.5}$$

for which we assume that within each run k, $(\epsilon_k(t_1), \ldots, \epsilon_k(t_n)) \mid \mathbf{X}$ are stationary, and

$$E(\boldsymbol{\epsilon} \mid \mathbf{X}) = \mathbf{0}, \quad \mathrm{cov}(\boldsymbol{\epsilon}, \boldsymbol{\epsilon} \mid \mathbf{X}) = \boldsymbol{\Sigma} = \mathbf{I}_{\mathrm{Run}} \otimes \Sigma_n, \tag{1.6}$$

where $\mathbf{I}_k$ denotes a $k \times k$ identity matrix, $\otimes$ denotes the Kronecker product (Golub and Van Loan 1996). Here, $\Sigma_n = \sigma^2 R_n$, where $\sigma^2 = \text{var}\{\epsilon_k(t_i) \mid \mathbf{X}\}$, R_n is the autocorrelation matrix of $\boldsymbol{\epsilon}_{\text{run}\,k} \mid \mathbf{X}$, namely, $R_n(i, j) = \text{cov}(\epsilon_k(t_i), \epsilon_k(t_j) \mid \mathbf{X})/\sigma^2$, for each $k = 1, \ldots, \text{Run}$. Clearly, if Run $= 1$ with single-run, then model (1.4) reduces to the classical PLM (1.2). As a comparison, model (1.2) suits better with a larger number n of time points, whereas model (1.4) adapts to a smaller number n of time points but allows multiple-run. Such features of the multiple-run PLM arise naturally from biomedical data in, for example, brain fMRI, where $n = 185$ and Run $= 6$ in our real data in Section 5.

Typically, the parametric component $\boldsymbol{\beta}_{\text{o}}$ is of primary interest, while the nonparametric components $\{\eta_{\text{o};k}(\cdot)\}_{k=1}^{\text{Run}}$ serve as nuisance functions. An important application of the multiple-run PLM to brain fMRI data was introduced in Zhang and Yu (2008) for detecting activated brain voxels in response to external stimuli. There, $\boldsymbol{\beta}_{\text{o}}$ corresponds to the part of hemodynamic response function (HRF) values which is the object of primary interest to neuroscientists; $\{\eta_{\text{o};k}(\cdot)\}_{k=1}^{\text{Run}}$ are the slowly drifting baseline of time. Determining whether a voxel is activated or not can be formulated as testing for the linear form of hypotheses,

$$H_0 : A\boldsymbol{\beta}_{\text{o}} = \boldsymbol{g}_0 \qquad \text{versus} \qquad H_1 : A\boldsymbol{\beta}_{\text{o}} \neq \boldsymbol{g}_0, \tag{1.7}$$

where A is a given $\boldsymbol{k} \times d$ full row rank matrix, and $\boldsymbol{g}_0$ is a known $\boldsymbol{k} \times 1$ vector. To detect brain regions of activation, multiple testing procedures will be applied for large-scale simultaneous inference (Efron 2007, 2010).

Three issues are addressed in this article for the multiple-run PLM.

Issue 1: We develop a computationally effective procedure for estimating $\boldsymbol{\beta}_{\text{o}}$ and $\boldsymbol{\eta}_{\text{o}}^*$ in the multiple-run PLM (1.5) by $\widehat{\boldsymbol{\beta}}$ and $\widehat{\boldsymbol{\eta}}^*$, which incorporate the covariance matrix of stationary time series errors, without constraining the distribution of $\boldsymbol{\epsilon} \mid \mathbf{X}$ to be parametric or Gaussian.

Issue 2: We develop theoretical and empirical criterions for the selection of bandwidth b in estimating $\boldsymbol{\beta}_{\text{o}}$ and $\boldsymbol{\eta}_{\text{o}}^*$. For the former, we propose to minimize $\text{MSE}(\widehat{\boldsymbol{\beta}} \mid \mathbf{X})$ and derive its explicit form in Proposition 1, where MSE is the mean-squared error. For the latter, we propose and compare two criterions: one is based on minimizing the covariance penalty motivated from Efron (2004), and another is to minimize $\text{MSE}(\widehat{\boldsymbol{\eta}}^* \mid \mathbf{X})$, which is explicitly given in Proposition 2. It is interesting to note that all three criterions characterize the reliance on b and Σ_n, but do not depend on $\boldsymbol{\beta}_{\text{o}}$.

Issue 3: Simulation studies reveal that when testing an individual null hypothesis (1.7), ignoring the error correlation will yield inaccurate detection. The resulting test statistics, when testing multiple sequences of hypotheses (1.7) simultaneously, will lead to more false detection and thus increase the false discovery rate, even if test statistics under true null hypotheses are independent. In contrast, the proposed method which integrates the temporal correlation performs comparably well with the oracle method.

Computationally efficient algorithms play a vital role in statistical analysis of massive fMRI data. This article aims to develop a stepwise algorithm for estimating $\boldsymbol{\beta}_{\text{o}}$ and $\boldsymbol{\eta}_{\text{o}}^*$ in the multiple-run PLM, bringing computational simplicity, flexibility, and efficiency to model fitting and estimation. The validity and applicability of this full scheme of numerical tasks also rely on analytical justification (some of which given in other work) for statistical properties of procedures involved in the algorithm and during subsequent analyses. For example, Step 2 requires a consistent $n \times n$ matrix estimator $\widehat{R}_n$. Among other options, difference-based estimators $\widehat{R}_n$ and $\widehat{R}_n^{-1}$ are computationally transparent with explicit rates of stochastic convergence derived in Guo and Zhang (2013), which assumed that R_n is g_n-banded with data-driven selection of g_n. Likewise, performing a subsequent significance test, such as (1.7), after estimating $\boldsymbol{\beta}_{\text{o}}$ and $\boldsymbol{\eta}_{\text{o}}^*$ typically involves a test statistic which is asymptotically distribution free under the null hypothesis. A χ^2-type test statistic $\mathbb{K}_{bc}$ (given in (4.2)) was examined in Zhang and Yu (2011), which assumed that R_n is g-banded with $g = 2$.

The rest of the article proceeds as follows. Section 2 proposes a stepwise procedure for estimating $\boldsymbol{\beta}_{\text{o}}$ and $\boldsymbol{\eta}_{\text{o}}^*$ in the multiple-run PLM. Section 3 justifies the validity of each step. Section 4 presents numerical evaluation of the proposed method. Section 5 illustrates real fMRI data application.

2. PROPOSED METHODOLOGY FOR ESTIMATING $\boldsymbol{\beta}_{\text{O}}$ AND $\boldsymbol{\eta}^*_{\text{O}}$

We first introduce some necessary notation. Define $V_n = R_n^{-1}$, and $\mathbf{V} = \mathbf{I}_{\text{Run}} \otimes R_n^{-1} = \mathbf{I}_{\text{Run}} \otimes V_n$. For any matrix $\mathrm{A} \in \mathbb{R}^{n\times m}$, we use $\mathcal{D}_{\text{order}}\mathrm{A}$ or $\mathcal{D}_{\text{order}}(\mathrm{A})$ for an $(n$ -order$) \times m$ matrix, where $\mathcal{D}_{\text{order}}$ denotes the difference operator, with order $\in \{1, 2, \ldots\}$. For example, the commonly used first-order and second-order differences correspond to using

$$\mathcal{D}_1 = \begin{pmatrix} -1 & 1 & 0 & 0 & \cdots & 0 & 0 \\ 0 & -1 & 1 & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \cdots & -1 & 1 \end{pmatrix}_{(n-1)\times n};$$

$$\mathcal{D}_2 = \begin{pmatrix} 1 & -2 & 1 & 0 & \cdots & 0 & 0 & 0 \\ 0 & 1 & -2 & 1 & \cdots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \cdots & 1 & -2 & 1 \end{pmatrix}_{(n-2)\times n}.$$

Let S_b be an $n \times n$ smoothing matrix using nonparametric local-linear regression (Fan and Gijbels 1996), associated with the time points $\{t_1, \ldots, t_n\}$, with the (i, j)th entry equal to

$$S_b(i, j) = (1, 0)\{\mathbf{T}(t_i)^T \mathbf{W}(t_i)\mathbf{T}(t_i)\}^{-1}(1, t_j - t_i)^T K_b(t_j - t_i), \tag{2.1}$$

where $K(t)$ is a kernel function, $b > 0$ is a bandwidth parameter, $K_b(t) = K(t/b)/b$,

$$\mathbf{T}(t) = \begin{pmatrix} 1 & t_1 - t \\ \vdots & \vdots \\ 1 & t_n - t \end{pmatrix}, \text{ and } \mathbf{W}(t) = \mathrm{diag}\{K_b(t_1 - t), \ldots, K_b(t_n - t)\}.$$

Define $M_b = (\mathbf{I}_n - S_b^T)V_n(\mathbf{I}_n - S_b)$, and $\mathbf{A}_b = \mathbf{X}^T(\mathbf{I}_{\text{Run}} \otimes M_b)$. Unless otherwise stated, $\|\cdot\|$ denotes the L_2 norm of a vector.

Sections 2.1 and 2.2 will introduce the proposed method for estimating the parametric and nonparametric components. For illustrative simplicity, justification of each step will be given in Section 3.

2.1 PROPOSED METHODOLOGY FOR ESTIMATING $\boldsymbol{\beta}_{\mathrm{o}}$

Step 1: For the kth run, $k = 1, \ldots, \mathrm{Run}$, compute the lag-one differences $\mathcal{D}_1(\mathbf{y}_{\mathrm{run}\,k})$ and $\mathcal{D}_1(\mathbf{X}_{\mathrm{run}\,k})$. Combine all runs to obtain an initial estimator of $\boldsymbol{\beta}_{\mathrm{o}}$ via

$$\widehat{\boldsymbol{\beta}}_{\mathrm{init}} = \{\mathrm{diff}(\mathbf{X}, 1)^T \mathrm{diff}(\mathbf{X}, 1)\}^{-1} \mathrm{diff}(\mathbf{X}, 1)^T \mathrm{diff}(\mathbf{y}, 1), \tag{2.2}$$

where

$$\mathrm{diff}(\mathbf{X}, 1) = \begin{pmatrix} \mathcal{D}_1(\mathbf{X}_{\mathrm{run}\,1}) \\ \vdots \\ \mathcal{D}_1(\mathbf{X}_{\mathrm{run}\,\mathrm{Run}}) \end{pmatrix}, \quad \mathrm{diff}(\mathbf{y}, 1) = \begin{pmatrix} \mathcal{D}_1(\mathbf{y}_{\mathrm{run}\,1}) \\ \vdots \\ \mathcal{D}_1(\mathbf{y}_{\mathrm{run}\,\mathrm{Run}}) \end{pmatrix}.$$

Compute the vector of residuals, defined as $\mathbf{res}(\widehat{\boldsymbol{\beta}}_{\mathrm{init}})$, where

$$\mathbf{res}(\widehat{\boldsymbol{\beta}}) = \mathbf{y} - \mathbf{X}\widehat{\boldsymbol{\beta}} = \begin{pmatrix} \mathbf{res}_{\mathrm{run}\,1}(\widehat{\boldsymbol{\beta}}) \\ \vdots \\ \mathbf{res}_{\mathrm{run}\,\mathrm{Run}}(\widehat{\boldsymbol{\beta}}) \end{pmatrix}.$$

Step 2: For the kth run, $k = 1, \ldots, \mathrm{Run}$, compute $\widehat{\mathbf{e}}_{\mathrm{run}\,k}(\widehat{\boldsymbol{\beta}}_{\mathrm{init}}) = \mathcal{D}_2(\mathbf{res}_{\mathrm{run}\,k}(\widehat{\boldsymbol{\beta}}_{\mathrm{init}}))$. Based on $\{\widehat{\mathbf{e}}_{\mathrm{run}\,1}(\widehat{\boldsymbol{\beta}}_{\mathrm{init}}), \ldots, \widehat{\mathbf{e}}_{\mathrm{run}\,\mathrm{Run}}(\widehat{\boldsymbol{\beta}}_{\mathrm{init}})\}$ from all runs, obtain estimators $\widehat{\sigma}^2$, $\widehat{R}_n$ and $\widehat{V}_n$. Thus $\widehat{\Sigma}_n = \widehat{\sigma}^2 \widehat{R}_n$, and $\widehat{\mathbf{V}} = \mathbf{I}_{\mathrm{Run}} \otimes \widehat{V}_n$.

Step 3: For the kth run, $k = 1, \ldots, \mathrm{Run}$, note from (1.4) that $\mathbf{res}_{\mathrm{run}\,k}(\boldsymbol{\beta}_{\mathrm{o}}) \equiv \mathbf{y}_{\mathrm{run}\,k} - \mathbf{X}_{\mathrm{run}\,k}\boldsymbol{\beta}_{\mathrm{o}} = \boldsymbol{\eta}_{\mathrm{o;run}\,k} + \boldsymbol{\epsilon}_{\mathrm{run}\,k}$. This motivates us to estimate the nonparametric component $\boldsymbol{\eta}_{\mathrm{o;run}\,k}$ by $\widehat{\boldsymbol{\eta}}_{\mathrm{run}\,k}(\mathrm{b}_{1k}; \widehat{\boldsymbol{\beta}}_{\mathrm{init}})$, where

$$\widehat{\boldsymbol{\eta}}_{\mathrm{run}\,k}(b; \widehat{\boldsymbol{\beta}}) = S_b(\mathbf{y}_{\mathrm{run}\,k} - \mathbf{X}_{\mathrm{run}\,k}\widehat{\boldsymbol{\beta}}). \tag{2.3}$$

Here, the optimal bandwidth parameter b_{1k} in the smoothing matrix $S_{\mathrm{b}_{1k}}$ is chosen to minimize the covariance-penalty $\mathrm{cov\_pen}_{\mathrm{run}\,k}(b; \widehat{\boldsymbol{\beta}}_{\mathrm{init}})$ as a function of b, where

$$\mathrm{cov\_pen}_{\mathrm{run}\,k}(b; \widehat{\boldsymbol{\beta}}) = \|\mathbf{res}_{\mathrm{run}\,k}(\widehat{\boldsymbol{\beta}}) - S_b\, \mathbf{res}_{\mathrm{run}\,k}(\widehat{\boldsymbol{\beta}})\|^2 + 2\sigma^2\, \mathrm{tr}(S_b\, R_n). \tag{2.4}$$

For unknown σ^2 and R_n, their estimates from Step 2 will be used in $\mathrm{cov\_pen}_{\mathrm{run}\,k}(b; \widehat{\boldsymbol{\beta}}_{\mathrm{init}})$.

