

# Semi-parametric inference for large-scale data with temporally dependent noise

Chunming Zhang

*Department of Statistics, University of Wisconsin, Madison, WI 53706, USA*  
e-mail: [cmzhang@stat.wisc.edu](mailto:cmzhang@stat.wisc.edu)

Xiao Guo

*Department of Statistics and Finance, School of Management, University of Science and Technology of China, Hefei, Anhui 230026, China*  
e-mail: [xiaoguo@ustc.edu.cn](mailto:xiaoguo@ustc.edu.cn)

Min Chen

*School of Mathematical Sciences, Shanxi University, Taiyuan 030006, China; Academy of Mathematics and Systems Science, Chinese Academy of Sciences, Beijing 100190, China*  
e-mail: [mchen@amss.ac.cn](mailto:mchen@amss.ac.cn)

and

Xinze Du

*Department of Statistics and Finance, School of Management, University of Science and Technology of China, Hefei, Anhui 230026, China*  
e-mail: [duxinze@mail.ustc.edu.cn](mailto:duxinze@mail.ustc.edu.cn)

**Abstract:** Temporal dependence is frequently encountered in large-scale structured noisy data, arising from scientific studies in neuroscience and meteorology, among others. This challenging characteristic may not align with existing theoretical frameworks or data analysis tools. Motivated by multi-session fMRI time series data, this paper introduces a novel semi-parametric inference procedure suitable for a broad class of “non-stationary, non-Gaussian, temporally dependent” noise processes in time-course data. It develops a new test statistic based on a tapering-type estimator of the large-dimensional noise auto-covariance matrix and establishes its asymptotic chi-squared distribution. Our method not only relaxes the consistency requirement for the noise covariance matrix estimator but also avoids direct matrix inversion without sacrificing detection power. It adapts well to both stationary and a wider range of temporal noise processes, making it particularly effective for handling challenging scenarios involving very large scales of data and large dimensions of noise covariance matrices. We demonstrate the efficacy of the proposed procedure through simulation evaluations and real fMRI data analysis.

**MSC2020 subject classifications:** Primary 62F35; secondary 62F30, 62F12.

**Keywords and phrases:** Auto-covariance matrix, false discovery rate, fMRI, neuroimaging, partially-linear model.

Received August 2022.

## Contents

1	Introduction . . . . .	2963
1.1	Motivation of this work . . . . .	2964
1.2	Outline of this work . . . . .	2965
2	Semi-parametric model with temporally dependent noise . . . . .	2967
2.1	Semi-parametric model . . . . .	2967
2.2	Assumptions on the noise process . . . . .	2968
2.3	Implications of the noise assumptions . . . . .	2970
3	Semi-parametric inference with temporally dependent noise . . . . .	2970
4	Auto-covariance matrix estimation for temporally dependent noise . . . . .	2972
5	Theoretical and methodological issues . . . . .	2972
5.1	Asymptotic distributions of the proposed semi-parametric test statistics . . . . .	2973
5.2	Choice of the tapering matrix $\mathbf{T}$ in (4.1) . . . . .	2975
5.3	Selection of tuning parameters $b$ and $g_n$ . . . . .	2977
5.3.1	Data-driven choice of $g_n$ using a given $b$ . . . . .	2977
5.3.2	Data-driven choice of $b$ . . . . .	2978
6	Simulation study . . . . .	2979
6.1	Set-up . . . . .	2979
6.2	Inference for 1D temporally correlated data . . . . .	2980
6.3	Inference for 2D temporally and spatially correlated data . . . . .	2984
6.4	Inference for 3D temporally and spatially correlated data . . . . .	2986
7	Real brain fMRI analysis . . . . .	2986
8	Discussion . . . . .	2988
A	Proofs of main results . . . . .	2989
	Acknowledgments . . . . .	3004
	Funding . . . . .	3004
	References . . . . .	3004

## 1. Introduction

The advancement of high-throughput technology has transformed research in many fields. Large-scale structured data often exhibit complex forms of dependence, non-stationarity, and non-Gaussianity. Examples include datasets from functional magnetic resonance imaging (fMRI) in neuroinformatics, interferometric synthetic aperture radar images from Earth-orbiting satellites in geophysics [1, 8], and air quality measurements from thousands of weather stations in meteorology [25, 37, 11], among others. These challenging features may not align with existing theoretical frameworks or data analysis tools. Ignoring such dependencies and non-Gaussian non-stationarity can lead to inaccurate statistical estimations and inference procedures for large-scale imaging data. This paper aims to develop some useful statistical methodologies and inference tools to effectively address temporal dependence.

### 1.1. Motivation of this work

In recent years, the study of human cognition has greatly benefited from innovations in fMRI, an exciting technique aimed at delineating task-activated regions in the human brains. A typical fMRI dataset for a single subject and session can be viewed as a huge collection of cuboid elements, typically ranging from  $10^4$  to  $10^5$  voxels  $v$ , each associated with a time series of responses denoted as  $\{Y^{(v)}(t_i) : i = 1, \dots, n; v \in \mathcal{V}\}$ , where the series length,  $n$ , can extend to the order of  $10^3$ . Analyzing fMRI data commonly involves two stages:

- |                                   |   |
|-----------------------------------|---|
| Stage-I (voxelwise analysis):     | modeling and making inferences<br>for fMRI signals at each voxel. |
| Stage-II (whole-brain inference): | detecting regions of brain activity<br>across the entire brain.   |

A significant challenge in voxelwise inference lies in the fact that the observed signals are contaminated with highly **correlated** noise in the time domain. Moreover, several factors may induce non-Gaussian **non-stationarity** in the noise process. For instance, background memory processes or extraneous auditory and visual stimuli can lead to changes in the noise's variance [17, 26]. Additionally, abrupt movements by the subject represent another source of non-stationarity [9]. Thus, it is essential to adequately account for non-stationary, non-Gaussian temporal dependence to accurately detect activation in fMRI.

In the literature, several statistical modeling and analysis methods have been developed for single-voxel fMRI, including the general linear model (GLM) [19, 20, 39] and the semi-parametric model approach [46], each with its own constraints.

(i) GLM approaches usually assume that the noise process is a stationary time series, such as AR(1) [4], AR( $m$ ) [39], and ARMA(1, 1) [31], among others. Some variations of non-stationarity, such as independent heterogeneous processes [26, 29], AR combined with independent heterogeneous noise [9], and long-range dependent  $1/f$ -like processes [26], have been considered within the GLM framework. However, GLM approaches often require restrictive assumptions on model parameters, such as specific parametric forms for the hemodynamic response function (HRF) and the temporal drift function [39]. These assumptions can lead to biased estimators of true responses and test statistics with limited power for detection.

(ii) Semi-parametric models of the form:

$$\mathbf{Y}^{(v)} = \mathbf{S} \mathbf{h}^{(v)} + \mathbf{d}^{(v)} + \boldsymbol{\varepsilon}^{(v)}, \quad v \in \mathcal{V} \quad (1.1)$$

were developed in [46] for time-course data  $\mathbf{Y}^{(v)} = (Y^{(v)}(t_1), \dots, Y^{(v)}(t_n))^T$  at voxel  $v$ , with external stimuli in the matrix  $\mathbf{S}$  common to all voxels. These models provide more flexibility in modeling assumptions compared to the GLM approach but are restricted to a stationary  $g$ -dependent noise process  $\boldsymbol{\varepsilon}^{(v)} = (\varepsilon^{(v)}(t_1), \dots, \varepsilon^{(v)}(t_n))^T$ . Detailed descriptions can be found in Section 2.1. [15] considered two types of stationary noise processes: a  $g_n$ -dependent sequence and a linear process with a dependence structure related to the number  $n$  of time

points. [27] modeled the noise process as a fractional Brownian motion process. These approaches require the noise covariance matrix estimator  $\widehat{\Sigma} \in \mathbb{R}^{n \times n}$  to be consistent to the true  $\Sigma_n$  and involve taking the inverse of  $\widehat{\Sigma}$ , posing significant challenges for statistical inference with large-scale time-course data.

In the literature, semi-parametric inference for stationary time series includes [10], [16], [22], and [23], among others. Specifically, [10] studied the partially linear single-index model; [16] proposed the partial-linear single-index model with ARMA noise; and [22] examined the partially time-varying coefficient model. However, these mentioned semi-parametric models differ from the partially linear model presented in our paper. [23] considered the partially linear model and investigated the simultaneous inference of the nonparametric part with stationary noise, but their results do not apply to non-stationary noise. Semi-parametric models for non-stationary time series have also been explored in the literature; see, e.g., [35], [38], and [13]. [35] introduced the semi-parametric factor model, which is distinct from the partially linear model in our work. [38] conducted semi-parametric inference for panel data with non-stationary noise. However, they assumed that the noise terms are independent, so their results are not applicable to non-stationary temporally dependent noise. [13] examined the semi-parametric model, which encompasses the partially linear model. Although the response in their model is a non-stationary time series, they assumed that the noise term is stationary. Consequently, their results cannot be used to analyze our model with non-stationary noise. Moreover, if the model in [13] simplifies to the partially linear model, as it does in our paper, their estimation method corresponds to ours, except that we employ the local-linear method while they opt for the local-constant method. Consequently, there is no need to compare our method with the approaches mentioned above.

In stark contrast, very few works have explored the semi-parametric model for fMRI data in the context of non-stationary, non-Gaussian, temporally dependent noise. Given the inherently time-varying nature of fMRI noises, the assumption of stationarity may not be realistically appropriate. Therefore, it is desirable to investigate how the identification of activated brain voxels can benefit from semi-parametric modeling, including parameter learning and inference under a broader range of temporally dependent noise processes.

## 1.2. Outline of this work

To detect activated brain regions, we first conduct voxelwise inference in Stage-I using the data  $\{\mathbf{Y}^{(v)} : v \in \mathcal{V}\}$ , and subsequently, in Stage-II, we conduct multiple testing using the sequence of voxelwise  $p$ -values  $\{p^{(v)} : v \in \mathcal{V}\}$ . This paper mainly focuses on Stage-I for theoretical analysis, while both stages are involved in the numerical studies. Regarding Stage-II, exploring alternative approaches, such as the recent multivariate tensor method, could be an interesting topic for future research.

In this paper, we will conduct a novel semi-parametric inference procedure that accommodates serially correlated noise, which can be non-stationary and

non-Gaussian. This has applications in fMRI and other large-scale imaging data with time series noise. We will investigate three key issues.

Concerning the structure of temporally dependent noise, we will initially consider short-range dependence, building on the framework presented in [47] and [45]. This framework encompasses various time series models, including stationary models (e.g., ARMA model [4, 39, 31], stationary  $g$ -dependent process [46]) and non-stationary models (e.g., independent heterogeneous process [26, 29] and AR+independent heterogeneous noise model [9]) as special cases. See Section 2 for details. Extensions to other types of dependence will be explored in Section 8.

Inspired by the detection of brain activity within a voxel, we will explore the significance test of the parameter vector in the semi-parametric model (2.1) with temporally dependent noise. Due to non-stationarity and the unique structure of the Toeplitz design matrix, some existing forms of semi-parametric test statistics, particularly well-suited for stationary noise, are no longer applicable in non-stationary scenarios. To accommodate non-stationarity, we will develop a new semi-parametric test statistic by utilizing an appropriate estimator  $\widehat{\Sigma}$  of  $\Sigma_n$ . Unlike existing results, our method circumvents the consistency requirement and avoids direct inversion of  $\widehat{\Sigma}$  in high dimensions while maintaining the size without compromising detection power. This approach strikes a balance between statistical guarantees and computational efficiency, especially in practical large-scale imaging data scenarios. See Section 3 for details.

