Statistica Sinica **25** (2015), 475-498
doi:http://dx.doi.org/10.5705/ss.2013.110

ESTIMATION OF THE ERROR AUTOCORRELATION MATRIX IN SEMIPARAMETRIC MODEL FOR FMRI DATA

Xiao Guo and Chunming Zhang

University of Wisconsin-Madison

Abstract: In statistical analysis of functional magnetic resonance imaging (fMRI), dealing with the temporal correlation is a major challenge in assessing changes within voxels. This paper aims to address this issue by considering a semiparametric model for single-voxel fMRI. For the error process in the semiparametric model with autocorrelation matrix $\mathbf{R}$, we adopt the difference-based method to construct a banded estimate $\widehat{\mathbf{R}}$ of $\mathbf{R}$, and propose a refined estimate $\widehat{\mathbf{R}}_*^{-1}$ of $\mathbf{R}^{-1}$. Under mild regularity conditions, we establish consistency of $\widehat{\mathbf{R}}$ and $\widehat{\mathbf{R}}_*$ with explicit convergence rates. We also demonstrate convergence of $\widehat{\mathbf{R}}_*^{-1}$ in mean square under the L_∞ norm, though this convergence property does not hold for $\widehat{\mathbf{R}}^{-1}$. Data-driven procedures for choosing the banding parameter and refining the estimate are developed, and simulation studies reveal their satisfactory performance. Numerical results suggest that $\widehat{\mathbf{R}}_*^{-1}$ performs well when applied to the semiparametric test statistics for detecting brain activity.

Key words and phrases: Autocorrelation matrix, difference-based method, fMRI, inverse, semiparametric model.

1. Introduction

Functional magnetic resonance imaging (fMRI) is a powerful technique to study the neural underpinnings of human cognition. It has been widely used to detect the areas of the brain responsible for certain motor or sensory tasks. An fMRI data set for a single scan on a single subject typically consists of a time series of responses for each of the voxels throughout the brain, where a voxel is a three dimensional rectangular cuboid. It is well known that fMRI data are highly temporally correlated within the same voxel and spatially correlated between different voxels (Lange (1996); Zarahn, Aguirre, and D'Esposito (1997)). A common strategy for fMRI data analysis has two major components. The first assesses the task-associated signal changes for each voxel, where adequate accommodation of the within-voxel temporal correlation is required. The second determines the regions of activation over the human brain based on the voxel-wise assessment, where the spatial correlation across voxels should be taken into consideration. We focus our attention on dealing with the temporal correlation, an essential and challenging issue for the first component.

Different statistical modeling and analysis approaches have been studied for single-voxel fMRI, including the general linear model (GLM) approach (Lange (1996); Lazar et al. (2001); Worsley et al. (2002)) and the semiparametric model approach (Zhang and Yu (2008)). For the GLM approach, ignoring the temporal correlation can result in biased test statistics for hypothesis testing of the model parameters and thus lead to inaccurate detection of the brain activity. Several methods have been proposed to handle the serial correlation with some variant of the general linear model, that can be classified into three groups (Monti (2011)): the pre-coloring method (Friston et al. (1995); Worsley and Friston (1995)), the pre-whitening method with the error term of the model assumed to be AR(1) (Bullmore et al. (1996)), AR(m) (Worsley et al. (2002)), and ARMA$(1, 1)$ (Purdon et al. (2001)), among others, and the explicit noise modeling method (Lund et al. (2006)).

Here, we consider the semiparametric approach that yields more accurate prediction of the time-course behavior of neuronal responses than the GLM approach. For semiparametric inference, detection of the brain activity in a certain voxel can be achieved via tests of significance of the hemodynamic response function (HRF). In constructing the semiparametric test statistics, estimation of $\mathbf{R}^{-1}$ plays an important role, where $\mathbf{R}$ is the correlation matrix of the temporally correlated error process. A difference-based approach (Zhang et al. (2008)) was developed to estimate the autocorrelation matrix of a stationary g-dependent error process. Since the fMRI noise containing neural and nonneural sources has long-range correlations (Friston et al. (2000)), it is desirable to explore whether estimation of $\mathbf{R}$, under a more general assumption of the error process, can benefit from the difference-based method, which is computationally fast. Moreover, asymptotic properties of the estimates of $\mathbf{R}$ and $\mathbf{R}^{-1}$ need to be established for the semiparametric inference of fMRI.

We consider two classes of error processes. One is a linear process with dependence structure related to the sample size; it covers a broad range of stationary time series models, for example, the stationary ARMA(p, q) model. The other is a stationary g_n-dependent process with g_n being bounded or diverging; it includes the g-dependent process (Zhang et al. (2008); Zhang and Yu (2008)) as a special case. Under these assumptions, estimation of the Toeplitz error autocorrelation matrix $\mathbf{R}$ is a challenging task. Directly applying the difference-based approach leads to a badly behaved estimate of $\mathbf{R}$, because too many autocovariances are estimated. Several regularization methods (e.g., banding, tapering, and thresholding) have been studied to yield consistent estimators for large covariance matrices in areas of high-dimensional inference (Bickel and Levina (2008a,b); Cai, Zhang, and Zhou (2010); El Karoui (2008); Furrer and Bengtsson (2007)) and time series analysis (McMurry and Politis (2010); Wu and Pourahmadi (2009);

Xiao and Wu (2012)). We adopt the regularization idea of banding and employ the difference-based approach to develop an estimate $\widehat{\mathbf{R}}$ of $\mathbf{R}$. For $\mathbf{R}^{-1}$, we propose a refined estimate $\widehat{\mathbf{R}}_*^{-1}$.

Cai, Ren, and Zhou (2012), Wu and Pourahmadi (2009), and Xiao and Wu (2012) have studied the banding method for large Toeplitz covariance matrix estimation and obtained convergence rates of the estimates. Their results are not readily translated into results for our methods and the asymptotic properties of $\widehat{\mathbf{R}}$ and $\widehat{\mathbf{R}}_*^{-1}$ are needed. Under some mild regularity conditions, we establish explicit convergence rates of $\widehat{\mathbf{R}}$ and $\widehat{\mathbf{R}}_*$ and demonstrate convergence of $\widehat{\mathbf{R}}_*^{-1}$ in mean square under the L_∞ norm. For practical guidance, data-driven approaches to choosing the banding parameter and refining the estimate of $\mathbf{R}^{-1}$ are developed. Our contribution to the analysis of brain fMRI is in providing a computationally efficient and consistent estimate of $\mathbf{R}^{-1}$ for semiparametric inference.

The rest of this paper is organized as follows. Section 2 reviews the semiparametric model for fMRI in Zhang and Yu (2008), and develops the banded estimate $\widehat{\mathbf{R}}$. Section 3 obtains explicit convergence rates of $\widehat{\mathbf{R}}$ and $\widehat{\mathbf{R}}_*$, and establishes consistency of $\widehat{\mathbf{R}}_*^{-1}$ under the L_∞ norm. Section 4 provides the data-driven methods for choosing the banding parameter and refining the estimate of $\mathbf{R}^{-1}$. Section 5 presents simulation evaluations of $\widehat{\mathbf{R}}_*^{-1}$ in applying it to the semiparametric test statistics in Zhang and Yu (2008). Section 6 is illustrated with fMRI data. The Appendix includes the conditions and proofs for the main results; it is available as an online supplement.

We introduce some notation. For a matrix A, $\lambda_{\min}(A)$ and $\lambda_{\max}(A)$ denote the smallest and largest eigenvalues of A, respectively, and $A \succ 0$ signifies that A is positive definite. For $A = \{a_{ij}\}$, the L_1, L_2 and L_∞ norms are $\|A\|_1 = \max_j \sum_i |a_{ij}|$, $\|A\| \equiv \|A\|_2 = \{\lambda_{\max}(A^T A)\}^{1/2}$ and $\|A\|_\infty = \max_i \sum_j |a_{ij}|$, respectively. If A is symmetric, $\|A\|_1 = \|A\|_\infty \geq \|A\|_2$; see, e.g., Golub and Van Loan (1989). For a vector $\boldsymbol{v} = (v_1, \ldots, v_p)^T$, the L_r norm is $\|\boldsymbol{v}\|_r = (\sum_{i=1}^p |v_i|^r)^{1/r}$ for $r \geq 1$. Denote by $\boldsymbol{z}_{p,q}$ the pth column of the $q \times q$ identity matrix. The banding operator is defined as $M_k(A) = \{a_{ij}\, \mathrm{I}(|i-j| \leq k)\}$ (Bickel and Levina (2008b)), where k is a positive integer. In the following, c and C denote generic positive constants which may vary from place to place, and do not depend on n.

2. Model and Method

2.1. Semiparametric model for single-voxel fMRI (Zhang and Yu (2008))

We briefly review the semiparametric model proposed by Zhang and Yu (2008). For single-voxel fMRI, the observations are $\{y(t_i), s_1(t_i), \ldots, s_\ell(t_i)\}_{i=1}^n$,

where $y(\cdot)$ is the fMRI signal, $s_j(t_i)$ is the external input of the jth type of stimulus at time t_i, and ℓ is the number of stimulus or event types. The stimulus-related variation of the signal in a particular voxel is characterized by the hemodynamic response function (HRF). Denote by $h_j(t)$ the HRF corresponding to the jth event type at time t after neural activity. Zhang and Yu (2008) assumed that $h_j(t) = 0$ for $t > t_m$, $j = 1, \ldots, \ell$, where m is far below n. The semiparametric model for single-voxel fMRI is expressed as

$$\mathbf{y} = \mathbf{S}\mathbf{h} + \mathbf{d} + \boldsymbol{\epsilon}, \tag{2.1}$$

where $\mathbf{y} = (y(t_1), \ldots, y(t_n))^T$, $\mathbf{S} = (\mathbf{S}_1, \ldots, \mathbf{S}_\ell)$, $\mathbf{h} = (\mathbf{h}_1^T, \ldots, \mathbf{h}_\ell^T)^T$, $\mathbf{d} = (d(t_1), \ldots, d(t_n))^T$, $\boldsymbol{\epsilon} = (\epsilon(t_1), \ldots, \epsilon(t_n))^T$, $t_i = i/n$ for $i = 1, \ldots, n$; $\mathbf{S}_j$ and $\mathbf{h}_j$ denote the $n \times m$ Toeplitz design matrix and the $m \times 1$ HRF vector of the jth stimulus type, respectively, for $j = 1, \ldots, \ell$,

$$\mathbf{S}_j = \begin{pmatrix} s_j(0) & 0 & \cdots & 0 \\ s_j(t_2 - t_1) & s_j(0) & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ s_j(t_m - t_1) & s_j(t_m - t_2) & \cdots & s_j(0) \\ \vdots & \vdots & \cdots & \vdots \\ s_j(t_n - t_1) & s_j(t_n - t_2) & \cdots & s_j(t_n - t_m) \end{pmatrix}_{n \times m};$$

$\mathbf{h}_j = (h_j(t_1), \ldots, h_j(t_m))^T$, $d(\cdot)$ is a smooth drift function, and $\epsilon(t)$ is a zero-mean error process with correlation matrix $\mathbf{R}$.

