SEMIPARAMETRIC DETECTION OF SIGNIFICANT
ACTIVATION FOR BRAIN FMRI

BY CHUNMING ZHANG* AND TAO YU†

University of Wisconsin-Madison

Functional magnetic resonance imaging (fMRI) aims to locate activated regions in human brains when specific tasks are performed. The conventional tool for analyzing fMRI data applies some variant of the linear model, which is restrictive in modeling assumptions. To yield more accurate prediction of the time-course behavior of neuronal responses, the semi-parametric inference for the underlying hemodynamic response function is developed to identification of significantly activated voxels. Under mild regularity conditions, we demonstrate that a class of the proposed semi-parametric test statistics, based on the local-linear estimation technique, follow $\chi^2$ distributions under null hypotheses for a number of useful hypotheses. Furthermore, the asymptotic power functions of the constructed tests are derived under the fixed and contiguous alternatives. Simulation evaluations and real fMRI data application endorse that the semiparametric inference procedure delivers more efficient detection of activated brain areas than popular imaging analysis tools AFNI and FSL.

1. Introduction. Neuroscience is a scientific discipline dedicated to studying the structure, function, and pathology of the brain and nervous system, and furthermore lies at the forefront of investigation of the brain and mind. Functional magnetic resonance imaging (fMRI) emerges as a new and exciting non-invasive imaging technique that aims at the localization of functional brain areas in a living human brain, that is, the detection of areas or regions that are responsible for the processing of certain stimuli. Adequate statistical modeling and analysis of the massive spatio-temporal datasets generated by fMRI pose significant challenges to conventional statistical methods. First, a typical fMRI dataset for a single subject contains a temporally highly correlated time series of measurements,
taken every 2 seconds or so for about an hour, on each of, say 64 × 64 × 30 voxels (for volume element in three dimensional space) throughout the brain. Accordingly, the data sets are so enormous that properly accommodating both temporal and spatial correlation is needed. Second, models relating fMRI signals to neural changes are complex. The standard tool for analyzing fMRI data is some variant of the linear model, usually fitted separately by least-squares to each voxel (Worsley and Friston 1995). After that, test fitted model parameters for statistical significance and draw colors on top of significant voxels. This comprises the major procedure of statistical parametric mapping (SPM), popularly used in neuroimage study (Friston, et al. 1997). Recent reviews of the statistical issues in fMRI for brain imaging and the statistical methods for analyzing fMRI data can be found in Lange (1996), Lazar, et al. (2001), Fahrmeir and Gössl (2002) and Worsley, et al. (2002), among others.

In this paper, we aim to develop voxel-wise semi-parametric inference for the underlying hemodynamic response function (HRF), which is the part of primary interest to neuroscientists. For instance, identifying whether a particular voxel is activated when a subject performs certain motor, sensory or cognitive tasks can be expressed in terms of a statistical test of the hypothesis that HRF is zero. In order to generate brain activation maps, statistical inference must be drawn from voxelwise estimates of HRF. We will first develop a semi-parametric modeling and estimation approach to obtain statistically more efficient estimates of the underlying HRF associated with fMRI experiments. Compared with the general linear model approach in previous studies, our approach has the advantage that we do not specify any a priori shape for HRF, and that neither do we assume any particular form of the temporal drift function. Taking full advantage of these flexibilities will help reduce the bias due to model misspecification and enhance the power of detection.

Addressing the issue of semiparametric inference for brain fMRI is a non-trivial task, however. Existing parametric statistical inference procedures for fMRI are not immediately applicable to our approach in which the HRFs are estimated semi-parametrically. The work on the generalized likelihood ratio test (Fan, Zhang and Zhang 2001) sheds light on nonparametric inference, based on function estimation under nonparametric models with independent errors, and in the meanwhile, is not readily translated into results from other models. Moreover, as emphasized in Section 3, some standard results of semi-parametric models are not directly applicable to the context of fMRI data, due to the distinctive feature of the Toeplitz design matrix and the complicated dependence structure of the error process. Hence a rigorous in-
vestigation of the semiparametric inference applied to the important area
of research in fMRI is required. This paper fills that gap in the literature.
Under mild regularity conditions, we show that a class of the proposed semi-
parametric test statistics follow \( \chi^2 \) distributions under null hypotheses for a
number of useful hypotheses. To yield improved finite sample performance
of the proposed test statistic, we further explore its bias-corrected version
and derive the corresponding asymptotic distribution. Moreover, the asympto-
tic power functions of the constructed tests are derived under the fixed
and contiguous alternatives. These results are not only important in gaining
theoretical insight into semi-parametric inference applied to a much broader
range of scientific problems, but also helpful in offering valuable practical
guidance for implementation of these techniques.

The rest of the paper is arranged as follows. Section 2 reviews statistical
models for single-voxel and single-run fMRI. Section 3 describes the
semi-parametric estimation of HRF, based on the local-linear nonparametric
smoothing technique. Section 4 establishes the asymptotic distribution of
the proposed test statistics. Section 5 presents simulation evaluations
and compares the activated brain regions using popular imaging analysis
tools AFNI (at http://afni.nimh.nih.gov/afni/) (Cox 1996) and FSL
(at http://www.fmrib.ox.ac.uk/fsl/) (Smith, et al. 2004 and Woolrich,
et al. 2001). Section 6 applies the semiparametric inference to a real fMRI
dataset. Technical conditions and detailed proofs are relegated to the Ap-
pendix.

2. Statistical Models for Single-Voxel and Single-Run fMRI. We begin with a brief overview of the convolution model popularly used
in fMRI study. The BOLD (Blood Oxygenation Level-Dependent) signal
response to neuronal activity is heavily lagged and damped by the hemody-
namic response. Following Ward (2001) and Worsley, et al. (2002), a single-
voxel fMRI time-series \( \{s(t_i), y(t_i)\}_{i=1}^{n} \), for a given scan and a given subject,
can be captured by the convolution model,

\[
y(t) = s \ast h(t) + d(t) + \epsilon(t), \quad t = t_1, \ldots, t_n,
\]

where \( \ast \) denotes the convolution operator, \( y(t) \) is the measured noisy fMRI
signal, \( s(t) \) is the external input stimulus at time \( t \), which could be from a
design either block or event-related and \( s(t) = 1 \) or 0 indicates the presence
or absence of a stimulus, \( h(t) \) is the hemodynamic response function (HRF)
at time \( t \) after neural activity, \( d(t) \) is a slowly drifting baseline of time \( t \),
and \( \epsilon(t) \) is a zero-mean error process, consisting of non-neural noise (due to
respiration and blood flow pulsations through the cardiac cycle) and “white
noise” (from random/thermal currents in the body and the scanner).
2.1. **Existing methods for modeling HRF, drift and error.** In neuroimaging studies, most existing methods model $h(\cdot)$ as the difference of two gamma functions or a linear combination of gamma functions or a linear combination of a gamma function and its Taylor expansion (Worsley, et al. 2002; Lange and Zeger 1997; Josephs and Henson 1999). Genovese (2000) constructed $h(\cdot)$ as a “bell” function with cubic splines. As a nuisance component in (2.1), the temporal drift $d(\cdot)$ is usually approximated by a quadratic or higher-order polynomial (Worsley, et al. 2002) or polynomial splines (Genovese 2000). Note that restrictive assumptions on HRF and drift may produce biased estimates of the true hemodynamic responses. Goutte, Nielsen and Hansen (2000) estimated $h(\cdot)$ using smooth FIR filters and reported that some subtle details of HRF can be revealed by the filters but not by previous approaches based on gamma functions. The errors $\epsilon(t_i)$ are well-known to be temporally auto-correlated. Genovese (2000) assumed independent errors for computational convenience. Other assumptions like the AR($p$) structure, most commonly AR(1), are used in Worsley, et al. (2002). As an illustration, Table 1 tabulates the HRF, drift and error implemented in software AFNI and FSL.

<table>
<thead>
<tr>
<th>Table 1</th>
<th>HRF, drift, and error implemented in AFNI and FSL</th>
</tr>
</thead>
<tbody>
<tr>
<td>AFNI (tool 3dDeconvolve)</td>
<td>FSL (tool FEAT)</td>
</tr>
<tr>
<td>$h(t)$ finite impulse response filter</td>
<td>difference of two gamma functions, which is the canonical form</td>
</tr>
<tr>
<td>$d(t)$ quadratic polynomial</td>
<td>removed in the preprocessing, using high pass temporal filtering (Gaussian weighted LSF straight line fitting)</td>
</tr>
<tr>
<td>$\epsilon(t)$ i.i.d.</td>
<td>autocorrelation estimated by Tukey tapering of the spectrum of the residuals</td>
</tr>
</tbody>
</table>

3. **Semi-parametric Estimation of HRF.** Estimating HRF in (2.1) is a deconvolution problem. Ideally, HRF is a high-dimensional smooth function and is non-identically zero if the voxel responds to the stimuli. We will describe a semiparametric method for characterizing properties of the hemodynamic response in the presence of unknown smooth drift. Such characterization is essential for accurate prediction of time-course behavior of neuronal responses.

Typically, the peak value of HRF $h(\cdot)$ is reached after a short delay of the stimulus and drops quickly to zero. A typical example of $h(\cdot)$ given in Glover (1999) is plotted in Figure 1. Clearly, the region $\{t : h(t) \neq 0\}$ is sparse in its temporal domain. Thus, to obtain statistically efficient estimates
of HRF associated with event-related fMRI experiments, the \textbf{sparsity} of HRF needs to be taken into account. We thus suppose that $h(t) = 0$ for $t > t_m$ and focus on estimating the first $m$ values of $h(t_i)$, where $m$ is fewer than $n$, the length of the fMRI time series. Similar to the regularization technique discussed in Bickel and Li (2006), such \textit{qualitative} assumption aims at obtaining well-behaved solutions to overparametrized estimation problems, thus is particularly appealing for dimension reduction with high-dimensional problems. The semiparametric modeling and inference in the paper are applicable to all $m < n$. Data-driven selection of $m$ can be made via a change-point approach or other model-selection criteria. To facilitate discussion, we assume that $y(\cdot)$ and $s(\cdot)$ have equal time resolution of 1 second. Letting $\mathbf{y} = (y(t_1), \ldots, y(t_n))^T$, 

$$
\mathbf{S} = \begin{bmatrix}
s(0) & 0 & \cdots & 0 \\
s(t_2 - t_1) & s(0) & \cdots & 0 \\
\vdots & \vdots & \ddots & \vdots \\
s(t_m - t_1) & s(t_m - t_2) & \cdots & s(0) \\
\vdots & \vdots & \ddots & \vdots \\
s(t_n - t_1) & s(t_n - t_2) & \cdots & s(t_n - t_m)
\end{bmatrix},
$$

$h = (h(t_1), \ldots, h(t_m))^T$, $\mathbf{d} = (d(t_1), \ldots, d(t_n))^T$ and $\mathbf{e} = (e(t_1), \ldots, e(t_n))^T$, model (2.1) can be re-expressed as $\mathbf{y} = \mathbf{S}h + \mathbf{d} + \mathbf{e}$, where $\mathbf{S}$ is a Toeplitz matrix.

In general, for multiple types of stimuli, model (2.1) can be extended to be

$$
y(t) = s_1 * h_1(t) + \cdots + s_r * h_r(t) + d(t) + e(t), \quad t = t_1, \ldots, t_n.
$$

Corresponding to the $j$th type of stimulus, denote by $s_j(\cdot)$ the time-varying stimulus function, $\mathbf{S}_j$ the $n \times m$ Toeplitz design matrix, and $\mathbf{h}_j$ the $m \times 1$
vector of HRF. Then model (3.1) can be rewritten as

$$\mathbf{y} = \mathbf{S}_1 \mathbf{h}_1 + \cdots + \mathbf{S}_r \mathbf{h}_r + \mathbf{d} + \epsilon = \mathbf{Sh} + \mathbf{d} + \epsilon$$

where $\mathbf{S} = [\mathbf{S}_1 \ldots \mathbf{S}_r]$ and $\mathbf{h} = [\mathbf{h}_1^T, \ldots, \mathbf{h}_r^T]^T$. To accommodate fMRI data with multiple runs, we only need to supplement the matrix $\mathbf{S}$ by adding the Toeplitz design matrix arising from each run.