Regarding the estimation of  $\Sigma_n$  for temporally dependent noise, several regularization methods have been studied in the context of estimating large covariance matrices. These methods have been explored in areas of high-dimensional inference [3, 5, 7, 18, 44] and stationary time series analysis [28, 40, 41, 43]. They include techniques like banding, tapering, and thresholding. We will adopt the tapering-based method [3, 43] for two reasons: **(1)** Most penalization methods require a diverging (approaching infinity) number of i.i.d. replicates, whereas a typical fMRI experiment involves only a finite number of replicates (e.g., sessions, scans, or runs), making penalization methods unsuitable. **(2)** The entries of  $\Sigma_n$  decay as they move away from the diagonal, and the tapering-based estimator is positive semidefinite, further ensuring positive-definiteness by adding  $\kappa_n \mathbf{I}_n$  with a shrinking  $\kappa_n > 0$ . The tapering method is more suitable than the thresholding method due to the “tapering” structure of  $\Sigma_n$ , where elements far from the diagonal exhibit decay. In contrast, the banding method does not guarantee positive-semidefiniteness. Therefore, the tapering method is the preferred choice for estimating  $\Sigma_n$ . See Section 4 for further details.

Section 5 explores the asymptotic properties of the new inference procedure and develops practical computational strategies. Simulation studies in Section 6, which includes 1D, 2D, and 3D datasets, support these theoretical properties and demonstrate the better performance of the new procedure compared to existing ones. Section 7 analyzes a real fMRI dataset, and Section 8 provides a brief conclusion.

Details of technical derivations and proofs are relegated to the Appendix. We introduce some notations below. Denote  $\mathbf{0}_q = (0, \dots, 0)^T \in \mathbb{R}^q$ , and  $\mathbf{1}_q =$

$(1, \dots, 1)^T \in \mathbb{R}^q$  for  $q \in \mathbb{Z}^+$ . Let  $\mathbf{I}_q \in \mathbb{R}^{q \times q}$  be the identity matrix, and  $\mathbf{e}_{i;q}$  be the  $i$ th column of  $\mathbf{I}_q$  for  $i = 1, \dots, q$ . For a square matrix  $A \in \mathbb{R}^{q \times q}$ ,  $A \succ 0$  and  $A \succeq 0$  mean that  $A$  is positive definite and positive semidefinite, respectively. Denote by  $\lambda_{\max}(A)$  the maximum eigenvalue of  $A$ . For a generic matrix  $B \in \mathbb{R}^{q \times q'}$ , let  $B(i, j)$  be the element of  $B$  in the  $i$ th row and  $j$ th column for  $i = 1, \dots, q$  and  $j = 1, \dots, q'$ . The  $L_1$ ,  $L_2$ , and  $L_\infty$  norms of  $B$  are denoted as  $\|B\|_1 = \max_{1 \leq j \leq q'} \sum_{i=1}^q |B(i, j)|$ ,  $\|B\|_2 = \{\lambda_{\max}(B^T B)\}^{1/2}$ , and  $\|B\|_\infty = \max_{1 \leq i \leq q} \sum_{j=1}^{q'} |B(i, j)|$ , respectively. For a vector  $\mathbf{v} = (v_1, \dots, v_q)^T \in \mathbb{R}^q$ , the  $L_1$  and  $L_2$  norms are represented as  $\|\mathbf{v}\|_1 = \sum_{i=1}^q |v_i|$  and  $\|\mathbf{v}\|_2 = (\sum_{i=1}^q v_i^2)^{1/2}$ , respectively. For any scalar random variable  $X$  and constant  $a \in (0, \infty)$ , define  $\|X\|_a = \{E(|X|^a)\}^{1/a}$ . In the following,  $C$  and  $c$  represent generic finite constants that are independent of the sample size, denoted as  $n$ .

## 2. Semi-parametric model with temporally dependent noise

### 2.1. Semi-parametric model

We begin with a brief introduction to the semi-parametric partially-linear model (PLM):

$$Y_i = \mathbf{S}_i^T \mathbf{h} + d(t_i) + \varepsilon(t_i), \quad i = 1, \dots, n. \quad (2.1)$$

Here, the response is denoted as  $Y_i$  and the covariates are represented as  $(t_i, \mathbf{S}_i)$ , where  $t_i \in [0, 1]$ . In this model, we aim to estimate the unknown parameter vector  $\mathbf{h}$  and the non-parametric function  $d(\cdot)$ , while  $\varepsilon(t_i)$  represents the noise terms with  $E\{\varepsilon(t_i)\} = 0$ . We impose the assumption  $t_i \in [0, 1]$  to ensure that  $d(\cdot)$  is compactly supported, which is necessary for consistent non-parametric estimation of  $d(t)$ . It's worth noting that, in the case of real fMRI data, the duration between consecutive observations remains unchanged with  $n$ .

For temporal data measured at time points  $\{t_i\}_{i=1}^n$  (e.g., equally spaced  $t_i = i/n$  relevant to fMRI and [8]), denoting  $\mathbf{S} = (\mathbf{S}_1, \dots, \mathbf{S}_n)^T$ , the PLM (2.1) can be re-expressed as:

$$\mathbf{Y} = \mathbf{S} \mathbf{h} + \mathbf{d} + \boldsymbol{\varepsilon} = \sum_{j=1}^{\ell} \mathbf{S}_j \mathbf{h}_j + \mathbf{d} + \boldsymbol{\varepsilon}, \quad (2.2)$$

with

- $\mathbf{Y} = (Y(t_1), \dots, Y(t_n))^T \in \mathbb{R}^{n \times 1}$  for the observed responses.
- $\mathbf{S} = (\mathbf{S}_1, \dots, \mathbf{S}_\ell) \in \mathbb{R}^{n \times (\ell m)}$  for the design matrix. E.g.,

$$\mathbf{S}_j = \begin{pmatrix} s_j(0) & 0 & \cdots & 0 \\ s_j(t_2 - t_1) & s_j(0) & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ s_j(t_m - t_1) & s_j(t_m - t_2) & \cdots & s_j(0) \\ \vdots & \vdots & \ddots & \vdots \\ s_j(t_n - t_1) & s_j(t_n - t_2) & \cdots & s_j(t_n - t_m) \end{pmatrix} \in \mathbb{R}^{n \times m}, \quad j = 1, \dots, \ell,$$

is a Toeplitz design matrix associated with  $s_j(t_i)$ , representing the input of the  $j$ th type of observed external stimulus at time  $t_i$ . Please refer to [46] for the derivation of  $\mathbf{S}$  from the convolution model.

- $\mathbf{h} = (\mathbf{h}_1^T, \dots, \mathbf{h}_\ell^T)^T \in \mathbb{R}^{(\ell m) \times 1}$  for the vector of unknown regression parameters (relevant to brain HRF) of primary research interest. For instance,  $\mathbf{h}_j = (h_j(t_1), \dots, h_j(t_m))^T \in \mathbb{R}^{m \times 1}$  is the parameter vector associated with the  $j$ th type of observed external stimulus.
- $\mathbf{d} = (d(t_1), \dots, d(t_n))^T \in \mathbb{R}^{n \times 1}$  for the unknown smooth functions  $d(\cdot)$  (related to temporal drift) at time points  $\{t_i\}_{i=1}^n$ .
- $\boldsymbol{\varepsilon} = (\varepsilon(t_1), \dots, \varepsilon(t_n))^T \in \mathbb{R}^{n \times 1}$  for the zero-mean noise process, independent of  $\mathbf{S}$ .

Model (2.2) corresponds to model (1.1) with the superscript  $(v)$  omitted for notational brevity.

For fMRI data,  $\ell$  is the number of external stimuli, and  $m$  indicates the length of each HRF parameter vector  $\mathbf{h}_j$ ,  $j = 1, \dots, \ell$ . In real fMRI data, the HRF is typically sparse, with  $m$  usually specified as a small integer much less than  $n$ , for instance, a common choice is 18 as a rule of thumb. The design matrix  $\mathbf{S}$  in fMRI data depends on how external stimuli are presented to the subjects [39]. For the sake of presentation, we assume that  $s_j(\cdot)$  starts at the same time for all  $j = 1, \dots, \ell$ . However, our method can also accommodate design matrices that allow different starting times for different types of stimuli, e.g.,  $s_j(t_i) = 0$  for some  $j = 1, \dots, \ell$  and all  $i = 1, \dots, n_0$ , with a fixed integer  $n_0 > 1$ . While our method is developed for fMRI data with the Toeplitz design matrix  $\mathbf{S}_j$ , our theoretical arguments do not rely on the Toeplitz structure and hold for general design matrices that satisfy Conditions A4 and A10. It's worth noting that model (2.2) also has the flexibility to accommodate multiple sessions (i.e., runs or segments). However, for brevity, the discussions in Sections 2–5 will concentrate on a single run.

The smooth function  $d(t)$  describes the slowly drifting baseline in the fMRI signal, which can vary between different voxels. The HRF characterizes the stimulus-related variation in the signal for a specific voxel. For each voxel, the HRF is commonly assumed to be identical when the same type of stimulus is presented at different times, as discussed in [39].

This work aims to detect the brain activities by testing the significance of the HRF parameters. While inference for the nonparametric part  $d(\cdot)$  is feasible (using methods such as [42]), it falls outside the scope of this paper because  $d(\cdot)$  is not directly relevant to event-related brain activities.

## 2.2. Assumptions on the noise process

The noise terms  $\{\varepsilon(t_i)\}_{i=1}^n$  in either (2.1) or (2.2) exhibit serial correlation, with the covariance structure characterized by a positive-definite auto-covariance matrix,

$$\Sigma_n = \text{cov}(\boldsymbol{\varepsilon}, \boldsymbol{\varepsilon}) \in \mathbb{R}^{n \times n}.$$

Most of the current literature relies on the assumption of stationarity or Gaussianity. Motivated from the framework in [45], we incorporate non-Gaussian nonstationarity into the sequence of noise terms  $\{\varepsilon(t_i)\}$ , as follows:

$$\varepsilon(t_i) = G(t_i; \mathcal{F}_i), i = 1, \dots, n, \quad (2.3)$$

where  $G : [0, 1] \times \mathbb{R}^\infty \rightarrow \mathbb{R}$  is a measurable function, and  $\mathcal{F}_i = (\dots, e_{i-1}, e_i)$  represents a shift process associated with a sequence of i.i.d. random variables  $\{e_k\}_{k=-\infty}^\infty$ . This formulation ensures that  $\varepsilon(t_i)$  is well-defined and maintains  $E\{\varepsilon(t_i)\} = 0$ . In (2.3),  $G(t; \cdot)$  describes the noise-generating mechanism, which can be time-varying with respect to  $t$ . Conversely, if  $G(t; \cdot)$  does not depend on  $t$ , then  $\{\varepsilon(t_i)\}$  reduces to a stationary process. It's important to note that the noise mechanism in (2.3) is applicable to various types of time points  $t_i \in [0, 1]$ , including  $t_i = i/n$ . Model (2.3) encompasses a wide range of non-stationary processes and naturally extends many existing stationary time series models to their non-stationary counterparts. For instance, it includes non-stationary linear processes  $G(t; \mathcal{F}_i) = \sum_{j=0}^\infty a_j(t)\epsilon_{i-j}$ , where  $\epsilon_i$  represents i.i.d. random variables, and  $a_j(\cdot)$ ,  $j = 1, 2, \dots$ , are functions that ensure  $G(t; \mathcal{F}_i)$  is well defined for all  $t \in [0, 1]$ . Furthermore, model (2.3) also includes non-stationary nonlinear time series:  $G(t; \mathcal{F}_i) = R(t, G(t; \mathcal{F}_{i-1}), \epsilon_i)$ , where  $R$  is a measurable function. As discussed in the fMRI literature, the noise is time-varying non-stationary; see, e.g., [24] and [26], while our assumption includes common non-stationary linear processes, e.g., time-varying AR and time-varying MA, as special cases. For detailed descriptions and examples of the noise mechanism, please refer to Section 2 of [45] and Section 4 of [47].