Estimation of $\mathbf{R}^{-1}$ is essential in semiparametric inference for fMRI. Consider the hypotheses for detecting the brain activity in a certain voxel:

$$H_0 : \mathbf{U}\mathbf{h} = \mathbf{0} \qquad \text{versus} \qquad H_1 : \mathbf{U}\mathbf{h} \neq \mathbf{0},$$

with $\mathbf{U}$ a full row rank matrix of size $K_0 \times \ell m$ and K_0 a positive integer. Zhang and Yu (2008) proposed the semiparametric test statistic $\mathbb{K} = (\mathbf{U}\widehat{\mathbf{h}})^T\{\mathbf{U}(\widetilde{\mathbf{S}}^T\widehat{\mathbf{R}}^{-1}\widetilde{\mathbf{S}})^{-1}\mathbf{U}^T\}^{-1}(\mathbf{U}\widehat{\mathbf{h}})/\{\widehat{\mathbf{r}}^T\widehat{\mathbf{R}}^{-1}\widehat{\mathbf{r}}/(n - \ell m)\}$ and the bias-corrected version $\mathbb{K}_{\mathrm{bc}} = (\mathbf{U}\widehat{\mathbf{h}}_{\mathrm{bc}})^T\{\mathbf{U}(\widetilde{\mathbf{S}}^T\widehat{\mathbf{R}}^{-1}\widetilde{\mathbf{S}})^{-1}\mathbf{U}^T\}^{-1}(\mathbf{U}\widehat{\mathbf{h}}_{\mathrm{bc}})/\{\widehat{\mathbf{r}}_{\mathrm{bc}}^T\widehat{\mathbf{R}}^{-1}\widehat{\mathbf{r}}_{\mathrm{bc}}/(n - \ell m)\}$, where $\widehat{\mathbf{R}}^{-1}$ is the estimate of $\mathbf{R}^{-1}$. They further demonstrated that under H_0, $\mathbb{K} \xrightarrow{\mathcal{D}} \chi^2_{K_0}$ and $\mathbb{K}_{\mathrm{bc}} \xrightarrow{\mathcal{D}} \chi^2_{K_0}$, assuming some regularity conditions of which one is

$$E(\|\widehat{\mathbf{R}}^{-1} - \mathbf{R}^{-1}\|_\infty^2) = o(1); \tag{2.2}$$

see Condition A8 therein. We develop a consistent estimate of $\mathbf{R}$ and a refined estimate of $\mathbf{R}^{-1}$ that fulfills (2.2).

2.2. Assumption of the error process

It's known that the error process consisting of physical and physiological processes (Lund et al. (2006)) is highly temporally correlated. Assumptions on the error process have included AR(1) (Bullmore et al. (1996)), AR(m) (Worsley et al. (2002)), ARMA(1, 1) (Purdon et al. (2001)), and g-dependence (Zhang et al. (2008); Zhang and Yu (2008)). We consider linear processes and g_n-dependent processes for the error.

Suppose $\{\epsilon(t_i)\}$ is a zero-mean linear process (see, e.g., Shumway and Stoffer (2011)) with dependence structure related to n. As in Condition A1 in the Appendix,

$$\epsilon(t_i) = \sum_{j=-\infty}^{\infty} \phi_{n;j} w_{i-j},$$

where (i) $\phi_{n;0} \equiv 1$; (ii) for any $n \geq 1$, there exists $1 \leq g_n \leq n-1$ and $\alpha_n > 4$, such that

$$|\phi_{n;j}| \leq \begin{cases} C, & \text{if } |j| \leq \frac{g_n}{2}, \\ C\,|2j|^{-\alpha_n}, & \text{if } |j| > \frac{g_n}{2}, \end{cases}$$

with a constant $C > 0$; (iii) $\{w_i\}_{i=-\infty}^{\infty}$ is a sequence of independent white noises (WN) with $E(w_i) = 0$, $E(w_i^2) = \sigma_w^2$, and $\sup_i E(w_i^4) < \infty$.

Assumption (ii) illustrates the decay rate of the coefficients, and $\alpha_n > 4$ guarantees that $\sum_{j=-\infty}^{\infty} |\phi_{n;j}| \leq C g_n$. From (iii), the error process has zero mean and a finite fourth moment for any fixed n, which is weaker than the Gaussian tail (Cai, Ren, and Zhou (2012)) and higher order of polynomial tail (Xiao and Wu (2012)) assumptions. Condition A1 is satisfied by, for example, the zero-mean stationary ARMA(p, q) model with a finite fourth moment and independent white noises, for any $1 \leq g_n \leq n-1$ and any fixed $\alpha_n > 4$.

Motivated by the g-dependence assumption in Zhang et al. (2008) and Zhang and Yu (2008), we consider another class of error process with $\boldsymbol{\epsilon}$ a zero-mean stationary g_n-dependent sequence, where g_n can be bounded or diverge with n, see Condition B1 in the Appendix. This assumption includes the g-dependent process (Zhang et al. (2008); Zhang and Yu (2008)) as a special case. There is overlap between A1 and B1, for example the zero-mean MA(g_n) process with a finite fourth moment and independent white noises satisfies both.

Under A1 or B1, the true autocorrelation matrix of $\{\epsilon(t_i)\}$ is Toeplitz,

$$\mathbf{R} = \begin{pmatrix} 1 & \rho(1) & \cdots & \rho(n-1) \\ \rho(1) & 1 & \cdots & \rho(n-2) \\ \vdots & \vdots & \ddots & \vdots \\ \rho(n-1) & \rho(n-2) & \cdots & 1 \end{pmatrix},$$

where $\rho(k) = \gamma(k)/\gamma(0)$, for $k = 1, \ldots, n-1$ and $\gamma(k) = \text{cov}\{\epsilon(t_i), \epsilon(t_{i+k})\}$ is the autocovariance.

Remark 1. Under Condition A1,

$$\gamma(k) = \sigma_w^2 \sum_{j=-\infty}^{\infty} \phi_{n;j}\phi_{n;j+|k|}, \quad k = 0, \pm 1, \ldots, \pm(n-1). \tag{2.3}$$

From (2.3), we can show that $\sigma_w^2 \leq \gamma(0) \leq Cg_n$, and

$$|\gamma(k)| \leq \begin{cases} C(g_n + 1 - k), & \text{if } 0 \leq k \leq g_n, \\ C(2k - g_n)^{1-\alpha_n}, & \text{if } g_n < k \leq n-1. \end{cases}$$

Therefore, A1 is significantly different from the assumptions in Cai, Ren, and Zhou (2012) and Xiao and Wu (2012), where the dependence structures of the time series are not related to n. Under B1, $\gamma(k) = 0$ for $k > g_n$.

Under A1 then,

$$|\rho(k)| \leq \begin{cases} \dfrac{C(g_n + 1 - k)}{\gamma(0)}, & \text{if } 0 \leq k \leq g_n, \\ \dfrac{C(2k - g_n)^{1-\alpha_n}}{\gamma(0)}, & \text{if } g_n < k \leq n-1, \end{cases}$$

so $\mathbf{R}$ is a "bandable" matrix where the entries decay as they move away from the diagonal. The decay rate of $|\rho(k)|$ for $k > g_n$ is much faster than that for $k \leq g_n$. Thus, g_n could be regarded as the natural banding parameter of $\mathbf{R}$. Under B1, $\mathbf{R}$ is a banded matrix with g_n the exact banding parameter.

2.3. Estimation of the error autocorrelation matrix

Several methods have been proposed to estimate large Toeplitz covariance matrices, see Cai, Ren, and Zhou (2012), McMurry and Politis (2010), Wu and Pourahmadi (2009), and Xiao and Wu (2012), among others. Essentially, they consider regularizing the sample covariance matrix, which is directly constructed from the observed time series. However, for fMRI data, the error process is unobservable. To obtain the error autocovariance estimates, we adopt the difference-based method (Zhang et al. (2008)), that is computationally fast. Since $\mathbf{R}$ is "bandable" under either of the two conditions for the error process, we construct a g_n-banded estimate.

At (2.1), we take the second-order difference of $(\mathbf{y} - \mathbf{Sh})$ and get

$$\mathbf{e} \equiv \mathbf{D}_2(\mathbf{y} - \mathbf{Sh}) = \mathbf{D}_2(\boldsymbol{\epsilon} + \mathbf{d}), \tag{2.4}$$

where $\mathbf{D}_2$ is the $(n-2)\times n$ matrix

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

Here, $e(t_i) = \{\epsilon(t_i) - 2\epsilon(t_{i-1}) + \epsilon(t_{i-2})\} + \{d(t_i) - 2d(t_{i-1}) + d(t_{i-2})\}$, for $i = 3, \ldots, n$, where $e(t_i) = \boldsymbol{z}_{i-2,n-2}^T \mathbf{e}$ and $\boldsymbol{z}_{p,q}$ is the pth column of the $q \times q$ identity matrix. Let $\gamma_e(k) = \operatorname{cov}\{e(t_i), e(t_{i+k})\}$ for $0 \le k \le n-3$. Then

$$\gamma_e(k) = \gamma(k-2) - 4\gamma(k-1) + 6\gamma(k) - 4\gamma(k+1) + \gamma(k+2), \quad k = 0, \ldots, n-3,$$

which implies

$$\begin{pmatrix} \gamma_e(0) \\ \vdots \\ \gamma_e(k) \end{pmatrix} = A_k \begin{pmatrix} \gamma(0) \\ \vdots \\ \gamma(k) \end{pmatrix} + B_k \begin{pmatrix} \gamma(k+1) \\ \gamma(k+2) \end{pmatrix}, \quad k = 0, \ldots, n-3, \tag{2.5}$$

where A_k is a $(k+1)\times(k+1)$ matrix

$$A_k = \begin{pmatrix} 6 & -8 & 2 & 0 & 0 & \cdots & 0 & 0 & 0 & 0 \\ -4 & 7 & -4 & 1 & 0 & \cdots & 0 & 0 & 0 & 0 \\ 1 & -4 & 6 & -4 & 1 & \cdots & 0 & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & 0 & \cdots & 0 & 1 & -4 & 6 \end{pmatrix}_{(k+1)\times(k+1)},$$

and B_k is a $(k+1)\times 2$ matrix, satisfying

$$B_0 = (-8,\ 2), \quad B_1 = \begin{pmatrix} 2 & 0 \\ -4 & 1 \end{pmatrix}, \quad B_k^T = \begin{pmatrix} 0 & \cdots & 0 & 1 & -4 \\ 0 & \cdots & 0 & 0 & 1 \end{pmatrix},\ k \ge 2.$$

The basic idea of the difference-based method is to first obtain the empirical estimates of $\gamma_e(k)$ for $k = 0, \ldots, g_n$, and then, by (2.5), acquire the estimates of $\gamma(k)$ for $k = 0, \ldots, g_n$, which enables us to construct the error autocorrelation matrix estimate. A two-step procedure describes in detail how to build a g_n-banded estimate of $\mathbf{R}$.