The vector of nonparametric components is then estimated by $\widehat{\boldsymbol{\eta}}^*(\mathrm{b}_{11}, \ldots, \mathrm{b}_{1\mathrm{Run}}; \widehat{\boldsymbol{\beta}}_{\mathrm{init}})$, where

$$\widehat{\boldsymbol{\eta}}^*(b_1, \ldots, b_{\mathrm{Run}}; \widehat{\boldsymbol{\beta}}) = (\widehat{\boldsymbol{\eta}}^T_{\mathrm{run}\,1}(b_1; \widehat{\boldsymbol{\beta}}), \ldots, \widehat{\boldsymbol{\eta}}^T_{\mathrm{run}\,\mathrm{Run}}(b_{\mathrm{Run}}; \widehat{\boldsymbol{\beta}}))^T. \tag{2.5}$$

Step 4: For the kth run, $k = 1, \ldots, \mathrm{Run}$, and any bandwidth parameter $b > 0$, notice from (1.4) that $(\mathbf{I}_n - S_b)\mathbf{y}_{\mathrm{run}\,k} = (\mathbf{I}_n - S_b)\mathbf{X}_{\mathrm{run}\,k}\boldsymbol{\beta}_{\mathrm{o}} + (\mathbf{I}_n - S_b)\boldsymbol{\eta}_{\mathrm{o;run}\,k} + (\mathbf{I}_n - S_b)\boldsymbol{\epsilon}_{\mathrm{run}\,k}$, or rewritten as

$$\widetilde{\mathbf{y}}_{\mathrm{run}\,k} = \widetilde{\mathbf{X}}_{\mathrm{run}\,k}\boldsymbol{\beta}_{\mathrm{o}} + \widetilde{\boldsymbol{\eta}}_{\mathrm{o;run}\,k} + \widetilde{\boldsymbol{\epsilon}}_{\mathrm{run}\,k}.$$

Thus

$$\begin{pmatrix} \widetilde{\mathbf{y}}_{\text{run }1} \\ \vdots \\ \widetilde{\mathbf{y}}_{\text{run Run}} \end{pmatrix} = \begin{pmatrix} \widetilde{\mathbf{X}}_{\text{run }1} \\ \vdots \\ \widetilde{\mathbf{X}}_{\text{run Run}} \end{pmatrix} \boldsymbol{\beta}_{\text{o}} + \begin{pmatrix} \widetilde{\boldsymbol{\eta}}_{\text{o;run }1} \\ \vdots \\ \widetilde{\boldsymbol{\eta}}_{\text{o;run Run}} \end{pmatrix} + \begin{pmatrix} \widetilde{\boldsymbol{\epsilon}}_{\text{run }1} \\ \vdots \\ \widetilde{\boldsymbol{\epsilon}}_{\text{run Run}} \end{pmatrix}, \tag{2.6}$$

namely,

$$\widetilde{\mathbf{y}} = \widetilde{\mathbf{X}}\boldsymbol{\beta}_{\text{o}} + \widetilde{\boldsymbol{\eta}}_{\text{o}}^{*} + \widetilde{\boldsymbol{\epsilon}}, \tag{2.7}$$

where $\widetilde{\mathbf{y}} = \{\mathbf{I}_{\text{Run}} \otimes (\mathbf{I}_n - S_b)\}\mathbf{y}$, $\widetilde{\mathbf{X}} = \{\mathbf{I}_{\text{Run}} \otimes (\mathbf{I}_n - S_b)\}\mathbf{X}$, $\widetilde{\boldsymbol{\eta}}_{\text{o}}^{*} = \{\mathbf{I}_{\text{Run}} \otimes (\mathbf{I}_n - S_b)\}\boldsymbol{\eta}_{\text{o}}^{*}$, and $\widetilde{\boldsymbol{\epsilon}} = \{\mathbf{I}_{\text{Run}} \otimes (\mathbf{I}_n - S_b)\}\boldsymbol{\epsilon}$. Then the parameter $\boldsymbol{\beta}_{\text{o}}$ is estimated by

$$\widehat{\boldsymbol{\beta}}(b) = \left(\widetilde{\mathbf{X}}^T\mathbf{V}\widetilde{\mathbf{X}}\right)^{-1}\widetilde{\mathbf{X}}^T\mathbf{V}\widetilde{\mathbf{y}} = (\mathbf{A}_b\mathbf{X})^{-1}\mathbf{A}_b\mathbf{y}. \tag{2.8}$$

The optimal bandwidth parameter b used in S_b for obtaining $\widehat{\boldsymbol{\beta}}(b)$ is chosen to minimize

$$\text{MSE}\{\widehat{\boldsymbol{\beta}}(b) \mid \mathbf{X}\};$$

see Proposition 1 for the explicit expression. Call such optimal parameter b_2. For unknown $(\sigma^2, R_n, V_n; \boldsymbol{\eta}_{\text{o}}^{*})$, their estimates from Steps 2–3 will be used in $\widehat{\boldsymbol{\beta}}(b)$ and $\text{MSE}\{\widehat{\boldsymbol{\beta}}(b) \mid \mathbf{X}\}$.

Step 5: Using b_2 chosen in Step 4, obtain the smoothing matrix S_{b_2} and estimate $\boldsymbol{\beta}_{\text{o}}$ by

$$\widehat{\boldsymbol{\beta}}(b_2).$$

This completes the procedure for estimating $\boldsymbol{\beta}_{\text{o}}$.

2.2 Proposed Methodology for Estimating η^*_{O}

Recall that the nonparametric estimator in Step 3 is rough, since $\widehat{\boldsymbol{\beta}}_{\text{init}}$ is some initial estimator of the parametric component. It is thus natural to improve the performance of the nonparametric estimator after obtaining a more efficient estimator of $\boldsymbol{\beta}_{\text{o}}$.

Step 6: For the kth run, estimate the nonparametric component $\boldsymbol{\eta}_{\text{o;run }k}$ in a way similar to that in Step 3, except that $\widehat{\boldsymbol{\beta}}_{\text{init}}$ is replaced by $\widehat{\boldsymbol{\beta}}(b_2)$ obtained from Step 5, leading to $\widehat{\boldsymbol{\eta}}_{\text{run }k}(b_{1k}; \widehat{\boldsymbol{\beta}}(b_2))$. The optimal choice of the bandwidth parameter b_{1k} minimizes $\text{cov\_pen}_{\text{run }k}(b; \widehat{\boldsymbol{\beta}}(b_2))$. Then the estimator of $\boldsymbol{\eta}_{\text{o}}^{*}$ is $\widehat{\boldsymbol{\eta}}^{*}(b_{11}, \ldots, b_{1\text{Run}}; \widehat{\boldsymbol{\beta}}(b_2))$.

We will make remarks on alternative approaches for estimating $\boldsymbol{\eta}_{\text{o;run }k}$. Options include, $\widehat{\boldsymbol{\eta}}_{\text{run }k}(b_{1k}; \widehat{\boldsymbol{\beta}}(b_2))$ or $\widehat{\boldsymbol{\eta}}_{\text{run }k}(b_{1k}; \widehat{\boldsymbol{\beta}}(b_{1k}))$, where b_{1k} minimizes

$$\text{MSE}\left\{\widehat{\boldsymbol{\eta}}_{\text{run }k}\left(b; \widehat{\boldsymbol{\beta}}(b_2)\right)\middle|\mathbf{X}\right\}, \text{ or } \text{MSE}\left\{\widehat{\boldsymbol{\eta}}_{\text{run }k}\left(b; \widehat{\boldsymbol{\beta}}(b)\right)\middle|\mathbf{X}\right\}, \tag{2.9}$$

respectively; see Proposition 2 for the closed-form expressions of (2.9). It can be shown that minimizing (2.9) with respect to b is asymptotically equivalent to minimizing (2.4). In finite-sample cases, the covariance penalty criterion (2.4) depends on (σ^2, R_n), thus gains computational simplicity and is applicable to both Step 3 and Step 6, whereas the MSE criterion (2.9) relies on $(\sigma^2, R_n, V_n; \boldsymbol{\eta}_{\text{o}}^{*})$, thus is computationally more involved and only applicable to Step 6.

For simulation studies, if we wish to obtain optimal constant bandwidths, one for estimating $\boldsymbol{\beta}_{\text{o}}$ and one for estimating $\boldsymbol{\eta}_{\text{o}}^{*}$, the criterions can be based on minimizing $\text{MSE}\{\widehat{\boldsymbol{\beta}}(b)\} =$

$E[\text{MSE}\{\widehat{\boldsymbol{\beta}}(b) \mid \mathbf{X}\}]$ for the former, and $\text{MSE}\{\widehat{\boldsymbol{\eta}}_{\text{run}\,k}(b; \widehat{\boldsymbol{\beta}}(b))\} = E[\text{MSE}\{\widehat{\boldsymbol{\eta}}_{\text{run}\,k}(b; \widehat{\boldsymbol{\beta}}(b)) \mid \mathbf{X}\}]$ for the latter, where the expectations can be approximated by averages across simulations.

2.3 Multiple-Run With Identical Nonparametric Components

For multiple-run with Run ≥ 2, if $\boldsymbol{\eta}_{\text{o;run}\,1} = \cdots = \boldsymbol{\eta}_{\text{o;run Run}} \equiv \boldsymbol{\eta}_{\text{o}}$, then the estimation in Step 3 and Step 6 can be further improved as follows. The estimator of $\boldsymbol{\eta}_{\text{o}}$ common to all runs is given by $\widehat{\boldsymbol{\eta}}(b; \widehat{\boldsymbol{\beta}}(b_2))$, where

$$\widehat{\boldsymbol{\eta}}(b; \widehat{\boldsymbol{\beta}}) = S_b\, \mathbf{res}_{\cdot}(\widehat{\boldsymbol{\beta}}), \tag{2.10}$$

with $\mathbf{res}_{\cdot}(\widehat{\boldsymbol{\beta}}) = \mathbf{y}_{\cdot} - \mathbf{X}_{\cdot}\widehat{\boldsymbol{\beta}}$, $\mathbf{y}_{\cdot} = (1/\text{Run}) \sum_{k=1}^{\text{Run}} \mathbf{y}_{\text{run}\,k} = (Y_{\cdot}(t_1), \ldots, Y_{\cdot}(t_n))^T$, and $\mathbf{X}_{\cdot} = (1/\text{Run}) \sum_{k=1}^{\text{Run}} \mathbf{X}_{\text{run}\,k}$. The optimal bandwidth b is chosen to minimize $\text{cov\_pen}(b; \widehat{\boldsymbol{\beta}}(b_2))$, where $\widehat{\boldsymbol{\beta}}(b_2)$ is from Step 5, and

$$\text{cov\_pen}(b; \widehat{\boldsymbol{\beta}}) = \left\| \mathbf{res}_{\cdot}(\widehat{\boldsymbol{\beta}}) - S_b\, \mathbf{res}_{\cdot}(\widehat{\boldsymbol{\beta}}) \right\|^2 + \frac{1}{\text{Run}} 2\sigma^2 \operatorname{tr}(S_b\, R_n). \tag{2.11}$$

Similarly, alternative approaches use either $\widehat{\boldsymbol{\eta}}(b; \widehat{\boldsymbol{\beta}}(b_2))$ or $\widehat{\boldsymbol{\eta}}(b; \widehat{\boldsymbol{\beta}}(b))$ to estimate $\boldsymbol{\eta}_{\text{o}}$ common to all runs. In that case, the optimal choice of b minimizes $\text{MSE}\{\widehat{\boldsymbol{\eta}}(b; \widehat{\boldsymbol{\beta}}(b_2)) \mid \mathbf{X}\}$ or $\text{MSE}\{\widehat{\boldsymbol{\eta}}(b; \widehat{\boldsymbol{\beta}}(b)) \mid \mathbf{X}\}$, respectively, where

$$\text{MSE}\{\widehat{\boldsymbol{\eta}}(b_1; \widehat{\boldsymbol{\beta}}(b)) \mid \mathbf{X}\} = I_3(b_1; b) + I_4(b_1; b), \tag{2.12}$$

with

$$\begin{aligned}
I_3(b_1; b) &= \left\| \left(\mathbf{I}_n - S_{b_1}\right)\boldsymbol{\eta}_{\text{o}} + S_{b_1}\, \mathbf{X}_{\cdot}(\mathbf{A}_b\mathbf{X})^{-1}\left(\mathbf{A}_b\boldsymbol{\eta}_{\text{o}}^*\right) \right\|^2, \\
I_4(b_1; b) &= \sigma^2 \operatorname{tr}\Big(S_{b_1}\, [(1/\text{Run})R_n - \mathbf{X}_{\cdot}(\mathbf{A}_b\mathbf{X})^{-1}(R_n M_b \mathbf{X}_{\cdot})^T - (R_n M_b \mathbf{X}_{\cdot})(\mathbf{A}_b\mathbf{X})^{-1}\mathbf{X}_{\cdot}^T \\
&\qquad + \mathbf{X}_{\cdot}(\mathbf{A}_b\mathbf{X})^{-1}\mathbf{X}^T\{\mathbf{I}_{\text{Run}} \otimes (M_b R_n M_b)\}\mathbf{X}(\mathbf{A}_b\mathbf{X})^{-1}\mathbf{X}_{\cdot}^T]S_{b_1}^T\Big).
\end{aligned}$$

Derivations of (2.10), (2.11), and (2.12) are given in the Appendix. For the real fMRI data in Section 5, which consists of six runs, there is little difference between options of assuming nonparametric components to be identical or not in estimating the parametric component.

3. JUSTIFICATION OF EACH STEP IN SECTION 2

As observed from Section 2, the implementation of $\widehat{\boldsymbol{\beta}}(b_2)$ in Step 5 hinges on estimating $(\sigma^2, R_n, V_n; \boldsymbol{\eta}_{\text{o}}^*)$ used in Step 4. Steps 1–2 together serve for estimating (σ^2, R_n, V_n), whereas Step 3 aims to estimate $\boldsymbol{\eta}_{\text{o}}^*$.

3.1 Least-Squares Estimation in Step 1

For the kth run, we observe from model (1.4) that

$$\begin{aligned}
Y_k(t_{i+1}) &= \boldsymbol{X}_{k;i+1}^T\boldsymbol{\beta}_{\text{o}} + \eta_{\text{o};k}(t_{i+1}) + \epsilon_k(t_{i+1}), \\
Y_k(t_i) &= \boldsymbol{X}_{k;i}^T\boldsymbol{\beta}_{\text{o}} + \eta_{\text{o};k}(t_i) + \epsilon_k(t_i).
\end{aligned}$$

The first-order difference thus gives

$$Y_k(t_{i+1}) - Y_k(t_i) = (\boldsymbol{X}_{k;i+1} - \boldsymbol{X}_{k;i})^T \boldsymbol{\beta}_{\text{o}} + O(n^{-1}) + \{\epsilon_k(t_{i+1}) - \epsilon_k(t_i)\},\ i = 1, \ldots, n-1,$$

namely,

$$\mathcal{D}_1(\mathbf{y}_{\text{run}\,k}) \approx \mathcal{D}_1(\mathbf{X}_{\text{run}\,k})\boldsymbol{\beta}_{\text{o}} + \mathcal{D}_1(\boldsymbol{\epsilon}_{\text{run}\,k}), \quad k = 1, \ldots, \text{Run},$$

provided that $\eta_{\text{o};k}(\cdot)$ has the bounded derivative. Hence, $\boldsymbol{\beta}_{\text{o}}$ can be estimated by an ordinary least-squares regression method applied to the set of differenced data from all runs.