To further accommodate non-stationary short-range dependence, we introduce  $e'_i$  as an i.i.d. replicate of  $e_i$ , and define

$$\mathcal{F}'_{i;k} = \begin{cases} (\dots, e_{k-1}, e'_k, e_{k+1}, \dots, e_i), & \text{if } k \leq i, \\ \mathcal{F}_i, & \text{if } k > i. \end{cases}$$

The dependence of  $G(t; \mathcal{F}_i)$  on  $e_0$  across  $t \in [0, 1]$  can be quantified by

$$\Delta_{i,\omega} = \sup_{t \in [0,1]} \|G(t; \mathcal{F}_i) - G(t; \mathcal{F}'_{i;0})\|_\omega,$$

where  $\omega > 0$  is a constant. The cumulative dependence of  $\{G(t; \mathcal{F}_i)\}_{i=q}^\infty$  on  $e_0$  is represented by

$$\Theta_{q,\omega} = \sum_{i=q}^\infty \Delta_{i,\omega} \quad (2.4)$$

with a non-negative integer  $q$ . As in Condition A3 in Section 5.1, we consider the short-range dependence condition, as expressed by the quantity  $\Theta_{q,\omega}$  in (2.4), i.e.,

$$\Theta_{0,\omega} < \infty \quad \text{for a constant } \omega \in [4, \infty). \quad (2.5)$$

As a remark, the assumption  $\omega \geq 4$  is necessary to derive the error bound for the quadratic form of the noise in Lemma 4. This same condition was considered for both stationary and non-stationary time series analysis, as seen in, e.g., [41] and [45], respectively.

### 2.3. Implications of the noise assumptions

Utilizing assumption (2.5), we demonstrate in Lemma 1 some properties of the noise process  $\{\varepsilon(t_i)\}$ . One noteworthy implication is that the absolute auto-covariance of the noise diminishes as the time-lag  $q$  increases:

$$|\Sigma_n(i+q, i)| = |\text{cov}\{\varepsilon(t_{i+q}), \varepsilon(t_i)\}| \leq C \Theta_{q,\omega}.$$

These properties further motivate us to develop a tapering-based estimator for the noise auto-covariance matrix  $\Sigma_n$  in Section 4.

**Lemma 1.** *Assume that  $\{\varepsilon(t_i)\}$  follows a “non-stationary short-range dependence” as described in (2.3) and (2.5). Then,  $\sup_{i=1,\dots,n} \|\varepsilon(t_i)\|_\omega \leq \Theta_{0,\omega}$  and  $\sup_{i=1,\dots,n-q} |\text{E}\{\varepsilon(t_{i+q})\varepsilon(t_i)\}| \leq C \Theta_{q,\omega}$  for any  $q \in \{1, \dots, n-1\}$ , where  $C \in (0, \infty)$  is a constant.*

Under the short-range dependence condition (2.5), the noise mechanism (2.3) encompasses a wide range of stationary and non-stationary time series noise models as special cases. These include stationary ARMA models [4, 39, 31] and  $g$ -dependent processes [46], as well as non-stationary i.n.i.d. processes [26, 29] and AR + i.n.i.d. processes [9].

Lemma 1 only necessitates moment conditions (i.e., Condition A3) for the noise terms  $\{\varepsilon(t_i)\}$ , without imposing specific distributional assumptions. Therefore, our method is generally applicable to non-Gaussian scenarios. Additionally, the numerical results in Section 6 demonstrate the robust performance of our method under non-Gaussian conditions.

### 3. Semi-parametric inference with temporally dependent noise

The detection of brain activity can be achieved by testing the significance of the parametric part  $\mathbf{h}$  (associated with HRF) in model (2.2). If a specific voxel is activated by a certain type of stimulus, then the corresponding HRF is non-identically zero; otherwise, the HRF is zero. Moreover, checking the equivalence between HRFs activated by different types of stimuli is also of interest to neuroscientists. This inspires us to consider the linear form of hypotheses on  $\mathbf{h}$ :

$$H_0 : \mathbf{A}\mathbf{h} = \mathbf{0}_k \quad \text{versus} \quad H_1 : \mathbf{A}\mathbf{h} \neq \mathbf{0}_k. \quad (3.1)$$

Here,  $\mathbf{A} \in \mathbb{R}^{k \times (\ell m)}$  is a full row rank matrix of coefficients, and  $k$  is a fixed positive integer. In particular, if  $\mathbf{A} = \mathbf{I}_{\ell m}$ , then (3.1) can be employed to detect the activity at a single voxel. If  $\mathbf{A} = (\mathbf{I}_m, -\mathbf{I}_m, \mathbf{0}_{m \times m}, \dots, \mathbf{0}_{m \times m})$ , we can test the equivalence of two HRFs,  $\mathbf{h}_1$  and  $\mathbf{h}_2$ . Other choices of  $\mathbf{A}$  are also applicable, depending on the specific structures of the HRFs being tested.

To fit the semi-parametric model (2.2) and carry out the semi-parametric inference (3.1) for  $\mathbf{h}$ , we first develop a semi-parametric estimation for the non-parametric part  $\mathbf{d}$  in (2.2) using the local-linear estimation method [12]. Denote

by  $S_b \in \mathbb{R}^{n \times n}$  the local-linear smoothing matrix associated with the design points  $\{t_1, \dots, t_n\}$ , with entries given by:

$$S_b(i, j) = (1, 0) \{ \mathbf{X}(t_i)^T \mathbf{W}(t_i) \mathbf{X}(t_i) \}^{-1} (1, t_j - t_i)^T \frac{1}{b} K\left(\frac{t_j - t_i}{b}\right), \quad (3.2)$$

for  $i, j = 1, \dots, n$ , where  $K(\cdot)$  is a kernel function,  $b > 0$  is a bandwidth parameter, and

$$\mathbf{X}(t) = \begin{pmatrix} 1 & \cdots & 1 \\ t_1 - t & \cdots & t_n - t \end{pmatrix}^T, \quad \mathbf{W}(t) = \text{diag} \left\{ \frac{1}{b} K\left(\frac{t_1 - t}{b}\right), \dots, \frac{1}{b} K\left(\frac{t_n - t}{b}\right) \right\}.$$

Applying the local-linear estimation method to model (2.2), we obtain  $S_b \mathbf{Y} = S_b(\mathbf{S} \mathbf{h} + \mathbf{d} + \boldsymbol{\varepsilon})$ , which leads to:

$$\tilde{\mathbf{Y}} = \tilde{\mathbf{S}} \mathbf{h} + \tilde{\mathbf{d}} + \tilde{\boldsymbol{\varepsilon}}, \quad (3.3)$$

where  $\tilde{\mathbf{Y}} = (\mathbf{I}_n - S_b) \mathbf{Y}$ ,  $\tilde{\mathbf{S}} = (\mathbf{I}_n - S_b) \mathbf{S}$ ,  $\tilde{\mathbf{d}} = (\mathbf{I}_n - S_b) \mathbf{d}$  and  $\tilde{\boldsymbol{\varepsilon}} = (\mathbf{I}_n - S_b) \boldsymbol{\varepsilon}$ , with  $\mathbf{I}_n$  representing an identity matrix.

Due to the smoothness of the function  $d(\cdot)$  (see Condition A1 in Section 5.1), the part  $\tilde{\mathbf{d}}$  in (3.3) can be negligible. Consequently, model (3.3) takes the approximate form,

$$\tilde{\mathbf{Y}} \approx \tilde{\mathbf{S}} \mathbf{h} + \tilde{\boldsymbol{\varepsilon}}. \quad (3.4)$$

The ordinary least squares method provides the estimator for the parametric part  $\mathbf{h}$  as

$$\hat{\mathbf{h}} = (\tilde{\mathbf{S}}^T \tilde{\mathbf{S}})^{-1} \tilde{\mathbf{S}}^T \tilde{\mathbf{Y}}. \quad (3.5)$$

Similarly, the non-parametric part  $\mathbf{d}$  is estimated as  $\hat{\mathbf{d}} = S_b(\mathbf{Y} - \mathbf{S} \hat{\mathbf{h}})$ .

The asymptotic normality of  $\hat{\mathbf{h}}$  is demonstrated in Lemma 5 in the Appendix:  $n^{1/2}(\hat{\mathbf{h}} - \mathbf{h}) \xrightarrow{D} N(\mathbf{0}, M_1^{-1} M_2 M_1^{-1})$  if  $n^{-1} \tilde{\mathbf{S}}^T \tilde{\mathbf{S}} \xrightarrow{P} M_1 \succ 0$  and  $n^{-1} \tilde{\mathbf{S}}^T \Sigma_n \tilde{\mathbf{S}} \xrightarrow{P} M_2 \succ 0$  under certain regularity conditions. This result allows us to propose a new Wald-type test statistic for conducting the significance test of (3.1):

$$\mathbb{W}(\Sigma_n; b) = (\mathbf{A} \hat{\mathbf{h}})^T \{ \mathbf{A} (\tilde{\mathbf{S}}^T \tilde{\mathbf{S}})^{-1} \tilde{\mathbf{S}}^T \Sigma_n \tilde{\mathbf{S}} (\tilde{\mathbf{S}}^T \tilde{\mathbf{S}})^{-1} \mathbf{A}^T \}^{-1} (\mathbf{A} \hat{\mathbf{h}}), \quad (3.6)$$

which incorporates the noise auto-covariance matrix  $\Sigma_n = \text{cov}(\boldsymbol{\varepsilon}, \boldsymbol{\varepsilon}) \in \mathbb{R}^{n \times n}$ , rather than  $\Sigma_n^{-1}$ , due to the presence of  $\Sigma_n$  in  $n^{-1} \tilde{\mathbf{S}}^T \Sigma_n \tilde{\mathbf{S}}$  in the asymptotic covariance of  $n^{1/2} \hat{\mathbf{h}}$ . In [46],  $R_n^{-1}$  (the inverse of the noise auto-correlation matrix) is used in their test statistics because they estimate  $\mathbf{h}$  using the weighted least-squares estimator, whose asymptotic covariance matrix includes  $R_n^{-1}$ . Moreover, we could directly use  $\Sigma_n$  in (3.6) instead of  $\text{cov}(\tilde{\boldsymbol{\varepsilon}}, \tilde{\boldsymbol{\varepsilon}})$  because the auto-correlation among  $S_b \boldsymbol{\varepsilon}$  is negligible, as shown in Lemma 2.

We wish to emphasize that, in comparison to certain existing methods like (5.1) (combined with the estimator  $\hat{\mathbf{h}}$  in (5.2)), which assume stationarity and necessitate the computation of the inverse covariance matrix  $\Sigma_n^{-1}$  for parameter estimation or significance testing, the new test statistic (3.6) (associated

with the estimator  $\widehat{\mathbf{h}}$  in (3.5) leverages  $\Sigma_n$  directly, without requiring the computation of its inverse  $\Sigma_n^{-1}$ . In Section 5, we provide statistical guarantees for the broad applicability of  $\mathbb{W}(\Sigma_n; b)$ . Moreover, the advantages in numerical efficiency become especially evident when analyzing large-scale imaging data. In cases with very large  $n$ , such as  $n = 2800$  in the real fMRI data discussed in Section 7, developing a valid empirical estimator for  $\Sigma_n^{-1}$  presents more challenges than estimating  $\Sigma_n$  itself. However, our proposed test statistic (4.2) only necessitates the computation of the inverse of fixed-dimensional matrices, making it computationally more efficient.