Model (3.2) is conceivably a semiparametric regression model, with a vector $\mathbf{h}$ of length $rm$ for parametric components and a vector $\mathbf{d}$ of length $n$ for nonparametric components. The parametric components (related to the unknown HRF) are of our primary interest, whereas the nonparametric components (related to the unknown temporal drift) serve as nuisance effects, and the noise components $\epsilon$ are serially correlated. We wish to emphasize that due to the special structure of the design matrix $\mathbf{S}$ associated with fMRI design, some commonly used assumptions, such as independence between rows of a design matrix, fail to hold. In addition, the unobservable true correlation structure of $\epsilon$ is often complicated. Thus some standard results of semi-parametric models are not directly applicable to the current fMRI data.

We now describe the semi-parametric estimation of both HRF and the nonparametric drift function in (3.2). Let $\mathbf{S}_d$ be an $n \times n$ local-linear smoothing matrix, associated with the design points $\{t_1, \ldots, t_n\}$, with the $(i,j)$th entry equal to

$$S_d(i,j) = (1,0)\{X(t_i)^T W(t_i) X(t_i)\}^{-1}(1,t_j-t_i)^T K((t_j-t_i)/b)/b.$$  

where $K$ is a kernel function, $b > 0$ is a bandwidth parameter,

$$X(t) = \begin{bmatrix} 1 & t_1-t \cdots & t_n-t \end{bmatrix}, \text{ and } W(t) = \text{diag}\{K((t_1-t)/b), \ldots, K((t_n-t)/b)\}.$$  

See Fan and Gijbels (1996) which provides a comprehensive account of the local-linear and local-polynomial regression techniques. (For expositional simplicity, this paper is confined to the local-linear method.) Note that the matrix $\mathbf{S}_d$ carries information about the design points, kernel $K$ and bandwidth $b$, but does not rely on the configuration of the response variables. Refer to Zhang (2003, Sec. 2.3) for further discussion on finite-sample and asymptotic properties of the smoothing matrix. Notice that smoothing the entries of $\mathbf{y}$ via local-linear method is equivalent to applying $\mathbf{S}_d$ to $\mathbf{y}$. We observe from (3.2) that

$$\bar{\mathbf{y}} = \bar{\mathbf{Sh}} + \bar{\mathbf{d}} + \bar{\epsilon},$$
where \( \tilde{y} = (I - S_d)y \), \( \tilde{S} = (I - S_d)S \), \( \tilde{d} = (I - S_d)d \), \( \tilde{e} = (I - S_d)e \), and \( I \) denotes an identity matrix. Ignoring \( \tilde{d} \), model (3.4) can be regarded as a general linear model. Denote by \( R \) the true correlation matrix of \( e \), namely, \( \text{cov}(e, e) = \sigma^2 R \), with variance \( \sigma^2 \). Let \( \hat{R} \) be an estimate of \( R \). By the weighted least squares method, an estimate of \( h \) is produced by

\[
\hat{h} = (\tilde{S}^T \hat{R}^{-1} \tilde{S})^{-1} \tilde{S}^T \hat{R}^{-1} \tilde{y},
\]

which in turn supplies estimates of the drift components formed by

\[
\hat{d} = S_d(y - \hat{S}h).
\]

4. Semi-parametric Hypothesis Test for HRF. Identification of a particular brain region with a specific function has become a central theme in neuroscience. In this section, we consider constructing test statistics to test whether a particular voxel is activated by the stimuli and whether HRFs activated by different types of stimuli really differ. They correspond to testing the following hypotheses, \( H_0 : h = 0 \) versus \( H_1 : h \neq 0 \), and \( H_0 : h_{j_1} = h_{j_2} \) versus \( H_1 : h_{j_1} \neq h_{j_2} \) where \( j_1 \neq j_2 \). Under the semiparametric model (3.2), all these testing problems can be formulated in a more general form,

\[
H_0 : Ah = 0 \quad \text{versus} \quad H_1 : Ah \neq 0,
\]

where \( A \) is a full row rank matrix with rank(\( A \)) = \( k \).

4.1. Asymptotic null distributions. Motivated from the parametric \( F \)-statistic in linear regression models and the justification of power comparison (Zhang and Dette, 2004) between nonparametric tests for regression curves based on kernel smoothing techniques, we first examine the following semi-parametric test statistic, represented by

\[
\mathbb{K} = \frac{(Ah)^T \{A(\tilde{S}^T \hat{R}^{-1} \tilde{S})^{-1}A^T\}^{-1}(Ah)}{\tilde{r}^T \hat{R}^{-1} \tilde{r} / (n - rm)},
\]

where \( \tilde{r} = \tilde{y} - \hat{S} \). Theorem 4.1 below establishes the asymptotic null distribution of \( \mathbb{K} \).

**Theorem 4.1.** Assume Condition A in the Appendix. Then under \( H_0 \) in (4.1) where \( A \) is a \( k \times rm \) matrix with rank(\( A \)) = \( k \), it follows that \( \mathbb{K} \overset{\mathcal{L}}{\to} \chi^2_k \), where \( \mathcal{L} \) denotes converges in distribution.
Our simulation evaluation in Section 5 demonstrates that the finite sampling distribution of $K$ is reasonably well approximated by its asymptotic $\chi^2$ distribution, whereas when the noise level decreases, the approximation may become less accurate. See Figure 2 (right panel). Technically, as manifested in the proof of Theorem 4.1, the asymptotic $\chi^2$ distribution of $K$ follows from the asymptotic normality of $\hat{h}$ shown in Lemma 7, which relies on the fact that a term $J_1$ (associated with the drift vector $d$) is stochastically dominated by a term $J_2$ (associated with the error vector $e$). Practically, in finite-sample situations, low noise levels do not necessarily guarantee that $J_1$ is stochastically negligible compared with $J_2$. Consequently, the finite sampling distributions of $h$ and $K$ may appear biased towards the normality and $\chi^2$ distribution respectively. In these situations, we adopt the bias corrected version of $K$, defined as

$$K_{bc} = \frac{(A\hat{h}_{bc})^T \{A(S^T \hat{R}^{-1} \hat{S})^{-1}A^T\}^{-1}(A\hat{h}_{bc})}{\hat{r}_{bc}^T \hat{r}_{bc}/(n - rm)},$$

where $\hat{h}_{bc} = \hat{h} - (\hat{S}^T \hat{R}^{-1} \hat{S})^{-1} \hat{S}^T \hat{R}^{-1} \bar{d}$, $\hat{r}_{bc} = \hat{r} - \bar{d}$, $\bar{d} = S_d(y - \hat{h})$ and $\bar{d} = (I - S_d)d$. Note that as the sequence length $n$ grows, $\bar{d}$ is negligible but practically adjusts for the bias caused by $J_1$ due to the ignorance of $\bar{d}$ in (3.4). Theorem 4.2 below reveals that $K_{bc}$ and $K$ have the same asymptotic null distributions.

**Theorem 4.2.** Assume Condition $A$ in the Appendix. Then under $H_0$ in (4.1) where $A$ is a $k \times rm$ matrix with rank($A$) = $k$, it follows that $K_{bc} \xrightarrow{d} \chi^2_k$.

We would like to make some remarks on the derivations of Theorems 4.1–4.2. First, it is tempting to try to show that $n^{-1} S^T \hat{R}^{-1} S \overset{p}{\rightarrow} M$ for some positive definite matrix $M$, where $\overset{p}{\rightarrow}$ denotes converges in probability. Nonetheless, for fMRI data, since the $n \times n$ correlation matrix $R$ of $e$ is generally far more complicated than the diagonal matrix of independent errors, deriving an explicit form of $M$ is nearly intractable. To overcome this technical difficulty, we have demonstrated that it suffices to verify that $R$ satisfies

$$\text{var}\{n^{-1} \xi_{j_1, \ell_1}^T (I - S_d) R^{-1} (I - S_d) \xi_{j_2, \ell_2}\} \rightarrow 0$$

for all $j_1, j_2 = 1, \ldots, r$ and all $\ell_1, \ell_2 = 1, \ldots, m$, where $\xi_{j, \ell}$ is the $\ell$th column vector of $S_j$, and $\hat{R}$ fulfills Condition $A8$ in the Appendix,

$$\|\hat{R}^{-1} - R^{-1}\|_\infty = o(1)$$
where $\|B\|_\infty = \max_{1 \leq i \leq n} \sum_{j=1}^n |B(i, j)|$ denotes the $\infty$-norm of a matrix $B$. See Lemma 6 and Corollary 2. Namely, the explicit form of $M$ is indeed not needed in deriving the asymptotic null distributions of $K$ and $K_{bc}$. Second, Condition A8 together with $\|B\|_2 \leq \{\|B\|_1 \|B\|_\infty\}^{1/2}$ (Golub and Van Loan 1996) and the symmetry of $\hat{R}$ and $R$ guarantees $\|\hat{R}^{-1} - R^{-1}\|_2 = o(1)$, which is typically interpreted as the “consistency” of large covariance matrix estimators (Bickel and Levina, 2006).

**Remark 1.** In real world applications, each fMRI sequence length is not very long. For instance, $n$ is 185 for each run in the real fMRI dataset described in Section 6. This indicates that the “mixing assumptions”, commonly made in the asymptotic studies of nonlinear time series (Bosq, 1998; Fan and Yao, 2003), may not hold for fMRI data. Therefore, the sampling properties of $K$ and $K_{bc}$ are studied using the more realistic error assumption A3 of Condition A in the Appendix, which could possibly be weakened.

**Remark 2.** Throughout numerical work in the paper, parametric estimation of the error covariance matrix adopts a computationally fast and effective scheme developed in Zhang, Lu, Johnstone, Oakes and Davidson (2006), which assumes Condition A3 in the Appendix with $g = 2$. Moreover, since an fMRI dataset contains time-course measurements over voxels, the number of which is typically of the order of $10^4 \sim 10^5$, the conventional false discovery rate (FDR) approach (Benjamini and Hochberg, 1995; Storey, 2002) can be adopted to deal with multiple testing procedure. Other useful and elaborate procedures for covariance matrix estimation and multiple comparison may also be employed. Particularly, Zhang, et al. (2006) illustrated numerical evidence that the existing FDR approach tends to find activation in tiny scattered regions of the brain which are more likely to be false discoveries, and carefully devised a new FDR approach, which gains efficiency over existing FDR approach.

4.2. Asymptotic power functions. To appreciate the discriminating power of the proposed tests in assessing the significance of activated areas, the asymptotic power is analyzed. Theorem 4.3 demonstrates that both $K$ and $K_{bc}$ are consistent against all fixed deviations from the null model.

**Theorem 4.3.** Assume Condition A in the Appendix and $n^{-1} \hat{S}^T R^{-1} \hat{S} \overset{P}{\to} M$ where $M$ is positive definite. Then under the fixed alternative $H_1$ in (4.1),

\[
\begin{align*}
n^{-1} K & \overset{P}{\to} (Ah)^T (AM^{-1}A^T)^{-1} Ah / \sigma^2 > 0, \\
n^{-1} K_{bc} & \overset{P}{\to} (Ah)^T (AM^{-1}A^T)^{-1} Ah / \sigma^2 > 0.
\end{align*}
\]
The results in Theorem 4.3 indicate that under the fixed alternative \( H_1 \),
\[
\mathbb{K} \xrightarrow{P} +\infty, \quad \text{and} \quad \mathbb{K}_{bc} \xrightarrow{P} +\infty,
\]
at the common rate \( n \). Hence the test statistics \( \mathbb{K} \) and \( \mathbb{K}_{bc} \) have power functions tending to one against fixed alternatives.

Consider a sequence of local alternatives, defined by
\[
H_{1n} : A h = \delta_n c,
\]
where \( \delta_n = n^{-1/2} \) and \( c = (c_1, \ldots, c_k)^T \neq 0 \). Theorem 4.4 explores the asymptotic distributions of \( \mathbb{K} \) and \( \mathbb{K}_{bc} \) under the local alternatives \( H_{1n} \).

**Theorem 4.4.** Assume Condition A in the Appendix and \( n^{-1} S^T R^{-1} S \xrightarrow{P} \mathbf{M} \) where \( \mathbf{M} \) is positive definite. Then under the local alternative \( H_{1n} \) in (4.2), \( \mathbb{K} \xrightarrow{D} \chi_1^2(\tau^2) \) and \( \mathbb{K}_{bc} \xrightarrow{D} \chi_k^2(\tau^2) \), with noncentrality parameter \( \tau^2 = c^T (AM^{-1}A^T)^{-1} c / \sigma^2 \).