3.2 Variance and Covariance Matrix Estimation in Step 2

Recall that within each run, the error covariance matrix $\Sigma_n = \sigma^2 R_n$ is an $n \times n$ matrix. Unlike most of existing approaches for estimating large covariance matrices (e.g., Bickel and Levina 2008a,b) in which the number of replicates diverges to infinity, the number of runs in our application is either 1 or at most finite. As an illustration, we will adopt the banded covariance matrix estimator in Guo and Zhang (2013), which has derived the explicit rate of convergence under a wide range of stationary time series error models, although other types of consistent estimators may also exist and perform well.

As for the order of difference, Fan and Zhang (2003) demonstrated that in diffusion models for financial time series data, a higher order difference will escalate the asymptotic variance of nonparametric function estimation. Thus, we adopt the second-order difference.

3.3 Bandwidth Selection Method in Step 3

Recall that in Step 3, for any kth run, $k = 1, \ldots, \text{Run}$, estimating $\boldsymbol{\eta}_{\text{o;run}\,k}$ corresponds to the nonparametric estimation of the signals in a signal plus noise model. More generally, we consider the nonparametric regression model,

$$Y_i = \mu(\boldsymbol{X}_i) + \epsilon_i, \quad i = 1, \ldots, n,$$

where $E(Y_i \mid \boldsymbol{X}_1, \ldots, \boldsymbol{X}_n) = \mu(\boldsymbol{X}_i) = \mu_i$ and $\text{cov}(\epsilon_i, \epsilon_j \mid \boldsymbol{X}_1, \ldots, \boldsymbol{X}_n) = \Sigma_n(i, j)$. It is thus natural to choose the bandwidth parameter which minimizes the prediction errors, by using the covariance penalty approach in Efron (2004).

For a future test sample $(\boldsymbol{X}_i^o, Y_i^o)$, which is an iid copy of $(\boldsymbol{X}_i, Y_i)$ in the set $\mathcal{T}_n = \{(\boldsymbol{X}_i, Y_i) : i = 1, \ldots, n\}$ of training samples, suppose that $\widehat{\mu}_i = \widehat{\mu}_i(\mathcal{T}_n)$ is the estimate of μ_i. The true (random) predictive error of using $\widehat{\mu}_i$ to predict Y_i^o is

$$\begin{aligned}
\text{Err}_i &= E\{(Y_i^o - \widehat{\mu}_i)^2 | \mathcal{T}_n\} \\
&= E[\{(Y_i^o - \mu_i) - (\widehat{\mu}_i - \mu_i)\}^2 | \mathcal{T}_n] \\
&= E\{(Y_i^o - \mu_i)^2 - 2(Y_i^o - \mu_i)(\widehat{\mu}_i - \mu_i) + (\widehat{\mu}_i - \mu_i)^2 | \mathcal{T}_n\} \\
&= E\{(Y_i^o - \mu_i)^2 + (\widehat{\mu}_i - \mu_i)^2 | \mathcal{T}_n\} \\
&= E\{(Y_i^o - \mu_i)^2 | \mathcal{T}_n\} + E\{(\widehat{\mu}_i - \mu_i)^2 | \mathcal{T}_n\} \\
&= E\{(Y_i^o - \mu_i)^2\} + (\widehat{\mu}_i - \mu_i)^2 = E\{(Y_i - \mu_i)^2\} + (\widehat{\mu}_i - \mu_i)^2, \qquad (3.1)
\end{aligned}$$

where we use the fact that

$$E\big\{\big(Y_i^o - \mu_i\big)\big(\widehat{\mu}_i - \mu_i\big)\big|\mathcal{T}_n\big\} = \big(\widehat{\mu}_i - \mu_i\big)E\big\{\big(Y_i^o - \mu_i\big)\big|\mathcal{T}_n\big\}$$
$$= \big(\widehat{\mu}_i - \mu_i\big)E\big(Y_i^o - \mu_i\big) = \big(\widehat{\mu}_i - \mu_i\big)E\big(Y_i - \mu_i\big) = 0.$$

Then (3.1) together with an identity $(y-\mu)^2 + (\widehat{\mu}-\mu)^2 = (y-\widehat{\mu})^2 + 2(y-\mu)(\widehat{\mu}-\mu)$ implies

$$\begin{aligned} E(\mathrm{Err}_i) &= E\{(Y_i-\mu_i)^2\} + E\{(\widehat{\mu}_i-\mu_i)^2\} \\ &= E\{(Y_i-\mu_i)^2 + (\widehat{\mu}_i-\mu_i)^2\} \\ &= E\{(Y_i-\widehat{\mu}_i)^2 + 2(Y_i-\mu_i)(\widehat{\mu}_i-\mu_i)\} \\ &= E(\mathrm{err}_i) + 2E\{(Y_i-\mu_i)(\widehat{\mu}_i-\mu_i)\} \\ &= E(\mathrm{err}_i) + 2E\{\mathrm{cov}(Y_i,\widehat{\mu}_i \mid \boldsymbol{X}_1,\ldots,\boldsymbol{X}_n)\}, \end{aligned}$$

where $\mathrm{err}_i = (Y_i - \widehat{\mu}_i)^2$. Define

$$\widehat{\mathrm{Err}}_i = \mathrm{err}_i + 2\widehat{\mathrm{cov}}\big(Y_i,\widehat{\mu}_i \mid \boldsymbol{X}_1,\ldots,\boldsymbol{X}_n\big)$$

to be the estimator of Err_i, where $\widehat{\mathrm{cov}}(Y_i,\widehat{\mu}_i \mid \boldsymbol{X}_1,\ldots,\boldsymbol{X}_n)$ is an estimator of $\mathrm{cov}(Y_i,\widehat{\mu}_i \mid \boldsymbol{X}_1,\ldots,\boldsymbol{X}_n)$ when it contains unobserved quantities. Then the total predictive error is estimated by

$$\begin{aligned} \sum_{i=1}^n \widehat{\mathrm{Err}}_i &= \sum_{i=1}^n \mathrm{err}_i + 2\sum_{i=1}^n \widehat{\mathrm{cov}}(Y_i,\widehat{\mu}_i \mid \boldsymbol{X}_1,\ldots,\boldsymbol{X}_n) \\ &= \sum_{i=1}^n (Y_i-\widehat{\mu}_i)^2 + 2\sum_{i=1}^n \widehat{\mathrm{cov}}(Y_i,\widehat{\mu}_i \mid \boldsymbol{X}_1,\ldots,\boldsymbol{X}_n) \\ &= \|\mathbf{y}-\widehat{\boldsymbol{\mu}}\|^2 + 2\sum_{i=1}^n \widehat{\mathrm{cov}}(Y_i,\widehat{\mu}_i \mid \boldsymbol{X}_1,\ldots,\boldsymbol{X}_n), \end{aligned} \tag{3.2}$$

where $\mathbf{y} = (Y_1,\ldots,Y_n)^T$ and $\widehat{\boldsymbol{\mu}} = (\widehat{\mu}_1,\ldots,\widehat{\mu}_n)^T$.

For any linear predictor $\widehat{\boldsymbol{\mu}}$, that is, $\widehat{\boldsymbol{\mu}} = M\mathbf{y}$, where entries $M(i,j)$ of M depend only on $(\boldsymbol{X}_1,\ldots,\boldsymbol{X}_n)$, but not on $(Y_1,\ldots,Y_n)$, we obtain

$$\begin{aligned} \mathrm{cov}(Y_i,\widehat{\mu}_i \mid \boldsymbol{X}_1,\ldots,\boldsymbol{X}_n) &= \mathrm{cov}\left(Y_i, \sum_{j=1}^n M(i,j)Y_j \,\middle|\, \boldsymbol{X}_1,\ldots,\boldsymbol{X}_n\right) \\ &= \sum_{j=1}^n M(i,j)\,\mathrm{cov}(Y_i,Y_j \mid \boldsymbol{X}_1,\ldots,\boldsymbol{X}_n) \\ &= \sum_{j=1}^n M(i,j)\Sigma_n(i,j) = \sum_{j=1}^n M(i,j)\Sigma_n(j,i) = (M\Sigma_n)(i,i), \end{aligned}$$

$$\sum_{i=1}^n \mathrm{cov}(Y_i,\widehat{\mu}_i \mid \boldsymbol{X}_1,\ldots,\boldsymbol{X}_n) = \sum_{i=1}^n (M\Sigma_n)(i,i) = \mathrm{tr}(M\Sigma_n).$$

In this case, (3.2) becomes the covariance-penalty,

$$\text{cov\_pen} = \left\|\mathbf{y} - \widehat{\boldsymbol{\mu}}\right\|^2 + 2\,\text{tr}(M\Sigma_n). \tag{3.3}$$

For Step 3, the vector $\mathbf{res}_{\text{run}\,k}(\boldsymbol{\beta}_{\text{o}})$ is retrieved by the nonparametric linear predictor $S_b\,\mathbf{res}_{\text{run}\,k}(\boldsymbol{\beta}_{\text{o}})$. Such correspondence applied to (3.3) leads to the covariance-penalty, $\|\mathbf{res}_{\text{run}\,k}(\boldsymbol{\beta}_{\text{o}}) - S_b\,\mathbf{res}_{\text{run}\,k}(\boldsymbol{\beta}_{\text{o}})\|^2 + 2\sigma^2\,\text{tr}(S_b\,R_n)$. This criterion, when $\boldsymbol{\beta}_{\text{o}}$ is estimated by an estimator $\widehat{\boldsymbol{\beta}}$, coincides with (2.4).

3.4 Bandwidth Selection Method in Step 4 and Step 5

Ideally, we wish to choose the bandwidth b to minimize $\text{SSE}(b) = \|\widehat{\boldsymbol{\beta}}(b) - \boldsymbol{\beta}_{\text{o}}\|^2$. But this is practically infeasible, since $\boldsymbol{\beta}_{\text{o}}$ is unknown. Proposition 1 indicates that $\text{MSE}\{\widehat{\boldsymbol{\beta}}(b) \mid \mathbf{X}\}$ is free of $\boldsymbol{\beta}_{\text{o}}$. This criterion will offer a practically more useful and effective approach.

Proposition 1. For $\widehat{\boldsymbol{\beta}}(b)$ in (2.8), and any $b > 0$, we obtain

$$\text{MSE}\{\widehat{\boldsymbol{\beta}}(b) \mid \mathbf{X}\} = \|(\mathbf{A}_b\mathbf{X})^{-1}(\mathbf{A}_b\boldsymbol{\eta}_{\text{o}}^*)\|^2 + \sigma^2\,\text{tr}[(\mathbf{A}_b\mathbf{X})^{-1}\mathbf{X}^T\{\mathbf{I}_{\text{Run}} \otimes (M_b R_n M_b)\}\mathbf{X}(\mathbf{A}_b\mathbf{X})^{-1}],$$

where

$$\begin{aligned}
M_b &= \left(\mathbf{I}_n - S_b^T\right)V_n(\mathbf{I}_n - S_b) \in \mathbb{R}^{n\times n},\\
\mathbf{A}_b &= \mathbf{X}^T(\mathbf{I}_{\text{Run}} \otimes M_b),\\
\mathbf{A}_b\mathbf{X} &= \sum_{k=1}^{\text{Run}} \mathbf{X}_{\text{run}\,k}^T M_b \mathbf{X}_{\text{run}\,k} \in \mathbb{R}^{d\times d},\\
\mathbf{A}_b\boldsymbol{\eta}_{\text{o}}^* &= \sum_{k=1}^{\text{Run}} \mathbf{X}_{\text{run}\,k}^T M_b \boldsymbol{\eta}_{\text{o;run}\,k} \in \mathbb{R}^{d\times 1},\\
\mathbf{X}^T\{\mathbf{I}_{\text{Run}} \otimes (M_b R_n M_b)\}\mathbf{X} &= \sum_{k=1}^{\text{Run}} \mathbf{X}_{\text{run}\,k}^T M_b R_n M_b \mathbf{X}_{\text{run}\,k} \in \mathbb{R}^{d\times d}.
\end{aligned}$$

In practice, the true values of σ^2, R_n, V_n, and $\boldsymbol{\eta}_{\text{o}}^*$ in Proposition 1 are unknown and need to be estimated. This motivates us to substitute $(\sigma^2, R_n, V_n; \boldsymbol{\eta}_{\text{o}}^*)$ by their estimates $(\widehat{\sigma}^2, \widehat{R}_n, \widehat{V}_n, \widehat{\boldsymbol{\eta}}^*)$ to form $\widehat{\text{MSE}}(\widehat{\boldsymbol{\beta}} \mid \mathbf{X})$, an empirical plug-in estimate of $\text{MSE}(\widehat{\boldsymbol{\beta}} \mid \mathbf{X})$, and then select the bandwidth parameter which minimizes $\widehat{\text{MSE}}(\widehat{\boldsymbol{\beta}} \mid \mathbf{X})$. Among the four estimates, $(\widehat{\sigma}^2, \widehat{R}_n, \widehat{V}_n)$ are obtained from Step 2; $\widehat{\boldsymbol{\eta}}^*$ is obtained from Step 3.

3.5 Bandwidth Selection Method in Step 6

Similar to Proposition 1, Proposition 2 states that the MSE of nonparametric estimators is free of $\boldsymbol{\beta}_{\mathrm{o}}$.