#### 4. Auto-covariance matrix estimation for temporally dependent noise

In practical applications, the true noise auto-covariance matrix  $\Sigma_n$  is unknown, leading to the need for estimating  $\Sigma_n \in \mathbb{R}^{n \times n}$  before applying the Wald-type test statistic in (3.6). We propose estimating  $\Sigma_n$  using the tapering-based estimator:

$$\widehat{\Sigma} = \{(\widetilde{\mathbf{Y}} - \widetilde{\mathbf{S}}\widehat{\mathbf{h}})(\widetilde{\mathbf{Y}} - \widetilde{\mathbf{S}}\widehat{\mathbf{h}})^T\} \circ \mathbf{T} + \kappa_n \mathbf{I}_n. \quad (4.1)$$

Here, the tapering matrix  $\mathbf{T} \in \mathbb{R}^{n \times n}$  is positive-semidefinite and Toeplitz. The operator “ $\circ$ ” represents the Schur matrix product [33] (or entrywise product), and  $\kappa_n > 0$  is a shrinking parameter. The inclusion of  $\kappa_n \mathbf{I}_n$  in (4.1) ensures the positivity of  $\widehat{\Sigma}$ . Theoretically,  $\kappa_n$  approaches zero, and it’s worth noting that our theoretical results remain valid even if  $\kappa_n = 0$ . In practice, the choice of  $\kappa_n$  has a negligible impact on numerical results, provided  $\kappa_n > 0$  is sufficiently small. We can set  $\kappa_n$  as a user-defined small constant, e.g.,  $\kappa_n = 10^{-5}$ , or adopt a data-driven choice like  $\kappa_n = \|\widetilde{\mathbf{Y}} - \widetilde{\mathbf{S}}\widehat{\mathbf{h}}\|_2^2/n^\nu$  with a user-defined constant  $\nu \geq 2$ , as motivated by [28] (see Section 3 therein). By substituting the estimator  $\widehat{\Sigma}$  into (3.6), we obtain a practical form of the semi-parametric test statistic:

$$\mathbb{W}(\widehat{\Sigma}; b) = (\mathbf{A}\widehat{\mathbf{h}})^T \{ \mathbf{A}(\widetilde{\mathbf{S}}^T \widehat{\Sigma} \widetilde{\mathbf{S}})^{-1} \widetilde{\mathbf{S}}^T \widehat{\Sigma} \widetilde{\mathbf{S}} (\widetilde{\mathbf{S}}^T \widehat{\Sigma} \widetilde{\mathbf{S}})^{-1} \mathbf{A}^T \}^{-1} (\mathbf{A}\widehat{\mathbf{h}}). \quad (4.2)$$

#### 5. Theoretical and methodological issues

The tapering-based method is commonly employed for large covariance matrix estimation in high-dimensional inference [3, 7] and stationary time series analysis [28, 43]. This approach effectively ensures the positive-semidefiniteness of the resulting estimator. However, in the presence of non-stationary noise, directly verifying the consistency of the tapering-based estimator  $\widehat{\Sigma}$  in (4.1) appears challenging. Therefore, conducting the asymptotic analysis of the proposed semi-parametric test statistic  $\mathbb{W}(\widehat{\Sigma}; b)$  with non-stationary dependent noise is a nontrivial undertaking. Nonetheless, in Section 5.1, we demonstrate that the proposed  $\mathbb{W}(\widehat{\Sigma}; b)$  follows an asymptotic  $\chi_k^2$  distribution under the null hypothesis. Furthermore, we derive the asymptotic power functions of the proposed test statistic under both fixed and contiguous alternatives. Section 5.2 addresses the selection of the tapering matrix  $\mathbf{T}$  in the estimator  $\widehat{\Sigma}$ , and in Section 5.3, we develop data-driven methods for tuning parameter selection.

### 5.1. Asymptotic distributions of the proposed semi-parametric test statistics

This section rigorously examines the asymptotic properties of  $\mathbb{W}(\Sigma_n; b)$  and  $\mathbb{W}(\widehat{\Sigma}; b)$  in the presence of non-Gaussian temporally dependent noise. The technical assumptions for the theoretical analysis are provided below:

#### Condition A

- A1. The second derivative of the drift function  $d(t)$  is continuous and bounded for  $t \in [0, 1]$ .
- A2. The kernel  $K(x)$  is a Lipschitz continuous and bounded probability density function, symmetric with respect to  $x = 0$  and supported on  $[-L, L]$ , where  $L \in (0, \infty)$  is a constant.
- A3. The noise process  $\{\varepsilon(t_i)\}$  satisfies (2.3), and  $\Theta_{0,\omega} < \infty$  for a constant  $\omega \in [4, \infty)$ .
- A4. Assume  $\{s_j(\cdot)\}$  is independent of  $\varepsilon$ ,  $E\{s_j(t)\} = p_j$  for  $j = 1, \dots, \ell$ , and  $|s_j(t)| < C$  for any  $j = 1, \dots, \ell$  and  $t \in [0, 1]$ . Suppose  $\{s_j(t_1)\}_{j=1, \dots, \ell}^{\ell}$ ,  $\{s_j(t_n)\}_{j=1}^{\ell}$  are i.i.d..
- A5.  $b \rightarrow 0$  and  $nb \rightarrow \infty$  as  $n \rightarrow \infty$ .
- A6.  $t_i = i/n$  for  $i = 1, \dots, n$ .
- A7.  $\kappa_n > 0$  and  $\lim_{n \rightarrow \infty} \kappa_n = 0$ .
- A8.  $0 < c \leq \lambda_{\min}(\Sigma_n) \leq \lambda_{\max}(\Sigma_n) \leq C < \infty$ .
- A9. Assume  $\mathbf{T} \succeq 0$  is an  $n \times n$  symmetric and Toeplitz matrix satisfying  $1 = \mathbf{T}(1, 1) \geq \mathbf{T}(2, 1) \geq \dots \geq \mathbf{T}(n, 1) \geq 0$ , and  $\|\mathbf{T}\|_1 = o(n^{1/2})$ . There exists a sequence of positive integers  $\{x_n\}_{n=1}^{\infty}$  such that  $\lim_{n \rightarrow \infty} x_n = \infty$  and  $\lim_{n \rightarrow \infty} \mathbf{T}(x_n, 1) = 1$ .
- A10.  $E\{\{\mathbf{S} - E(\mathbf{S})\}^T \{\mathbf{S} - E(\mathbf{S})\}\} / n \succ c \mathbf{I}_{\ell m}$  for some constant  $c > 0$ .

The asymptotic distributions of  $\mathbb{W}(\Sigma_n; b)$  and  $\mathbb{W}(\widehat{\Sigma}; b)$  under  $H_0$  are demonstrated in Theorem 1 below.

**Theorem 1.** *Assume Condition A in Section 5.1. Under  $H_0$  in (3.1) where A is a  $k \times (\ell m)$  matrix and  $\text{rank}(\mathbf{A}) = k$ , we have that  $\mathbb{W}(\Sigma_n; b) \xrightarrow{D} \chi_k^2$ , and  $\mathbb{W}(\widehat{\Sigma}; b) \xrightarrow{D} \chi_k^2$ .*

Theorem 1 demonstrates that the null distributions of  $\mathbb{W}(\Sigma_n; b)$  and  $\mathbb{W}(\widehat{\Sigma}; b)$  asymptotically follow  $\chi_k^2$ . To prove the results for  $\mathbb{W}(\widehat{\Sigma}; b)$  in Theorem 1, a challenging aspect involving showing that  $n^{-1} \widetilde{\mathbf{S}}^T \widehat{\Sigma} \widetilde{\mathbf{S}} - n^{-1} \widetilde{\mathbf{S}}^T \Sigma_n \widetilde{\mathbf{S}} = o_P(1)$  (see Lemma 6 in the Appendix). In [46], for stationary noise processes, a similar condition  $n^{-1} \widetilde{\mathbf{S}}^T \widehat{\mathbf{R}}_n^{-1} \widetilde{\mathbf{S}} - n^{-1} \widetilde{\mathbf{S}}^T \mathbf{R}_n^{-1} \widetilde{\mathbf{S}} = o_P(1)$  (see (A.16) therein) is demonstrated to derive the asymptotic distributions of their proposed test statistics  $\mathbb{K}$  and its bias-corrected version  $\mathbb{K}_{bc}$ . This is done by assuming that  $E(\|\widehat{\mathbf{R}}_n^{-1} - \mathbf{R}_n^{-1}\|_{\infty}^2) = o(1)$  (see Condition A8 therein), where  $\widehat{\mathbf{R}}_n$  is an estimator of the noise auto-

correlation matrix  $R_n$ . There,

$$\mathbb{K} = \frac{(\mathbf{A}\tilde{\mathbf{h}})^T \{ \mathbf{A}(\tilde{\mathbf{S}}^T \hat{\mathbf{R}}_n^{-1} \tilde{\mathbf{S}})^{-1} \mathbf{A}^T \}^{-1} (\mathbf{A}\tilde{\mathbf{h}})}{\tilde{\mathbf{r}}^T \hat{\mathbf{R}}_n^{-1} \tilde{\mathbf{r}} / (n - \ell m)}, \quad (5.1)$$

with  $\tilde{\mathbf{r}} = \tilde{\mathbf{Y}} - \tilde{\mathbf{S}}\tilde{\mathbf{h}}$  and

$$\tilde{\mathbf{h}} = (\tilde{\mathbf{S}}^T \hat{\mathbf{R}}_n^{-1} \tilde{\mathbf{S}})^{-1} \tilde{\mathbf{S}}^T \hat{\mathbf{R}}_n^{-1} \tilde{\mathbf{Y}}. \quad (5.2)$$

[15] demonstrated that a banded estimator of  $R_n$  satisfies Condition A8 in [46] under a stationary  $g_n$ -dependent assumption for  $\varepsilon$ . In our work, it appears that a straightforward approach to prove Lemma 6 is to show  $E(\|\hat{\Sigma} - \Sigma_n\|_\infty^2) = o(1)$ . However, under the non-stationary situation, demonstrating the consistency of  $\hat{\Sigma}$  is extremely challenging due to the large number of parameters estimated. Here, we establish Lemma 6 by leveraging the special structure of the tapering-based estimator  $\hat{\Sigma}$  and the assumption of short-range dependence for the noise process  $\{\varepsilon(t_i)\}$ . Our assumption regarding the bandwidth,  $b \rightarrow 0$  and  $nb \rightarrow \infty$  in Condition A5 is milder than those required for analyzing the partially linear model in [32] and [21]. This is because we considered fixed design points  $t_1, \dots, t_n$  for the nonparametric component, while they used random design. Fixed design allows us to disregard the estimation error of the conditional expectation of  $\mathbf{S}$  given the design points of the nonparametric component. Nevertheless, controlling this estimation error necessitates stringent conditions on  $b$  under random design; refer to the first paragraph of Section 3 in [32] and the fourth paragraph on page 278 in [21] for further details.