Step I. An initial estimate of $\mathbf{h}$ is constructed based on the first-order difference of $\mathbf{y}$.

Let $\mathbf{D}_1$ be an $(n-1)\times n$ matrix

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

The first-order difference of $\mathbf{y}$ yields

$$\mathbf{D}_1\mathbf{y} = \mathbf{D}_1\mathbf{S}\mathbf{h} + \mathbf{D}_1\mathbf{d} + \mathbf{D}_1\boldsymbol{\epsilon}. \tag{2.6}$$

By smoothness of the drift function in A3, $d(t_i) - d(t_{i-1}) = d'(t_{i-1})n^{-1} + O(n^{-2}) = O(n^{-1})$, and the term $\mathbf{D}_1\mathbf{d}$ in (2.6) can be ignored. Ordinary least-squares method provides an initial estimate of the HRF,

$$\widehat{\mathbf{h}}_{\mathrm{DBE}} = \{(\mathbf{D}_1\mathbf{S})^T(\mathbf{D}_1\mathbf{S})\}^{-1}(\mathbf{D}_1\mathbf{S})^T(\mathbf{D}_1\mathbf{y}).$$

Step II. Construct the g_n-banded estimate of $\mathbf{R}$ based on the empirical estimates of $\gamma_e(k)$, for $k = 0, \ldots, g_n$.

Substituting $\widehat{\mathbf{h}}_{\mathrm{DBE}}$ for $\mathbf{h}$ in (2.4), we get $\widehat{\mathbf{e}} \equiv \mathbf{D}_2(\mathbf{y} - \mathbf{S}\widehat{\mathbf{h}}_{\mathrm{DBE}})$. The empirical estimate of $\gamma_e(k)$ is

$$\widehat{\gamma}_e(k) = \frac{1}{n}\sum_{i=3}^{n-k} \widehat{e}(t_i)\widehat{e}(t_{i+k}), \quad k = 0, \ldots, g_n, \tag{2.7}$$

where $\widehat{e}(t_i) = \boldsymbol{z}_{i-2,n-2}^T\widehat{\mathbf{e}}$, for $i = 3, \ldots, n$. It is also used by Wu and Pourahmadi (2009). By taking $k = g_n$ in (2.5),

$$\boldsymbol{\gamma}_e = A_{g_n}\boldsymbol{\gamma} + B_{g_n}\begin{pmatrix}\gamma(g_n+1) & \gamma(g_n+2)\end{pmatrix}^T, \tag{2.8}$$

where $\boldsymbol{\gamma}_e = (\gamma_e(0), \ldots, \gamma_e(g_n))^T$ and $\boldsymbol{\gamma} = (\gamma(0), \ldots, \gamma(g_n))^T$. Since $\gamma(g_n+1)$ and $\gamma(g_n+2)$ are negligible, as shown in Remark 1, we drop the second term on the right hand side of (2.8) and estimate $\boldsymbol{\gamma}$ by

$$\widehat{\boldsymbol{\gamma}} = A_{g_n}^{-1}\widehat{\boldsymbol{\gamma}}_e, \tag{2.9}$$

where $\widehat{\boldsymbol{\gamma}}_e = (\widehat{\gamma}_e(0), \ldots, \widehat{\gamma}_e(g_n))^T$. The estimate of $\rho(k)$ is derived as

$$\widehat{\rho}(k) = \begin{cases} \dfrac{\widehat{\gamma}(k)}{\widehat{\gamma}(0)}, & \text{if } 0 \le k \le g_n, \\ 0, & \text{if } g_n < k \le n-1, \end{cases}$$

where $\widehat{\gamma}(k) = \boldsymbol{z}_{k+1,g_n+1}^T\widehat{\boldsymbol{\gamma}}$, and the g_n-banded estimate of $\mathbf{R}$ is

$$\widehat{\mathbf{R}} = \begin{pmatrix} 1 & \widehat{\rho}(1) & \cdots & \widehat{\rho}(n-1) \\ \widehat{\rho}(1) & 1 & \cdots & \widehat{\rho}(n-2) \\ \vdots & \vdots & \ddots & \vdots \\ \widehat{\rho}(n-1) & \widehat{\rho}(n-2) & \cdots & 1 \end{pmatrix}. \tag{2.10}$$

This estimation procedure is for single-run fMRI. If we have a fixed number of replications, the estimate of the autocorrelation matrix can be acquired as follows.

Let $(\mathbf{y}^1, \mathbf{S}^1), \ldots, (\mathbf{y}^N, \mathbf{S}^N)$ be the observations, where $\mathbf{y}^i$ and $\mathbf{S}^i$ are the vector of response and the design matrix for the ith realization, and N is a fixed integer denoting the number of replications. From Step I above, we obtain the initial estimates of the HRFs, $\widehat{\mathbf{h}}^i_{\mathrm{DBE}} = \{(\mathbf{D}_1\mathbf{S}^i)^T\ (\mathbf{D}_1\mathbf{S}^i)\}^{-1}(\mathbf{D}_1\mathbf{S}^i)^T(\mathbf{D}_1\mathbf{y}^i)$ for $i = 1, \ldots, N$. The empirical estimates $\widehat{\boldsymbol{\gamma}}^i_e$ is based on $\mathbf{D}_2(\mathbf{y}^i - \mathbf{S}^i\widehat{\mathbf{h}}^i_{\mathrm{DBE}})$ following Step II. We estimate $\boldsymbol{\gamma}_e$ by

$$\widehat{\boldsymbol{\gamma}}_e = \frac{1}{N}\sum_{i=1}^{N}\widehat{\boldsymbol{\gamma}}^i_e. \tag{2.11}$$

Thereafter, the same approaches as in Step II lead to $\widehat{\mathbf{R}}$.

The estimate (2.11) parallels the one proposed in Cai, Ren, and Zhou (2012). In the following, we focus on single-run fMRI.

3. Theoretical Results

We establish an explicit convergence rate of $\widehat{\mathbf{R}}$, develop a positive-definite estimate of $\mathbf{R}^{-1}$, and demonstrate consistency of the estimate under the L_∞ norm. The conditions and proofs are given in the Appendix, available on the online supplement.

Theorem 1. *Assume* A2−A5 *in the Appendix. Under either of the following two assumptions:*

1. A1 *in the Appendix holds,* $n = o(g_n^{8+2\alpha_n} + 2^{\alpha_n})$ *and* $g_n^{14}/n = o(1)$,
2. B1 *in the Appendix holds and* $g_n^{14}/n = o(1)$,

for $\widehat{\mathbf{R}}$ *in* (2.10), $\|\widehat{\mathbf{R}} - \mathbf{R}\|_\infty = O_P(g_n^7/n^{1/2})$.

In Theorem 1, the banded estimate $\widehat{\mathbf{R}}$ is consistent under the L_∞ norm with convergence rate $g_n^7/n^{1/2}$. Recall that $\boldsymbol{\gamma}$ is estimated by $\widehat{\boldsymbol{\gamma}} = A_{g_n}^{-1}\widehat{\boldsymbol{\gamma}}_e$, as in (2.9). In Lemma 1 in the Appendix, we show that $A_{g_n}^{-1}$ exists for any $g_n \geq 1$ and $\|A_{g_n}^{-1}\|_1 = O(g_n^4)$ as $n \to \infty$, the major reason for the power of g_n in the convergence rate. It is the price we have to pay for the difference-based method in order to gain computational efficiency. From Theorem 1, we obtain the same convergence rate for $\widehat{\mathbf{R}}$ under either the linear process assumption or the g_n-dependence assumption for the error process.

For the ARMA(p, q) model that satisfies A1, we can take g_n to satisfy the rate assumptions in Theorem 1 with any constant-valued $\alpha_n > 4$.

Remark 2. In Xiao and Wu (2012), the optimal convergence rate of the banded autocovariance matrix estimate of stationary process is obtained. We compare our convergence rate with theirs under the following conditions. Suppose that $\{\epsilon(t_i)\}$ are observed and $\epsilon(t_i) = \sum_{j=0}^{\infty}\phi_j w_{i-j}$, where (i) $\phi_0 = 1$, (ii) $|\phi_j| \leq$

$C|2j|^{-\alpha_n}$, for $j \neq 0$, with $\alpha_n \equiv \alpha_0 + 1$, $\alpha_0 > 3$, (iii) $\{w_i\}_{i=-\infty}^{\infty}$ are i.i.d. white noises with $E(w_i) = 0$, $E(w_i^2) = \sigma_w^2$ and $E(w_i^p) < \infty$ for a constant $p > 4$.

In Xiao and Wu (2012), the convergence rate of the banded estimate under the L_2 norm is $\{\log(n)/n\}^{\alpha_0/(2\alpha_0+1)}$. For our convergence rate, the assumption $n = o(g_n^{8+2\alpha_n} + 2^{\alpha_n})$ implies $(1/n)^{(\alpha_0-2)/(2\alpha_0+10)} = o(g_n^7/n^{1/2})$. Since $\alpha_0/(2\alpha_0 + 1) > (\alpha_0 - 2)/(2\alpha_0 + 10)$, the convergence rate in Xiao and Wu (2012) is sharper. One important reason is that they considered the matrix L_2 norm in order to make use of the spectral density of the stationary process, while we use the L_∞ norm which is not smaller than the L_2 norm.

Proposition 1. *Assume* A1–A5 *in the Appendix. If* $\alpha_n g_n \to \infty$ *and* $g_n^{14}/n = o(1)$, *for* $\widehat{\mathbf{R}}$ *in* (2.10), $\|\widehat{\mathbf{R}} - \mathbf{R}\|_\infty = o_P(1)$.

Proposition 2. *Assume* A2−A6 *in the Appendix. Under either of the following two assumptions:*

1. A1 *in the Appendix holds,* $\alpha_n g_n \to \infty$ *and* $g_n^{14}/n = o(1)$,
2. B1 *in the Appendix holds and* $g_n^{14}/n = o(1)$,

for $\widehat{\mathbf{R}}$ *in* (2.10), $\lim_{n\to\infty} \mathrm{P}(\widehat{\mathbf{R}} \succ 0) = 1$.