Proposition 2. For $\widehat{\boldsymbol{\beta}}(b)$ in (2.8), $\widehat{\boldsymbol{\eta}}^*(b_1, \ldots, b_{\mathrm{Run}}; \widehat{\boldsymbol{\beta}})$ in (2.5), and any $b_1 > 0, \ldots, b_{\mathrm{Run}} > 0$ and $b > 0$, we obtain

$$\mathrm{MSE}\{\widehat{\boldsymbol{\eta}}^*(b_1, \ldots, b_{\mathrm{Run}}; \widehat{\boldsymbol{\beta}}(b)) | \mathbf{X}\} = \sum_{k=1}^{\mathrm{Run}} \mathrm{MSE}\{\widehat{\boldsymbol{\eta}}_{\mathrm{run}\,k}(b_k; \widehat{\boldsymbol{\beta}}(b)) | \mathbf{X}\},$$

$$\mathrm{MSE}\{\widehat{\boldsymbol{\eta}}_{\mathrm{run}\,k}(b_k; \widehat{\boldsymbol{\beta}}(b)) | \mathbf{X}\} = I_{3;k}(b_k; b) + I_{4;k}(b_k; b),$$

where

$$\begin{aligned}
I_{3;k}(b_k; b) &= \left\| (\mathbf{I}_n - S_{b_k}) \boldsymbol{\eta}_{\mathrm{o;run}\,k} + S_{b_k} \mathbf{X}_{\mathrm{run}\,k} (\mathbf{A}_b \mathbf{X})^{-1} (\mathbf{A}_b \boldsymbol{\eta}_{\mathrm{o}}^*) \right\|^2, \\
I_{4;k}(b_k; b) &= \sigma^2 \operatorname{tr}\big(S_{b_k} \big[R_n - \mathbf{X}_{\mathrm{run}\,k} (\mathbf{A}_b \mathbf{X})^{-1} (R_n M_b \mathbf{X}_{\mathrm{run}\,k})^T - (R_n M_b \mathbf{X}_{\mathrm{run}\,k}) (\mathbf{A}_b \mathbf{X})^{-1} \mathbf{X}_{\mathrm{run}\,k}^T \\
&\quad + \mathbf{X}_{\mathrm{run}\,k} (\mathbf{A}_b \mathbf{X})^{-1} \mathbf{X}^T \{\mathbf{I}_{\mathrm{Run}} \otimes (M_b R_n M_b)\} \mathbf{X} (\mathbf{A}_b \mathbf{X})^{-1} \mathbf{X}_{\mathrm{run}\,k}^T \big] S_{b_k}^T \big),
\end{aligned}$$

and M_b and $\mathbf{A}_b$ are as defined in Proposition 1.

4. SIMULATION STUDY

We conduct simulation studies to evaluate the stepwise estimation procedure in Section 2 for the multiple-run PLM, with Run $= 2$ and $n = 300$. The data-generating process mimics that from the fMRI experiment (Glover 1999). The true parameter vector $\boldsymbol{\beta}_{\mathrm{o}} = (\beta_{\mathrm{o};1}, \ldots, \beta_{\mathrm{o};d})^T$ is quantified by

$$\beta_{\mathrm{o};j} = \left\{ \frac{g_1(1.5(j-1) - t_s)}{g_1(a_1 b_1)} - c \frac{g_2(1.5(j-1) - t_s)}{g_2(a_2 b_2)} \right\} \mathrm{I}(1.5(j-1) - t_s > 0), \tag{4.1}$$

for $j = 1, \ldots, d$, where $g_1(t) = t^{a_1} \exp(-t/b_1)$ and $g_2(t) = t^{a_2} \exp(-t/b_2)$, with $d = 20$, $a_1 = 5$, $b_1 = 0.9$, $a_2 = 12$, $b_2 = 0.7$, $c = 0.4$, $t_s = 5.5$ and $\mathrm{I}(\cdot)$ is an indicator function. Within the kth run,

$$\mathbf{X}_{\mathrm{run}\,k} = \begin{pmatrix}
s_{k;1}(0) & 0 & \cdots & 0 \\
s_{k;1}(t_2 - t_1) & s_{k;1}(0) & \cdots & 0 \\
\vdots & \vdots & \ddots & \vdots \\
s_{k;1}(t_d - t_1) & s_{k;1}(t_d - t_2) & \cdots & s_{k;1}(0) \\
\vdots & \vdots & \cdots & \vdots \\
s_{k;1}(t_n - t_1) & s_{k;1}(t_n - t_2) & \cdots & s_{k;1}(t_n - t_d)
\end{pmatrix},$$

where $\{s_{k;1}(t_i)\}_{i=1}^n \overset{\mathrm{iid}}{\sim} \mathrm{Bernoulli}(1/2)$, and the nonparametric function $\eta_{\mathrm{o};k}(t_i) = 10 \sin\{\pi(t_i - 0.21)\}$, at $t_i = i/n$. The errors $\{\epsilon_k(t_i)\}_{i=1}^n$ are generated from the fourth-order moving average (MA(4)) process (Cryer and Chan 2008),

$$\epsilon_k(t_i) = z(t_i) + \theta_1 z(t_{i-1}) + \theta_2 z(t_{i-2}) + \theta_3 z(t_{i-3}) + \theta_4 z(t_{i-4}),$$

where $\theta_1 = 0.75$, $\theta_2 = 0.5$, $\theta_3 = 0.25$, and $\theta_4 = 0.35$; $\{z(t_i)\} \overset{\text{iid}}{\sim} N(0, \sigma_Z^2)$; $\sigma_Z^2 = 0.4786^2/1$ and $0.4786^2/8$ for noise levels ranging from large to small. Local-linear regression method combined with Epanechnikov kernel (Silverman 1986) supported on $[-1, 1]$ is used for nonparametric estimation. Unless otherwise stated, Monte Carlo simulations are replicated for 500 times.

4.1 Estimation of Parametric Component

First, we assess the performance of a parametric estimator $\widehat{\boldsymbol{\beta}}$ by $\text{SSE}(\widehat{\boldsymbol{\beta}}) = \|\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_{\text{o}}\|^2$. Comparison is made between the following methods for $\widehat{\boldsymbol{\beta}}$.

"iid-V-b-0.1": $\widehat{\boldsymbol{\beta}}(b)$ given in (2.8) with $\mathbf{V}$ incorrectly set to $\mathbf{I}_{\text{Run}} \otimes \mathbf{I}_n$ and $b = 0.1$;

"iid-V-b-0.5": $\widehat{\boldsymbol{\beta}}(b)$ given in (2.8) with $\mathbf{V}$ incorrectly set to $\mathbf{I}_{\text{Run}} \otimes \mathbf{I}_n$ and $b = 0.5$;

"iid-V-b-dd": $\widehat{\boldsymbol{\beta}}(b)$ given in (2.8) with $\mathbf{V}$ incorrectly set to $\mathbf{I}_{\text{Run}} \otimes \mathbf{I}_n$ and b chosen via data-driven method to minimize $\widehat{\text{MSE}}\{\widehat{\boldsymbol{\beta}}(b) \mid \mathbf{X}\}$, which replaces $(\sigma^2, R_n, V_n; \boldsymbol{\eta}_{\text{o}}^*)$ in $\text{MSE}\{\widehat{\boldsymbol{\beta}}(b) \mid \mathbf{X}\}$ by setting $V_n = \mathbf{I}_n$ and $R_n = \mathbf{I}_n$ and estimators $(\widehat{\sigma}^2; \widehat{\boldsymbol{\eta}}^*)$ in Steps 1–3;

"est-V-b-0.1": $\widehat{\boldsymbol{\beta}}(b)$ given in (2.8) with $\mathbf{V}$ estimated by $\widehat{\mathbf{V}}$ and $b = 0.1$;

"est-V-b-0.5": $\widehat{\boldsymbol{\beta}}(b)$ given in (2.8) with $\mathbf{V}$ estimated by $\widehat{\mathbf{V}}$ and $b = 0.5$;

"full-data-driven": $\widehat{\boldsymbol{\beta}}(b)$ given in (2.8) with $\mathbf{V}$ estimated by $\widehat{\mathbf{V}}$ and b chosen via data-driven method to minimize $\widehat{\text{MSE}}\{\widehat{\boldsymbol{\beta}}(b) \mid \mathbf{X}\}$, which replaces $(\sigma^2, R_n, V_n; \boldsymbol{\eta}_{\text{o}}^*)$ in $\text{MSE}\{\widehat{\boldsymbol{\beta}}(b) \mid \mathbf{X}\}$ by their estimators in Steps 1–3;

"oracle-random": $\widehat{\boldsymbol{\beta}}(b)$ given in (2.8) with true $\mathbf{V}$ and b chosen to minimize $\text{MSE}\{\widehat{\boldsymbol{\beta}}(b) \mid \mathbf{X}\}$;

"oracle-constant": $\widehat{\boldsymbol{\beta}}(b)$ given in (2.8) with true $\mathbf{V}$ and (constant) b chosen to minimize $\text{MSE}\{\widehat{\boldsymbol{\beta}}(b)\}$, approximated from 100 sets of simulated samples.

Boxplots of SSE using different methods are compared in Figure 1; Figure 2 compares boxplots of estimated parameters resulting from those methods. We observe the following aspects. (i) For a fixed bandwidth, ignoring the correlation structure will tend to degrade the estimation of parametric component, compared with that of incorporating the correlation. See "iid-V-b-0.1" versus "est-V-b-0.1," and "iid-V-b-0.5" versus "est-V-b-0.5." (ii) A smaller bandwidth will tend to reduce SSE, whereas an excessively large bandwidth has an adverse affect on the parametric estimator; the data-driven counterpart is preferable. This is seen from "iid-V-b-0.1" versus "iid-V-b-0.5" versus "iid-V-b-dd," and "est-V-b-0.1" versus "est-V-b-0.5" versus "full-data-driven." (iii) For the two versions of oracle methods, the random choice of bandwidth performs similarly to the constant choice, but none of them can be practically implemented. (iv) The "full-data-driven" procedure compares well with the oracle counterpart, but is practically feasible, and is thus recommended for real data application.

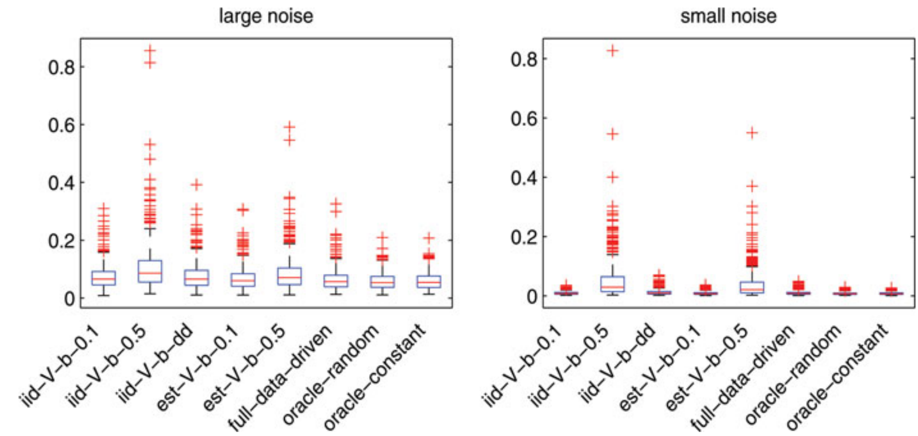

Figure 1. Boxplots of SSE($\widehat{\boldsymbol{\beta}}$) using different methods for $\widehat{\boldsymbol{\beta}}$, indicated below each boxplot. Left panel: with large noise level; right panel: with small noise level.

Second, we examine the null distribution of test statistics for testing $H_0 : \boldsymbol{\beta}_{\mathrm{o}} = \mathbf{0}$. We adopt the semiparametric test statistic $\mathbb{K}_{bc}$ (given in Zhang and Yu (2008) for testing $H_0 : A\boldsymbol{\beta}_{\mathrm{o}} = \boldsymbol{g}_0$) in the form,

$$\frac{\left(A\widehat{\boldsymbol{\beta}}_* - \boldsymbol{g}_0\right)^T \left\{A\left(\widetilde{\mathbf{X}}^T \mathbf{V} \widetilde{\mathbf{X}}\right)^{-1} A^T\right\}^{-1} \left(A\widehat{\boldsymbol{\beta}}_* - \boldsymbol{g}_0\right)}{\widehat{\mathbf{r}}_*^T \mathbf{V} \widehat{\mathbf{r}}_* \big/ (\mathrm{Run} \times n - d)}, \tag{4.2}$$

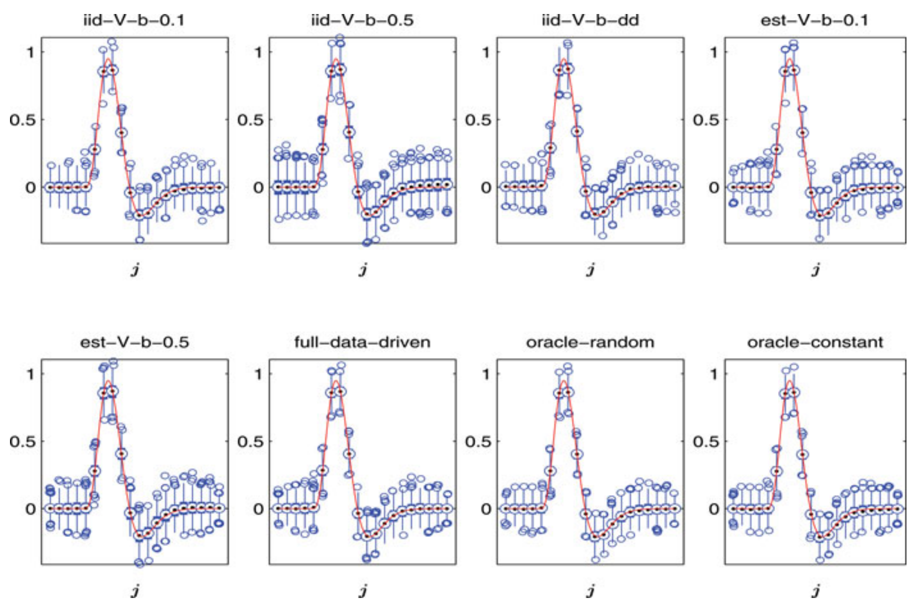

Figure 2. Boxplots of $\widehat{\beta}_j$, $j = 1, \ldots, d$ (from left to right within each panel), using different methods for $\widehat{\boldsymbol{\beta}}$, with large noise level. Solid curves connect true values of parameters.

by setting $A = \mathbf{I}_d$ and $\boldsymbol{g}_0 = \mathbf{0}$, where $\mathbb{K}_{bc}$ corresponds to using $\widehat{\boldsymbol{\beta}}_* = \widehat{\boldsymbol{\beta}}_{bc}(\widehat{\boldsymbol{\beta}}(b))$ and $\widehat{\mathbf{r}}_* = \widehat{\mathbf{r}}_{bc}(\widehat{\boldsymbol{\beta}}(b))$, with

$$\widehat{\boldsymbol{\beta}}_{bc}(\widehat{\boldsymbol{\beta}}) = \widehat{\boldsymbol{\beta}}(b) - (\mathbf{A}_b\mathbf{X})^{-1}\left\{\sum_{k=1}^{\text{Run}} \mathbf{X}_{\text{run}\,k}^T M_b \widehat{\boldsymbol{\eta}}_{\text{run}\,k}(b;\widehat{\boldsymbol{\beta}})\right\},$$

$$\widehat{\mathbf{r}}_{bc}(\widehat{\boldsymbol{\beta}}) = \{\mathbf{I}_{\text{Run}} \otimes (\mathbf{I}_n - S_b)\}\widehat{\mathbf{r}}(\widehat{\boldsymbol{\beta}}).$$

Zhang an Yu (2008) showed that under H_0, $\mathbb{K}_{bc} \xrightarrow{\mathcal{D}} \chi^2_d$. Figure 3 compares the cubic root transform of empirical (1st to 99th) percentiles of $\mathbb{K}_{bc}$, based on parametric estimators $\widehat{\boldsymbol{\beta}}(b)$ and associated $\mathbf{V}$ described at the start of Section 4.1, versus cubic root transform of percentiles of the χ^2_d distribution, where the data are generated as before except that $\boldsymbol{\beta}_{\text{o}} = \mathbf{0}$ is set in accordance with H_0. We observe from "iid-V-b-0.1" and "iid-V-b-0.5" that the sampling null distributions of test statistics, when ignoring the correlation structure, depart substantially from the χ^2_d distribution, causing the testing procedures to detect many false significance (and thus inaccurate detection of brain activity in the fMRI data). The data-driven counterpart "iid-V-b-dd" does not ameliorate the discrepancy either. Among the methods that do incorporate the error correlation structure, both "est-V-b-0.1" and "est-V-b-0.5" improve but specify the bandwidths in an ad hoc way. In contrast, the "full-data-driven" method is genuinely practical and compares reasonably well with the two oracle counterparts.