*Remark 1.* The primary technical tools employed in the proof of Theorem 1 to address non-stationarity involve  $a_n$ -dependent and martingale approximations for linear and quadratic forms. These tools only necessitate the short-range dependence outlined in Condition A3. While Lipschitz continuity is often required, for instance, to approximate  $\Sigma_n(i, j)$  using a specific covariance function for a time-varying model as demonstrated in [45], it is not needed in our paper because we directly employ  $\Sigma_n$  in the test statistic (3.6) without any approximation.

Next, we delve into the asymptotic powers of  $\mathbb{W}(\Sigma_n; b)$  and  $\mathbb{W}(\hat{\Sigma}; b)$  against the fixed alternative in (3.1).

**Theorem 2.** *Assume Condition A in Section 5.1, with  $n^{-1}\tilde{\mathbf{S}}^T\tilde{\mathbf{S}} \xrightarrow{P} M_1 \succ 0$  and  $n^{-1}\tilde{\mathbf{S}}^T\Sigma_n\tilde{\mathbf{S}} \xrightarrow{P} M_2 \succ 0$ . Under the fixed alternative  $H_1$  in (3.1), we have that*

$$\begin{aligned} n^{-1}\mathbb{W}(\Sigma_n; b) &\xrightarrow{P} (\mathbf{A}\mathbf{h})^T (\mathbf{A}M_1^{-1}M_2M_1^{-1}\mathbf{A}^T)^{-1} (\mathbf{A}\mathbf{h}) > 0, \\ n^{-1}\mathbb{W}(\hat{\Sigma}; b) &\xrightarrow{P} (\mathbf{A}\mathbf{h})^T (\mathbf{A}M_1^{-1}M_2M_1^{-1}\mathbf{A}^T)^{-1} (\mathbf{A}\mathbf{h}) > 0. \end{aligned}$$

Theorem 2 suggests that, under any fixed alternative  $H_1$ ,  $\mathbb{W}(\Sigma_n; b)$  and  $\mathbb{W}(\hat{\Sigma}; b)$  diverge to infinity with a rate of  $n$ . Hence, the power functions of the proposed test statistics approach one under  $H_1$  as  $n \rightarrow \infty$ . As demonstrated in the proof of Theorem 1,  $n^{-1}\tilde{\mathbf{S}}^T\Sigma_n\tilde{\mathbf{S}} - E(n^{-1}\tilde{\mathbf{S}}^T\Sigma_n\tilde{\mathbf{S}}) = o_P(1)$  and  $n^{-1}\tilde{\mathbf{S}}^T\tilde{\mathbf{S}} -$

$E(n^{-1}\tilde{\mathbf{S}}^T\tilde{\mathbf{S}}) = o_P(1)$ , which implies that  $M_1 = \lim_{n \rightarrow \infty} E(n^{-1}\tilde{\mathbf{S}}^T\Sigma_n\tilde{\mathbf{S}})$  and  $M_2 = \lim_{n \rightarrow \infty} E(n^{-1}\tilde{\mathbf{S}}^T\tilde{\mathbf{S}})$ , provided the limits exist. Given that the limits of  $\mathbb{W}(\Sigma_n; b)$  and  $\mathbb{W}(\hat{\Sigma}; b)$  are infinity and do not depend on the explicit forms of  $M_1$  and  $M_2$ , this prompts us to explore the possibility of relaxing the conditions  $n^{-1}\tilde{\mathbf{S}}^T\tilde{\mathbf{S}} \xrightarrow{P} M_1 \succ 0$  and  $n^{-1}\tilde{\mathbf{S}}^T\Sigma_n\tilde{\mathbf{S}} \xrightarrow{P} M_2 \succ 0$  in Theorem 2. As in the proof of Theorem 1, we can demonstrate that under Condition A, for any subsequence  $\{n_l\}_{l=1}^\infty$ , there exists a further subsequence  $\{n_{l_j}\}_{j=1}^\infty$  such that  $n_{l_j}^{-1}\tilde{\mathbf{S}}^T\tilde{\mathbf{S}} \xrightarrow{P} M_1^* \succ 0$  and  $n_{l_j}^{-1}\tilde{\mathbf{S}}^T\Sigma_{n_{l_j}}\tilde{\mathbf{S}} \xrightarrow{P} M_2^* \succ 0$  as  $j \rightarrow \infty$ . This result further enables us to establish the divergence of  $\mathbb{W}(\Sigma_n; b)$  and  $\mathbb{W}(\hat{\Sigma}; b)$  under  $H_1$  by assuming relaxed conditions.

**Corollary 1.** *Assume Condition A in Section 5.1. Under the fixed alternative  $H_1$  in (3.1), we have that  $\mathbb{W}(\Sigma_n; b) \xrightarrow{P} \infty$  and  $\mathbb{W}(\hat{\Sigma}; b) \xrightarrow{P} \infty$ .*

In addition to the fixed alternative  $H_1$  in (3.1), we also examine the shrinking contiguous alternatives within an  $n^{-1/2}$  neighborhood of  $H_0$  in (3.1), defined as:

$$H_{1n} : \mathbf{A}\mathbf{h} = n^{-1/2}\mathbf{c}, \quad (5.3)$$

where  $\mathbf{c} = (c_1, \dots, c_k)^T \neq \mathbf{0}_k$ . Theorem 3 below establishes nontrivial power functions of  $\mathbb{W}(\Sigma_n; b)$  and  $\mathbb{W}(\hat{\Sigma}; b)$  against the local alternatives in (5.3).

**Theorem 3.** *Assume Condition A in Section 5.1, with  $n^{-1}\tilde{\mathbf{S}}^T\tilde{\mathbf{S}} \xrightarrow{P} M_1 \succ 0$  and  $n^{-1}\tilde{\mathbf{S}}^T\Sigma_n\tilde{\mathbf{S}} \xrightarrow{P} M_2 \succ 0$ . Then, under the local alternative  $H_{1n}$  in (5.3), we have that  $\mathbb{W}(\Sigma_n; b) \xrightarrow{D} \chi_k^2(\tau^2)$  and  $\mathbb{W}(\hat{\Sigma}; b) \xrightarrow{D} \chi_k^2(\tau^2)$ , where the constant  $\tau^2 = \mathbf{c}^T(\mathbf{A}M_1^{-1}M_2M_1^{-1}\mathbf{A}^T)^{-1}\mathbf{c} > 0$ .*

Theorem 3 asserts that under local alternatives, our proposed test statistics follow an asymptotic noncentral  $\chi_k^2$  distribution, where the noncentrality parameter  $\tau^2$  depends on  $M_1$  and  $M_2$ .

## 5.2. Choice of the tapering matrix $\mathbf{T}$ in (4.1)

The asymptotic properties of the proposed  $\mathbb{W}(\hat{\Sigma}; b)$  in (4.2) depend on certain technical assumptions (Condition A9 in Section 5.1) concerning the tapering matrix  $\mathbf{T}$  used in (4.1) to construct  $\hat{\Sigma}$ . A wide class of tapering matrices satisfies these conditions. In our numerical experiments, we employ  $\mathbf{T}_{g_n}$ , defined as:

$$\mathbf{T}_{g_n}(i, j) = \begin{cases} 1 - |i - j|/g_n, & \text{if } |i - j| < g_n, \\ 0, & \text{if } |i - j| \geq g_n, \end{cases} \quad \text{for } i, j = 1, \dots, n, \quad (5.4)$$

where the banding parameter  $g_n \in \{1, \dots, n\}$  depends on  $n$ . This triangular tapering matrix  $\mathbf{T}_{g_n}$ , as used in [43] for estimating the autocovariance matrix in stationary time series, has been shown to be rate-optimal. Therefore, we

adopt (5.4) for our method. Various other tapering methods are available in the literature. For instance, [3] considered

$$T_{\tau_n}(i, j) = \exp(-|i - j|/\tau_n)$$

(see Assumption A on page 207 therein), which satisfies our Condition A9 under proper assumptions on  $\tau_n$ . Another tapering matrix used in both [6] (see equation (5) on page 2121 therein) and [28] (see Equation (4) therein) is defined as:

$$T_{k_n}(i, j) = \begin{cases} 1, & \text{if } |i - j| \leq k_n/2, \\ 2 - 2|i - j|/k_n, & \text{if } k_n/2 < |i - j| < k_n, \\ 0, & \text{if } |i - j| \geq k_n, \end{cases} \quad \text{for } i, j = 1, \dots, n.$$

While [6] demonstrated that this tapering estimator is rate optimal for estimating high-dimensional covariance matrices with i.i.d. observations, it is important to note that  $T_{k_n}$  does not guarantee positive semidefiniteness. In Proposition 1 below, we discuss the conditions on  $g_n$  under which  $T_{g_n}$  fulfills Condition A9 as outlined in Section 5.1.

**Proposition 1.** *If  $\lim_{n \rightarrow \infty} g_n = \infty$  and  $g_n = o(n^{1/2})$ , then the tapering matrix  $T_{g_n}$  defined in (5.4) satisfies Condition A9 in Section 5.1.*

For brevity, let's denote the noise auto-covariance matrix estimator in (4.1) as:

$$\widehat{\Sigma}_{b;g_n} = \{(\widetilde{Y} - \widetilde{S}\widehat{h})(\widetilde{Y} - \widetilde{S}\widehat{h})^T\} \circ T_{g_n} + \kappa_n \mathbf{I}_n, \quad (5.5)$$

calculated with the bandwidth parameter  $b$  and banding parameter  $g_n$  of the tapering matrix  $T_{g_n}$ . Together with Proposition 1, the conclusions in Theorems 1–3 and Corollary 1 still hold for

$$\mathbb{W}(\widehat{\Sigma}_{b;g_n}; b) = (A\widehat{h})^T \{A(\widetilde{S}^T \widetilde{S})^{-1} \widetilde{S}^T \widehat{\Sigma}_{b;g_n} \widetilde{S}(\widetilde{S}^T \widetilde{S})^{-1} A^T\}^{-1} (A\widehat{h}),$$

i.e., the semi-parametric test statistic calculated with the true  $\Sigma_n$  substituted by the estimator  $\widehat{\Sigma}_{b;g_n}$ .

In [46], semi-parametric test statistics  $\mathbb{K}$  and  $\mathbb{K}_{bc}$  were developed based on a  $g$ -banded estimator of the noise auto-correlation matrix  $R_n$ , with  $g$  as a fixed integer. [15] studied a  $g_n$ -banded estimator  $\widehat{R}_n$  of  $R_n$  for  $\mathbb{K}$  and  $\mathbb{K}_{bc}$ , demonstrating the consistency of  $\widehat{R}_n$  by assuming  $g_n = o(n^{1/14})$ . Despite the different structures and assumptions, it is important to clarify the distinction between the divergence rate of the banding parameter  $g_n$  for  $\widehat{R}_n$  in [15] and that for  $\widehat{\Sigma}_{b;g_n}$  in Proposition 1. On the one hand,  $\widehat{R}_n$  is a banded estimator whereas  $\widehat{\Sigma}_{b;g_n}$  is a tapering-based estimator. On the other hand, [15] aims to prove the consistency of  $\widehat{R}_n$ , which requires stronger conditions on  $g_n$ , whereas the consistency of  $\widehat{\Sigma}_{b;g_n}$  is not required in our paper.

### 5.3. Selection of tuning parameters $b$ and $g_n$

This section develops the data-driven methods for selecting the bandwidth parameter  $b$  for the local-linear smoothing matrix  $S_b$  in (3.2) and the banding parameter  $g_n$  for the tapering matrix  $T_{g_n}$  in (5.4). In practical implementation, we proceed by first calculating the data-driven bandwidth  $\hat{b}$  of  $b$ , as explained in Section 5.3.2 below. We then compute the data-driven choice  $\hat{g}(\hat{b})$  for  $g_n$ , as discussed in Section 5.3.1. Additional simulation evaluations (omitted in the paper) demonstrate that the proposed method for selecting the tuning parameters  $b$  and  $g_n$  is robust against a small percentage of outlier contamination.