The results in Theorem 4.4 indicate that the tests have non-trivial local power detecting local alternatives approaching the null at the rate \( n^{-1/2} \). A simple calculation shows that the asymptotic power of the tests against local misspecification (4.2) equals
\[
\int_0^\infty \frac{\exp\left\{-\left(x + \frac{\tau^2}{2}\right)/2\right\}}{2^{k/2}} \sum_{j=0}^{\infty} \frac{x^{k/2+j-1} \tau^{2j}}{\Gamma(k/2 + j) 2^j j!} \, dx,
\]
where \( \chi_{k,1-\alpha}^2 \) is the \( 1 - \alpha \) quantile of \( \chi_k^2 \) distribution and \( \Gamma(\cdot) \) denotes the gamma function.

**5. Simulation Study.** Throughout the numerical work, we use the Epanechnikov kernel function (Silverman, 1986) supported on \([-1,1]\). A complete copy of Matlab codes is available upon request.

**5.1. Hypothesis test of HRF at a single voxel.** As an illustration, the hypothesis testing for \( H_0 : h = 0 \) versus \( H_1 : h \neq 0 \) is undertaken. This is used for testing whether the brain activity in a voxel is triggered or not. To check the agreement between the \( \chi^2 \) distribution with finite sampling distributions of \( \mathbb{K} \) and \( \mathbb{K}_{bc} \) under \( H_0 \), the fMRI data are simulated as follows. We simulate an fMRI experiment with a single run and a single type of stimulus, where \( n = 400 \) and 500 realizations are conducted. (I) The time-varying stimuli are generated from independent Bernoulli trials such that \( P\{s(t_i) = 1\} = .5 \). (II) The HRF is \( h(t_i) = 0 \), \( i = 1, \ldots, 18 \) (so that
m = 18). (III) The drift function is $d(t_i) = 10 \sin\{\pi(t_i - .21)\}$, $i = 1, \ldots, n$.

(IV) The noise process $\epsilon$ is the sum of independent noise processes $\epsilon_1$ and $\epsilon_2$ (see Purdon, Solo, Weissko and Brown, 2001); $\{\epsilon_1(t_i)\}$ are i.i.d. normal with mean zero and variance $.5216^2$, $.3689^2$, $.2608^2$ and $.1844^2$ respectively; $\epsilon_2$ is AR(1), i.e., $\epsilon_2(t_i) = \rho \epsilon_2(t_{i-1}) + z(t_i)$ with $\rho = .638$ and $z(t_i)$ follows the normal distribution with mean zero and variance $.5216^2$, $.3689^2$, $.2608^2$ and $.1844^2$ respectively. These choices give the noise lag-one auto-correlation equal to .4 and the signal-to-noise-ratio (SNR) about 1, 2, 4 and 8, where $\text{SNR} = \text{variance}(\text{Sh})/\text{variance}(\epsilon)$.

![Graphs showing empirical quantiles of test statistics $K$ and $K_{bc}$](image)

**Figure 2.** Empirical Quantiles (on the y-axis) of Test Statistics $K$ and $K_{bc}$ (Where the Top Panel Uses the True $R$ and the Optimal Smoothing Parameters, and the Bottom Panel Uses the Estimated $R$ and Data-Driven Smoothing Parameters) versus Quantiles (on the x-axis) of $\chi^2_m$ Distribution. Solid line: the 45-degree reference line.

The QQ plots of the (1st up to 99th) percentiles of $K$ and $K_{bc}$ against those of the $\chi^2_m$ distribution are displayed in Figure 2. In the top panel, $K$ and $K_{bc}$ utilize the true covariance matrix and fix the smoothing parameters at their theoretically optimal values (minimizing the mean squared errors of estimators) for estimating HRF and drift in each simulation. For the sake of clarity, only the cases of SNR equal to 1 and 8 are presented; the former is the “large noise level” case whereas the latter is the “small noise level” case. In either case, we observe that the finite sampling distributions of $K$ and $K_{bc}$, at the realistic sample size 400, agree reasonably well with the $\chi^2$ distribution. The QQ plots also lend support that $K_{bc}$ is better than or at least as good as the bias un-corrected counterpart $K$.

For a more realistic comparison, $K$ and $K_{bc}$ in the bottom panel of Figure
use the estimated covariance matrices and data-driven smoothing
parameters. The results are similar in spirit to the ones in the top panel and continue
to support the bias-correction procedure.

5.2. Detection of activated brain regions. We simulate a whole brain
fMRI dataset, with aim of mimicking the true brain activity, to the maxi-
mum extent feasible. The experiment design, timings and size are exactly
the same as those of the real fMRI dataset in Section 6. An HRF profile
is extracted from an voxel which shows the strongest responses in the real
dataset. For each voxel, the simulated drift is obtained from an adequate
smoothing to the time series for the corresponding voxel of the real dataset.
The simulated noise variance profile is determined from a variance map,
which is made by a $5 \times 5 \times 5$ spatially median smoothing on median values
of squared residuals of the real time series subtracting the simulated drift
profile as mentioned before. The noise process $\epsilon(t)$ is generated in a fashion
similar to that of Section 5.1. Specifically, the variances of $\epsilon_1(t)$ and $z(t)$ are
chosen to be equal such that $\text{var}\{\epsilon(t)\}$ is one fifth of the variance map. The
HRF profile, in accordance with the stimuli in the experiment, are added to
two regions which are postulated to be truly active. In these two zones, the
HRFs have been rescaled to about 17% and 12% of the amplitude of the
original HRF profiles. The purpose of rescaling HRFs and noise variance is
to amplify the drift effect and weaken the HRF response so that the estima-
tion of HRF is more challenging. Figure 3 gives nine different slices which
highlight the two activated brain regions. Note that throughout the paper
we apply the same registration transform from the real brain data to the T1
high resolution image of the subject’s brain.

The gain in efficiency obtained by the semiparametric inference pro-
cedure is illustrated by comparing the activated brain regions identified by
our approach with methods offered by popular softwares AFNI and FSL.
The conventional FDR approach is performed at the FDR level .05. Inspec-
tion of Figure 4 reveals that $K$ and $K_{bc}$ are capable of locating both active
regions. In contrast, both AFNI and FSL fail to locate an activated brain
area, and that the other region, though correctly detected, has appreciably
reduced size relative to the actual size. This detection bias suggests that
the stringent modeling assumptions in Table 1 be relaxed to ameliorate the
effects of misspecification. Furthermore, as evidenced in Figure 5, all 4 meth-
ods, when applying the new FDR approach in Zhang, et al. (2006), achieve
more accurate detections than their counterparts in Figure 4, with $K$ and
$K_{bc}$ continuing to outperform AFNI and FSL. Therefore, for applications
to the real fMRI dataset in Section 6, we will only employ the new FDR

6. Real 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. Therefore, there were 6 types of trial (i.e., 6 types of stimuli): negative–enhance (neg-enh), negative–attend (neg-att), negative–suppress (neg-sup), positive–enhance (pos-enh), positive–attend (pos-att), positive–suppress (pos-sup). The sequence of trials was randomized. The time between successive trials also varied. There were 24 trials each of the neg-enh, neg-sup, pos-enh, and pos-sup; there were 11 trials each of the neg-att and pos-att.

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, thus $TR = 2$ secs and the total length is 1110. In contrast, the length of stimuli is 2220; the timing of the stimuli has a time resolution of 1 second, and thus each HRF output will also be sampled at 1 second. Hence, the odd rows of the design matrix $S$ in (3.2) suffice for analysis. The study aims to estimate the BOLD response to each of the trial types for 1–18 seconds following the image onset. We analyze the fMRI dataset containing one subject. The length of the estimated HRF is set equal to 18.

A comparison of the activated brain regions detected by $\mathcal{K}$, $\mathcal{K}_{bc}$, AFNI and FSL is visualized in Figure 6. Again, HRF in FSL is specified as the
Fig 4. Comparison of Activated Brain Regions Discovered for the Simulated fMRI Dataset. Top panel: $\mathcal{K}$ (on the left) and $\mathcal{K}_{loc}$ (on the right). Bottom panel: AFNI (on the left) and FSL (on the right). The conventional FDR approach is used. The FDR level is .05.
Fig 5. The Captions are the Same as Those in Figure 4, Except That the new FDR Approach in Zhang, et al. (2006) is used.
difference of two gamma functions, and the drift term in AFNI is specified as a quadratic polynomial. We use FDR at level .001 to carry out the multiple comparisons. This level is set to avoid excessive discoveries, most of which are thought to be false. Our detected regions are closer to those afforded by AFNI, but our methods find activation in much more clustered regions of the brain. For example, our results do not have the holes as seen in the detected regions on the first slice of AFNI and FSL. AFNI gives more tiny scattered findings, which are more likely to be false discoveries. FSL detects very scattered regions which are difficult to interpret. In addition, the volumes of the detected regions by FSL are substantially smaller than those of AFNI and our methods.

APPENDIX A: PROOFS OF MAIN RESULTS

We first impose some technical assumptions, which are not the weakest possible. Throughout the proof, $C$ is used as a generic finite constant.

**Condition A:**

A1. The drift function $d(t)$ has a bounded continuous second derivative.

A2. The kernel $K$ is a symmetric probability density function with compact support, say, $[-L, L]$, Lipschitz continuous, and $\sup_t K(t) \leq C$ for some constant $C \in (0, \infty)$.

A3. Assume that $\{\epsilon(t_i)\}$ is a stationary $g$-dependent sequence with $E\{\epsilon(t_1)\} = 0$, $E\{\epsilon^2(t_1)\} = \sigma^2$ and $E\{\epsilon^4(t_1)\} < \infty$. Furthermore, the eigenvalues of $R$, the true correlation matrix of $\epsilon$, are uniformly bounded away from zero and infinity.

A4. In model (3.1), $\{s_j(\cdot), j = 1, \ldots, r\}$, are independent of $\{\epsilon(\cdot)\}$. For the RPER design, $s_j(t)$ is stationary and $P\{s_j(t) = 1\} = p_j \in (0, 1), j = 1, \ldots, r$, and $\sum_{j=1}^r p_j < 1$. Assume that $s_{j_1}(t_u)$ and $s_{j_2}(t_v)$ are independent at any $t_u \neq t_v$.

A5. $n \to \infty$, $b \to 0$ and $nb \to \infty$.

A6. $t_i = i/n, i = 1, \ldots, n$.

A7. $\text{cov}(S^T, S^T) > 0$.

A8. $\|\hat{R}^{-1} - R^{-1}\|_\infty = o(1)$.

We next introduce some necessary notations and definition.

**Notations:** For the kernel $K$ and bandwidth $b > 0$, define $K_b(t) = K(t/b)/b$. Denote by $e_j$ the $j$th column of an identity matrix. Set vectors $1 = (1, \ldots, 1)^T$ and $0 = (0, \ldots, 0)^T$. Define a matrix $H$ with entries $H(i, j) = n^{-1}K_b(t_j - t_i), 1 \leq i, j \leq n$. Define $V = R^{-1}$ and $\rho(l)$ the noise auto-correlation coefficient. Denote by $\xi_{j, \ell}$ the $\ell$th column vector of $S_j$, i.e. $\xi_{j, \ell} = S_j e_\ell$. Throughout the proof, $\| \cdot \|$ refers to the $L_2$-norm unless otherwise stated.
Fig 6. Comparison of Activated Brain Regions Discovered for the Real fMRI Dataset. Top panel: K (on the left) and K_{bc} (on the right). Bottom panel: AFNI (on the left) and FSL (on the right). The new FDR approach in Zhang, et al. (2006) is used. The FDR level is .001.
Definition 1. An \( n \times n \) matrix \( B \) is called “Row Absolute Value Uniformly Summable” (RAVUS) if there exists \( C > 0 \) such that
\[
\sup_{n \geq 1} \sup_{1 \leq j \leq n} \sum_{i=1}^{n} |B(i, j)| \leq C.
\]
Likewise, \( B \) is called “Column Absolute Value Uniformly Summable” (CAVUS) if there exists \( C > 0 \) such that \( \sup_{n \geq 1} \sup_{1 \leq i \leq n} \sum_{j=1}^{n} |B(i, j)| \leq C \). Moreover, if a matrix is both RAVUS and CAVUS, it is called “Absolute Value Uniformly Summable” (AVUS).

Before proving the main results of the paper, we need Lemmas 1–8.

Lemma 1. If both matrices \( B_1 \in \mathbb{R}^{n \times n} \) and \( B_2 \in \mathbb{R}^{n \times n} \) are AVUS, where the notion AVUS is defined in Definition 1 above, then \( B_1 B_2 \) is AVUS.