Third, we evaluate the impact of ignoring error correlation on testing multiple sets of null hypotheses $H_0 : \boldsymbol{\beta}_{\text{o}} = \mathbf{0}$ simultaneously. As an illustration, the Benjamini-Hochberg multiple testing procedure (Benjamini and Hochberg 1995) is adopted to determine the threshold of p-values calculated from test statistics $\mathbb{K}_{bc}$. We generate 2000 sets of data

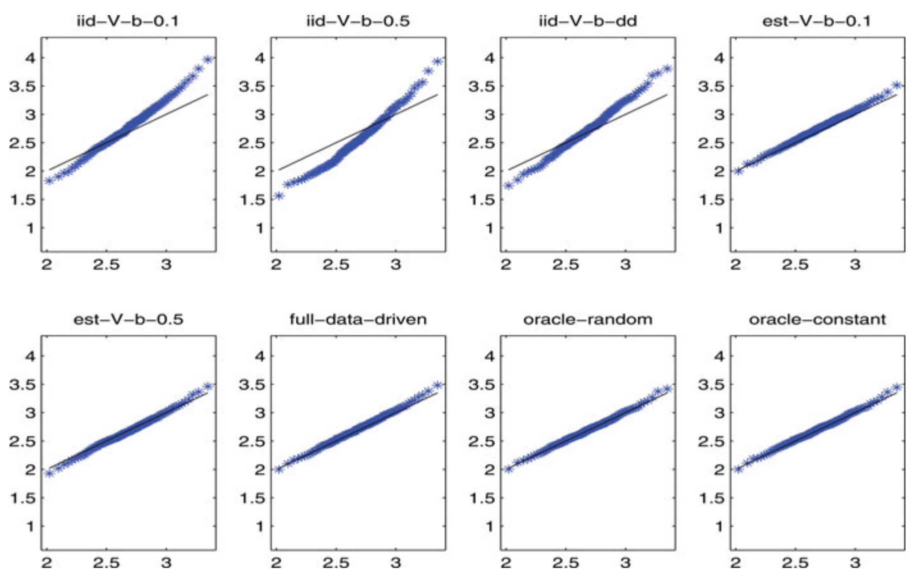

Figure 3. Cubic root transform of empirical percentiles (on the y axis) of $\mathbb{K}_{bc}$ versus cubic root transform of percentiles (on the x axis) of the χ^2 distribution, with large noise level. Solid line: the 45-degree reference line.

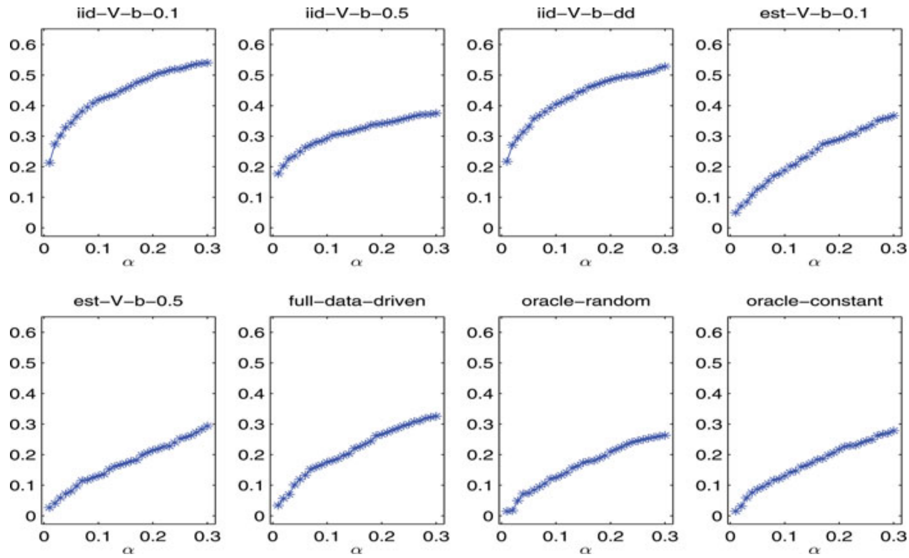

Figure 4. Compare the calculated FDP of $\mathbb{K}_{bc}$, with large noise level.

independently according to the multiple-run PLM as before, where 20% of them are associated with an alternative hypothesis ($\boldsymbol{\beta}_{\rm o}$ in (4.1)) and the remaining 80% are null ($\boldsymbol{\beta}_{\rm o} = \mathbf{0}$). The calculated false discovery proportion (FDP) are given in Figure 4, where the control level α varies from 0.01 to 0.30 in increments of 0.01. Clearly, all three methods "iid-V-b-0.1," "iid-V-b-0.5" and "iid-V-b-dd," which ignore the error correlation structure, fail to control the FDR at the desired control level. As seen, they inflate the FDP, particularly serious when α is small. The performance is improved by "est-V-b-0.1," "est-V-b-0.5" and "full-data-driven" which integrate the error correlation. Among them, the "full-data-driven" method compares well with two oracle counterparts.

4.2 Estimation of Nonparametric Component

As all runs share a common nonparametric component, we evaluate the nonparametric estimator $\widehat{\eta}(b; \widehat{\boldsymbol{\beta}})$ in (2.10). Similar comparison can be made without assuming nonparametric components to be identical. In general, the optimal bandwidths for estimating parametric and nonparametric components typically disagree, since the criterions for estimating parametric and nonparametric components do not match. For example, Figure 5 illustrates that the optimal bandwidth that minimizes MSE$\{\widehat{\boldsymbol{\beta}}(b)\}$ is typically larger than that minimizes MSE$\{\widehat{\eta}(b; \widehat{\boldsymbol{\beta}}(b))\}$. Estimators of the nonparametric component using 8 types of bandwidth parameters are compared below.

Method 1: $\widehat{\eta}(b; \widehat{\boldsymbol{\beta}}(b))$ in (2.10), where $\widehat{\boldsymbol{\beta}}(b)$ is given in (2.8) with $\mathbf{V}$ incorrectly set to $\mathbf{I}_{\rm Run} \otimes \mathbf{I}_n$ and $b = 0.1$;

Method 2: $\widehat{\eta}(b; \widehat{\boldsymbol{\beta}}(b))$ in (2.10), where $\widehat{\boldsymbol{\beta}}(b)$ is given in (2.8) with $\mathbf{V}$ incorrectly set to $\mathbf{I}_{\rm Run} \otimes \mathbf{I}_n$ and $b = 0.5$;

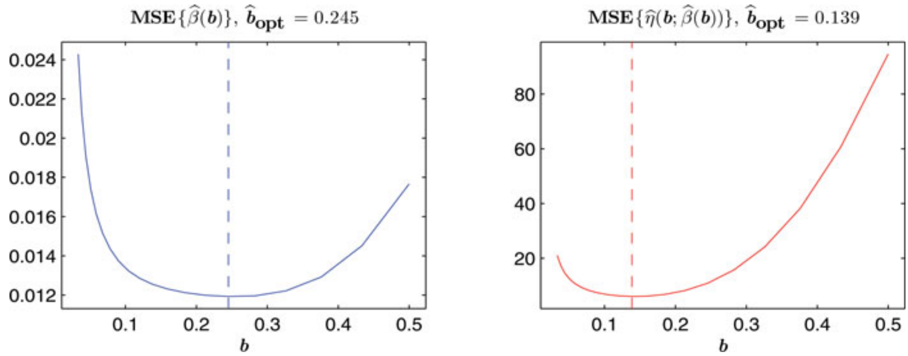

Figure 5. Plots of MSE$\{\widehat{\boldsymbol{\beta}}(b)\}$ and MSE$\{\widehat{\boldsymbol{\eta}}(b; \widehat{\boldsymbol{\beta}}(b))\}$ versus the bandwidth parameter b, with large noise level. Dashed vertical lines indicate the location (on the x axis) of the minimum.

Method 3: $\widehat{\boldsymbol{\eta}}(b; \widehat{\boldsymbol{\beta}}(b_2))$ in (2.10), where b minimizes $\widehat{\text{cov\_pen}}(b; \widehat{\boldsymbol{\beta}}(b_2))$ in (2.11), which replaces (σ^2, R_n) in (2.4) by their estimates in Steps 1–2, and b_2 minimizes $\widehat{\text{MSE}}\{\widehat{\boldsymbol{\beta}}(b) \mid \mathbf{X}\}$; $\widehat{\boldsymbol{\beta}}(b_2)$ is given in (2.8) with $\mathbf{V}$ estimated by $\widehat{\mathbf{V}}$;

Method 4: $\widehat{\boldsymbol{\eta}}(b; \widehat{\boldsymbol{\beta}}(b_2))$ in (2.10), where b minimizes $\text{cov\_pen}(b; \widehat{\boldsymbol{\beta}}(b_2))$ in (2.11), and b_2 minimizes $\text{MSE}\{\widehat{\boldsymbol{\beta}}(b) \mid \mathbf{X}\}$; $\widehat{\boldsymbol{\beta}}(b_2)$ is given in (2.8) with true $\mathbf{V}$;

Method 5: $\widehat{\boldsymbol{\eta}}(b; \widehat{\boldsymbol{\beta}}(b))$ in (2.10), where b minimizes $\widehat{\text{cov\_pen}}(b; \widehat{\boldsymbol{\beta}}(b))$ in (2.11); $\widehat{\boldsymbol{\beta}}(b)$ is given in (2.8) with $\mathbf{V}$ estimated by $\widehat{\mathbf{V}}$;

Method 6: $\widehat{\boldsymbol{\eta}}(b; \widehat{\boldsymbol{\beta}}(b))$ in (2.10), where b minimizes $\text{cov\_pen}(b; \widehat{\boldsymbol{\beta}}(b))$ in (2.11); $\widehat{\boldsymbol{\beta}}(b)$ is given in (2.8) with true $\mathbf{V}$;

Method 7: $\widehat{\boldsymbol{\eta}}(b; \widehat{\boldsymbol{\beta}}(b_2))$ in (2.10), where b minimizes $\widehat{\text{MSE}}\{\widehat{\boldsymbol{\eta}}(b; \widehat{\boldsymbol{\beta}}(b_2)) \mid \mathbf{X}\}$ in (2.12), which replaces $(\sigma^2, R_n; \boldsymbol{\eta}_o^*)$ in (2.9) by their estimates in Steps 1–3, and b_2 minimizes $\widehat{\text{MSE}}\{\widehat{\boldsymbol{\beta}}(b) \mid \mathbf{X}\}$; $\widehat{\boldsymbol{\beta}}(b_2)$ is given in (2.8) with $\mathbf{V}$ estimated by $\widehat{\mathbf{V}}$;

Method 8: $\widehat{\boldsymbol{\eta}}(b; \widehat{\boldsymbol{\beta}}(b_2))$ in (2.10), where b minimizes $\text{MSE}\{\widehat{\boldsymbol{\eta}}(b; \widehat{\boldsymbol{\beta}}(b_2)) \mid \mathbf{X}\}$ in (2.12), and b_2 minimizes $\text{MSE}\{\widehat{\boldsymbol{\beta}}(b) \mid \mathbf{X}\}$; $\widehat{\boldsymbol{\beta}}(b_2)$ is given in (2.8) with true $\mathbf{V}$;

Method 9: $\widehat{\boldsymbol{\eta}}(b; \widehat{\boldsymbol{\beta}}(b))$ in (2.10), where b minimizes $\widehat{\text{MSE}}\{\widehat{\boldsymbol{\eta}}(b; \widehat{\boldsymbol{\beta}}(b)) \mid \mathbf{X}\}$ in (2.12), which replaces $(\sigma^2, R_n; \boldsymbol{\eta}_o^*)$ in (2.9) by their estimates in Steps 1–3; $\widehat{\boldsymbol{\beta}}(b)$ is given in (2.8) with $\mathbf{V}$ estimated by $\widehat{\mathbf{V}}$;

Method 10: $\widehat{\boldsymbol{\eta}}(b; \widehat{\boldsymbol{\beta}}(b))$ in (2.10), where b minimizes $\text{MSE}\{\widehat{\boldsymbol{\eta}}(b; \widehat{\boldsymbol{\beta}}(b)) \mid \mathbf{X}\}$ in (2.12); $\widehat{\boldsymbol{\beta}}(b)$ is given in (2.8) with true $\mathbf{V}$;

Method 11: $\widehat{\boldsymbol{\eta}}(b; \widehat{\boldsymbol{\beta}}(b_2))$ in (2.10), where b minimizes $\widehat{\text{MSE}}\{\widehat{\boldsymbol{\beta}}(b) \mid \mathbf{X}\}$, and b_2 minimizes $\widehat{\text{MSE}}\{\widehat{\boldsymbol{\beta}}(b) \mid \mathbf{X}\}$; $\widehat{\boldsymbol{\beta}}(b_2)$ is given in (2.8) with $\mathbf{V}$ estimated by $\widehat{\mathbf{V}}$;

Method 12: $\widehat{\boldsymbol{\eta}}(b; \widehat{\boldsymbol{\beta}}(b_2))$ in (2.10), where b minimizes $\text{MSE}\{\widehat{\boldsymbol{\beta}}(b) \mid \mathbf{X}\}$, and b_2 minimizes $\text{MSE}\{\widehat{\boldsymbol{\beta}}(b) \mid \mathbf{X}\}$; $\widehat{\boldsymbol{\beta}}(b_2)$ is given in (2.8) with true $\mathbf{V}$;

Figure 6 reveals that all methods, except "Method 2," perform comparably well in estimating nonparametric components. This indicates that estimation of the nonparametric