#### 5.3.1. Data-driven choice of $g_n$ using a given $b$

Recall that two tuning parameters,  $b$  and  $g_n$ , serve different roles in the estimator  $\hat{\Sigma}_{b;g_n}$  in (5.5). We first develop a data-driven method to choose  $g_n$  for any given bandwidth parameter  $b$ .

In the literature on estimating high-dimensional auto-covariance or auto-correlation matrices of stationary time series [40, 15], the risk-minimization method is used to select the banding parameter. Estimating the risk functions in these works involves using the sub-sampling technique [30], which entails “overlapped” data splitting.

In cases involving non-stationary noise processes, the method of selecting  $g_n$  by directly minimizing  $E(\|\hat{\Sigma}_{b;g_n} - \Sigma_n\|_2)$  may not perform well, as the estimator  $\hat{\Sigma}_{b;g_n}$  may not be consistent to  $\Sigma_n$  for non-stationary noise processes. However, the result  $n^{-1}\tilde{\mathbf{S}}^T\hat{\Sigma}_{b;g_n}\tilde{\mathbf{S}} - n^{-1}\tilde{\mathbf{S}}^T\Sigma_n\tilde{\mathbf{S}} = o_P(1)$  (see Lemma 6) enables us to establish a more appropriate approach of selecting  $g_n$  by minimizing the proposed risk function,

$$r(g_n; b) = E\{\|\tilde{\mathbf{S}}^T(\hat{\Sigma}_{b;g_n} - \Sigma_n)\tilde{\mathbf{S}}\|_2\}. \quad (5.6)$$

To enhance computational efficiency (supported by Proposition 1),  $g_n$  is chosen from  $\{1, \dots, C_0\}$  rather than the entire set  $\{1, \dots, n\}$ , where  $C_0$  is a predetermined integer much smaller than  $n$ .

We will apply the concept of sub-sampling to estimate the proposed risk function in (5.6). Despite the non-stationary nature of the noise process, where the noise covariance structure may vary among sub-samples, a common decaying pattern persists across all sub-samples due to the short-range dependence. Therefore, the sub-sampling method can be employed to estimate the shared banding parameter for all sub-samples. Let  $\widetilde{\mathbf{res}} = \tilde{\mathbf{Y}} - \tilde{\mathbf{S}}\hat{\mathbf{h}}$  represent the vector of residuals from model (3.4). We use positive integers  $V$  and  $B$  to denote the total number of blocks (each comprising sub-samples) and the number of sub-samples within each block, respectively. Define  $G_\nu = \{(\nu - 1)\lfloor(n - B)/(V - 1)\rfloor + 1, \dots, (\nu - 1)\lfloor(n - B)/(V - 1)\rfloor + B\}$  to collect indices  $i \in \{1, \dots, n\}$  associated with sub-samples in the  $\nu$ th block, for  $\nu = 1, \dots, V$ , where  $\lfloor \cdot \rfloor$  represents the floor operator. Extract rows corresponding to indices  $i \in G_\mu$  from  $\widetilde{\mathbf{res}}$  and  $\tilde{\mathbf{S}}$ , denoted as  $\widetilde{\mathbf{res}}^{(\mu)}$  and  $\tilde{\mathbf{S}}^{(\mu)}$ , respectively. To estimate (5.6) using sub-samples,

replace the unknown  $\Sigma_n$  with the matrix  $\Sigma_n^{(\nu)} = \{\widetilde{\mathbf{res}}^{(\nu)}\}\{\widetilde{\mathbf{res}}^{(\nu)}\}^T \in \mathbb{R}^{B \times B}$  for  $\nu \in \{1, \dots, V\}$ . Similarly, replace  $\widetilde{\mathbf{S}}$  and  $\widehat{\Sigma}_{b;g_n}$  with  $\widetilde{\mathbf{S}}^{(\mu)}$  and

$$\widehat{\Sigma}_{b;g_n}^{(\mu)} = \Sigma_n^{(\mu)} \circ \mathbf{T}_{g_n}(1 : B, 1 : B) + \kappa_n \mathbf{I}_B,$$

respectively, using blocks  $\mu \in \{1, \dots, V\}$ , where  $\mathbf{T}_{g_n}(1 : B, 1 : B)$  denotes the upper left  $B \times B$  submatrix of  $\mathbf{T}_{g_n}$ . Consequently, the sub-sample estimator of  $r(g_n; b)$  takes the form:

$$\widehat{r}(g_n; b) = \frac{1}{V(V-1)} \sum_{\nu=1}^V \sum_{\mu=1: \mu \neq \nu}^V \left\| \widetilde{\mathbf{S}}^{(\mu)T} \{ \widehat{\Sigma}_{b;g_n}^{(\mu)} - \Sigma_n^{(\nu)} \} \widetilde{\mathbf{S}}^{(\mu)} \right\|_2.$$

For any given bandwidth  $b$ , this leads to the data-driven choice of the banding parameter as follows:

$$\widehat{g}(b) = \arg \min_{g_n \in \{1, \dots, C_0\}} \widehat{r}(g_n; b). \tag{5.7}$$

*Remark 2.* The implementation of the data-driven method for selecting  $g_n$  involves choices of  $C_0$ ,  $V$ , and  $B$ . In our numerical evaluations, we have found that reasonable choices are  $C_0 = \lfloor 3 \log(10n) \rfloor$ ,  $V = 20$ , and  $B = \lfloor 8n^{1/3} \rfloor$  for the proposed test statistic  $\mathbb{W}(\widehat{\Sigma}_{\widehat{b};\widehat{g}(\widehat{b})}; \widehat{b})$  across various settings.

### 5.3.2. Data-driven choice of $b$

Recall that the estimator  $\widehat{\mathbf{h}}$  in (3.5) of the parameter vector  $\mathbf{h}$  relies on the bandwidth  $b$  of the local-linear smoothing matrix  $S_b$ . It is thus natural to choose  $b$  to minimize the mean squared error (MSE) of  $\widehat{\mathbf{h}}$ . In the presence of non-stationary noise, this is expressed as

$$\begin{aligned} \text{MSE}(\widehat{\mathbf{h}} \mid \mathbf{S}) &= \|(\widetilde{\mathbf{S}}^T \widetilde{\mathbf{S}})^{-1} \widetilde{\mathbf{S}}^T \widetilde{\mathbf{d}}\|_2^2 \\ &+ \text{tr}\{(\widetilde{\mathbf{S}}^T \widetilde{\mathbf{S}})^{-1} \widetilde{\mathbf{S}}^T (\mathbf{I}_n - S_b) \Sigma_n (\mathbf{I}_n - S_b)^T \widetilde{\mathbf{S}} (\widetilde{\mathbf{S}}^T \widetilde{\mathbf{S}})^{-1}\}. \end{aligned} \tag{5.8}$$

Ideally, if  $\mathbf{d}$  and  $\Sigma_n$  are known, then we can calculate the oracle-constant bandwidth  $b_{\text{oracle}}$ , which minimizes  $\text{E}\{\text{MSE}(\widehat{\mathbf{h}} \mid \mathbf{S})\}$ .

For realistic applications with unknown  $\mathbf{d}$  and  $\Sigma_n$ , we select the data-driven choice of  $b$  by minimizing the plug-in estimate  $\widehat{\text{MSE}}(\widehat{\mathbf{h}} \mid \mathbf{S})$ , which replaces  $\mathbf{d}$  and  $\Sigma_n$  in  $\text{MSE}(\widehat{\mathbf{h}} \mid \mathbf{S})$  by their estimates. Detailed procedures are described in Steps 1–3 below.

**Step 1:** Obtain an initial estimate of  $\mathbf{h}$  by some preliminary procedure. E.g., taking the first-order difference of (2.2) yields  $D_1 \mathbf{Y} = D_1 \mathbf{S} \mathbf{h} + D_1 \mathbf{d} + D_1 \boldsymbol{\varepsilon}$ , where  $D_1$  is an  $(n-1) \times n$  matrix defined as

$$D_1 = \begin{pmatrix} -1 & 1 & 0 & \cdots & 0 & 0 \\ 0 & -1 & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & -1 & 1 \end{pmatrix}.$$

Since  $d(\cdot)$  is smooth under Condition A1,  $D_1\mathbf{d}$  is ignorable. The ordinary least squares method supplies the initial difference-based estimate of  $\mathbf{h}$ ,

$$\widehat{\mathbf{h}}_{\text{DBE}} = \{(D_1\mathbf{S})^T(D_1\mathbf{S})\}^{-1}(D_1\mathbf{S})^T(D_1\mathbf{Y}).$$

**Step 2:** Select an initial bandwidth parameter  $\widehat{b}_{\text{init}}$  by minimizing the covariance-penalty  $\|(\mathbf{I}_n - S_b)(\mathbf{Y} - \mathbf{S}\widehat{\mathbf{h}}_{\text{DBE}})\|_2^2 + 2\text{tr}\{S_b\widehat{\Sigma}_{b_0;\widehat{g}(b_0)}\}$ , where  $\widehat{\mathbf{h}}_{\text{DBE}}$  is obtained in **Step 1**, and  $\widehat{g}(b_0)$  is a banding parameter selected by (5.7) using a predetermined bandwidth  $b_0 = 0.5$ .

**Step 3:** The data-driven choice of the bandwidth  $\widehat{b}$  is obtained by minimizing the plug-in estimate  $\widehat{\text{MSE}}(\widehat{\mathbf{h}} | \mathbf{S})$ , with  $\mathbf{d}$  and  $\Sigma_n$  in (5.8) replaced by their estimates  $S_{\widehat{b}_{\text{init}}}(\mathbf{Y} - \mathbf{S}\widehat{\mathbf{h}}_{\text{DBE}})$  and  $\widehat{\Sigma}_{\widehat{b}_{\text{init}};\widehat{g}(\widehat{b}_{\text{init}})}$ , respectively, where  $\widehat{g}(\widehat{b}_{\text{init}})$  is the choice of the banding parameter selected by (5.7) using the bandwidth parameter  $\widehat{b}_{\text{init}}$  in **Step 2**.

Our simulation studies indicate that the choice of  $b_0$  in **Step 2** doesn't appreciably affect the performance of the proposed semi-parametric test statistic. In practice, other proper choices of  $b_0$  can also be used.

## 6. Simulation study

### 6.1. Set-up

This section assesses the empirical performance of the proposed semi-parametric test statistics in addressing inference questions pertinent to the fMRI study of neural activity:

$$H_0 : \mathbf{h} = \mathbf{0}_{\ell m} \quad \text{versus} \quad H_1 : \mathbf{h} \neq \mathbf{0}_{\ell m}. \quad (6.1)$$

The temporally correlated data at time points  $t_i = i/n$  are generated from model (2.2) for 2 runs, with  $n = 300$  observations within each run. For the external stimuli, we consider both a single event type ( $\ell = 1$ ) with  $m = 16$  and  $\{s_1(t_i)\}_{i=1}^n \stackrel{\text{i.i.d.}}{\sim} \text{Bernoulli}(0.5)$ , and two event types ( $\ell = 2$ ) with  $m = 12$  and  $\{s_1(t_i), s_2(t_i), 1 - s_1(t_i) - s_2(t_i)\}_{i=1}^n \stackrel{\text{i.i.d.}}{\sim} \text{Multinomial}(1; 1/3, 1/3, 1/3)$ . The non-parametric time drift function is defined as  $d(t) = 10 \sin\{\pi(t - 0.21)\}$ ,  $t \in [0, 1]$ . Non-parametric estimation utilizes local-linear estimation combined with the Epanechnikov kernel function [34], which is supported on  $[-1, 1]$ . The tapering-based covariance matrix estimator employs  $\kappa_n = 10^{-5}$  in (5.5).