Proof. By the definition, there exists \( C > 0 \) such that
\[
\sup_{n \geq 1} \sup_{1 \leq j \leq n} \sum_{i=1}^{n} |B_1(i, j)| \leq C, \quad \sup_{n \geq 1} \sup_{1 \leq i \leq n} \sum_{j=1}^{n} |B_2(i, j)| \leq C.
\]
We observe that
\[
\sup_{n \geq 1} \sup_{1 \leq j \leq n} \sum_{i=1}^{n} |(B_1 B_2)(i, j)| \leq \sup_{n \geq 1} \sup_{1 \leq i \leq n} \sum_{l=1}^{n} |B_1(i, l) B_2(l, j)|
\]
\[
= \sup_{n \geq 1} \sup_{1 \leq j \leq n} \sum_{l=1}^{n} |B_2(l, j)| \sum_{i=1}^{n} |B_1(i, l)|
\]
\[
\leq C_1 \sup_{n \geq 1} \sup_{1 \leq j \leq n} \sum_{l=1}^{n} |B_2(l, j)| \leq C^2.
\]
Thus \( B_1 B_2 \) is RAVUS and similarly is CAVUS. Hence \( B_1 B_2 \) is AVUS.

Lemma 2. Assume Condition A3. Then there exist constants \( C \in (0, \infty) \) and \( \lambda \in (0, 1) \) such that \( |V(i, j)| \leq C \lambda^{i-j} \) for all \( 1 \leq i, j \leq n \) and \( n \geq 1 \).

Proof. Under Condition A3, \( R \) is positive definite, centered and \( 2g \)-banded. Let \( a_n \) and \( b_n \) be the minimum and maximum eigenvalues of \( R \) and set \( r_n = b_n/a_n \). Applying Demko, Moss and Smith (1984, Prop. 2.2) gives that \(|V(i, j)| \leq C_n \lambda_n^{i-j}|, \) where \( C_n = \max\{a_n^{-1}, (1 + r_n^{1/2})^2/(2a_n r_n)\} \) and \( \lambda_n = (r_n^{1/2} - 1)/(r_n^{1/2} + 1)^{1/2} \). From Condition A3, \( a_n \) and \( b_n \) are bounded away from zero and infinity; it follows that \( r_n \) is bounded away from zero and infinity. Thus, there exist \( C \in (0, \infty) \) and \( \lambda \in (0, 1) \) such that \( C_n < C \) and \( \lambda_n < \lambda \). Hence \( |V(i, j)| \leq C_n \lambda_n^{i-j} \leq C \lambda^{i-j} \).
Lemma 3. Assume Conditions A2 and A5.

1. Let $f$ be Lipschitz continuous and bounded on an interval $[d_1, d_2]$ where $d_1 < d_2$. Set $u_j = d_1 + (d_2 - d_1)j/n$, $j = 1, \ldots, n$. Then, uniformly in $\tau$,

$$
\frac{1}{n} \sum_{j=1}^{n} K_b(u_j - \tau)f(u_j) = \frac{1}{d_2 - d_1} \int_{d_1}^{d_2} K_b(u - \tau)f(u)du + O\left(\frac{1}{nb}\right).
$$

2. Let $\{\epsilon_j\}_{j=1}^{\infty}$ be a sequence of $g$-dependent and identically distributed random variables. Assume $E(\epsilon_j^2) < \infty$. Then for $t_j = j/n$, $j = 1, \ldots, n$,

$$
\sup_{t \in [bL, 1-bL]} E\left[\left\{\frac{1}{n} \sum_{j=1}^{n} \epsilon_j K_b(t_j - t) - E(\epsilon_1)\right\}^2\right] = O\left(\frac{1}{nb}\right).
$$

Proof. We first show (A.1). By the assumptions, there exists a constant $C > 0$ such that $|K(s) - K(t)| \leq C|s - t|$, $K(s) \leq C$, $|f(s) - f(t)| \leq C|s - t|$ and $|f(s)| \leq C$ for any $s$ and $t$. Define $\mathcal{J} = \{ j \in \mathbb{Z} : n(bL + \tau - d_1)/(d_2 - d_1) \leq j \leq n(bL + \tau - d_1)/(d_2 - d_1)\} = \{l_1, \ldots, l_2\}$. Clearly $\#\mathcal{J} \leq 2nbL/(d_2 - d_1) + 2$, $K_b(u_j - \tau) = 0$ for any $j \notin \mathcal{J}$, and $K_b(u - \tau) = 0$ for $u \leq u_{l_1} - 1$ or $u \geq u_{l_2 + 1}$. It follows that

$$
\left| \frac{1}{n} \sum_{j=1}^{n} K_b(u_j - \tau)f(u_j) - \frac{1}{d_2 - d_1} \int_{d_1}^{d_2} K_b(u - \tau)f(u)du \right|
$$

$$
= \frac{1}{d_2 - d_1} \left| \sum_{j \in \mathcal{J}} \int_{u_{j-1}}^{u_j} K_b(u_j - \tau)f(u_j)du - \sum_{j \in \mathcal{J}} \int_{u_{j-1}}^{u_j} K_b(u - \tau)f(u)du \right|
$$

$$
- \int_{u_{l_2}}^{u_{l_2+1}} K_b(u - \tau)f(u)du
$$

$$
\leq \frac{1}{d_2 - d_1} \sum_{j \in \mathcal{J}} \left\{ \int_{u_{j-1}}^{u_j} \frac{C(u_j - u)|f(u_j)|}{b^2} du + \int_{u_{j-1}}^{u_j} K_b(u - \tau)C(u_j - u)du \right\} + \frac{C^2}{nb}
$$

$$
\leq \left( \frac{nbL}{d_2 - d_1} + 1 \right) \left( \frac{C^2}{b^2} + \frac{C^2}{b} \right) \frac{d_2 - d_1}{n^2} + \frac{C^2}{nb} = O\left(\frac{1}{nb}\right).
$$
We now show (A.2). Following (A.1),

\[
\sup_{t \in [bL,1-bL]} E \left[ \left\{ \frac{1}{n} \sum_{j=1}^{n} \epsilon_j K_b(t_j - t) - E(\epsilon_1) \right\}^2 \right] \\
= \sup_{t \in [bL,1-bL]} \text{var} \left\{ \frac{1}{n} \sum_{j=1}^{n} \epsilon_j K_b(t_j - t) \right\} \\
+ \sup_{t \in [bL,1-bL]} \left\{ \frac{1}{n} \sum_{j=1}^{n} K_b(t_j - t) - 1 \right\}^2 \{E(\epsilon_1)\}^2 \\
\leq \frac{C^2 g}{n^2 b^2} (2nbL + 2)\text{var}(\epsilon_1) + O\left( \frac{1}{n^2 b^2} \right) = O\left( \frac{1}{nb} \right).
\]

\[\square\]

**Corollary 1.** Assume Conditions A2, A5 and A6.
1. For any \( l = 0, 1, 2, \ldots \), we have that uniformly in \( t \in [0, 1] \),

\[ \frac{1}{n} \sum_{j=1}^{n} K_b(t_j - t)(t_j - t)^l = b^l \left\{ \int_{-t/b}^{(1-t)/b} K(u)u^l du + O\left( \frac{1}{nb} \right) \right\}, \]

and thus, uniformly in \( i \in [nbL, \ldots, n - nbL] \),

\[ \frac{1}{n} \sum_{j=1}^{n} K_b(t_j - t_i)(t_j - t_i)^l = b^l \left\{ \int_{-L}^{L} K(u)u^l du + O\left( \frac{1}{nb} \right) \right\}. \]

2. There exists \( C > 0 \) such that for all \( n = 1, 2, \ldots \), \( i \in \{1, \ldots, n\} \) and \( j \in \{1, \ldots, n\} \),

\[ nb|S_d(i, j)| \leq C. \]

Moreover, \( S_d \) is AVUS. Furthermore, for all \( n = 1, 2, \ldots \), \( i \in [nbL, \ldots, n - nbL] \) and \( j \in \{1, \ldots, n\} \),

\[ S_d(i, j) = H(i, j)(1 + c_{n, i}), \]

where \( \sup_{nbL \leq i \leq n - nbL} |c_{n, i}| = O\{1/(nb)\} \).

3. Let \( \{\epsilon_j\}_{j=1}^{\infty} \) be a sequence of \( g \)-dependent and identically distributed random variables. Assume \( E(\epsilon_1) = 0 \) and \( E(\epsilon_1^2) < \infty \). Let \( Y_i = n^{-1} \sum_{j=1}^{n} \epsilon_j K_b(t_j - t_i) \). Denote \( \epsilon = (\epsilon_1, \ldots, \epsilon_n)^T \) and \( y = (Y_1, \ldots, Y_n)^T \). Assume that \( B \in \mathbb{R}^{n \times n} \) is AVUS. Then

\[ n^{-1}E(\|By\|^2) = o(1), \]

and

\[ n^{-1}E(\|BS_d\epsilon\|^2) = o(1). \]
PROOF. Part 1. Following the proof of (A.1),
\[
\left| \frac{1}{n} \sum_{j=1}^{n} K_b(t_j - t) \left( \frac{t_j - t}{b} \right)^l - \int_{-t/b}^{(1-t)/b} K(u)u' du \right|
\]
\[
= \left| \sum_{j \in \mathbb{J}} \int_{t_{j-1}}^{t_j} K_b(t_j - t) \left( \frac{t_j - t}{b} \right)^l du
- \sum_{j \in \mathbb{J}} \int_{t_{j-1}}^{t_j} K_b(u - t) \left( \frac{u - t}{b} \right)^l du - \int_{t_{j+1}}^{t_{j+2}} K_b(u - t) \left( \frac{u - t}{b} \right)^l du \right|
\]
\[
\leq \sum_{j \in \mathbb{J}} \int_{t_{j-1}}^{t_j} \left| K_b(t_j - t) - K_b(u - t) \right| \left( \frac{t_j - t}{b} \right)^l \left( \frac{u - t}{b} \right)^l du
+ \sum_{j \in \mathbb{J}} \int_{t_{j-1}}^{t_j} \left( \frac{t_j - t}{b} \right)^l \left( \frac{u - t}{b} \right)^l du + \frac{1}{nb}.
\]
Note that for \( j \in \mathbb{J}, \left| (t_j - t)/b \right| \leq L \), and that for \( j \in \mathbb{J} \) and \( t_{j-1} \leq u \leq t_j, \left| ((t_j - t)/b) - ((u - t)/b) \right| \leq C \left| (t_j - u)/b \right| \). Applying the same argument for (A.1) completes the proof for (A.3) and in turn (A.4).

We then show Part 2. Define \( S_{n,t}(t) = n^{-1} \sum_{j=1}^{n} K_b(t_j - t)(t_j - t)^l \), \( l = 0, 1, 2 \), and \( S_n(t) = X(t)^TW(t)X(t) \). Then
\[
S_n(t) = \begin{bmatrix} S_{n,0}(t) & S_{n,1}(t) \\ S_{n,1}(t) & S_{n,2}(t) \end{bmatrix},
\]
\[
\{S_n(t)\}^{-1} = \frac{1}{n[S_{n,0}(t)S_{n,2}(t) - (S_{n,1}(t))^2]} \begin{bmatrix} S_{n,2}(t) & -S_{n,1}(t) \\ -S_{n,1}(t) & S_{n,0}(t) \end{bmatrix}.
\]
According to (A.3), uniformly in \( t \in [0, 1] \),
\[
\{S_n(t)\}^{-1} = \frac{1}{nb^2[f(t) + O\{1/(nb)\}]}
\times \begin{bmatrix} b^2[a_2(t) + O\{1/(nb)\}] & -b[a_1(t) + O\{1/(nb)\}] \\ -b[a_1(t) + O\{1/(nb)\}] & a_0(t) + O\{1/(nb)\} \end{bmatrix},
\]
where \( a_l(t) = \int_{-t/b}^{(1-t)/b} K(u)u^l du, l = 0, 1, 2 \), are all uniformly bounded in \( t \), and \( f(t) = a_0(t)a_2(t) - \{a_1(t)\}^2 \) is minimized at \( t = 0 \) with \( f(0) = .25\text{var}(|U|) > 0 \) for a random variable \( U \) with density \( K(u) \). It is seen from the definition of \( S_d \) in (3.3) that
\[
S_d(i,j) = \frac{K_b(t_j - t_i)/n}{f(t_i) + O\{1/(nb)\}} \left[ \{a_2(t_i) + O\left(\frac{1}{nb}\right)\} - \frac{t_j - t_i}{b} \{a_1(t_i) + O\left(\frac{1}{nb}\right)\} \right].
\]
Note that when \( j \notin [i-nbL, \ldots, i+nbL] \), \( K_b(t_j-t_i) = 0 \) implies \( S_d(i,j) = 0 \). Note also that when \( j \in [i-nbL, \ldots, i+nbL] \), \( |t_j-t_i|/b \leq L \). Thus, there exists \( C > 0 \) uniformly in \( i \in \{1, \ldots, n\} \) and \( j \in [i-nbL, \ldots, i+nbL] \), such that \( |S_d(i,j)| \leq Cn^{-1}K_b(t_j-t_i) \leq C \sup_i K(t)/nb \). Thus for some \( C > 0 \), uniformly in \( n = 1, 2, \ldots, i = 1, \ldots, n \) and \( j = 1, \ldots, n \),

\[
|S_d(i,j)| = \begin{cases} 
0, & \text{if } |j-i| > nbL, \\
\leq C/(nb), & \text{if } |j-i| \leq nbL,
\end{cases}
\]

which implies (A.5) and also implies that \( S_d \) is AVUS. Similar arguments for (A.5) combined with (A.4) and (3.3) yield (A.6).