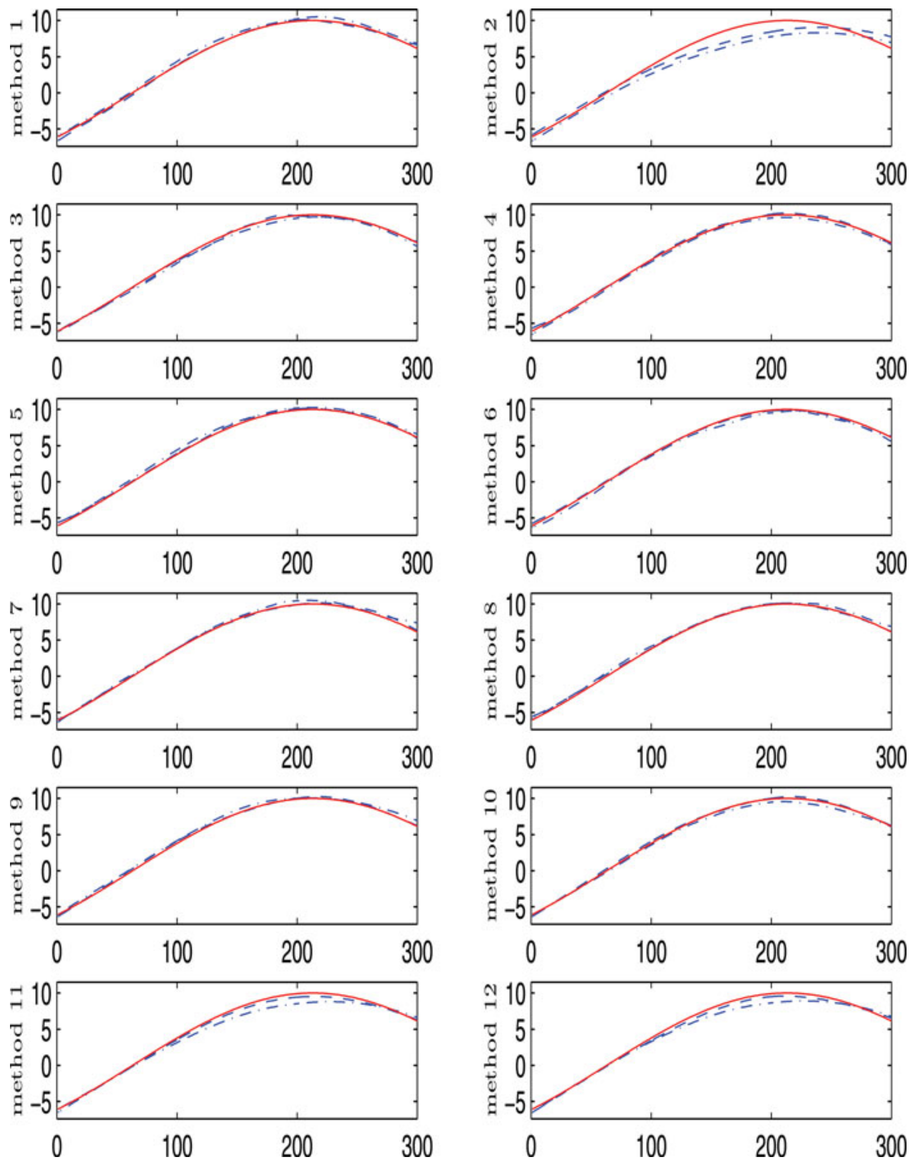

Figure 6. Compare estimators of nonparametric components, with large noise level. Solid curves denote the true nonparametric function, and the estimated curves from two typical samples are presented corresponding to the 25th (the dashed curve) and the 75th (the dash-dotted curve) percentiles among the ASE–ranked values, where $\mathrm{ASE} = \sum_{i=1}^{n}\{\widehat{\eta}(t_i) - \eta_o(t_i)\}^2/n$.

component is relatively less sensitive to the specification of error correlation structures than estimation of the parametric component. The larger biases from "Methods 11–12" are because the optimal bandwidths minimizing $\mathrm{MSE}\{\widehat{\boldsymbol{\beta}}(b) \mid \mathbf{X}\}$ are larger and tend to oversmooth than that minimizing $\mathrm{MSE}\{\widehat{\boldsymbol{\eta}}(b; \widehat{\boldsymbol{\beta}}(b)) \mid \mathbf{X}\}$, evidenced in Figure 5. Among all data-driven approaches, "Method 3" performs slightly better (and faster) than the others.

5. REAL DATA APPLICATION

In an emotional control study, subjects saw a series of negative or positive emotional images, and were asked to either suppress or enhance their emotional responses to the image, or to simply attend to the image. Thus, there were six types of trial (i.e., six types of stimuli). The sequence of trials was randomized. The time between successive trials also varied. The size of the whole brain dataset is $64 \times 64 \times 30$. At each voxel, the time series has 6 runs, each containing 185 observations with a time resolution of 2 sec. The study aims to estimate the BOLD response to each of the trial types for 1–18 sec following the image onset. The length of the estimated HRF parameters is set equal to 18.

The multiple-run PLM (1.4) is used to describe the data, with Run $= 6$ and $n = 185$. The parametric component

$$\boldsymbol{\beta}_{\mathrm{o}} = \left(\boldsymbol{\beta}_{\mathrm{o};1}^T, \ \ldots, \ \boldsymbol{\beta}_{\mathrm{o};6}^T\right)^T \in \mathbb{R}^d$$

is the vector of true HRF coefficients at 18 time points, where $\boldsymbol{\beta}_{\mathrm{o};1} = (h_1(1), \ldots, h_1(18))^T$, $\ldots$, $\boldsymbol{\beta}_{\mathrm{o};6} = (h_6(1), \ldots, h_6(18))^T$ are associated with six types of stimuli, thus $d = 6 \times 18$, and $\boldsymbol{\eta}_{\mathrm{o}}^*$ is the true drift function. Two voxels at coordinates (24, 32, 7) and (49, 41, 10) are declared to be activated using the test statistics $\mathbb{K}_{bc}$. Figures 7–8 present the estimates of HRF parameters (on the top panels) for each of six stimuli, and estimates of the drift function (on the bottom panels). As a comparison, the first two columns use "full-data-driven" method as in Section 4.1 for the parametric component, and "Method 3" as in Section 4.2 for the nonparametric components, whereas the last two columns use "iid-b-dd" method (specifying $R_n = \mathbf{I}_n$) as in Section 4.1, and "Method 3" (specifying $R_n = \mathbf{I}_n$), respectively. The second and fourth columns assume that all runs share a common drift function; the first and third columns remove this constraint. It is seen that there is negligible difference

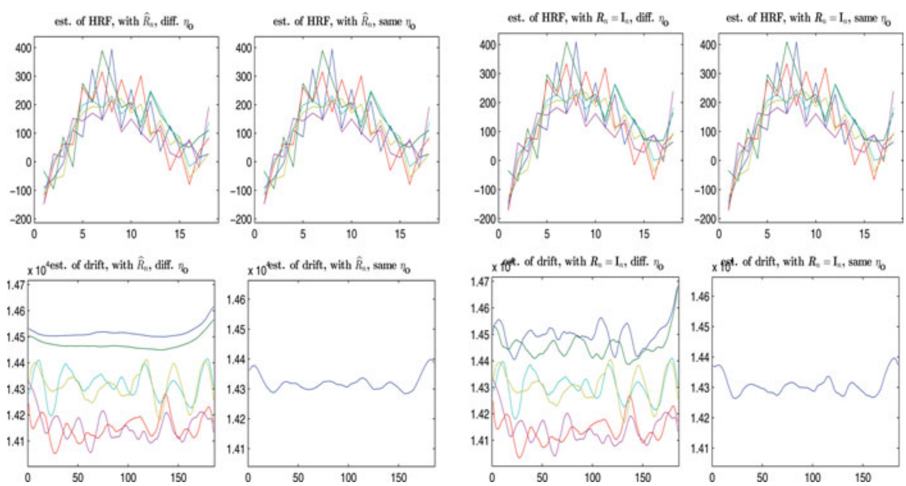

Figure 7. Top panels: compare estimates of HRF values $(h_r(1), \ldots, h_r(18))$ (connected within the rth stimulus) for each of six stimuli; bottom panels: compare estimates of the drift function for each of six runs. The first and third columns: without assuming all runs sharing identical nonparametric components; the second and fourth columns: assuming all runs sharing identical nonparametric components. The voxel is at (24, 32, 7).

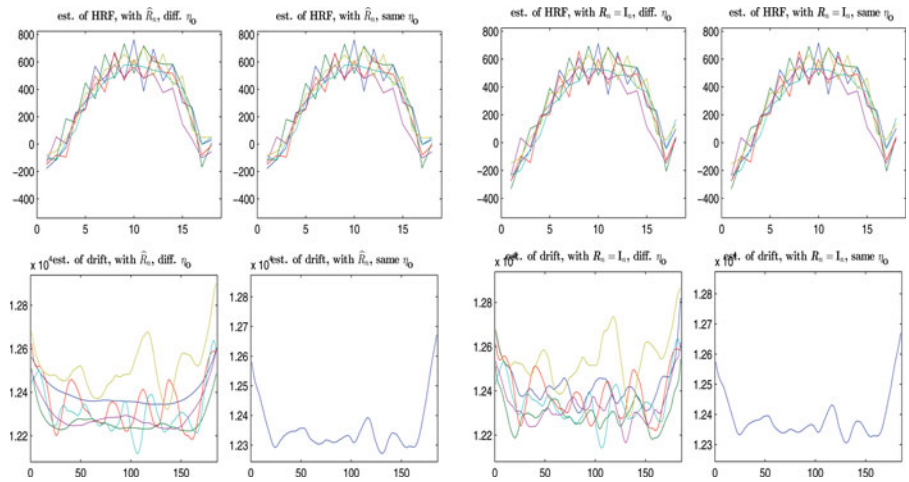

Figure 8. The caption is identical to that for Figure 7, except that the voxel is at (49, 41, 10).

between assuming the nonparametric functions to be either identical or not in estimating the parametric component. Due to the lack of space, comparisons performed at other voxels are omitted.

6. DISCUSSION

We have developed a method for estimating parametric and nonparametric components of the PLM with multiple-runs, in the presence of temporally correlated error terms. We have devised a stepwise algorithm (with justification in each step) for implementing our approach in practice, which is applicable to the analysis of large datasets, such as fMRI data. Matlab implementation of the algorithm is available in the online supplement.

This approach can be extended in several directions. For example, in Section 5, values $\{h_r(1), \ldots, h_r(18)\}$ may follow a smooth curve $h_r(t)$, for some $r \in \{1, \ldots, 6\}$. How do we incorporate some shape information of $\boldsymbol{\beta}_o$ (based on results of prior investigations) into the modeling and to what extent will this affect the computation complexity and estimation efficiency? It is also desirable to develop estimators $\widehat{R}_n$ with faster convergence rates without losing computational efficiency. We leave these issues for future research.

SUPPLEMENTARY MATERIALS

Online appendix: The appendix collects detailed derivations of Proposition 1, Proposition 2, (2.10), (2.11), and (2.12). (JCGS_online_appendix.pdf, pdf file).

Matlab package: All the Matlab script files (along with a readme file) used for simulation studies in the article. (JCGS_online_Matlab_codes.zip, zipped file).

ACKNOWLEDGMENTS

The authors thank the Associate Editor and an anonymous referee for insightful comments. Zhang's research is supported by the NSF grants DMS–1106586, DMS–1308872, and DMS–1521761, and Wisconsin Alumni Research Foundation. Han's research is supported by the Scientific Research Foundation of Northeast Dianli University (No. BSJXM-201216).

[Received May 2014. Revised September 2014.]

REFERENCES

Andrews, D. (1994), "Asymptotics for Semiparametric Econometric Models via Stochastic Equicontinuity," *Econometrica*, 62, 43–72. [123]

Benjamini, Y., and Hochberg, Y. (1995), "Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing," *Journal of the Royal Statistical Society*, Series B, 57, 289–300. [136]

Bickel, P. J., and Levina, E. (2008a), "Covariance Regularization by Thresholding," *The Annals of Statistics*, 36, 2577–2604. [130]

——— (2008b), "Regularized Estimation of Large Covariance Matrices," *The Annals of Statistics*, 36, 199–227. [130]

Cryer, J. D., and Chan, K.-S. (2008), *Time Series Analysis: With Applications in R*, (2nd ed.), Springer Texts in Statistics, New York: Springer Science. [134]

Efron, B. (2004), "The Estimation of Prediction Error: Covariance Penalties and Cross-Validation" (with discussion)," *Journal of the American Statististical Association*, 99, 619–642. [125,130]

——— (2007), "Size, Power and False Discovery Rates," *Annals of Statistics*, 35, 1351–1377. [125]

——— (2010), *Large-Scale Inference: Empirical Bayes Methods for Estimation, Testing, and Prediction*, IMS Monograph, Cambridge, UK: Cambridge University Press. [125]

Fan, J., and Gijbels, I. (1996), *Local Polynomial Modeling and Its Applications*, London: Chapman and Hall. [126]

Fan, J., and Huang, T. (2005), "Profile Likelihood Inferences on Semiparametric Varying-Coefficient Partially Linear Models," *Bernoulli*, 11, 1031–1057. [124]

Fan, J., and Zhang, C. M. (2003), "A Reexamination of Diffusion Estimators With Applications to Financial Model Validation," *Journal of the American Statististical Association*, 98, 118–134. [130]

Glover, G. H. (1999), "Deconvolution of Impulse Response in Event-Related BOLD fMRI," *NeuroImage*, 9, 416–429. [133]

Golub, G. H., and Van Loan, C. F. (1996), *Matrix Computations* (3rd ed.), Baltimore, MD: Johns Hopkins University Press. [125]

Guo, X., and Zhang, C. M. (2015), "Estimation of the Error Autocorrelation Matrix in Semiparametric Model for fMRI Data," *Statistica Sinica*, 25, 475–498. [126,130]

Hart, J. D. (1991), "Kernel Regression Estimation With Time Series Errors," *Journal of the Royal Statistical Society*, Series B, 53, 173–187. [124]

Lian, H., Liang, H., and Wang, L. (2014), "Generalized Additive Partial Linear Models for Clustered Data With Diverging Number of Covariates Using GEE," *Statistica Sinica*, 24, 173–196. [124]

Robinson, P. M. (1988), "Root-N-Consistent Semiparametric Regression," *Econometrica*, 56, 931–954. [123]

Silverman, B. W. (1986), *Density Estimation for Statistics and Data Analysis*, London: Chapman and Hall. [134]

Speckman, P. (1988), "Kernel Smoothing in Partial Linear Models," *Journal of the Royal Statistical Society*, Series B, 50, 413–436. [123]

Xiao, Z. J., Linton, O. B., Carroll, R. J., and Mammen, E. (2003), "More Efficient Local Polynomial Estimation in Nonparametric Regression With Autocorrelated Errors," *Journal of the American Statistical Association*, 98, 980–992. [124]

Yatchew, A. (1997), "An Elementary Estimator of the Partial Linear Model," *Economics Letters*, 57, 135–143. [123]