While Proposition 2 has $\widehat{\mathbf{R}}$ positive definite with probability tending to one, in finite-sample situations, $\widehat{\mathbf{R}}$ may not be. To obtain positive-definite estimates, several modification methods have been proposed (e.g., Cai and Zhou (2012); Wu and Xiao (2011); Zhang et al. (2008); Zhang and Yu (2008)). Zhang et al. (2008) and Zhang and Yu (2008) proposed a positive-definite estimate of $\mathbf{R}^{-1}$ by refining $\widehat{\mathbf{R}}^{-1}$. Their estimate is

$$\widehat{\mathbf{R}}_Z^{-1} = \begin{cases} \widehat{\mathbf{R}}^{-1}, & \text{if } \widehat{\mathbf{R}} \succ 0, \\ \mathbf{I}_n, & \text{otherwise,} \end{cases}$$

where $\mathbf{I}_n$ is the $n \times n$ identity matrix and $\widehat{\mathbf{R}}$ is constructed with $g_n = 2$. Motivated by Cai and Zhou (2012), we propose an estimate of $\mathbf{R}^{-1}$ by refining $\widehat{\mathbf{R}}_Z^{-1}$,

$$\widehat{\mathbf{R}}_*^{-1} = \begin{cases} \widehat{\mathbf{R}}^{-1}, & \text{if } \widehat{\mathbf{R}} \succ 0 \text{ and } \|\widehat{\mathbf{R}}^{-1}\|_\infty \le Dn^\omega, \\ \mathbf{I}_n, & \text{otherwise,} \end{cases} \tag{3.1}$$

where $0 < \omega < 1/2$ and $D > 0$ are constants.

When $\widehat{\mathbf{R}}$ is not positive definite, we estimate $\mathbf{R}^{-1}$ by $\mathbf{I}_n$. This is reasonable, because it is the true autocorrelation matrix when the error process is uncorrelated. Even if $\widehat{\mathbf{R}}$ is positive definite, it is possible that the eigenvalues of $\widehat{\mathbf{R}}$ are close to 0, resulting in a large scale of $\|\widehat{\mathbf{R}}^{-1}\|_\infty$. In order to exclude the case that $\widehat{\mathbf{R}}$ is almost singular, we set a threshold on $\|\widehat{\mathbf{R}}^{-1}\|_\infty$. Thus, $\widehat{\mathbf{R}}_*^{-1}$ is a positive-definite estimate of $\mathbf{R}^{-1}$ with the L_∞ norm bounded by $\max\{Dn^\omega, 1\}$. As an estimate of $\mathbf{R}$, $\widehat{\mathbf{R}}_*$ could be viewed as a refined version of $\widehat{\mathbf{R}}$ and $\widehat{\mathbf{R}}_Z$.

Proposition 3. *Under* A6 *in the Appendix and the conditions of Theorem* 1*,*

$$\|\widehat{\mathbf{R}}_Z - \mathbf{R}\|_\infty = O_P(g_n^7/n^{1/2}),\ \|\widehat{\mathbf{R}}_* - \mathbf{R}\|_\infty = O_P(g_n^7/n^{1/2}).$$

Theorem 2. *Assume* A2–A6 *in the Appendix. Under either of the following two assumptions:*

1. A1 *in the Appendix holds,* $\alpha_n g_n \to \infty$ *and* $g_n^{14}/n^{1-2\omega} = o(1)$*,*
2. B1 *in the Appendix holds and* $g_n^{14}/n^{1-2\omega} = o(1)$*,*

for $\widehat{\mathbf{R}}_*^{-1}$ *defined in* (3.1)*, we have*

$$\lim_{n\to\infty} E(\|\widehat{\mathbf{R}}_*^{-1} - \mathbf{R}^{-1}\|_\infty^2) = 0.$$

This result has $\widehat{\mathbf{R}}_*^{-1}$ fulfilling Condition A8 in Zhang and Yu (2008) under either of the two conditions for the error process. Hence, $\widehat{\mathbf{R}}_*^{-1}$ offers a satisfactory estimate of $\mathbf{R}^{-1}$ for the semiparametric inference for brain fMRI data. For $\widehat{\mathbf{R}}_Z^{-1}$, the convergence property described in Theorem 2 also holds under certain conditions, e.g., $0 < c \le \lambda_{\min}(\widehat{\mathbf{R}}_Z) \le \lambda_{\max}(\widehat{\mathbf{R}}_Z) < C$ or $\|\widehat{\mathbf{R}}_Z^{-1}\|_\infty \le C$. These conditions are relaxed here as $\|\widehat{\mathbf{R}}_*^{-1}\|_\infty$ is allowed to diverge with n.

Proposition 4. *Under* A6 *in the Appendix and the conditions of Theorem* 1*,*

$$\|\widehat{\mathbf{R}}^{-1} - \mathbf{R}^{-1}\|_\infty = O_P\Big(\frac{g_n^7}{n^{1/2}}\Big),\ \|\widehat{\mathbf{R}}_Z^{-1} - \mathbf{R}^{-1}\|_\infty = O_P\Big(\frac{g_n^7}{n^{1/2}}\Big),$$
$$\|\widehat{\mathbf{R}}_*^{-1} - \mathbf{R}^{-1}\|_\infty = O_P\Big(\frac{g_n^7}{n^{1/2}}\Big).$$

4. Data-driven Methods for g_n and $\widehat{\mathbf{R}}_*^{-1}$

4.1. Choosing the banding parameter g_n

Results in Section 3 give rates of g_n that lead to consistency of $\widehat{\mathbf{R}}$ and $\widehat{\mathbf{R}}_*^{-1}$, but they do not give much guidance for the choice of g_n in practice. To select the banding parameter, we utilize the idea of risk-minimization (Bickel and Levina (2008b)) and the technique of subsampling (Politis, Romano, and Wolf (1999)).

Since the estimates of $\gamma_e(k)$ and $\gamma(k)$ do not perform well when k is close to n, we select g_n from $\{0,\dots,T\}$ instead of $\{0,\dots,n-1\}$, where T is a predetermined integer much less than $(n-1)$. In the theoretical analysis, we took $1 \le g_n \le n-1$, but in applications, one needs to consider the case of $g_n = 0$. Let $\boldsymbol{\gamma}_{e,u} = (\gamma_e(0),\dots,\gamma_e(u),0,\dots,0)^T$ and $\boldsymbol{\gamma}_u = (\gamma(0),\dots,\gamma(u),0,\dots,0)^T$ be vectors of length $T+1$, for $u = 0,\dots,T$. Denote by $\widehat{\boldsymbol{\gamma}}_{e,u} = (\widehat{\gamma}_e(0),\dots,\widehat{\gamma}_e(u),0,\dots,0)^T$ the estimate of $\boldsymbol{\gamma}_{e,u}$, with $\widehat{\gamma}_e(k)$ calculated in (2.7) and by $\widehat{\boldsymbol{\gamma}}_u = (\widehat{\gamma}(0),\dots,\widehat{\gamma}(u),0,\dots,0)^T$ the estimate of $\boldsymbol{\gamma}_u$, satisfying $(\widehat{\gamma}(0),\dots,\widehat{\gamma}(u))^T = A_u^{-1}(\widehat{\gamma}_e(0),\dots,\widehat{\gamma}_e(u))^T$. Based on the risk-minimization method, we develop a two-step procedure to select g_n.

Step 1. The initial choice of the banding parameter g_e is that which minimizes the risk

$$r_e(g) = E(\|\widehat{\boldsymbol{\gamma}}_{e,g} - \boldsymbol{\gamma}_{e,T}\|_1), \tag{4.1}$$

for $g \in \{2, \ldots, T\}$.

We take g_e not less than 2 because $\mathbf{e}$, defined in (2.4), is the second-order difference of $\boldsymbol{\epsilon} + \mathbf{d}$. For the simplest case, if $\boldsymbol{\epsilon}$ is a sequence of white noises, $\mathbf{e}$ would be 2-dependent.

Step 2. Select the banding parameter as the minimizer of

$$r(g) = E(\|\widehat{\boldsymbol{\gamma}}_g - \boldsymbol{\gamma}_{g_e}\|_1), \tag{4.2}$$

for $g \in \{0, \ldots, g_e\}$.

Since the Toeplitz autocorrelation matrix is uniquely determined by the autocovariances, our risk functions (4.1) and (4.2) parallel the one in Wu and Pourahmadi (2009), which is constructed based on the Toeplitz autocovariance matrix and its estimate.

We subsample to estimate the risk functions. For $\{\widehat{e}(t_i)\}_{i=3}^n$, blocks of consecutive elements are used as legitimate subsamples. The length of each block is b, predetermined in practice. Therefore, there are $(n-b-1)$ blocks, and the νth one is $\{\widehat{e}(t_{\nu+2}), \widehat{e}(t_{\nu+3}), \ldots, \widehat{e}(t_{\nu+b+1})\}$, for $\nu = 1, \ldots, n-b-1$. We select V subsamples, denoted by $G_1, \ldots, G_V$, such that $G_\mu = \{\widehat{e}(t_{(\mu-1)\lfloor(n-b-2)/(V-1)\rfloor+3}), \ldots, \widehat{e}(t_{(\mu-1)\lfloor(n-b-2)/(V-1)\rfloor+b+2})\}$, for $\mu = 1, \ldots, V$, where $\lfloor\cdot\rfloor$ denotes the floor function. To enhance the computational efficiency, only $G_1, \ldots, G_V$ are used. Denote by $\widehat{\boldsymbol{\gamma}}_{e,u}^\nu = (\widehat{\gamma}_e^\nu(0), \ldots, \widehat{\gamma}_e^\nu(u), 0, \ldots, 0)^T$ and $\widehat{\boldsymbol{\gamma}}_u^\nu = (\widehat{\gamma}^\nu(0), \ldots, \widehat{\gamma}^\nu(u), 0, \ldots, 0)^T$ the estimates of $\boldsymbol{\gamma}_{e,u}$ and $\boldsymbol{\gamma}_u$ based on the subsample G_ν, respectively.