We now show Part 3. From (A.2), for any \( \epsilon > 0 \), there exists \( N \) such that \( n > N \) implies that \( E(Y_i^2) < \epsilon \) for all \( i \in [nbL, \ldots, n-nbL] \). It can also be shown that there exists \( C > 0 \) such that \( E(Y_i^2) \leq C \) for all \( i \in [1, \ldots, nbL] \cup [n-nbL, \ldots, n] \). Since \( B \) is AVUS, there exists \( C_1 > 0 \) such that \( \sup_{n \geq 1} \sup_{1 \leq j \leq n} \sum_{i=1}^{n} |B(i,j)| \leq C_1 \) and \( \sup_{n \geq 1} \sup_{1 \leq i \leq n} \sum_{j=1}^{n} |B(i,j)| \leq C_1 \). The proof of (A.7) is obtained as follows,

\[
n^{-1}E(||BY||^2) = \frac{1}{n} \sum_{i=1}^{n} \sum_{j=1}^{n} \sum_{k=1}^{n} B(i,j)B(i,k)E(Y_iY_k) \\
\leq \frac{1}{n} \sum_{i=1}^{n} \sum_{j=1}^{n} \sum_{k=1}^{n} |B(i,j)B(i,k)| \{E(Y_i^2)E(Y_k^2)\}^{1/2} \\
\leq \epsilon^{1/2}C^{1/2} \left( \frac{1}{n} \sum_{i=1}^{n} \left( \sum_{j=1}^{n} |B(i,j)| \right)^2 \right) \\
+ \frac{C}{n} \sum_{i=1}^{n} \sum_{j,k \in [1, \ldots, nbL] \cup [n-nbL, \ldots, n]} |B(i,j)B(i,k)| \\
\leq \epsilon^{1/2}C^{1/2}C_1^2 + 2C_1^2CbL + 2C_1^2C/n.
\]

Applying (A.5), (A.6) and similar arguments for (A.7) completes the proof of (A.8).

**Lemma 4.** Let \( \{X_n\}_{n=1}^{\infty} \) be a sequence of random variables such that every subsequence of \( X_n \) has a further subsequence converging in distribution to a same random variable \( X \). Then \( X_n \xrightarrow{d} X \).

**Proof.** Let \( \phi_Y \) denote the characteristic function of a random variable \( Y \). Since any subsequence \( \{n_i\}_{i=1}^{\infty} \) of \( \{1,2,\ldots\} \) has a further subsequence, \( \{n_{i_j}\}_{j=1}^{\infty} \), such that \( X_{n_{i_j}} \xrightarrow{d} X \), the Lévy-Cramér continuity theorem (Shao
2003, p. 56) implies that \( \phi_{X_{n j}}(t) \to \phi_X(t) \) as \( j \to \infty \) for any \( t \in \mathbb{R} \). This in turn indicates that \( \phi_{X_n}(t) \to \phi_X(t) \) as \( n \to \infty \) for any \( t \in \mathbb{R} \). Then we conclude that \( X_n \xrightarrow{D} X \) by repeated application of the Lévy-Cramér continuity theorem.

\[ \square \]

**Lemma 5.** Let \( \{\epsilon_i\}_{i=1}^n \) be a stationary \( g \)-dependent sequence with \( E(\epsilon_1) = 0 \) and \( E(\epsilon_1^2) < \infty \). Set \( x_{n,i} = \tau_{n,i} \epsilon_i \), \( i = 1, \ldots, n \), where \( \{\tau_{n,i}\} \) is independent of \( \{\epsilon_i\} \). Define \( \sigma_n^2(\{\tau_{n,i}\}) = E[(\sum_{i=1}^n x_{n,i})^2|\{\tau_{n,i}\}] \). If \( \sup_{n \geq 1} \sup_{1 \leq i \leq n} |\tau_{n,i}| \leq C \) and \( \sigma_n^2(\{\tau_{n,i}\}) = n \sigma_0^2(1 + o_P(1)) \) for some constants \( C > 0 \) and \( \sigma_0^2 \in (0, \infty) \), then \( \sigma_n^{-1} \sum_{i=1}^n x_{n,i} \xrightarrow{D} N(0,1) \).

**Proof.** The proof can be achieved by applying the blocking arguments and checking the Liapunov condition of central limit theorem.

**Lemma 6.** Assume model (3.2) and Condition A1–A7. Then

1. All three matrices \( \tilde{V}_R = V(I - S_d) \), \( \tilde{V}_L = (I - S_d)^T V \), and \( \tilde{V} = (I - S_d)^T V (I - S_d) \) are AVUS.
2. \( \text{var}(n^{-1} \xi_{j_1, \ell_1}^T \tilde{V} \xi_{j_2, \ell_2}) \to 0 \) for all \( j_1, j_2 = 1, \ldots, r \) and all \( \ell_1, \ell_2 = 1, \ldots, m \).
3. \( n^{-1} \tilde{S}^T R^{-1} \tilde{S} - E(n^{-1} \tilde{S}^T R^{-1} \tilde{S}) \to 0 \).
4. All entries of \( E(n^{-1} \tilde{S}^T R^{-1} \tilde{S}) \) are bounded.
5. All convergent subsequences of \( E(n^{-1} \tilde{S}^T R^{-1} \tilde{S}) \) are positive definite.

**Proof.** The proof of Part 1 can be obtained from applying Lemma 1, Lemma 2 and Part 2 of Corollary 1.

We next show Part 2. For \( \ell_1, \ell_2 = 1, \ldots, m \),

\[
n^{-1} \xi_{j_1, \ell_1}^T \tilde{V} \xi_{j_2, \ell_2} = n^{-1} \sum_{k_1=1}^n \sum_{k_2=1}^n s_{j_1}(k_1, \ell_1) \tilde{V}(k_1, k_2) s_{j_2}(k_2, \ell_2)
\]

\[
= n^{-1} \sum_{k_1=1}^n \sum_{k_2=1}^n s_{j_1}(k_1 - t_{\ell_1}) s_{j_2}(k_2 - t_{\ell_2}) \tilde{V}(k_1, k_2)
\]

\[
= I_{1,1} + I_{1,2} + I_{1,3} + I_{1,4},
\]
where
\[ I_{1,1} = n^{-1} p_j_1 p_j_2 \sum_{k_1 = \ell_1}^n \sum_{k_2 = \ell_2}^n \tilde{V}(k_1, k_2), \]
\[ I_{1,2} = n^{-1} p_j_2 \sum_{k_1 = \ell_1}^n \{ s_{j_1}(t_{k_1} - t_{\ell_1}) - p_j_1 \} \sum_{k_2 = \ell_2}^n \tilde{V}(k_1, k_2), \]
\[ I_{1,3} = n^{-1} p_j_1 \sum_{k_2 = \ell_2}^n \{ s_{j_2}(t_{k_2} - t_{\ell_2}) - p_j_2 \} \sum_{k_1 = \ell_1}^n \tilde{V}(k_1, k_2), \]
\[ I_{1,4} = n^{-1} \sum_{k_1 = \ell_1}^n \sum_{k_2 = \ell_2}^n \{ s_{j_1}(t_{k_1} - t_{\ell_1}) - p_j_1 \} \{ s_{j_2}(t_{k_2} - t_{\ell_2}) - p_j_2 \} \tilde{V}(k_1, k_2). \]

It is easily seen that
\[ \text{(A.9)} \quad \text{var}(I_{1,1}) = 0, \text{ var}(I_{1,2}) = O(n^{-1}), \text{ and var}(I_{1,3}) = O(n^{-1}). \]

For \( I_{1,4}, \)
\[ E(I_{1,4}^2) = \frac{1}{n^2} \sum_{k_1 = \ell_1}^n \sum_{k_2 = \ell_2}^n \sum_{k_3 = \ell_3}^n \sum_{k_4 = \ell_4}^n E\{ \{ s_{j_1}(t_{k_1} - t_{\ell_1}) - p_j_1 \} \{ s_{j_2}(t_{k_2} - t_{\ell_2}) - p_j_2 \} \tilde{V}(k_1, k_2) \tilde{V}(k_3, k_4), \}
\]
in which two situations will be discussed. In the situation of \( j_1 = j_2 = j, \)
the additive term above is nonzero only in the following four cases,
\[
I: \quad k_1 - \ell_1 = k_2 - \ell_2 = k_3 - \ell_1 = k_4 - \ell_2, \\
II: \quad \{ k_1 - \ell_1 = k_2 - \ell_2 \} \neq \{ k_3 - \ell_1 = k_4 - \ell_2 \}, \\
III: \quad \{ k_1 - \ell_1 = k_3 - \ell_1 \} \neq \{ k_2 - \ell_2 = k_4 - \ell_2 \}, \\
IV: \quad \{ k_1 - \ell_1 = k_4 - \ell_2 \} \neq \{ k_2 - \ell_2 = k_3 - \ell_1 \}.
\]

Thus, \( E(I_{1,4}^2) = E_I + E_{II} + E_{III} + E_{IV}, \) where
\[
E_I \leq n^{-2} [p_j(1 - p_j) \{ p_j^3 + (1 - p_j)^3 \}] \sum_{k_1 = 1}^n \{ \tilde{V}(k_1, k_1 + \ell_2 - \ell_1) \}^2 = O(n^{-2}),
\]
\[
E_{II} \leq n^{-2} [p_j(1 - p_j)^2] \sum_{k_1 = 1}^n \sum_{k_3 = 1}^n |\tilde{V}(k_1, k_1 + \ell_2 - \ell_1)| \cdot |\tilde{V}(k_3, k_3 + \ell_2 - \ell_1)| = O(n^{-2}),
\]
\[
E_{III} \leq n^{-2} [p_j(1 - p_j)^2] \sum_{k_1 = 1}^n \sum_{k_2 = 1}^n \{ \tilde{V}(k_1, k_2) \}^2 = O(n^{-2}),
\]
\[
E_{IV} \leq n^{-2} [p_j(1 - p_j)^2] \sum_{k_1 = 1}^n \sum_{k_2 = 1}^n |\tilde{V}(k_1, k_2)| \cdot |\tilde{V}(k_2 - \ell_2 + \ell_1, k_1 + \ell_2 - \ell_1)| = O(n^{-2}).
\]
Hence \( E(I_{1,4}^2) = O(n^{-2}) \) when \( j_1 = j_2 \). In the situation of \( j_1 \neq j_2 \), since \( s_{j_1}() \) are independent at different time points and similarly, \( s_{j_2}() \) are independent at different time points, \( E[\{s_{j_1}(t_{k_1} - t_{\ell_1}) - p_{j_1}\} \{s_{j_2}(t_{k_2} - t_{\ell_2}) - p_{j_2}\}] \) is nonzero only if \( k_1 = k_3 \) and \( k_2 = k_4 \). In this case,

\[
E(I_{1,4}^2) = \frac{1}{n^2} \sum_{k_1=\ell_1}^n \sum_{k_3=\ell_1}^n \sum_{k_2=\ell_2}^n \sum_{k_4=\ell_2}^n E[\{s_{j_1}(t_{k_1} - t_{\ell_1}) - p_{j_1}\} \{s_{j_2}(t_{k_2} - t_{\ell_2}) - p_{j_2}\} \tilde{V}(k_1, k_2)\tilde{V}(k_3, k_4)]
\]

\[
= \frac{1}{n^2} \sum_{k_1=\ell_1}^n \sum_{k_2=\ell_2}^n E[\{s_{j_1}(t_{k_1} - t_{\ell_1}) - p_{j_1}\}^2 \{s_{j_2}(t_{k_2} - t_{\ell_2}) - p_{j_2}\}^2] \times \{\tilde{V}(k_1, k_2)\}^2
\]

\[
= n^{-2} C \sum_{k_1=\ell_1}^n \sum_{k_2=\ell_2}^n \{\tilde{V}(k_1, k_2)\}^2 = O(n^{-2}).
\]

Thus in both situations, \( \text{var}(I_{1,4}) \to 0 \). This combined with (A.9) yields Part 2.