Zhang, C. M., and Yu, T. (2008), "Semiparametric Detection of Significant Activation for Brain fMRI," *The Annals of Statistics*, 36, 1693–1725. [125,135,136]

Online Supplement to "Accounting for time series errors in partially linear model with single- or multiple-run"

Appendix: Proofs of Main Results

Lemma 1 *For a random vector $\boldsymbol{X}$ and a sigma-field $\mathcal{F}$, $E(\|\boldsymbol{X}\|^2 \mid \mathcal{F}) = \|E(\boldsymbol{X} \mid \mathcal{F})\|^2 + \operatorname{tr}\{\operatorname{var}(\boldsymbol{X} \mid \mathcal{F})\}$.*

Proof:

$$
\begin{aligned}
E(\|\boldsymbol{X}\|^2 \mid \mathcal{F}) &= E(\|E(\boldsymbol{X} \mid \mathcal{F}) + \{\boldsymbol{X} - E(\boldsymbol{X} \mid \mathcal{F})\}\|^2 \mid \mathcal{F}) \\
&= E[\|E(\boldsymbol{X} \mid \mathcal{F})\|^2 + 2\{\boldsymbol{X} - E(\boldsymbol{X} \mid \mathcal{F})\}^T E(\boldsymbol{X} \mid \mathcal{F}) + \|\boldsymbol{X} - E(\boldsymbol{X} \mid \mathcal{F})\|^2 \mid \mathcal{F}] \\
&= \|E(\boldsymbol{X} \mid \mathcal{F})\|^2 + E\{\|\boldsymbol{X} - E(\boldsymbol{X} \mid \mathcal{F})\|^2 \mid \mathcal{F}\},
\end{aligned}
$$

where

$$
\begin{aligned}
E\{\|\boldsymbol{X} - E(\boldsymbol{X} \mid \mathcal{F})\|^2 \mid \mathcal{F}\} &= E(\operatorname{tr}[\{\boldsymbol{X} - E(\boldsymbol{X} \mid \mathcal{F})\}\{\boldsymbol{X} - E(\boldsymbol{X} \mid \mathcal{F})\}^T] \mid \mathcal{F}) \\
&= \operatorname{tr}(E[\{\boldsymbol{X} - E(\boldsymbol{X} \mid \mathcal{F})\}\{\boldsymbol{X} - E(\boldsymbol{X} \mid \mathcal{F})\}^T \mid \mathcal{F}]) \\
&= \operatorname{tr}\{\operatorname{var}(\boldsymbol{X} \mid \mathcal{F})\}.
\end{aligned}
$$

This completes the proof. ■

Proof of Proposition 1. For $\widehat{\boldsymbol{\beta}}(b)$ in (2.8), we now derive $\operatorname{MSE}\{\widehat{\boldsymbol{\beta}}(b) \mid \mathbf{X}\} = E\{\|\widehat{\boldsymbol{\beta}}(b) - \boldsymbol{\beta}_{\mathrm{o}}\|^2 \mid \mathbf{X}\}$. Lemma 1 indicates that

$$
\begin{aligned}
\operatorname{MSE}\{\widehat{\boldsymbol{\beta}}(b) \mid \mathbf{X}\} &= \|E\{\widehat{\boldsymbol{\beta}}(b) - \boldsymbol{\beta}_{\mathrm{o}} \mid \mathbf{X}\}\|^2 + \operatorname{tr}[\operatorname{var}\{\widehat{\boldsymbol{\beta}}(b) \mid \mathbf{X}\}] \\
&\equiv I_1(b) + I_2(b). \qquad \text{(A.1)}
\end{aligned}
$$

Recall $\widetilde{\mathbf{y}} = \{\mathbf{I}_{\mathrm{Run}} \otimes (\mathbf{I}_n - S_b)\}\mathbf{y}$ and $\widetilde{\mathbf{X}} = \{\mathbf{I}_{\mathrm{Run}} \otimes (\mathbf{I}_n - S_b)\}\mathbf{X}$. Then from (2.7)–(2.8) and (1.5),

$$
\begin{aligned}
\widehat{\boldsymbol{\beta}}(b) &= (\mathbf{A}_b\mathbf{X})^{-1}\mathbf{A}_b\mathbf{y} \\
&\equiv \boldsymbol{\beta}_{\mathrm{o}} + (\mathbf{A}_b\mathbf{X})^{-1}\mathbf{A}_b\boldsymbol{\eta}_{\mathrm{o}}^* + (\mathbf{A}_b\mathbf{X})^{-1}\mathbf{A}_b\boldsymbol{\epsilon},
\end{aligned}
$$

where

$$
\mathbf{A}_b = \widetilde{\mathbf{X}}^T\mathbf{V}\{\mathbf{I}_{\mathrm{Run}} \otimes (\mathbf{I}_n - S_b)\}
$$

$$
\begin{aligned}
&= \mathbf{X}^T\{\mathbf{I}_{\mathrm{Run}} \otimes (\mathbf{I}_n - S_b)^T\}(\mathbf{I}_{\mathrm{Run}} \otimes V_n)\{\mathbf{I}_{\mathrm{Run}} \otimes (\mathbf{I}_n - S_b)\} \\
&= \mathbf{X}^T[\mathbf{I}_{\mathrm{Run}} \otimes \{(\mathbf{I}_n - S_b)^T V_n (\mathbf{I}_n - S_b)\}] \\
&= \mathbf{X}^T(\mathbf{I}_{\mathrm{Run}} \otimes M_b) \in \mathbb{R}^{d\times(\mathrm{Run}\times n)}. \qquad \text{(A.2)}
\end{aligned}
$$

Hence

$$
\widehat{\boldsymbol{\beta}}(b) - \boldsymbol{\beta}_{\mathrm{o}} = (\mathbf{A}_b\mathbf{X})^{-1}\mathbf{A}_b\boldsymbol{\eta}_{\mathrm{o}}^* + (\mathbf{A}_b\mathbf{X})^{-1}\mathbf{A}_b\boldsymbol{\epsilon}. \qquad \text{(A.3)}
$$

Thus from (A.3) and (1.6),

$$
E\{\widehat{\boldsymbol{\beta}}(b) - \boldsymbol{\beta}_{\mathrm{o}} \mid \mathbf{X}\} = (\mathbf{A}_b\mathbf{X})^{-1}\mathbf{A}_b\boldsymbol{\eta}_{\mathrm{o}}^*. \qquad \text{(A.4)}
$$

Similarly,

$$
\begin{aligned}
\mathrm{var}\{\widehat{\boldsymbol{\beta}}(b) \mid \mathbf{X}\} &= \mathrm{var}\{(\mathbf{A}_b\mathbf{X})^{-1}\mathbf{A}_b\boldsymbol{\epsilon} \mid \mathbf{X}\} \\
&= (\mathbf{A}_b\mathbf{X})^{-1}\mathbf{A}_b\boldsymbol{\Sigma}\mathbf{A}_b^T(\mathbf{A}_b\mathbf{X})^{-1} \\
&= \sigma^2(\mathbf{A}_b\mathbf{X})^{-1}\mathbf{A}_b(\mathbf{I}_{\mathrm{Run}} \otimes R_n)\mathbf{A}_b^T(\mathbf{A}_b\mathbf{X})^{-1},
\end{aligned}
$$

where $\boldsymbol{\Sigma} = \mathbf{I}_{\mathrm{Run}} \otimes (\sigma^2 R_n) = \sigma^2(\mathbf{I}_{\mathrm{Run}} \otimes R_n)$, and from (A.2),

$$
\begin{aligned}
\mathbf{A}_b(\mathbf{I}_{\mathrm{Run}} \otimes R_n)\mathbf{A}_b^T &= \mathbf{X}^T(\mathbf{I}_{\mathrm{Run}} \otimes M_b)(\mathbf{I}_{\mathrm{Run}} \otimes R_n)(\mathbf{I}_{\mathrm{Run}} \otimes M_b)\mathbf{X} \\
&= \mathbf{X}^T\{\mathbf{I}_{\mathrm{Run}} \otimes (M_b R_n M_b)\}\mathbf{X}.
\end{aligned}
$$

Thus

$$
\mathrm{var}\{\widehat{\boldsymbol{\beta}}(b) \mid \mathbf{X}\} = \sigma^2(\mathbf{A}_b\mathbf{X})^{-1}\mathbf{X}^T\{\mathbf{I}_{\mathrm{Run}} \otimes (M_b R_n M_b)\}\mathbf{X}(\mathbf{A}_b\mathbf{X})^{-1}. \qquad \text{(A.5)}
$$

Collecting (A.4) and (A.5) to (A.1) gives

$$
\begin{aligned}
I_1(b) &= \|(\mathbf{A}_b\mathbf{X})^{-1}(\mathbf{A}_b\boldsymbol{\eta}_{\mathrm{o}}^*)\|^2, \\
I_2(b) &= \sigma^2\,\mathrm{tr}[(\mathbf{A}_b\mathbf{X})^{-1}\mathbf{X}^T\{\mathbf{I}_{\mathrm{Run}} \otimes (M_b R_n M_b)\}\mathbf{X}(\mathbf{A}_b\mathbf{X})^{-1}],
\end{aligned}
$$

in which (A.2) indicates that

$$
\begin{aligned}
\mathbf{A}_b\mathbf{X} = \mathbf{X}^T(\mathbf{I}_{\mathrm{Run}} \otimes M_b)\mathbf{X} &= \sum_{k=1}^{\mathrm{Run}} \mathbf{X}_{\mathrm{run}\,k}^T M_b \mathbf{X}_{\mathrm{run}\,k}, \\
\mathbf{A}_b\boldsymbol{\eta}_{\mathrm{o}}^* = \mathbf{X}^T(\mathbf{I}_{\mathrm{Run}} \otimes M_b)\boldsymbol{\eta}_{\mathrm{o}}^* &= \sum_{k=1}^{\mathrm{Run}} \mathbf{X}_{\mathrm{run}\,k}^T M_b \boldsymbol{\eta}_{\mathrm{o};\mathrm{run}\,k}, \\
\mathbf{X}^T\{\mathbf{I}_{\mathrm{Run}} \otimes (M_b R_n M_b)\}\mathbf{X} &= \sum_{k=1}^{\mathrm{Run}} \mathbf{X}_{\mathrm{run}\,k}^T M_b R_n M_b \mathbf{X}_{\mathrm{run}\,k}.
\end{aligned}
$$

This completes the proof. ■

Proof of Proposition 2. For $\widehat{\boldsymbol\eta}^* = (\widehat{\boldsymbol\eta}_{\mathrm{run}\,1}^T, \ldots, \widehat{\boldsymbol\eta}_{\mathrm{run}\,\mathrm{Run}}^T)^T$, note that $\mathrm{MSE}(\widehat{\boldsymbol\eta}^* \mid \mathbf{X}) = E(\|\widehat{\boldsymbol\eta}^* - \boldsymbol\eta_{\mathrm{o}}^*\|^2 \mid \mathbf{X}) = \sum_{k=1}^{\mathrm{Run}} E(\|\widehat{\boldsymbol\eta}_{\mathrm{run}\,k} - \boldsymbol\eta_{\mathrm{o;run}\,k}\|^2 \mid \mathbf{X}) = \sum_{k=1}^{\mathrm{Run}} \mathrm{MSE}(\widehat{\boldsymbol\eta}_{\mathrm{run}\,k} \mid \mathbf{X})$. It suffices to derive $\mathrm{MSE}(\widehat{\boldsymbol\eta}_{\mathrm{run}\,k} \mid \mathbf{X})$. From (2.3) and (1.4),

$$\begin{aligned}
\widehat{\boldsymbol\eta}_{\mathrm{run}\,k}(b_k; \widehat{\boldsymbol\beta}(b)) &= S_{b_k}\{\mathbf{y}_{\mathrm{run}\,k} - \mathbf{X}_{\mathrm{run}\,k}\widehat{\boldsymbol\beta}(b)\} \\
&= S_{b_k}\{\mathbf{X}_{\mathrm{run}\,k}\boldsymbol\beta_{\mathrm{o}} + \boldsymbol\eta_{\mathrm{o;run}\,k} + \boldsymbol\epsilon_{\mathrm{run}\,k} - \mathbf{X}_{\mathrm{run}\,k}\widehat{\boldsymbol\beta}(b)\} \\
&= S_{b_k}[\boldsymbol\eta_{\mathrm{o;run}\,k} + \boldsymbol\epsilon_{\mathrm{run}\,k} - \mathbf{X}_{\mathrm{run}\,k}\{\widehat{\boldsymbol\beta}(b) - \boldsymbol\beta_{\mathrm{o}}\}] \\
&= S_{b_k}\boldsymbol\eta_{\mathrm{o;run}\,k} + S_{b_k}[\boldsymbol\epsilon_{\mathrm{run}\,k} - \mathbf{X}_{\mathrm{run}\,k}\{\widehat{\boldsymbol\beta}(b) - \boldsymbol\beta_{\mathrm{o}}\}], \\
\widehat{\boldsymbol\eta}_{\mathrm{run}\,k}(b_k; \widehat{\boldsymbol\beta}(b)) - \boldsymbol\eta_{\mathrm{o;run}\,k} &= S_{b_k}\boldsymbol\eta_{\mathrm{o;run}\,k} - \boldsymbol\eta_{\mathrm{o;run}\,k} + S_{b_k}[\boldsymbol\epsilon_{\mathrm{run}\,k} - \mathbf{X}_{\mathrm{run}\,k}\{\widehat{\boldsymbol\beta}(b) - \boldsymbol\beta_{\mathrm{o}}\}]. \qquad (A.6)
\end{aligned}$$

From (A.6) and (A.4), we obtain

$$\begin{aligned}
& E\{\widehat{\boldsymbol\eta}_{\mathrm{run}\,k}(b_k; \widehat{\boldsymbol\beta}(b)) - \boldsymbol\eta_{\mathrm{o;run}\,k} \mid \mathbf{X}\} \\
= \;& S_{b_k}\boldsymbol\eta_{\mathrm{o;run}\,k} - \boldsymbol\eta_{\mathrm{o;run}\,k} - S_{b_k}\mathbf{X}_{\mathrm{run}\,k}E\{\widehat{\boldsymbol\beta}(b) - \boldsymbol\beta_{\mathrm{o}} \mid \mathbf{X}\} \\
= \;& S_{b_k}\boldsymbol\eta_{\mathrm{o;run}\,k} - \boldsymbol\eta_{\mathrm{o;run}\,k} - S_{b_k}\mathbf{X}_{\mathrm{run}\,k}(\mathbf{A}_b\mathbf{X})^{-1}\mathbf{A}_b\boldsymbol\eta_{\mathrm{o}}^*. \qquad (A.7)
\end{aligned}$$