For (4.1), alternatively, each of $G_1, \ldots, G_V$ is used to estimate $\boldsymbol{\gamma}_{e,T}$, which serves as the "target", and the other $(V-1)$ blocks are used to estimate $\boldsymbol{\gamma}_{e,g}$ for $g = 2, \ldots, T$. The "target" is a vector of length $T+1$, with T much smaller than n, so the number of estimated parameters in the "target" is much reduced and we use just one subsample to estimate the "target". The estimated risk is

$$\widehat{r}_e(g) = \frac{1}{V(V-1)} \sum_{\nu=1}^{V} \sum_{\mu=1, \mu\neq\nu}^{V} \|\widehat{\boldsymbol{\gamma}}_{e,g}^\mu - \widehat{\boldsymbol{\gamma}}_{e,T}^\nu\|_1.$$

The initial choice $\widehat{g}_e$ is the minimizer of $\widehat{r}_e(g)$ for $g \in \{2, \ldots, T\}$. Similarly, we estimate $r(g)$ by

$$\widehat{r}(g) = \frac{1}{V(V-1)} \sum_{\nu=1}^{V} \sum_{\mu=1, \mu\neq\nu}^{V} \|\widehat{\boldsymbol{\gamma}}_g^\mu - \widehat{\boldsymbol{\gamma}}_{\widehat{g}_e}^\nu\|_1,$$

and the banding parameter is selected as $\widehat{g} = \arg\min_{0 \le g \le \widehat{g}_e} \widehat{r}(g)$.

4.2. Choosing D and ω for $\widehat{\mathbf{R}}_*^{-1}$

To calculate $\widehat{\mathbf{R}}_*^{-1}$, we develop data-driven methods for the choices of $D > 0$ and $0 < \omega < 1/2$. After $\widehat{g}$ is selected, if $\widehat{\mathbf{R}}$ is not positive definite, from (3.1), $\mathbf{I}_n$ is used to estimate $\mathbf{R}^{-1}$ and there is no need to choose D and ω. Therefore, for the choices of D and ω, we only consider that $\widehat{\mathbf{R}} \succ 0$ which implies that $\widehat{\mathbf{R}}^{-1}$ exists. In finite-sample situations, we can fix ω to be $1/2$, since simulation results (not shown) reveal that for different values of $0 < \omega \le 1/2$, with a proper choice of D, the calculated $\widehat{\mathbf{R}}_*^{-1}$ are not significantly different. In the following, we will choose D from a proper interval on $\mathbb{R}^+$.

Following the procedure in Section 4.1, we take subsamples $G_1, \ldots, G_V$. Based on G_ν, $\nu = 1, \ldots, V$, denote by $\widehat{\mathbf{R}}_\nu$ the estimate of $\mathbf{R}$, and then the refined estimate of $\mathbf{R}^{-1}$ is

$$\widehat{\mathbf{R}}_{*;\nu,D}^{-1} = \begin{cases} \widehat{\mathbf{R}}_\nu^{-1}, & \text{if } \widehat{\mathbf{R}}_\nu \succ 0, \text{ and } \|\widehat{\mathbf{R}}_\nu^{-1}\|_\infty \le Dn^{1/2}, \\ \mathbf{I}_n, & \text{otherwise.} \end{cases}$$

Then, $\widehat{D}$ is selected as the minimizer of

$$\widehat{r}_*(D) = \frac{1}{V} \sum_{\nu=1}^{V} \|\widehat{\mathbf{R}}_{*;\nu,D}^{-1} - \widehat{\mathbf{R}}^{-1}\|_\infty,$$

by grid search method.

5. Numerical Results

We report on simulation studies to evaluate the data-driven methods in Section 4, check the performance of $\widehat{\mathbf{R}}_*^{-1}$ by applying it to the test statistics $\mathbb{K}$ and $\mathbb{K}_{\mathrm{bc}}$ (Zhang and Yu (2008)) and compare the performance of $\widehat{\mathbf{R}}_*$ and $\widehat{\mathbf{R}}_Z$ in Zhang and Yu (2008).

The hypotheses $H_0 : \mathbf{h} = \mathbf{0}$ versus $H_1 : \mathbf{h} \neq \mathbf{0}$ are considered to test whether there's neural activity in a voxel or not. We generated data for a single run with $n = 400$. According to the null hypothesis, $h_j(t_i) = 0$ for $i = 1, \ldots, m$, $j = 1, \ldots, \ell$. We used $d(t_i) = 10 \sin\{\pi(t_i - 0.21)\}$ as the drift function. Three types of errors were studied: MA(4), ARMA(1, 3), and the sum of AR(1) and a white noise. For each type of error, we considered both a single event type ($\ell = 1$) with $m = 20$, and two event types ($\ell = 2$) with $m = 15$. Stimuli were generated as follows: for $\ell = 1$, $\{s_1(t_i)\}$ are independent Bernoulli(0.5); for $\ell = 2$, $\{s_1(t_i), s_2(t_i), 1 - s_1(t_i) - s_2(t_i)\}$ are independent from the trinomial distribution (1; 1/3, 1/3, 1/3). For the local linear smoothing matrix S_d used in $\mathbb{K}$ and $\mathbb{K}_{\mathrm{bc}}$, we used the Epanechnikov kernel function; see (3.3) in Zhang and Yu (2008) for details. The bandwidth parameter of S_d was selected by the "pWPL" method proposed in Zhang et al. (2008).

The methods of choosing the banding parameter g_n and selecting D and ω for $\widehat{\mathbf{R}}_*^{-1}$ are illustrated in Section 4. In our simulation study, we took $V = 20$. For the block size b, as a general guide, Politis, Romano, and Wolf (1999) showed that b should grow to infinity while $b/n \to 0$ with a rate $n^{1/3}$; we took $b = \lfloor C_b n^{1/3} \rfloor$ where $C_b > 0$ is a constant. Simulation results (not shown) reveal that smaller values of C_b correspond to underestimating the banding parameter and larger values of C_b result in larger estimates. The choice of $C_b = 8$ has overall good performance in selecting the banding parameter for different models of the error process; we took $b = \lfloor 8n^{1/3} \rfloor$ in the calculation. The choice of T is also important in practice. One can use $T \le n/4$ as in Box and Jenkins (1976), but the default value in R is $T = 10\log(10n)$; we use $T = \lfloor C_T \log(10n) \rfloor$ with a positive constant C_T, provided $T < b$. Simulation results (not shown) show that the choice of C_T has no significant impact on the selection of the banding parameter. Taking computational efficiency into consideration, we fix $C_T = 3$ in practice, so $T = \lfloor 3\log(10n) \rfloor$.

To investigate the performance of $\widehat{\mathbf{R}}_*^{-1}$, we carried out 500 realizations in each simulation study and drew a QQ plot of the empirical (1st to 99th) percentiles of $\mathbb{K}$ and $\mathbb{K}_{\mathrm{bc}}$ versus the theoretical percentiles of the $\chi^2_{\ell m}$ distribution. We also compared the performance of $\widehat{\mathbf{R}}_*$ and $\widehat{\mathbf{R}}_Z$ in Zhang and Yu (2008).

5.1. Example 1: MA(4) error process

The error process was generated by $\epsilon(t_i) = z(t_i) + \theta_1 z(t_{i-1}) + \theta_2 z(t_{i-2}) + \theta_3 z(t_{i-3}) + \theta_4 z(t_{i-4})$, where $\theta_1 = 0.75$, $\theta_2 = 0.5$, $\theta_3 = 0.25$ and $\theta_4 = 0.35$. For a single event type, $\{z(t_i)\}$ were independent normal with mean 0 and variances 0.4786^2 and $0.4786^2/8$, respectively. These choices give the signal-to-noise ratios (SNRs) of about 1 and 8, respectively, where SNR $= \mathrm{var}(\mathbf{Sh})/\mathrm{var}(\boldsymbol{\epsilon})$. For two event types, $\{z(t_i)\}$ had the same distribution but with variances 0.4575^2 and $0.4575^2/8$, corresponding to SNRs about 1 and 8.

In order to see the importance of estimating the autocorrelation matrix $\mathbf{R}$, Figure 1 shows QQ plots of the percentiles of $\mathbb{K}$ and $\mathbb{K}_{\mathrm{bc}}$ against those of the $\chi^2_{\ell m}$ distribution. For the test statistics calculated in Figure 1, $\mathbf{R}$ was estimated by $\mathbf{I}_n$, which means the autocorrelation structure of the error process was ignored. The test statistics calculated in the top panels are for a single event type and those in the bottom panels are for two event types. From Figure 1, the finite sampling distributions of the test statistics calculated with $\mathbf{R}$ estimated by $\mathbf{I}_n$ do not show reasonable agreement with the $\chi^2_{\ell m}$ distribution; one should take the autocorrelation structure of the error process into consideration for fMRI data analysis.

In Figure 2, we plot $\widehat{r}_e(g)$ versus g in the left panels and $\widehat{r}(g)$ versus g in the right panels for the MA(4) error process with SNR = 1 based on a single

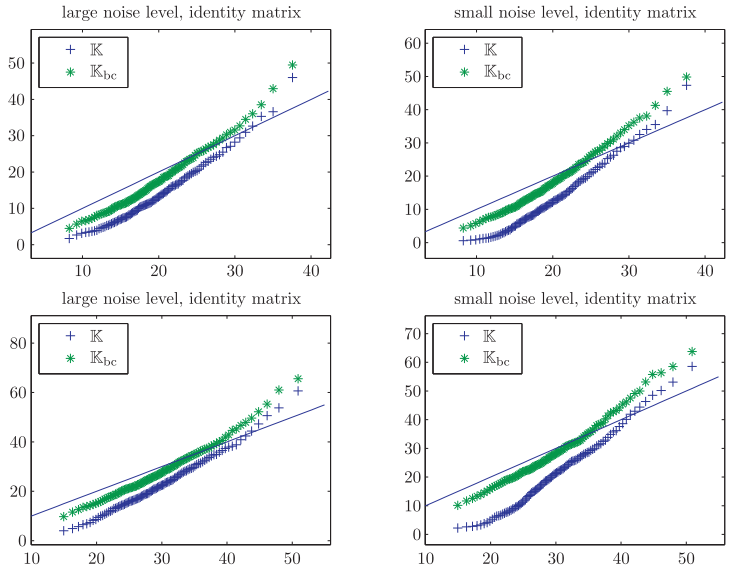

Figure 1. Empirical QQ plots of $\mathbb{K}$ and $\mathbb{K}_{\mathrm{bc}}$ for the MA(4) error process, statistics calculated with $\mathbf{R}$ estimated by $\mathbf{I}_n$ and data-driven method for bandwidth choice. Statistics in the top panels calculated for a single event type, those in the bottom panels for two event types. The solid line is the 45 degree straight line.