We then show Part 3. Recall that \( \tilde{S} = (I - S_d)S = [\tilde{S}_1, \ldots, \tilde{S}_r] \), where \( \tilde{S}_j = (I - S_d)S_j \). Then

\[
n^{-1}\tilde{S}^T R^{-1}\tilde{S} = n^{-1} \begin{bmatrix} \tilde{S}_1^T R^{-1}\tilde{S}_1 & \cdots & \tilde{S}_1^T R^{-1}\tilde{S}_r \\ \vdots & \ddots & \vdots \\ \tilde{S}_r^T R^{-1}\tilde{S}_1 & \cdots & \tilde{S}_r^T R^{-1}\tilde{S}_r \end{bmatrix}.
\]

It suffices to consider the block matrix \( n^{-1}\tilde{S}_j^T R^{-1}\tilde{S}_j \), whose \((\ell_1, \ell_2)\)th entry is \( C_{j_1,\ell_1,j_2,\ell_2} = n^{-1}E[\tilde{\xi}_{j_1,\ell_1} \tilde{V}\tilde{\xi}_{j_2,\ell_2}] \). By Part 2, \( \text{var}(C_{j_1,\ell_1,j_2,\ell_2}) \to 0 \), which in turn gives \( C_{j_1,\ell_1,j_2,\ell_2} = E(C_{j_1,\ell_1,j_2,\ell_2}) \overset{P}{\to} 0 \) and the conclusion of Part 3.

We now show Part 4, which can easily be derived from

\[
|n^{-1}\tilde{\xi}_{j_1,\ell_1} \tilde{V}\tilde{\xi}_{j_2,\ell_2}| \leq n^{-1} \sum_{k_1=1}^n \sum_{k_2=1}^n |\tilde{V}(k_1, k_2)| \leq C,
\]

since the entries of \( S_j \) are either 0 or 1.

Last, we show Part 5. For any \( \{n_k\}_{k=1}^\infty \) such that \( E(n_k^{-1}\tilde{S}^T R^{-1}\tilde{S}) \) converges to some limit \( M \), by Part 3, \( n_k^{-1}\tilde{S}^T R^{-1}\tilde{S} \overset{P}{\to} M \). Obviously, \( M \) is semi-positive definite. It remains to show that \( M \) is nonsingular. We now prove this by contradiction. Assume that there exists some \( c = (c_1^T, \ldots, c_r^T)^T \in \mathbb{R}^{rm} \), where \( c_j = (c_{j,1}, \ldots, c_{j,m})^T, j = 1, \ldots, r \), such that \( c \neq 0 \) and \( c^T M c = \).
0. Then $n_k^{-1}c^T \tilde{S}^T R^{-1} \tilde{S} c \overset{p}{\to} 0$. By the Schur decomposition, there exist $Q$ and $v_1, \ldots, v_{n_k}$, such that $V = Q^T \text{diag}(v_1, \ldots, v_{n_k})Q$ where $Q^T Q = I_{n_k}$ and $v_{n_k} \leq \cdots \leq v_1$. Furthermore, from Condition A3, $V$ is positive definite with eigenvalues bounded away from 0 and $\infty$. Thus, there exist constants $a$ and $b$ such that $0 < a < b < \infty$ and $0 < a \leq v_{n_k} \leq \cdots \leq v_1 \leq b < \infty$. Noting

$$n_k^{-1}c^T \tilde{S}^T R^{-1} \tilde{S} c = \tilde{c}^T \text{diag}(v_1, \ldots, v_{n_k}) \tilde{c} \geq a\|\tilde{c}\|^2,$$

where $\tilde{c} \equiv Q \tilde{S} c / \sqrt{n_k}$, we conclude that as $k \to \infty$,

$$(A.10) \quad \|(I - S_d) \tilde{S} c\|^2 / n_k \overset{p}{\to} 0.$$

Consider

$$(A.11) \quad \|(I - S_d) \tilde{S} c\|^2 / n = \|\{Sc - E(S) c\} - \{S_d Sc - E(S) c\}\|^2 / n$$

where $J_1 = \|Sc - E(S) c\|^2 / n$, $J_2 = \{Sc - E(S) c\}^T \{S_d Sc - E(S) c\} / n$ and $J_3 = \|S_d Sc - E(S) c\|^2 / n$. By the law of large numbers and block arguments, we can show that

$$(A.12) \quad J_1 \overset{p}{\to} \text{var} \left( \sum_{j=1}^r c_j^T s_{j,m} \right) \geq \left(1 - \sum_{j=1}^r p_j\right) \left(\sum_{j=1}^r p_j \|c_j\|^2\right) > 0,$$

where $s_{j,i} = S_j^T e_i$, $i = 1, \ldots, m$. For $J_3$, note

$$E(J_3) \leq o(1) + \frac{\|c\|^2}{n} \sum_{i=1}^n E \left[ \sum_{j=1}^r \sum_{k=1}^m \{ \sum_{\ell=1}^n S_d(i, \ell) s_j(t_\ell - t_k) - p_j \}^2 \right]$$

$$\leq o(1) + \frac{\|c\|^2}{n} \sum_{j=1}^r \sum_{k=1}^m (n - 2nbL) \sup_{i \in [nL, \ldots, n - nbL]} E \left[ \sum_{\ell=1}^n S_d(i, \ell) s_j(t_\ell - t_k) - p_j \right]^2$$

$$+ \frac{\|c\|^2}{n} \sum_{j=1}^r \sum_{k=1}^m 2nbL \sup_{i \in [1, \ldots, nL] \cup [n - nbL, \ldots, n]} E \left[ \sum_{\ell=1}^n S_d(i, \ell) s_j(t_\ell - t_k) - p_j \right]^2$$

$$= o(1) + \frac{\|c\|^2}{n} \sum_{j=1}^r \sum_{k=1}^m (n - 2nbL) \times \sup_{i \in [nL, \ldots, n - nbL]} E \left[ \{1 + o(1)\} \frac{1}{n} \sum_{\ell=1}^n K_b(t_\ell - t_i) s_j(t_\ell - t_k) - p_j \right]^2$$

$$+ \frac{\|c\|^2}{n} \sum_{j=1}^r \sum_{k=1}^m 2nbL \sup_{i \in [1, \ldots, nL] \cup [n - nbL, \ldots, n]} E \left[ \sum_{\ell=1}^n S_d(i, \ell) s_j(t_\ell - t_k) - p_j \right]^2$$

$$= o(1) + \|c\|^2 \left\{ \sum_{j=1}^r \sum_{k=1}^m o(1) \right\} + \|c\|^2 \left\{ \sum_{j=1}^r \sum_{k=1}^m O(b) \right\}.$$
In the last equality, the second term follows from (A.2), whereas the third term uses \(|s_j(t_e - t_k)| \leq 1\) and \(S_d\) being AVUS. Thus, \(E(J_3) \to 0\) and \(J_3 \geq 0\) imply that
\[(A.13) \quad J_3 = o_P(1).\]
By Cauchy-Schwartz inequality, \(|J_2| \leq 2(J_1J_3)^{1/2}\). Thus, \(J_2 = o_P(1)\). This together with (A.11), (A.12) and (A.13) shows that \(\|(I - S_d)\mathbf{S}\|_2^2/n \stackrel{P}{\to} \text{var}(\sum_{j=1}^r c_j^T s_{j,m})\), which contradicts (A.10).

**Lemma 7.** Assume Condition A. Suppose that \(n^{-1}\mathbf{S}_T\mathbf{R}^{-1}\mathbf{S} \xrightarrow{P} \mathbf{M}\), where \(\mathbf{M} \in \mathbb{R}^{m \times m}\) is positive definite. Then \(n^{-1}\mathbf{S}_T\mathbf{R}^{-1}\mathbf{S} \xrightarrow{P} \mathbf{M}\) and \(n^{1/2}(\mathbf{h} - \mathbf{h}) \xrightarrow{L} N(0, \sigma^2 \mathbf{M})\) and
\[(A.14) \quad n^{-1}\mathbf{S}_T(\mathbf{R}^{-1} - R^{-1})\mathbf{S} \xrightarrow{P} \mathbf{0}, \quad J_1^* - J_1 = o_P(1), \quad J_2^* - J_2 = o_P(1).\]

First, we will show \(J_1 = o_P(1)\). From (A.6), the \(i\)th entry of \((\mathbf{I} - S_d)\mathbf{d}\) is
\[
d(t_i) - \{1 + o(1)\} \frac{1}{n} \sum_{j=1}^n K_b(t_j - t_i)d(t_j)
= d(t_i) - \{1 + o(1)\} \int_{-t_i/b}^{(1+t_i)/b} K(u)d(t_i + ub)du + O\left(\frac{1}{nb}\right)
= \{1 + o(1)\} \int_{-L}^L \{ubK(u)d(t_i) + u^2b^2K(u)d''(\xi)/2\}du + o(1) + O\left(\frac{1}{nb}\right)
= o(1)
\]
uniformly in \(i \in [nbL, \ldots, n - nbL]\). When \(i \in \{1, \ldots, nbL\} \cup \{n - nbL, \ldots, n\}\),
\[
\left|d(t_i) - \{1 + o(1)\} \frac{1}{n} \sum_{j=1}^n K_b(t_j - t_i)d(t_j)\right|
\leq \sup_{i \in [0,1]} |d(t)|\left\{1 + \sup_{1 \leq \ell \leq n} \left|\frac{1}{n} \sum_{j=1}^n K_b(t_j - t_\ell)\right|\right\} \leq C,
\]
for some \(C > 0\). Thus,
\[(A.15) \quad \sup_{i \in \{nbL, \ldots, n - nbL\}} |e_i^T \mathbf{d}| = o(1), \quad \sup_{n \geq 1} \sup_{i \in \{1, \ldots, nbL\} \cup \{n - nbL, \ldots, n\}} |e_i^T \mathbf{d}| \leq C.\]
Consider the $j$th block vector of $J_1$: $J_{1,j} = n^{-1/2}S_j^TR^{-1}d$. Its $ith$ entry is $e_i^T J_{1,j} = n^{-1/2}d^TV(I - S_d)(\xi_{j,i} - p_j1)$. Then

$$E(e_i^T J_{1,j}) = n^{-1/2}d^TV(I - S_d)\{p_j(0_{i-1}, 1_{n-i+1}^T) - p_j1\}$$

$$= -p_jn^{-1/2}d^TV(I - S_d)(1_{i-1}^T, 0_{n-i+1}^T)$$

$$= -p_jn^{-1/2} \sum_{k=1}^n \sum_{\ell=1}^{i-1} d(t_k) \tilde{V}_R(k, \ell),$$

and thus $|E(e_i^T J_{1,j})| \leq p_jn^{-1/2} \sum_{k=1}^n \sum_{\ell=1}^{i-1} |\tilde{V}_R(k, \ell)|\{\sup_{1\leq t \leq n} |\tilde{d}(t)|\} = o(1)$ by (A.15) and AVUS of $\tilde{V}_R$. Moreover,

$$\text{var}(e_i^T J_{1,j}) = n^{-1}d^TV(I - S_d)\begin{bmatrix} 0 & 0 \\ 0 & p_j(1 - p_j)I \end{bmatrix}(I - S_d)^TVd$$

$$\leq n^{-1}p_j(1 - p_j)||\tilde{V}_Ld||^2,$$

which implies that $\text{var}(J_{1,j}) = 1_m1_m^T o(1)$ using similar derivations for (A.7).

Thus $J_{1,j} \overset{p}{\to} 0$ and hence $J_1 \overset{p}{\to} 0$.