We examine six noise models below, encompassing stationary (noise models 1 and 4), non-stationary (noise models 2, 3, 5, 6), and non-Gaussian (noise models 4, 5, 6) cases.

- **Noise models 1 and 4 (stationary noise process):** AR(1) + WN,

$$\varepsilon(t_i) = \zeta_1(t_i) + \zeta_2(t_i), \quad \text{for } i = 1, \dots, n,$$

where  $\zeta_1(t_i) = 0.3\zeta_1(t_{i-1}) + z_i$  is an AR(1) process,  $\zeta_2(t_i)$  is the white noise process,  $\{\zeta_2(t_i)\}_{i=1}^n$  is independent of  $\{z_i\}_{i=1}^n$ . In noise model 1,  $\{z_i\}_{i=1}^n \stackrel{\text{i.i.d.}}{\sim}$

$0.4N(0, 1)$ , and  $\{\zeta_2(t_i)\}_{i=1}^n \stackrel{\text{i.i.d.}}{\sim} 0.4N(0, 1)$ . In noise model 4,  $\{z_i\}_{i=1}^n \stackrel{\text{i.i.d.}}{\sim} c_1 t_4$ , and  $\{\zeta_2(t_i)\}_{i=1}^n \stackrel{\text{i.i.d.}}{\sim} c_2 t_4$ , with constants  $c_1$  and  $c_2$  adjusted to match variances with counterparts in noise model 1.

- **Noise models 2 and 5 (non-stationary noise process):** MA(3) with time-varying coefficients,

$$\varepsilon(t_i) = \theta_{0,i} z_i + \theta_{1,i} z_{i-1} + \theta_{2,i} z_{i-2} + \theta_{3,i} z_{i-3}, \text{ for } i = 1, \dots, n,$$

with coefficients  $\theta_{q,i} = \{0.5 + 0.5(i+2-q)/(n+2)\} \{I(q=0) + 0.3I(q \neq 0)\}$ , for  $q = 0, \dots, 3$  and  $i = 1, \dots, n$ . In noise model 2,  $\{z_i\}_{i=-2}^n \stackrel{\text{i.i.d.}}{\sim} N(0, 1)$ . In noise model 5,  $\{z_i\}_{i=-2}^n \stackrel{\text{i.i.d.}}{\sim} c_3 t_5$ , with the constant  $c_3$  adjusted to match the variance of the counterpart in noise model 2.

- **Noise models 3 and 6 (non-stationary noise process):** Time-varying AR(1),

$$\varepsilon(t_i) = \xi_i(t_i), \text{ for } i = 1, \dots, n,$$

where  $\xi_i(t) = \rho(t) \xi_{i-1}(t) + z_i$ , for  $t \in (0, 1)$ , with  $\rho(t) = -0.2\{I(0 \leq t \leq 1/4) + I(1/2 < t \leq 3/4)\} - 0.6\{I(1/4 < t \leq 1/2) + I(3/4 < t \leq 1)\}$ ;  $\{z_i\}_{i=1}^n$  are i.i.d., following the distribution  $N(0, 0.5^2)$  in noise model 3, and  $c_4 t_5$  in noise model 6, with the constant  $c_4$  adjusted to comply with the variances.

## 6.2. Inference for 1D temporally correlated data

The 1D temporally correlated data  $(\mathbf{S}, \mathbf{Y})$  are simulated from model (2.2) to mimic voxelwise fMRI time series. The plots depict empirical percentiles (from the 1st to the 99th) of the proposed semi-parametric test statistics,  $\mathbb{W}(\Sigma_n; b_{\text{oracle}})$  and  $\mathbb{W}(\widehat{\Sigma}_{\widehat{b}; \widehat{g}(\widehat{b})}; \widehat{b})$ , under  $H_0$  in (6.1) versus the theoretical percentiles of the  $\chi_{\ell m}^2$  distribution. These plots are displayed in Figures 1, 3, and 5, corresponding to temporal noise models 4, 5 and 6, respectively, based on Monte Carlo simulation replications 500 times for each setting. Here, the oracle-constant bandwidth,  $b_{\text{oracle}}$ , minimizes  $E\{\text{MSE}(\widehat{\mathbf{h}} \mid \mathbf{S})\}$ , where the expectation is approximated by an empirical average across 100 simulated samples. The data-driven choices,  $\widehat{b}$  and  $\widehat{g}(\widehat{b})$ , are as described in Section 5.3. For comparison, we also consider test statistics  $\mathbb{K}$  and  $\mathbb{K}_{\text{bc}}$  from [46], which assume stationarity with the auto-correlation matrix estimator  $\widehat{\mathbf{R}}_n$  by [15].

- Clearly, in both stationary and non-stationary cases, the finite-sample distributions of  $\mathbb{W}(\Sigma_n; b_{\text{oracle}})$  and  $\mathbb{W}(\widehat{\Sigma}_{\widehat{b}; \widehat{g}(\widehat{b})}; \widehat{b})$  agree reasonably well with the  $\chi_{\ell m}^2$  distribution. This provides support for the notion that the proposed test statistics  $\mathbb{W}(\widehat{\Sigma}_{\widehat{b}; \widehat{g}(\widehat{b})}; \widehat{b})$ , using the tapering-based estimator of  $\Sigma_n$  and data-driven choices of tuning parameters, are comparable to their oracle counterparts,  $\mathbb{W}(\Sigma_n; b_{\text{oracle}})$ , which use the true  $\Sigma_n$  and the oracle-constant bandwidth parameter. Thus, both  $\mathbb{W}(\widehat{\Sigma}_{\widehat{b}; \widehat{g}(\widehat{b})}; \widehat{b})$  and  $\mathbb{W}(\Sigma_n; b_{\text{oracle}})$  adapt conceivably well to the temporal dependence from either stationary or non-stationary noise.

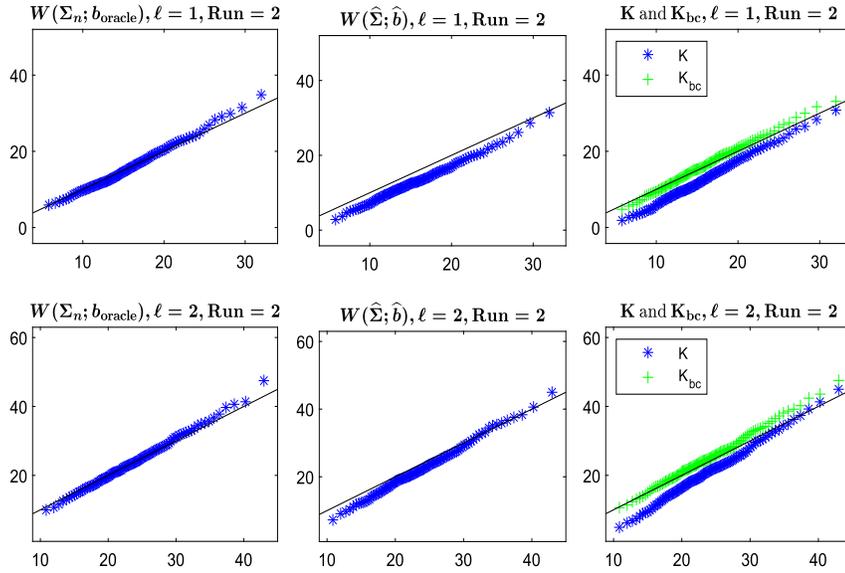


FIG 1. (Noise model 4) Empirical quantiles (on the y-axis) of  $W(\Sigma_n; b_{\text{oracle}})$ ,  $W(\widehat{\Sigma}_{\widehat{b}; \widehat{g}(\widehat{b})}; \widehat{b})$ ,  $\mathbb{K}$  and  $\mathbb{K}_{bc}$  versus quantiles (on the x-axis) of  $\chi^2$ . Top panels:  $\ell = 1$ ; bottom panels:  $\ell = 2$ . Left panels:  $W(\Sigma_n; b_{\text{oracle}})$ ; middle panels:  $W(\widehat{\Sigma}_{\widehat{b}; \widehat{g}(\widehat{b})}; \widehat{b})$ ; right panels:  $\mathbb{K}$  and  $\mathbb{K}_{bc}$ . The solid line represents the 45 degree straight line.

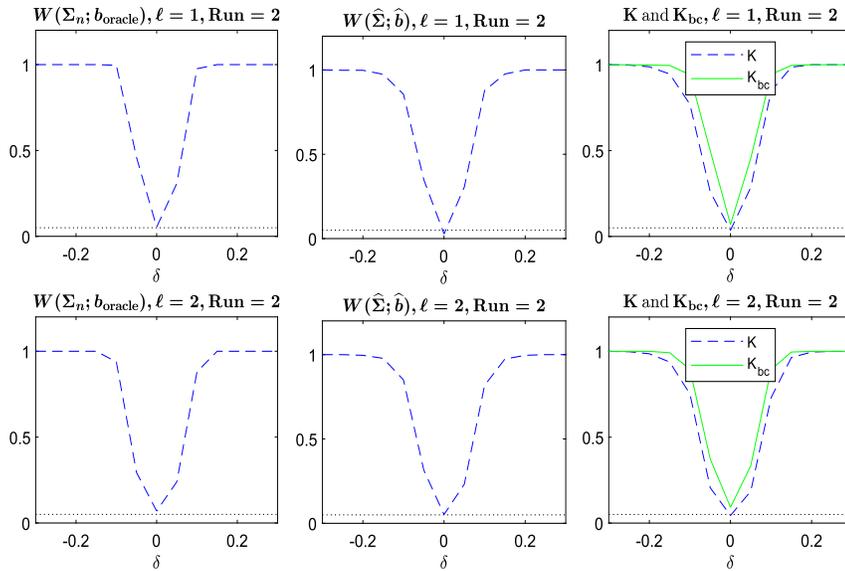


FIG 2. (Noise model 4) Empirical power functions of  $W(\Sigma_n; b_{\text{oracle}})$ ,  $W(\widehat{\Sigma}_{\widehat{b}; \widehat{g}(\widehat{b})}; \widehat{b})$ ,  $\mathbb{K}$  and  $\mathbb{K}_{bc}$ . Top panels:  $\ell = 1$ ; bottom panels:  $\ell = 2$ . Left panels:  $W(\Sigma_n; b_{\text{oracle}})$ ; middle panels:  $W(\widehat{\Sigma}_{\widehat{b}; \widehat{g}(\widehat{b})}; \widehat{b})$ ; right panels:  $\mathbb{K}$  and  $\mathbb{K}_{bc}$ . The dotted line represents the 5% significance level.