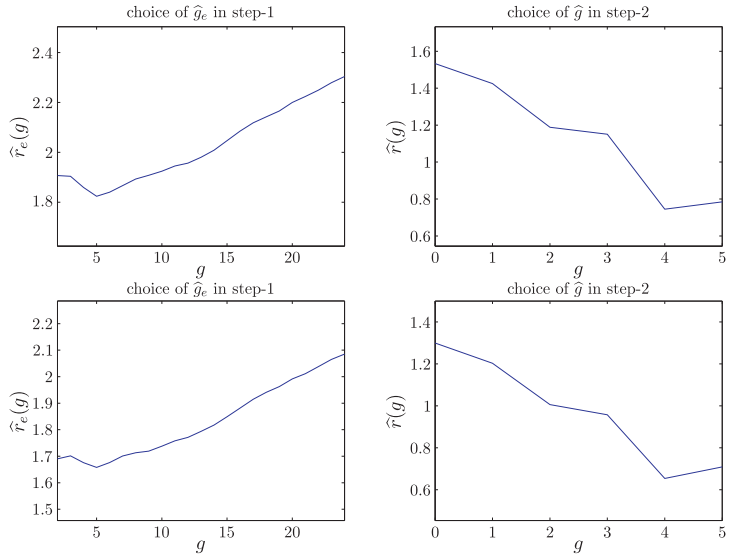

Figure 2. Plots of $\widehat{r}_e(g)$ versus g (left panels) and $\widehat{r}(g)$ versus g (right panels) for the MA(4) error process with SNR = 1 based on one realization. The top panels are for a single event type and the bottom panels correspond to two event types.

realization. The top panels correspond to a single type of stimulus, and the bottom panels are for two event types. Each plot shows a "V" shape with a global minimizer, which indicates that the data-driven method for choosing g_n is valid.

Figure 3 gives QQ plots of the percentiles of $\mathbb{K}$ and $\mathbb{K}_{\mathrm{bc}}$ against those of the $\chi^2_{\ell m}$ distribution for a single type of stimulus; Figure 4 shows QQ plots constructed for two event types. In each figure, statistics calculated with the true inverse autocorrelation matrix $\mathbf{R}^{-1}$ and the theoretical optimal choice of bandwidth (see, e.g., (3) in Zhang (2003)) are plotted in the top panels, while those using $\widehat{\mathbf{R}}_*^{-1}$ and data-driven method for the bandwidth choice are plotted in the bottom panels. The left and right panels correspond to large (SNR = 1) and small (SNR = 8) noise levels, respectively.

In the top panels of Figure 3, the finite sampling distributions of $\mathbb{K}$ and $\mathbb{K}_{\mathrm{bc}}$ show reasonable agreement with the $\chi^2_{\ell m}$ distribution for both large and small noise levels. For the plots in the bottom panels, we draw the same conclusion; the empirical distributions of $\mathbb{K}$ and $\mathbb{K}_{\mathrm{bc}}$ calculated with $\mathbf{R}^{-1}$ and $\widehat{\mathbf{R}}_*^{-1}$ are well approximated by the $\chi^2_{\ell m}$ distribution. In Figure 4, the QQ plots are constructed for two event types and the results are similar in spirit to the ones in Figure 3.

5.2. Example 2: ARMA(1, 3) error process

The error process was $\epsilon(t_i) = \eta\epsilon(t_{i-1}) + z(t_i) + \xi_1 z(t_{i-1}) + \xi_2 z(t_{i-2}) + \xi_3 z(t_{i-3})$, where $\eta = 0.1$, $\xi_1 = 0.9$, $\xi_2 = 0.7$, $\xi_3 = 0.25$ and $\{z(t_i)\}$ were independent $N(0, \sigma_z^2)$ and $N(0, \sigma_z^2/8)$. We took $\sigma_z^2 = 0.4079^2$ for $\ell = 1$, and $\sigma_z^2 = 0.3899^2$ for $\ell = 2$, such that the corresponding SNRs are about 1 and 8.

To check the validity of the data-driven method for choosing D, we plot $\widehat{r}_*(D)$ versus D for the ARMA(1, 3) error process with SNR = 1 based on a single realization, in Figure 5. The left panel is for a single event type and the right panel for two event types. Each plot is approximately "V"-shaped, which indicates that it is a reasonable way to select $\widehat{D}$ by minimizing $\widehat{r}_*(D)$.

The QQ plots for $\mathbb{K}$ and $\mathbb{K}_{\mathrm{bc}}$ against the $\chi^2_{\ell m}$ distribution are presented in Figure 6 for one event type and in Figure 7 for two event types. All plots show that both $\mathbb{K}$ and $\mathbb{K}_{\mathrm{bc}}$ have an approximated $\chi^2_{\ell m}$ distribution. Based on the conceivably good performance of $\widehat{\mathbf{R}}_*^{-1}$ in the semiparametric test statistics, we can conclude that $\widehat{\mathbf{R}}_*^{-1}$ serves as a satisfactory estimate of $\mathbf{R}^{-1}$.

5.3. Example 3: Sum of AR(1) and white noise error process

The error process consists of an AR(1) time series and a sequence of white noises, $\epsilon(t_i) = \zeta_1(t_i) + \zeta_2(t_i)$. Here $\{\zeta_1(t_i)\}$ was the white noise, independent $N(0, \sigma_\zeta^2)$ and $N(0, \sigma_\zeta^2/8)$, respectively; $\{\zeta_2(t_i)\}$ were generated from the AR(1) model $\zeta_2(t_i) = \rho\zeta_2(t_{i-1}) + z(t_i)$, where $\rho = 0.638$ and $z(t_i) \sim N(0, \sigma_z^2)$ and

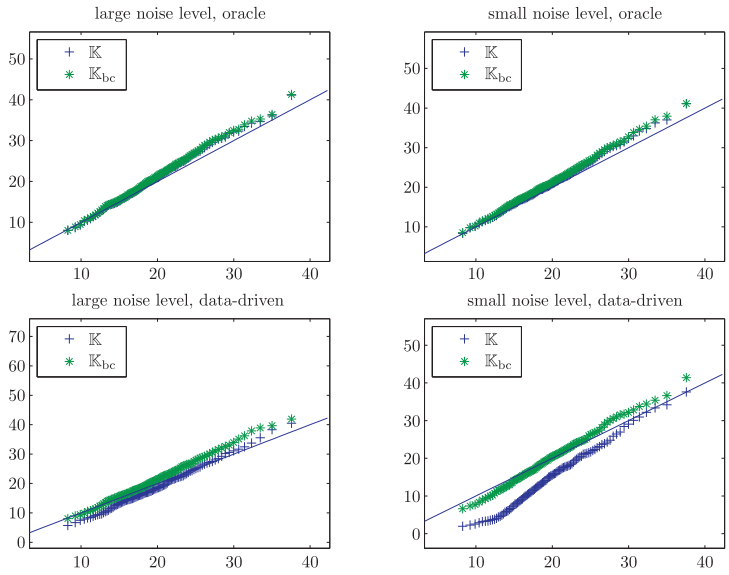

Figure 3. Empirical QQ plots of $\mathbb{K}$ and $\mathbb{K}_{\mathrm{bc}}$ for the MA(4) error process with a single event type, statistics in the top panels calculated with $\mathbf{R}^{-1}$ and theoretical optimal bandwidth choice, those in the bottom panels used $\widehat{\mathbf{R}}_*^{-1}$ and data-driven method for bandwidth choice. The solid line is the 45 degree straight line.

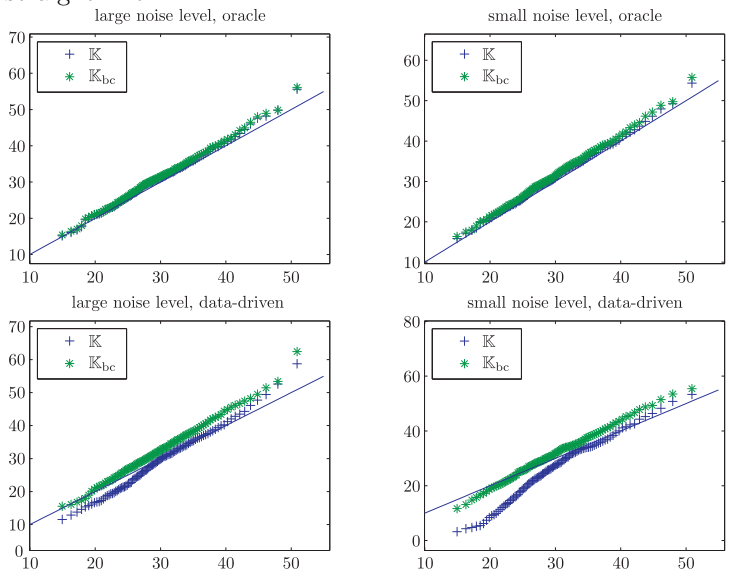

Figure 4. Empirical QQ plots of $\mathbb{K}$ and $\mathbb{K}_{\mathrm{bc}}$ for the MA(4) error process with two event types, statistics in the top panels calculated with $\mathbf{R}^{-1}$ and theoretical optimal bandwidth choice, those in the bottom panels used $\widehat{\mathbf{R}}_*^{-1}$ and data-driven method for bandwidth choice. The solid line is the 45 degree straight line.

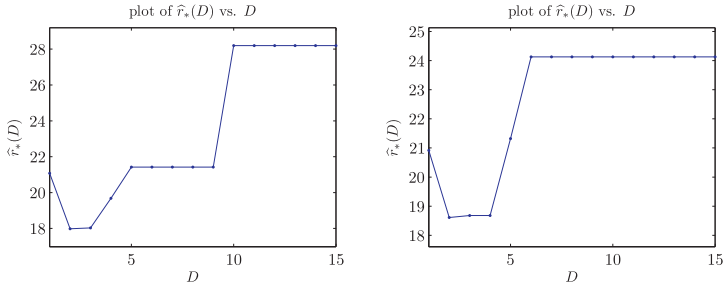

Figure 5. Plots of $\widehat{r}_*(D)$ versus D for the ARMA$(1,3)$ error process with SNR $= 1$ based on a single realization. The left panel corresponds to a single event type and the right panel is for two event types. The points in each plot correspond to the ones used for the grid search method.

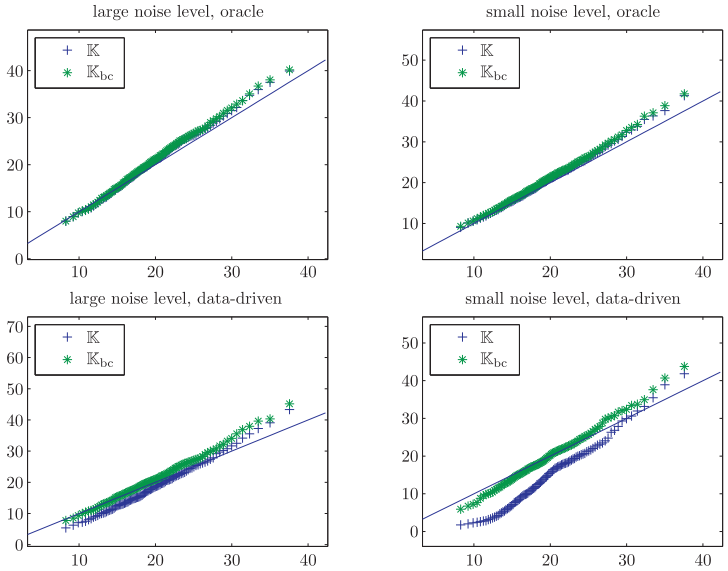

Figure 6. Empirical QQ plots of $\mathbb{K}$ and $\mathbb{K}_{\mathrm{bc}}$ for the ARMA$(1,3)$ error process with a single event type, statistics in the top panels calculated with $\mathbf{R}^{-1}$ and theoretical optimal bandwidth choice, those in the bottom panels used $\widehat{\mathbf{R}}_*^{-1}$ and data-driven method for bandwidth choice. The solid line is the 45 degree straight line.