Second, we will show that $J_2 \overset{p}{\to} N(0, \sigma^2 M)$. Since $\tilde{e} = \epsilon - S_d\epsilon$, $J_2 = n^{-1/2}S^TR^{-1}\epsilon - n^{-1/2}S_j^TR^{-1}S_d\epsilon \equiv J_{21} - J_{22}$. Consider the $j$th block vector of $J_{22}: J_{22,j} = n^{-1/2}S_j^TR^{-1}S_d\epsilon$. Its $ith$ entry is $e_i^T J_{22,j} = n^{-1/2}(S_d\epsilon)^T V(I - S_d)\xi_{j,i}$, thus

$$E(e_i^T J_{22,j}) = E[E\{n^{-1/2}(S_d\epsilon)^T V(I - S_d)\xi_{j,i}\}] = 0,$$

and

$$\text{var}(e_i^T J_{22,j}) = \text{var}(E(e_i^T J_{22,j}|\epsilon)) + E\{\text{var}(e_i^T J_{22,j}|\epsilon)\}$$

$$= n^{-1}p_j^2 \text{var}\{(S_d\epsilon)^T V(I - S_d)(1_{i-1}^T, 0_{n-i+1}^T)\}$$

$$+ n^{-1} E\{(S_d\epsilon)^T V(I - S_d)\begin{bmatrix} 0 & 0 \\ 0 & p_j(1 - p_j)I \end{bmatrix}(I - S_d)^TV(S_d\epsilon)\}$$

$$\leq n^{-1}p_j^2 \sigma^2 \|R^{1/2}S_d^TV(I - S_d)(1_{i-1}^T, 0_{n-i+1}^T)\|^2$$

$$+ n^{-1}p_j(1 - p_j) E(||\tilde{V}_LS_d\epsilon||^2)$$

$$= o(1) + o(1) = o(1).$$

In the last equality, the first $o(1)$ is from Lemma 2 and similar arguments for (A.8). The second $o(1)$ is from (A.8) and Lemma 6 PART 1. Thus, $J_{22,j} = o_P(1)$ and $J_{22} = o_P(1)$. For $J_{21}$, by the Cramer-Wold device, it suffices to
show that for any \( w = (w_1^T, \ldots, w_T^T)^T \in \mathbb{R}^{rm} \), \( w^T J_{21} \overset{c}{\rightarrow} N(0, \sigma^2 w^T M w) \), where \( w_j = (w_{j1}, \ldots, w_{jm})^T, j = 1, \ldots, r \). Note that \( w^T J_{21} = n^{-1/2} \sum_{i=1}^{n} \tau_{n,i} \epsilon_i \), where \( \tau_{n,i} = \sum_{j=1}^{r} \sum_{k=1}^{m} \sum_{\ell=1}^{m} w_{j \ell} \delta_j (t_k - t_\ell) \bar{V}_L(k, i) \). Thus

\[
|\tau_{n,i}| \leq r m \left( \max_{1 \leq j \leq r} \max_{1 \leq \ell \leq m} |w_{j \ell}| \right) \sum_{k=1}^{n} |\bar{V}_L(k, i)| \leq C,
\]

where the last inequality is from PART 1 of Lemma 6. Also, \( \sigma_n^2 \{\{\tau_{n,i}\}\} = n \operatorname{var}(w^T J_{21} | \{\tau_{n,i}\}) = \sigma^2 w^T \tilde{S} T^{-1} \tilde{S} w = n \sigma^2 w^T M w \{1 + o_P(1)\} = n \sigma^2 \{1 + o_P(1)\} \), where \( \sigma_n^2 = \sigma^2 w^T M w \). By Lemma 5, the result follows.

Third, to verify (A.14), it is sufficient to show

(A.16) \quad n^{-1} \tilde{S}^T (\tilde{R}^{-1} - R^{-1}) \tilde{S} = o_P(1),

(A.17) \quad n^{-1/2} \tilde{S}^T (\tilde{R}^{-1} - R^{-1}) \tilde{d} = o_P(1),

(A.18) \quad n^{-1/2} \tilde{S}^T (\tilde{R}^{-1} - R^{-1}) \tilde{e} = o_P(1).

Note that Condition A8 implies that \( \tilde{R}^{-1} - R^{-1} \) is \( \mathbf{AVUS} \). Similar arguments for Lemma 6, \( J_1 \) and \( J_2 \) complete the proofs for (A.16)–(A.18) respectively.

\[ \square \]

**Corollary 2.** Assume Condition A. Then

1. \( \hat{h} \overset{P}{\rightarrow} h \).
2. \( (A\hat{h} - Ah)^T \{A(\tilde{S}^T \tilde{R}^{-1} \tilde{S})^{-1} A^T\}^{-1} (A\hat{h} - Ah) \overset{L}{\rightarrow} \sigma^2 h_k^2 \).

**Proof.** By Lemma 6, for any subsequence \( \{n_{j1}\}_{j=1}^\infty \), there exists a further subsequence, \( \{n_{j2}\}_{j=1}^\infty \), such that \( n_{j1}^{-1} \tilde{S}^T \tilde{R}^{-1} \tilde{S} \overset{p}{\rightarrow} M_l \) for some positive definite matrix \( M_l \). For this \( \{n_{j2}\}_{j=1}^\infty \), an appeal to Lemma 7 gives \( n_{j2}^{-1} \tilde{S}^T \tilde{R}^{-1} \tilde{S} \overset{p}{\rightarrow} M_l \) and \( n_{j2}^{-1/2} (\hat{h} - h) \overset{L}{\rightarrow} N(0, \sigma^2 M_l^{-1}) \).

It follows that along \( \{n_{j1}\}_{j=1}^\infty \), \( \hat{h} \overset{P}{\rightarrow} h \), as \( j \to \infty \). Thus, for any subsequence of \( \hat{h} \), there exists a further subsequence along which \( \hat{h} \overset{P}{\rightarrow} h \). This gives \( \hat{h} \overset{P}{\rightarrow} h \), as \( n \to \infty \).

We now show the second part. Applying Slusky’s theorem gives that as \( j \to \infty \), \( A(\tilde{S}^T \tilde{R}^{-1} \tilde{S})^{-1} A^T \) has an asymptotic Gaussian distribution with mean vector zero and variance-covariance matrix \( \sigma^2 h_k^2 \), which implies that, for \( \{n_{j2}\}_{j=1}^\infty \),

\[
(\hat{h} - h)^T A^T \{A(\tilde{S}^T \tilde{R}^{-1} \tilde{S})^{-1} A^T\}^{-1} A(\hat{h} - h) \overset{L}{\rightarrow} \sigma^2 h_k^2, \text{ as } j \to \infty.
\]

Applying Lemma 4, we deduce that \( (\hat{h} - h)^T A^T \{A(\tilde{S}^T \tilde{R}^{-1} \tilde{S})^{-1} A^T\}^{-1} A(\hat{h} - h) \overset{L}{\rightarrow} \sigma^2 h_k^2, \text{ as } n \to \infty. \) \[ \square \]
LEMMA 8. Assume Condition A. Then \( \hat{\mathbf{r}}^T \hat{R}^{-1} \hat{\mathbf{r}} / n \xrightarrow{P} \sigma^2 \).

PROOF. By the definition of \( \hat{\mathbf{r}} \),
\[
(A.19) \quad \hat{\mathbf{r}} = \hat{\mathbf{y}} - \hat{\mathbf{S}} \hat{\mathbf{h}} = (\mathbf{I} - \mathbf{S}_d)(\mathbf{y} - \mathbf{S} \hat{\mathbf{h}}) = \hat{\mathbf{S}}(\mathbf{h} - \hat{\mathbf{h}}) + \hat{\mathbf{d}} + \hat{\epsilon}.
\]
Notice that \( n^{-1} \hat{\mathbf{r}}^T \hat{R}^{-1} \hat{\mathbf{r}} = n^{-1} \hat{\mathbf{r}}^T \hat{R}^{-1} \hat{\mathbf{r}} + n^{-1} \hat{\mathbf{r}}^T (\hat{R}^{-1} - R^{-1}) \hat{\mathbf{r}} \), in which
\[
\begin{align*}
n^{-1} \hat{\mathbf{r}}^T \hat{R}^{-1} \hat{\mathbf{r}} &= n^{-1} \| R^{-1/2} (\hat{\mathbf{S}}(\mathbf{h} - \hat{\mathbf{h}}) + \hat{\mathbf{d}}) \| ^2 \\
&= n^{-1} \| R^{-1/2} (\mathbf{S}(\mathbf{h} - \mathbf{h}) + \mathbf{d} - \mathbf{S}_d \epsilon) + R^{-1/2} \epsilon \| ^2 \\
&\equiv I_1 + 2I_2 + I_3,
\end{align*}
\]
where \( I_1 = n^{-1} \| R^{-1/2} (\hat{\mathbf{S}}(\mathbf{h} - \mathbf{h}) + \hat{\mathbf{d}} - \mathbf{S}_d \epsilon) \| ^2 \), \( I_2 = n^{-1} \| (\mathbf{S}(\mathbf{h} - \mathbf{h}) + \mathbf{d} - \mathbf{S}_d \epsilon) \| ^2 \) and \( I_3 = n^{-1} \| R^{-1/2} \epsilon \| ^2 \). The proof will complete by showing \( I_1 = o_P(1), I_2 = o_P(1), I_3 = \sigma^2 + o_P(1) \) and \( n^{-1} \hat{\mathbf{r}}^T (\hat{R}^{-1} - R^{-1}) \hat{\mathbf{r}} = o_P(1) \).

First, consider \( I_1 \). Note that
\[
\begin{align*}
I_1 &= n^{-1} \| R^{-1/2} (\hat{\mathbf{S}}(\mathbf{h} - \mathbf{h}) + \hat{\mathbf{d}} - \mathbf{S}_d \epsilon) \| ^2 \\
&\leq 3n^{-1} \| R^{-1/2} \hat{\mathbf{S}}(\mathbf{h} - \mathbf{h}) \| ^2 + 3n^{-1} \| R^{-1/2} \hat{\mathbf{d}} \| ^2 + 3n^{-1} \| R^{-1/2} \mathbf{S}_d \epsilon \| ^2.
\end{align*}
\]
The first term is \( o_P(1) \) by Lemma 6 and Corollary 2, the second and third terms are both \( o_P(1) \) by (A.8), (A.15) and similar derivations for (A.7). Thus \( I_1 = o_P(1) \).

Second, consider \( I_3 = n^{-1} \sum_{i=1}^{n} \sum_{j=1}^{n} \epsilon(t_i) \epsilon(t_j) V(i, j) \). Then
\[
\begin{align*}
E(I_3) &= n^{-2} E(\epsilon^T R^{-1} \epsilon) = n^{-2} \text{trace} \{ E(\epsilon \epsilon^T) R^{-1} \} = \sigma^2, \\
E(I_3^2) &= n^{-2} \sum_{k=1}^{n} \sum_{k=2}^{n} \sum_{k=3}^{n} \sum_{k=4}^{n} e(k_1, k_2, k_3, k_4) V(k_1, k_2) V(k_3, k_4),
\end{align*}
\]
where \( e(k_1, k_2, k_3, k_4) = E \{ \epsilon(t_{k_1}) \epsilon(t_{k_2}) \epsilon(t_{k_3}) \epsilon(t_{k_4}) \} \). There are only 4 possible cases in which \( e(k_1, k_2, k_3, k_4) \) is nonzero. In Case 1, for any \( i \in \{ k_1, k_2, k_3, k_4 \} \), there exists \( j \in \{ k_1, k_2, k_3, k_4 \}, i \neq j \), such that \( |i - j| \leq g \). Then
\[
\begin{align*}
&\left| n^{-2} \sum_{\text{Case 1}} \sum_{k=1}^{n} \sum_{k=2}^{n} \sum_{k=3}^{n} \sum_{k=4}^{n} e(k_1, k_2, k_3, k_4) V(k_1, k_2) V(k_3, k_4) \right| \\
&\leq n^{-2} \sum_{k_1=1}^{n} \sum_{k_2=1}^{n} \sum_{k_3=1}^{n} \sum_{k_4=1}^{n} E[ \{ \epsilon(t_1) \}^4 ] \times |V(k_1, k_2)| \cdot |V(k_3, k_4)| \\
&\leq n^{-2} C^2.
\end{align*}
\]
In Case 2, \(|k_1 - k_2| \leq g\) and \(|k_3 - k_4| \leq g\), but for any \(i \in \{k_1, k_2\}\) and \(j \in \{k_3, k_4\}\), \(|i - j| > g\). Then
\[
\begin{align*}
  & n^{-2}\sum_2 \sum_{k_1k_2} \sum_{k_3k_4} e(k_1k_2k_3k_4)V(k_1k_2)V(k_3k_4) \\
  = & n^{-2}\sigma^4\sum_{k_1k_2} \sum_{k_3k_4} \rho(|k_2 - k_1|)V(k_1k_2)\sum_{k_3k_4} \rho(|k_3 - k_4|)V(k_3k_4) \\
  & - n^{-2}\sigma^4\sum_1 \sum_{k_1k_2} \sum_{k_3k_4} \rho(|k_2 - k_1|)\rho(|k_3 - k_4|)V(k_1k_2)V(k_3k_4) \\
  = & \{E(I_3)\}^2 - O(n^{-2}).
\end{align*}
\]

In Case 3, \(|k_1 - k_3| \leq g\) and \(|k_2 - k_4| \leq g\), but for any \(i \in \{k_1, k_3\}\) and \(j \in \{k_2, k_4\}\), \(|i - j| > g\). Then
\[
\begin{align*}
  & n^{-2}\sum_3 \sum_{k_1k_2} \sum_{k_3k_4} e(k_1k_2k_3k_4)V(k_1k_2)V(k_3k_4) \\
  \leq & n^{-2}\sigma^4 \sum_{k_1=1}^{n} \sum_{k_3-k_1 \leq g}^{k_2=1} \sum_{k_4-k_2 \leq g}^{k_3=1} \rho(|k_3 - k_1|)\rho(|k_4 - k_2|) \\
  & \times |V(k_1k_2)| \cdot |V(k_3k_4)| \\
  \leq & n^{-1}C^2.
\end{align*}
\]

A similar result holds for Case 4, where \(|k_1 - k_4| \leq g\) and \(|k_2 - k_3| \leq g\), but for any \(i \in \{k_1, k_4\}\) and \(j \in \{k_2, k_3\}\), \(|i - j| > g\). Combining Cases 1–4, \(E(I_4^2) \rightarrow \{E(I_3)\}^2\), which leads to \(I_3 \overset{p}{\rightarrow} E(I_3) = \sigma^2\).