Similarly, from (A.6) and (A.5), we obtain

$$\begin{aligned}
& \mathrm{var}\{\widehat{\boldsymbol\eta}_{\mathrm{run}\,k}(b_k; \widehat{\boldsymbol\beta}(b)) \mid \mathbf{X}\} \\
= \;& S_{b_k}\mathrm{var}(\boldsymbol\epsilon_{\mathrm{run}\,k} - \mathbf{X}_{\mathrm{run}\,k}\{\widehat{\boldsymbol\beta}(b) - \boldsymbol\beta_{\mathrm{o}}\} \mid \mathbf{X})S_{b_k}^T \\
= \;& S_{b_k}[\mathrm{var}(\boldsymbol\epsilon_{\mathrm{run}\,k} \mid \mathbf{X}) \\
& -\mathrm{cov}(\boldsymbol\epsilon_{\mathrm{run}\,k}, \mathbf{X}_{\mathrm{run}\,k}\{\widehat{\boldsymbol\beta}(b) - \boldsymbol\beta_{\mathrm{o}}\} \mid \mathbf{X}) \\
& -\mathrm{cov}(\mathbf{X}_{\mathrm{run}\,k}\{\widehat{\boldsymbol\beta}(b) - \boldsymbol\beta_{\mathrm{o}}\}, \boldsymbol\epsilon_{\mathrm{run}\,k} \mid \mathbf{X}) \\
& +\mathrm{var}(\mathbf{X}_{\mathrm{run}\,k}\{\widehat{\boldsymbol\beta}(b) - \boldsymbol\beta_{\mathrm{o}}\} \mid \mathbf{X})]S_{b_k}^T \\
= \;& S_{b_k}\{I_{4;k;1} - I_{4;k;3}(b) - I_{4;k;2}(b) + I_{4;k;4}(b)\}S_{b_k}^T, \qquad (A.8)
\end{aligned}$$

where from (1.6),

$$I_{4;k;1} = \Sigma_n = \sigma^2 R_n;$$

from (A.5),

$$\begin{aligned}
I_{4;k;4}(b) &= \mathrm{var}(\mathbf{X}_{\mathrm{run}\,k}\{\widehat{\boldsymbol\beta}(b) - \boldsymbol\beta_{\mathrm{o}}\} \mid \mathbf{X}) \\
&= \mathbf{X}_{\mathrm{run}\,k}\mathrm{var}\{\widehat{\boldsymbol\beta}(b) - \boldsymbol\beta_{\mathrm{o}} \mid \mathbf{X}\}\mathbf{X}_{\mathrm{run}\,k}^T \\
&= \sigma^2\mathbf{X}_{\mathrm{run}\,k}(\mathbf{A}_b\mathbf{X})^{-1}\mathbf{X}^T\{\mathbf{I}_{\mathrm{Run}} \otimes (M_b R_n M_b)\}\mathbf{X}(\mathbf{A}_b\mathbf{X})^{-1}\mathbf{X}_{\mathrm{run}\,k}^T,
\end{aligned}$$

and

$$
\begin{aligned}
I_{4;k;3}(b) &= \mathrm{cov}(\boldsymbol{\epsilon}_{\mathrm{run}\,k}, \mathbf{X}_{\mathrm{run}\,k}\{\widehat{\boldsymbol{\beta}}(b) - \boldsymbol{\beta}_{\mathrm{o}}\} \mid \mathbf{X}) \\
&= \mathrm{cov}\{\boldsymbol{\epsilon}_{\mathrm{run}\,k}, \mathbf{X}_{\mathrm{run}\,k}(\mathbf{A}_b\mathbf{X})^{-1}\mathbf{A}_b\boldsymbol{\epsilon} \mid \mathbf{X}\} \\
&= \mathrm{cov}(\boldsymbol{\epsilon}_{\mathrm{run}\,k}, \boldsymbol{\epsilon} \mid \mathbf{X})\mathbf{A}_b^T(\mathbf{A}_b\mathbf{X})^{-1}\mathbf{X}_{\mathrm{run}\,k}^T \\
&= (\boldsymbol{e}_{k,\mathrm{Run}}^T \otimes \mathbf{I}_n)\mathrm{cov}(\boldsymbol{\epsilon}, \boldsymbol{\epsilon} \mid \mathbf{X})\mathbf{A}_b^T(\mathbf{A}_b\mathbf{X})^{-1}\mathbf{X}_{\mathrm{run}\,k}^T \\
&= (\boldsymbol{e}_{k,\mathrm{Run}}^T \otimes \mathbf{I}_n)\sigma^2(\mathbf{I}_{\mathrm{Run}} \otimes R_n)\mathbf{A}_b^T(\mathbf{A}_b\mathbf{X})^{-1}\mathbf{X}_{\mathrm{run}\,k}^T \\
&= (\boldsymbol{e}_{k,\mathrm{Run}}^T \otimes \mathbf{I}_n)\sigma^2(\mathbf{I}_{\mathrm{Run}} \otimes R_n)(\mathbf{I}_{\mathrm{Run}} \otimes M_b)\mathbf{X}(\mathbf{A}_b\mathbf{X})^{-1}\mathbf{X}_{\mathrm{run}\,k}^T \\
&= \sigma^2\{\boldsymbol{e}_{k,\mathrm{Run}}^T \otimes (R_n M_b)\}\mathbf{X}(\mathbf{A}_b\mathbf{X})^{-1}\mathbf{X}_{\mathrm{run}\,k}^T \\
&= \sigma^2(R_n M_b \mathbf{X}_{\mathrm{run}\,k})(\mathbf{A}_b\mathbf{X})^{-1}\mathbf{X}_{\mathrm{run}\,k}^T,
\end{aligned}
$$

by using the fact that $\boldsymbol{\epsilon}_{\mathrm{run}\,k} = (\boldsymbol{e}_{k,\mathrm{Run}}^T \otimes \mathbf{I}_n)\boldsymbol{\epsilon}$ with $\boldsymbol{e}_{k,N}$ denoting the kth column vector of a $N \times N$ identity matrix, (A.3) and (A.2), and similarly,

$$
\begin{aligned}
I_{4;k;2}(b) &= \mathrm{cov}(\mathbf{X}_{\mathrm{run}\,k}\{\widehat{\boldsymbol{\beta}}(b) - \boldsymbol{\beta}_{\mathrm{o}}\}, \boldsymbol{\epsilon}_{\mathrm{run}\,k} \mid \mathbf{X}) = \{I_{4;k;3}(b)\}^T \\
&= \sigma^2\mathbf{X}_{\mathrm{run}\,k}(\mathbf{A}_b\mathbf{X})^{-1}(R_n M_b \mathbf{X}_{\mathrm{run}\,k})^T.
\end{aligned}
$$

Applying Proposition 1 again gives

$$
\begin{aligned}
&\mathrm{MSE}\{\widehat{\boldsymbol{\eta}}_{\mathrm{run}\,k}(b_k; \widehat{\boldsymbol{\beta}}(b)) \mid \mathbf{X}\} \\
&= \|E\{\widehat{\boldsymbol{\eta}}_{\mathrm{run}\,k}(b_k; \widehat{\boldsymbol{\beta}}(b)) - \boldsymbol{\eta}_{\mathrm{o;run}\,k} \mid \mathbf{X}\}\|^2 + \mathrm{tr}[\mathrm{var}\{\widehat{\boldsymbol{\eta}}_{\mathrm{run}\,k}(b_k; \widehat{\boldsymbol{\beta}}(b)) \mid \mathbf{X}\}] \\
&= I_{3;k}(b_k; b) + I_{4;k}(b_k; b),
\end{aligned}
$$

where from (A.7),

$$
\begin{aligned}
I_{3;k}(b_k; b) &= \|S_{b_k}\boldsymbol{\eta}_{\mathrm{o;run}\,k} - \boldsymbol{\eta}_{\mathrm{o;run}\,k} - S_{b_k}\mathbf{X}_{\mathrm{run}\,k}(\mathbf{A}_b\mathbf{X})^{-1}\mathbf{A}_b\boldsymbol{\eta}_{\mathrm{o}}^*\|^2 \\
&= \|(\mathbf{I}_n - S_{b_k})\boldsymbol{\eta}_{\mathrm{o;run}\,k} + S_{b_k}\mathbf{X}_{\mathrm{run}\,k}(\mathbf{A}_b\mathbf{X})^{-1}\mathbf{A}_b\boldsymbol{\eta}_{\mathrm{o}}^*\|^2,
\end{aligned}
$$

and from (A.8),

$$
\begin{aligned}
I_{4;k}(b_k; b) &= \mathrm{tr}[S_{b_k}\{I_{4;k;1} - I_{4;k;3}(b) - I_{4;k;2}(b) + I_{4;k;4}(b)\}S_{b_k}^T] \\
&= \sigma^2\,\mathrm{tr}\Big(S_{b_k}\Big[R_n - I_{4;k;2}^*(b) - I_{4;k;3}^*(b) \\
&\quad +\mathbf{X}_{\mathrm{run}\,k}(\mathbf{A}_b\mathbf{X})^{-1}\mathbf{X}^T\{\mathbf{I}_{\mathrm{Run}} \otimes (M_b R_n M_b)\}\mathbf{X}(\mathbf{A}_b\mathbf{X})^{-1}\mathbf{X}_{\mathrm{run}\,k}^T\Big]S_{b_k}^T\Big),
\end{aligned}
$$

where

$$
\begin{aligned}
I_{4;k;2}^*(b) &= \mathbf{X}_{\mathrm{run}\,k}(\mathbf{A}_b\mathbf{X})^{-1}(R_n M_b \mathbf{X}_{\mathrm{run}\,k})^T, \\
I_{4;k;3}^*(b) &= \{I_{4;k;2}^*(b)\}^T.
\end{aligned}
$$

This completes the proof. ■

Proof of (2.10)**.** Assume that $\boldsymbol{\eta}_{\mathrm{o;run\,1}} = \cdots = \boldsymbol{\eta}_{\mathrm{o;run\,Run}} = \boldsymbol{\eta}_{\mathrm{o}}$. Then model (1.4) becomes $Y_k(t_i) - \boldsymbol{X}_{k;i}^T\boldsymbol{\beta}_{\mathrm{o}} = \eta_{\mathrm{o}}(t_i) + \epsilon_k(t_i)$, $i = 1, \ldots, n$ and $k = 1, \ldots, \mathrm{Run}$. The local-linear estimation of $\eta_{\mathrm{o}}(t)$ proceeds as follows. For $t_i \approx t$, Taylor expansion gives $\eta_{\mathrm{o}}(t_i) \approx \eta_{\mathrm{o}}(t) + \eta_{\mathrm{o}}'(t)(t_i - t) \equiv \boldsymbol{T}_i(t)^T(\eta_{\mathrm{o}}(t), \eta_{\mathrm{o}}'(t))$, where $\boldsymbol{T}_i(t) = (1, t_i - t)^T$. For any given $\widehat{\boldsymbol{\beta}}$, the local-linear estimation method in Fan and Gijbels (1996) estimates $\eta_{\mathrm{o}}(t)$ by

$$\begin{aligned} \widehat{\eta}(t) &= \boldsymbol{e}_{1,2}^T\Big[\arg\min_{\boldsymbol{\alpha}} \frac{1}{\mathrm{Run}} \sum_{k=1}^{\mathrm{Run}} \sum_{i=1}^{n} \{Y_k(t_i) - \boldsymbol{X}_{k;i}^T\widehat{\boldsymbol{\beta}} - \boldsymbol{T}_i(t)^T\boldsymbol{\alpha}\}^2 K_b(t_i - t)\Big] \\ &= \boldsymbol{e}_{1,2}^T F_n(\boldsymbol{\alpha}), \end{aligned}$$

where $\boldsymbol{e}_{1,2} = (1, 0)^T$ and $F_n(\boldsymbol{\alpha}) = (1/\mathrm{Run}) \sum_{k=1}^{\mathrm{Run}} \sum_{i=1}^{n} \{Y_k(t_i) - \boldsymbol{X}_{k;i}^T\widehat{\boldsymbol{\beta}} - \boldsymbol{T}_i(t)^T\boldsymbol{\alpha}\}^2 K_b(t_i - t)$. Note that

$$\begin{aligned} F_n'(\boldsymbol{\alpha}) &= -2\frac{1}{\mathrm{Run}} \sum_{k=1}^{\mathrm{Run}} \sum_{i=1}^{n} \boldsymbol{T}_i(t)\{Y_k(t_i) - \boldsymbol{X}_{k;i}^T\widehat{\boldsymbol{\beta}} - \boldsymbol{T}_i(t)^T\boldsymbol{\alpha}\} K_b(t_i - t) \\ &= -2\sum_{i=1}^{n} \boldsymbol{T}_i(t)\{Y_{\boldsymbol{\cdot}}(t_i) - \boldsymbol{X}_{\boldsymbol{\cdot};i}^T\widehat{\boldsymbol{\beta}} - \boldsymbol{T}_i(t)^T\boldsymbol{\alpha}\} K_b(t_i - t). \end{aligned}$$

Hence $\widehat{\eta}(\cdot)$ corresponds to the estimated nonparametric regression function, based on data $\{(t_i, Y_{\boldsymbol{\cdot}}(t_i) - \boldsymbol{X}_{\boldsymbol{\cdot};i}^T\widehat{\boldsymbol{\beta}})\}_{i=1}^n$. Such correspondence indicates that the estimator $(\widehat{\eta}(t_1), \ldots, \widehat{\eta}(t_n))^T$ of $(\eta_{\mathrm{o}}(t_1), \ldots, \eta_{\mathrm{o}}(t_n))^T$ is $S_b(\mathbf{y}_{\boldsymbol{\cdot}} - \mathbf{X}_{\boldsymbol{\cdot}}\widehat{\boldsymbol{\beta}}) = S_b\, \mathbf{res}_{\boldsymbol{\cdot}}(\widehat{\boldsymbol{\beta}})$, where S_b is given in (2.1). ■

Proof of (2.11)**.** Utilizing the derivation given in Section 3.3 for (2.4) and the fact $\mathrm{var}(\boldsymbol{\epsilon}_{\boldsymbol{\cdot}} \mid \mathbf{X}) = (1/\mathrm{Run})\sigma^2 R_n$, completes the proof. The details are omitted. ■

Proof of (2.12)**.** The proof completes by utilizing the proof for Proposition 2, along with $\mathrm{var}(\boldsymbol{\epsilon}_{\boldsymbol{\cdot}} \mid \mathbf{X}) = (1/\mathrm{Run})\sigma^2 R_n$ and $\mathrm{cov}(\boldsymbol{\epsilon}_{\boldsymbol{\cdot}}, \boldsymbol{\epsilon} \mid \mathbf{X}) = (1/\mathrm{Run})(\mathbf{1}_{\mathrm{Run}}^T \otimes \mathbf{I}_n)\mathrm{cov}(\boldsymbol{\epsilon}, \boldsymbol{\epsilon} \mid \mathbf{X})$, where $\mathbf{1}_k = (1, \ldots, 1)^T \in \mathbb{R}^k$. The details are omitted. ■