$N(0, \sigma_z^2/8)$, respectively. For a single event type, we took $\sigma_\zeta^2 = 0.2430^2$ and $\sigma_z^2 = 0.4861^2$, and for two event types, $\sigma_\zeta^2 = 0.2324^2$ and $\sigma_z^2 = 0.4647^2$, which give the SNRs of about 1 and 8.

Figure 8 and Figure 9 display the QQ plots of $\mathbb{K}$ and $\mathbb{K}_{\mathrm{bc}}$ versus the $\chi^2_{\ell m}$ distribution. The former figure is constructed for a single event type while the latter one is made for two event types. The test statistics calculated by $\mathbf{R}^{-1}$

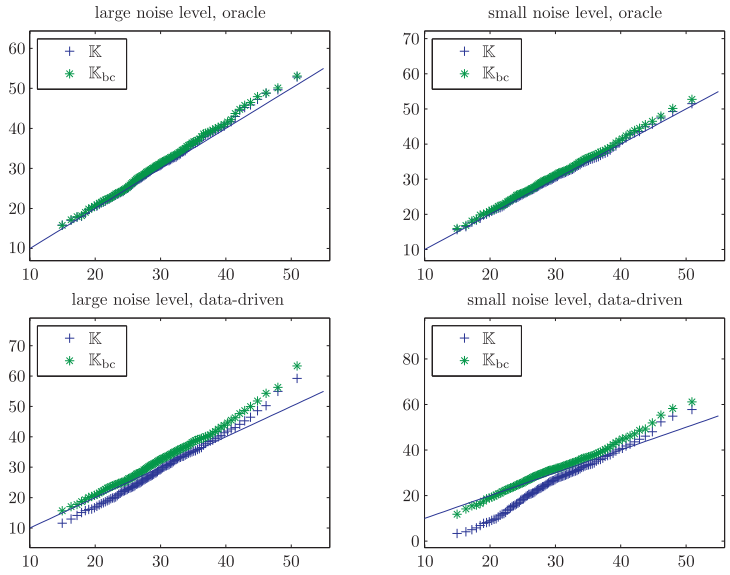

Figure 7. Empirical QQ plots of $\mathbb{K}$ and $\mathbb{K}_{\mathrm{bc}}$ for the ARMA$(1,3)$ error process with two event types, statistics in the top panels calculated with $\mathbf{R}^{-1}$ and theoretical optimal bandwidth choice, those in the bottom panels used $\widehat{\mathbf{R}}_*^{-1}$ and data-driven method for bandwidth choice. The solid line is the 45 degree straight line.

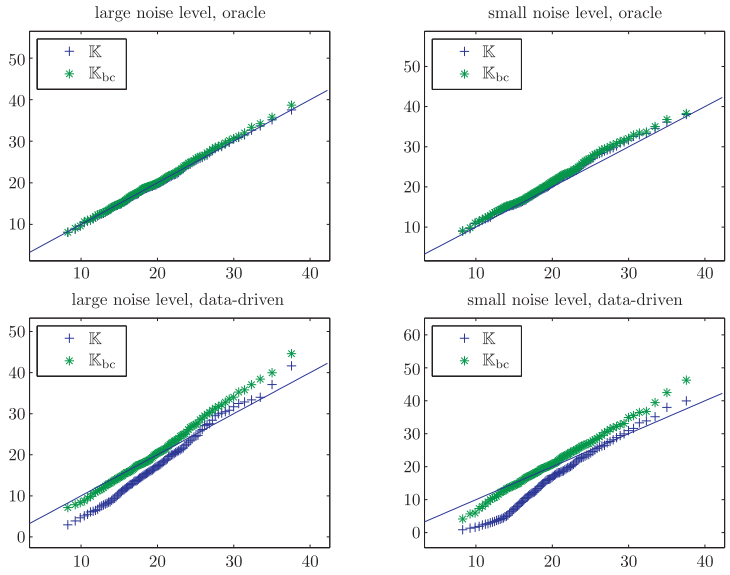

Figure 8. *Empirical QQ plots of* $\mathbb{K}$ *and* $\mathbb{K}_{\mathrm{bc}}$ *for the* AR(1) + WN *error process with a single event type, statistics in the top panels calculated with* $\mathbf{R}^{-1}$ *and theoretical optimal bandwidth choice, those in the bottom panels used* $\widehat{\mathbf{R}}_*^{-1}$ *and data-driven method for bandwidth choice. The solid line is the* 45 *degree straight line.*

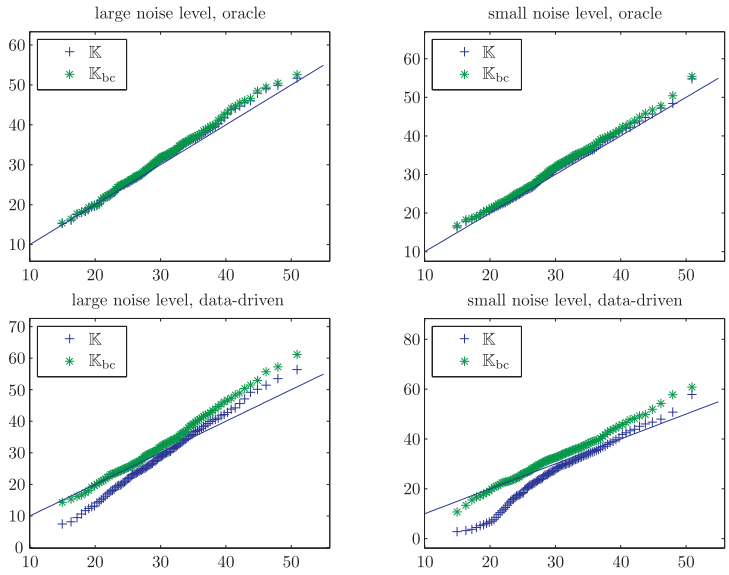

Figure 9. Empirical QQ plots of $\mathbb{K}$ and $\mathbb{K}_{\rm bc}$ for the AR(1) + WN error process with two event types, statistics in the top panels calculated with $\mathbf{R}^{-1}$ and theoretical optimal bandwidth choice, those in the bottom panels used $\widehat{\mathbf{R}}_*^{-1}$ and data-driven method for bandwidth choice. The solid line is the 45 degree straight line.

and $\widehat{\mathbf{R}}_*^{-1}$ have finite sampling distributions coinciding with the $\chi^2_{\ell m}$ distribution fairly well.

5.4. Comparison of $\widehat{\mathbf{R}}_*$ and $\widehat{\mathbf{R}}_Z$

The difference between $\widehat{\mathbf{R}}_*$ and the estimator $\widehat{\mathbf{R}}_Z$ in Zhang et al. (2008) and Zhang and Yu (2008), is that we develop data-driven methods to select the banding parameter g_n and choose ω and D for refining the estimator. The choice of the banding parameter for $\widehat{\mathbf{R}}_Z$ is $g_n = 2$, and hence $\widehat{\mathbf{R}}_Z$ is computationally faster than $\widehat{\mathbf{R}}_*$. In the following, we compare the two estimators based on the error processes of Sections 5.1−5.3. We compare the average selected banding parameter by the data-driven method with $g_n = 2$ used in Zhang et al. (2008) and Zhang and Yu (2008). The average loss of $\widehat{\mathbf{R}}_*$ and $\widehat{\mathbf{R}}_Z$ under the L_∞ norm, $\|\widehat{\mathbf{R}}_* - \mathbf{R}\|_\infty$ and $\|\widehat{\mathbf{R}}_Z - \mathbf{R}\|_\infty$, were also compared.

Table 1 presents g_0 (minimizer of the empirical version of (4.2)), the estimated banding parameter by the data-driven method $\widehat{g}$, and the losses of $\widehat{\mathbf{R}}_*$ and $\widehat{\mathbf{R}}_Z$ under the L_∞ norm. From Table 1, we see that the data-driven method picks the banding parameter close to g_0. For the MA(4) and ARMA(1, 3) models, the loss of $\widehat{\mathbf{R}}_*$ is smaller than that of $\widehat{\mathbf{R}}_Z$. For the AR(1) + WN model, where $g_0 = 2$,

Table 1. Comparison of g_0 and estimated banding parameter, losses of $\widehat{\mathbf{R}}_*$ and $\widehat{\mathbf{R}}_Z$.

		Banding parameter		Losses	
SNR	Model	g_0	$\widehat{g}$ (SD)	$\widehat{\mathbf{R}}_*$ (SD)	$\widehat{\mathbf{R}}_Z$ (SD)
		Single event type			
1	MA(4)	4	3.9 (0.04)	0.53 (0.03)	1.70 (0.01)
	ARMA(1, 3)	3	2.8 (0.02)	0.34 (0.02)	1.09 (0.03)
	AR(1) + WN	2	1.6 (0.03)	2.23 (0.02)	2.14 (0.01)
8	MA(4)	4	3.9 (0.04)	0.54 (0.03)	1.69 (0.01)
	ARMA(1, 3)	3	2.7 (0.02)	0.38 (0.02)	1.04 (0.03)
	AR(1) + WN	2	1.6 (0.03)	2.25 (0.02)	2.14 (0.01)
		Two event types			
1	MA(4)	4	3.9 (0.04)	0.52 (0.03)	1.70 (0.01)
	ARMA(1, 3)	3	2.8 (0.02)	0.36 (0.02)	1.11 (0.03)
	AR(1) + WN	2	1.6 (0.03)	2.24 (0.02)	2.14 (0.01)
8	MA(4)	4	3.9 (0.04)	0.53 (0.03)	1.70 (0.01)
	ARMA(1, 3)	3	2.8 (0.02)	0.38 (0.02)	1.05 (0.03)
	AR(1) + WN	2	1.7 (0.03)	2.24 (0.02)	2.14 (0.01)

$\widehat{\mathbf{R}}_Z$ performs slightly better since the choice of the banding parameter is exactly $g_n = 2$.