Third, consider \(I_2\). By Cauchy-Schwartz inequality, \(I_2 = o_P(1)\).

Fourth, to deduce \(n^{-1}\hat{\mathbf{r}}^T(\hat{R}^{-1} - R^{-1})\hat{\mathbf{r}} = o_P(1)\), it is sufficient to show that
\[
\begin{align*}
  (A.20) \quad & n^{-1}\{\tilde{\mathbf{S}}(\mathbf{h} - \hat{\mathbf{h}})^T(\hat{R}^{-1} - R^{-1})\tilde{\mathbf{S}}(\mathbf{h} - \hat{\mathbf{h}}) = o_P(1), \\
  (A.21) \quad & n^{-1}\tilde{\mathbf{d}}^T(\hat{R}^{-1} - R^{-1})\tilde{\mathbf{d}} = o_P(1), \\
  (A.22) \quad & n^{-1}\tilde{\mathbf{e}}^T(\hat{R}^{-1} - R^{-1})\tilde{\mathbf{e}} = o_P(1).
\end{align*}
\]

It is easy to see that (A.20) follows from (A.16) and Corollary 2, whereas (A.21)–(A.22) are obtained by similar arguments for (A.16)–(A.18). \(\square\)

**Lemma 9.** Assume Condition A. Then
1. \(n^{-1/2}\tilde{\mathbf{S}}^T\hat{R}^{-1}\tilde{\mathbf{d}} = o_P(1)\).
2. \((\hat{A}\hat{\mathbf{h}}_{bc} - \mathbf{A})^T\{A(\tilde{\mathbf{S}}^T\hat{R}^{-1}\tilde{\mathbf{S}})^{-1}A^T\}^{-1}(\hat{A}\hat{\mathbf{h}}_{bc} - \mathbf{A}) \overset{P}{\rightarrow} \sigma^2 \chi_k^2\).
3. \(n^{-1}\tilde{\mathbf{r}}_{bc}^T\hat{R}^{-1}\tilde{\mathbf{r}}_{bc} \overset{P}{\rightarrow} \sigma^2\).
\textbf{Proof.} To show the first part, note that \( \tilde{d} = (I - S_d)\hat{d} = (I - S_d)S_d(y - \hat{S}h) = S_d(I - S_d)(y - \hat{S}h) = S_d\tilde{r} \). Thus from (A.19),
\begin{equation}
(I - S_d)S_d\xi_{j,\ell} = (I - S_d)[\{\xi_{j,\ell} - E(\xi_{j,\ell})\} - \{p_j\mathbf{1} - E(\xi_{j,\ell})\}] \\
= (I - S_d)[S_d\{\xi_{j,\ell} - E(\xi_{j,\ell})\} - S_d(p_j\mathbf{1}_j - 1, 0_{n - j + 1})^T],
\end{equation}
where \( S_d = n^{-1/2}S^T R^{-1}\tilde{d} = I_1 + I_2 + I_3 \), where \( I_1 = n^{-1/2}S^T R^{-1}(I - S_d)S_d(h - \hat{h}) \), \( I_2 = n^{-1/2}S^T R^{-1}S_d\tilde{d} \) and \( I_3 = n^{-1/2}S^T R^{-1}(I - S_d)S_d\xi \). We now show that each term is \( o_P(1) \). For \( I_1 \), from Lemma 7, we have \( n^{1/2}(h - \hat{h}) = O_P(1) \), thus we only need to consider the matrix \( n^{-1/2}S^T R^{-1}(I - S_d)S_d\xi \).

For its block matrix \( n^{-1/2}S^T R^{-1}(I - S_d)S_dS_{j_2} \), the \((\ell_1, \ell_2)\)th entry satisfies
\[
|n^{-1}e^T_{\ell_1}S_{j_1}^TR^{-1}(I - S_d)S_dS_{j_2}e_{\ell_2}| \\
\leq n^{-1/2}\|R^{-1/2}S_{j_1}^TR^{-1}R^{-1/2}(I - S_d)S_d\xi_{j_2, \ell_2}\| \\
\leq I_1(I_1 + I_3),
\]
where \( I_{11} = n^{-1/2}\|R^{-1/2}S_{j_1}^T e_{\ell_1}\| \), \( I_{12} = n^{-1/2}\|R^{-1/2}(I - S_d)S_d\xi_{j_2, \ell_2} - E(\xi_{j_2, \ell_2})\| \) and \( I_{13} = n^{-1/2}\|R^{-1/2}(I - S_d)S_d(p_j\mathbf{1}_j^T - 1, 0_{n - j + 1})^T\| \). Then by Lemma 6, \( I_{11} = O_P(1) \). By (A.8), \( I_{12} = o(1) \), and similarly, \( I_{13} = o(1) \). Thus \( I_1 = o_P(1) \). For \( I_2 \), using the same procedures as in Lemma 7 for proving \( J_1 = o_P(1) \), we can show \( I_2 = o_P(1) \). For \( I_3 \), using the same procedures as in Lemma 7 for proving \( J_{22} = o_P(1) \), we obtain \( I_3 = o_P(1) \). Thus \( n^{-1/2}S^T R^{-1}\tilde{d} = o_P(1) \). It remains to show that \( n^{-1/2}S^T (\hat{R}^{-1} - R^{-1})\tilde{d} = o_P(1) \), whose proof is similar to that of (A.17).

To show the second part, recall that \( \hat{h}_{bc} = h - (n^{-1}S^T R^{-1}h)^{-1}(n^{-1}S^T R^{-1}\tilde{h}) \). Using the first part together with Lemma 6 and Corollary 2 leads to the second part.

To show the third part, note that \( \hat{r}_{bc} = \hat{r} - \tilde{d} \) and \( n^{-1}\hat{r}_{bc}^T R^{-1}\hat{r}_{bc} = n^{-1}\hat{r}_{bc}^T R^{-1}r_{bc} + n^{-1}\hat{r}_{bc}^T (\hat{R}^{-1} - R^{-1})r_{bc} \), in which
\[
n^{-1}\hat{r}_{bc}^T R^{-1}r_{bc} = n^{-1}\|R^{-1/2}(r - \tilde{d})\|^2 = J_1 - 2J_2 + J_3,
\]
where \( J_1 = n^{-1}\|R^{-1/2}r\|^2 \), \( J_2 = 2n^{-1}\hat{r}^T R^{-1}\tilde{d} \) and \( J_3 = n^{-1}\|R^{-1/2}\tilde{d}\|^2 \).
From Lemma 8, \( J_1 = \sigma^2 + o_P(1) \). From (A.23),
\[
J_3 = n^{-1} \| R^{-1/2}(I - S_d)S_d(h - \hat{h}) + S_d\hat{d} + (I - S_d)S_d\epsilon \|^2 \\
\leq 3n^{-1} \{ \| R^{-1/2}(I - S_d)S_d(h - \hat{h}) \|^2 + \| R^{-1/2}S_d\hat{d} \|^2 \\
+ \| R^{-1/2}(I - S_d)S_d\epsilon \|^2 \}.
\]

Using similar proofs for the numerator, we obtain \( J_3 = o_P(1) \). By using the Cauchy-Schwartz inequality, \( J_2 = o_P(1) \). Thus \( n^{-1}\hat{r}_T R^{-1}\hat{r}_bc \overset{P}{\to} \sigma^2 \). To show \( n^{-1}\hat{r}_T R^{-1}\hat{r}_bc = o_P(1) \), it is sufficient to show that
\[
(A.24) \quad n^{-1}\hat{r}_T(R^{-1} - R^{-1})\hat{r} = o_P(1),
\]
\[
(A.25) \quad n^{-1}\hat{d}_T(R^{-1} - R^{-1})\hat{d} = o_P(1),
\]

where (A.24) directly follows from the fourth step of the proof for Lemma 8 and (A.25) uses similar proofs for (A.21)–(A.22). \( \square \)

A.1. Proof of Theorem 4.1. From Corollary 2, under \( H_0 \) in (4.1), the numerator of \( \mathbb{K} \) converges in distribution to \( \sigma^2 \chi^2_1 \). This combined with Lemma 8 arrives at the desired result for \( \mathbb{K} \).

A.2. Proof of Theorem 4.2. Under \( H_0 \) in (4.1), the second and third parts of Lemma 9 complete the proof for \( \mathbb{K}_{bc} \).

A.3. Proof of Theorem 4.3. The numerator of \( \mathbb{K} \) can be decomposed into 3 additive terms,
\[
I_1 = (\hat{A}h - Ah)^T \{ A(\hat{S}^T R^{-1} S)^{-1} A^T \}^{-1} (\hat{A}h - Ah),
\]
\[
I_2 = 2n(Ah)^T \{ A(n^{-1}\hat{S}^T R^{-1} S)^{-1} A^T \}^{-1} (Ah - Ah),
\]
\[
I_3 = n(Ah)^T \{ A(n^{-1}\hat{S}^T R^{-1} S)^{-1} A^T \}^{-1} (Ah).
\]

Notice that \( I_1 \xrightarrow{\mathcal{L}} \sigma^2 \chi^2_1 \) following the second part of Corollary 2; \( I_3 = n(Ah)^T (AM^{-1}A^T)^{-1} Ah \{ 1 + o_P(1) \} \) by Lemma 7 and \( H_1 \) in (4.1); \( I_2 = O_P(\sqrt{n}) \) by Cauchy-Schwartz inequality. These along with Lemma 8 complete the proof for \( \mathbb{K} \). The proof for \( \mathbb{K}_{bc} \) is similar and omitted.

A.4. Proof of Theorem 4.4. Following Lemma 7, under \( H_{1n} \) in (4.2), \( n^{1/2}\hat{A}h \xrightarrow{\mathcal{L}} N(c, \sigma^2 AM^{-1}A^T) \). Thus
\[
\frac{\{ A(n^{-1}\hat{S}^T R^{-1} S)^{-1} A^T \}^{-1/2} n^{1/2} \hat{A}h}{(\hat{r}^T R^{-1}\hat{r})^{1/2}} \xrightarrow{\mathcal{L}} N((AM^{-1}A^T)^{-1/2} c/\sigma, I_k).
\]

This completes the proof for \( \mathbb{K} \). Similar proofs for \( \mathbb{K}_{bc} \) are omitted.
ACKNOWLEDGEMENTS

We thank Richard Davidson, Tom Johnstone, and Terry Oakes at the Waisman Laboratory for Brain Imaging and Behavior, University of Wisconsin-Madison, for helpful comments and providing the fMRI data set.

REFERENCES


DEPARTMENT OF STATISTICS
UNIVERSITY OF WISCONSIN
1300 UNIVERSITY AVENUE
MADISON, WI 53706
E-MAIL: cmzhang@stat.wisc.edu

DEPARTMENT OF STATISTICS
UNIVERSITY OF WISCONSIN
1300 UNIVERSITY AVENUE
MADISON, WI 53706
E-MAIL: yutao@stat.wisc.edu