6. Data Analysis

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. 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 secs. Refer to Zhang and Yu (2008) for details of the dataset and detection of activated brain regions. The study aims to estimate the BOLD (Blood Oxygenation Level-Dependent) response to each of the trial types for 1–18 seconds following the image onset. The length of the estimated HRF is set equal to 18.

The methodological difference between Zhang and Yu (2008) and this paper is between $\widehat{\mathbf{R}}_Z^{-1}$ and $\widehat{\mathbf{R}}_*^{-1}$ for estimating $\mathbf{R}^{-1}$, where $\widehat{\mathbf{R}}_*^{-1}$ selects g_n and D via data-driven methods. Our numerical results indicate that the banding parameter $\widehat{g}$ selected for $\widehat{\mathbf{R}}_*^{-1}$ is very close to $g_n = 2$ used for $\widehat{\mathbf{R}}_Z^{-1}$. They in turn yield nearly identical estimates of HRF and the corresponding test statistics, without altering the declared significance. For example, Table 2 reports the banding parameter and D selected for $\widehat{\mathbf{R}}_*^{-1}$ at two activated voxels. Figure 10 compares the estimates of HRF using the semi-parametric model based on $\widehat{\mathbf{R}}_*^{-1}$ and $\widehat{\mathbf{R}}_Z^{-1}$,

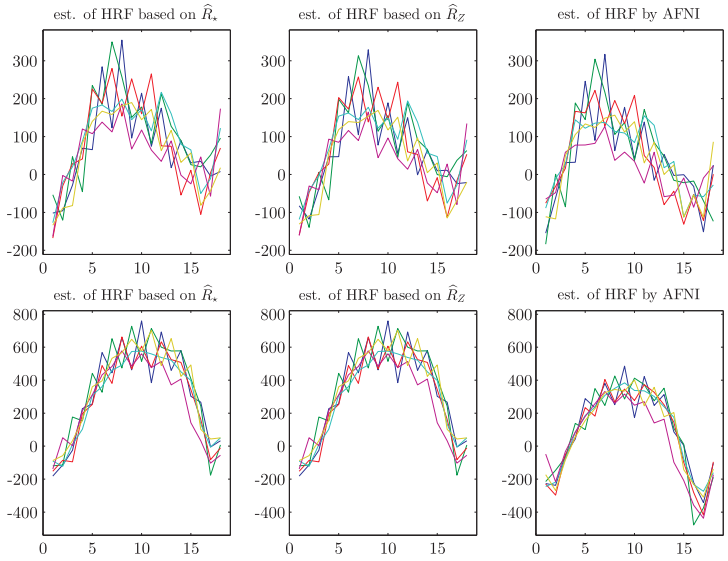

Figure 10. Compare estimates of HRF for each of six stimuli. Top panels: at voxel $(24, 32, 7)$; bottom panels: at voxel $(49, 41, 10)$.

Table 2. Compare banding parameters used for $\widehat{\mathbf{R}}_*^{-1}$ and $\widehat{\mathbf{R}}_Z^{-1}$ at two voxels.

voxel	for $\widehat{\mathbf{R}}_*^{-1}$	for $\widehat{\mathbf{R}}_Z^{-1}$
$(24, 32, 7)$	$\widehat{g} = 1$, $\widehat{g}_e = 2$, $\widehat{D} = 21$	$g_n = 2$
$(49, 41, 10)$	$\widehat{g} = 2$, $\widehat{g}_e = 2$, $\widehat{D} = 16$	$g_n = 2$

and the parametric model via AFNI software (Zhang et al. (2008)). Similar comparison can be performed at other voxels.

7. Discussion

Dealing with the temporally correlated error process is an important issue in fMRI data analysis. The existing methods for handling this problem require strong assumptions on the error process. We consider estimating the error autocorrelation matrix and its inverse under a more general assumption on the error process, that allows for a wider range of the error correlation structure. The existing methods for estimating covariance matrix are not applicable to our case because of the special structure of the semiparametric model for fMRI data. We adopt a difference-based method, and develop a banded estimate of $\mathbf{R}$ and a refined estimate of $\mathbf{R}^{-1}$. Compared with the smoothing-based method, the difference-based method enjoys good practical performance and computational efficiency. We demonstrate consistency of our proposed estimators and provide data-driven approaches for calculating the estimates in practice. Hence, we pro-

vide a computationally efficient and consistent estimate of $\mathbf{R}^{-1}$ for semiparametric inference.

The optimal convergence rate of the covariance matrix estimate is an important topic, studied in Cai, Zhang, and Zhou (2010), Cai, Ren, and Zhou (2012), Cai and Zhou (2012), and Xiao and Wu (2012). These results are not applicable to our case. It would be interesting to investigate the optimal convergence rate of the estimates of $\mathbf{R}$ and $\mathbf{R}^{-1}$ under our setting. Our proposed estimator is particularly useful for analyzing fMRI data under the semiparametric model, but may not be easily used to estimate the error correlation matrix for other datasets, or for fMRI data with different models.

Acknowledgements

The authors thank the Co-Editor, the Associate Editor and three referees for insightful comments and suggestions. The research is supported by the National Science Foundation grants DMS-1106586, DMS-1308872 and Wisconsin Alumni Research Foundation.

References

Bickel, P. J. and Levina, E. (2008a). Covariance regularization by thresholding. *Ann. Statist.* **36**, 2577-2604.

Bickel, P. J. and Levina, E. (2008b). Regularized estimation of large covariance matrices. *Ann. Statist.* **36**, 199-227.

Box, G. E. P. and Jenkins, G. M. (1976). *Time Series Analysis: Forecasting and Control.* Revised Edition. Holden-Day, San Francisco.

Bullmore, E., Brammer, M., Williams, S. C. R., Rabe-Hesketh, S., Janot, N., David, A., Mellers, J., Howard, R. and Sham, P. (1996). Statistical methods of estimation and inference for functional MR image analysis. *Magn. Reson. Med.* **35**, 261-277.

Cai, T. T., Ren, Z. and Zhou, H. H. (2012). Optimal rates of convergence for estimating Toeplitz covariance matrices. Published online by *Probab. Theory Relat. Fields.*

Cai, T. T., Zhang, C. and Zhou, H. H. (2010). Optimal rates of convergence for covariance matrix estimation. *Ann. Statist.* **38**, 2118-2144.

Cai, T. T. and Zhou, H. H. (2012). Minimax estimation of large covariance matrices under ℓ_1-norm. *Statistica Sinica* **22**, 1319-1378.

El Karoui, N. (2008). Operator norm consistent estimation of large-dimensional sparse covariance matrices. *Ann. Statist.* **36**, 2717-2756.

Friston, K. J., Holmes, A. P., Poline, J-B., Grasby, P. J., Williams, S. C. R. and Frackowiak, R. S. J. (1995). Analysis of fMRI time-series revisited. *NeuroImage* **2**, 45-53.

Friston, K. J., Josephs, O., Zarahn, E., Holmes, A. P., Rouquette, S. and Poline, J-B. (2000). To smooth or not to smooth? Bias and efficiency in fMRI time-series analysis. *NeuroImage* **12**, 196-208.

Furrer, R. and Bengtsson, T. (2007). Estimation of high-dimensional prior and posterior covariance matrices in Kalman filter variants. *J. Multivariate Anal.* **98**, 227-255.

Golub, G. H. and Van Loan, C. F. (1989). *Matrix Computations*, third edition. John Hopkins Univ. Press, Baltimore.

Hoskins, W. D. and Ponzo, P. J. (1972). Some properties of a class of band matrices. *Math. Comp.* **26**, 393-400.

Lange, N. (1996). Tutorial in biostatistics: Statistical approaches to human brain mapping by functional magnetic resonance imaging. *Statist. Med.* **15**, 389-428.

Lazar, N. A., Eddy, W. F., Genovese, C. R. and Welling, J. (2001). Statistical issues in fMRI for brain imaging. *Internat. Statist. Rev.* **69**, 105-127.

Lund, T. E., Madsen, K. H., Sidaros, K., Luo, W. L. and Nichols, T. E. (2006). Non-white noise in fMRI: does modelling have an impact? *NeuroImage* **29**, 54-66.

McMurry, T. L. and Politis, D. N. (2010). Banded and tapered estimates for autocovariance matrices and the linear process bootstrap. *J. Time Ser. Anal.* **31**, 471-482.

Miller, K. S. (1981). On the inverse of the sum of matrices. *Math. Magazine* **54**, 67-72.

Monti, M. M. (2011). Statistical analysis of fMRI time-series: a critical review of the GLM approach. *Front. Hum. Neurosci.* **5**, article 28.

Politis, D. N., Romano, J. P. and Wolf, M. (1999). *Subsampling*. Springer-Verlag, New York.

Purdon, P. L., Solo, V., Weisskoff, R. M. and Brown, E. N. (2001). Locally regularized spatiotemporal modeling and model comparison for functional MRI. *NeuroImage* **14**, 912-923.

Shumway, R. H. and Stoffer, D. S. (2011). *Time Series Analysis and Its Applications With R Examples*. 3rd edition. Springer.

Worsley, K. J. and Friston, K. J. (1995). Analysis of fMRI time-series revisited–again. *NeuroImage* **2**, 173-181.

Worsley, K. J., Liao, C. H., Aston, J., Petre, V., Duncan, G. H., Morales, F. and Evans, A. C. (2002). A general statistical analysis for fMRI data. *NeuroImage* **15**, 1-15.

Wu, W. B. and Pourahmadi, M. (2009). Banding sample autocovariance matrices of stationary processes. *Statist. Sinica* **19**, 1755-1768.

Wu, W. B. and Xiao, H. (2011). Covariance matrix estimation in time series. *Handbook of Statistics* **30**, 187-209.

Xiao, H. and Wu, W. B. (2012). Covariance matrix estimation for stationary time series. *Ann. Statist.* **40**, 466-493.

Zarahn, E., Aguirre, G. K. and D'Esposito, M. (1997). Empirical analyses of BOLD fMRI statistics. I. Spatially unsmoothed data collected under null-hypothesis conditions. *NeuroImage* **5**, 179-197.

Zhang, C. M. (2003). Calibrating the degrees of freedom for automatic data smoothing and effective curve checking. *J. Amer. Statist. Assoc.* **98**, 609-628.

Zhang, C. M., Lu, Y., Johnstone, T., Oakes, T. and Davidson, R. J. (2008). Efficient modeling and inference for event-related fMRI data. *Comput. Statist. Data Anal.* **52**, 4859-4871.

Zhang, C. M. and Yu, T. (2008). Semiparametric detection of significant activation for brain fMRI. *Ann. Statist.* **36**, 1693-1725.

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

E-mail: xguo@stat.wisc.edu

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

E-mail: cmzhang@stat.wisc.edu

(Received May 2013; accepted January 2014)