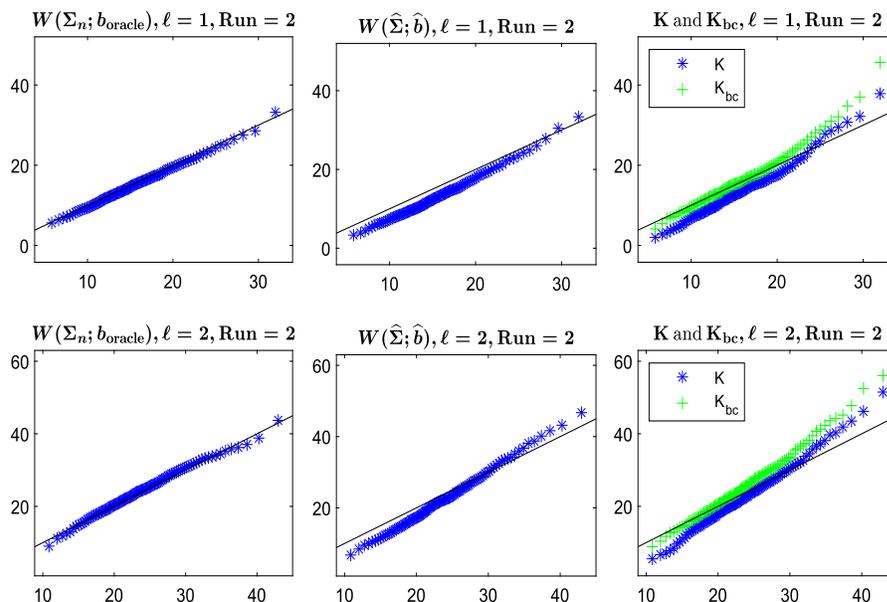


FIG 3. (**Noise model 5**) Empirical quantiles (on the  $y$ -axis) of  $\mathbb{W}(\Sigma_n; b_{\text{oracle}})$ ,  $\mathbb{W}(\widehat{\Sigma}_{\widehat{b}; \widehat{g}(\widehat{b})}; \widehat{b})$ ,  $\mathbb{K}$  and  $\mathbb{K}_{bc}$  versus quantiles (on the  $x$ -axis) of  $\chi^2$ . Top panels:  $\ell = 1$ ; bottom panels:  $\ell = 2$ . Left panels:  $\mathbb{W}(\Sigma_n; b_{\text{oracle}})$ ; middle panels:  $\mathbb{W}(\widehat{\Sigma}_{\widehat{b}; \widehat{g}(\widehat{b})}; \widehat{b})$ ; right panels:  $\mathbb{K}$  and  $\mathbb{K}_{bc}$ . The solid line represents the 45 degree straight line.

- (ii) In contrast, in the presence of non-stationary correlated noise, the finite sampling distributions of  $\mathbb{K}$  and  $\mathbb{K}_{bc}$  in [46] do not closely resemble the  $\chi_{\ell m}^2$  distribution, as observed in the right panels of Figures 3 and 5. The agreement with the chi-squared distribution is confirmed only in the stationary case, as displayed in the right panels of Figure 1.

To assess the test statistics' power, data  $(\mathbf{S}, \mathbf{Y})$  were generated from model (2.2) with a non-zero true parameter  $\mathbf{h} = \delta \mathbf{1}_{\ell m}$ , where  $\delta$  ranges from  $-0.3$  to  $0.3$  in increments of  $0.05$ . Empirical rejection rates of  $H_0$  in (6.1) for different  $\delta$  values are plotted in Figures 2, 4, and 6.

- (iii) As expected, when  $\delta = 0$ , the powers of  $\mathbb{W}(\Sigma_n; b_{\text{oracle}})$  and  $\mathbb{W}(\widehat{\Sigma}_{\widehat{b}; \widehat{g}(\widehat{b})}; \widehat{b})$  are nearly equal to the nominal significance level of  $0.05$ . As  $\delta$  deviates from  $0$ , the powers increase towards one. Hence, the proposed test statistics exhibit strong powers against the non-null  $\mathbf{h}$ , while maintaining control over the type-I error rate. Results for test statistics  $\mathbb{K}$  and  $\mathbb{K}_{bc}$  in [46], displayed in the right panels of Figure 2, also perform welly and show similar results to our proposed method due to the stationary nature of the noise process.
- (iv) However, unexpectedly, the right panels of Figures 4 and 6 reveal that the empirical powers of  $\mathbb{K}$  and  $\mathbb{K}_{bc}$  in [46] exceed the nominal level when

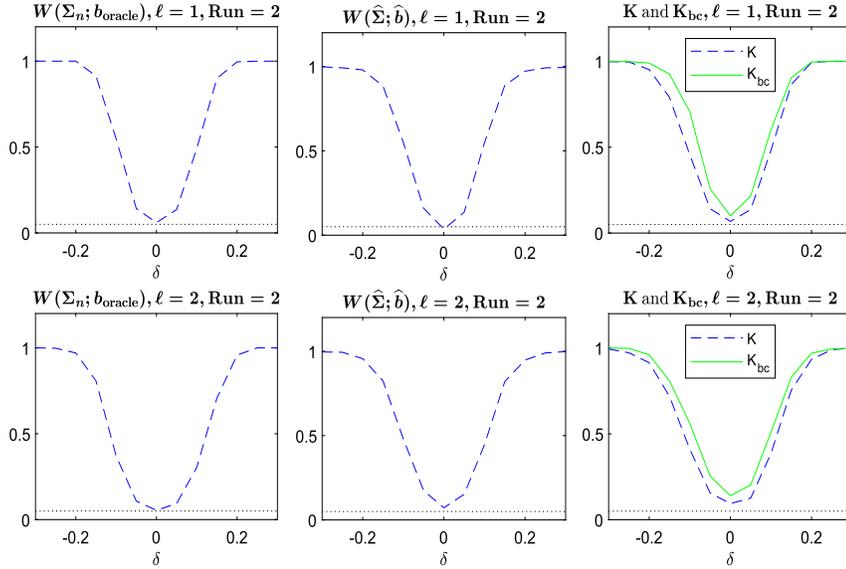


FIG 4. (*Noise model 5*) Empirical power functions of  $\mathbb{W}(\Sigma_n; b_{\text{oracle}})$ ,  $\mathbb{W}(\widehat{\Sigma}_{\widehat{b}; \widehat{g}(\widehat{b})}; \widehat{b})$ ,  $\mathbb{K}$  and  $\mathbb{K}_{bc}$ . Top panels:  $\ell = 1$ ; bottom panels:  $\ell = 2$ . Left panels:  $\mathbb{W}(\Sigma_n; b_{\text{oracle}})$ ; middle panels:  $\mathbb{W}(\widehat{\Sigma}_{\widehat{b}; \widehat{g}(\widehat{b})}; \widehat{b})$ ; right panels:  $\mathbb{K}$  and  $\mathbb{K}_{bc}$ . The dotted line represents the 5% significance level.

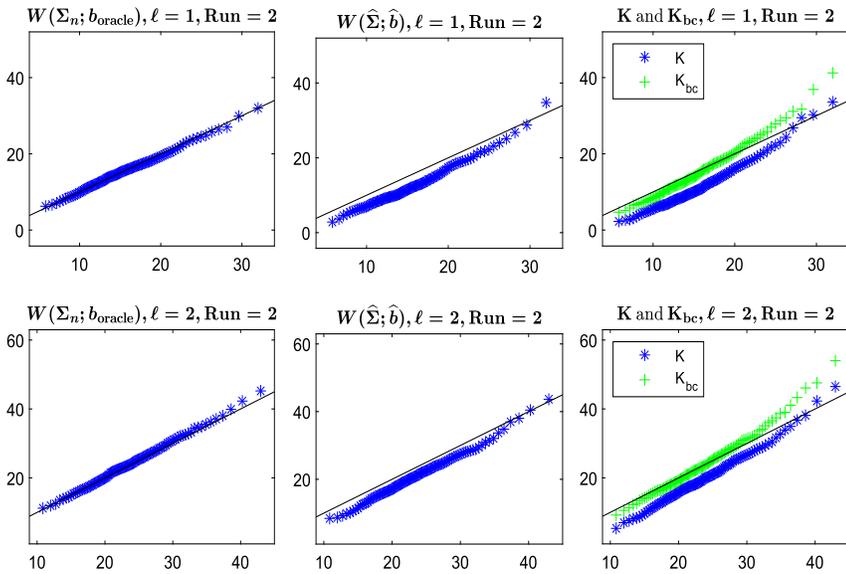


FIG 5. (*Noise model 6*) Empirical quantiles (on the y-axis) of  $\mathbb{W}(\Sigma_n; b_{\text{oracle}})$ ,  $\mathbb{W}(\widehat{\Sigma}_{\widehat{b}; \widehat{g}(\widehat{b})}; \widehat{b})$ ,  $\mathbb{K}$  and  $\mathbb{K}_{bc}$  versus quantiles (on the x-axis) of  $\chi^2$ . Top panels:  $\ell = 1$ ; bottom panels:  $\ell = 2$ . Left panels:  $\mathbb{W}(\Sigma_n; b_{\text{oracle}})$ ; middle panels:  $\mathbb{W}(\widehat{\Sigma}_{\widehat{b}; \widehat{g}(\widehat{b})}; \widehat{b})$ ; right panels:  $\mathbb{K}$  and  $\mathbb{K}_{bc}$ . The solid line represents the 45 degree straight line.

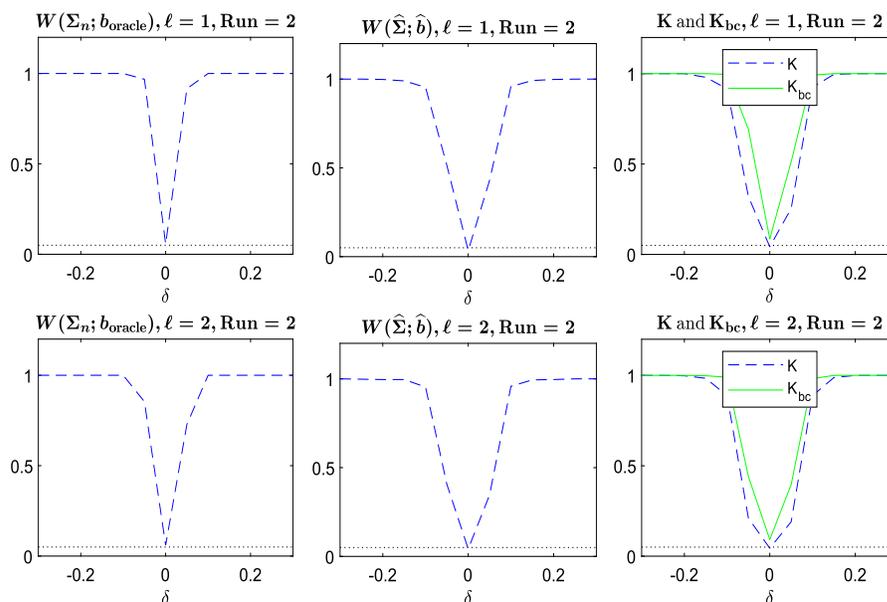


FIG 6. (**Noise model 6**) Empirical power functions of  $\mathbb{W}(\Sigma_n; b_{\text{oracle}})$ ,  $\mathbb{W}(\widehat{\Sigma}; \widehat{b})$ ,  $\mathbb{K}$  and  $\mathbb{K}_{bc}$ . Top panels:  $\ell = 1$ ; bottom panels:  $\ell = 2$ . Left panels:  $\mathbb{W}(\Sigma_n; b_{\text{oracle}})$ ; middle panels:  $\mathbb{W}(\widehat{\Sigma}_{\widehat{b}, \widehat{g}(\widehat{b})}; \widehat{b})$ ; right panels:  $\mathbb{K}$  and  $\mathbb{K}_{bc}$ . The dotted line represents the 5% significance level.

$\delta = 0$ , a consequence of the non-stationary temporally correlated noise process.

### 6.3. Inference for 2D temporally and spatially correlated data

This section illustrates the semi-parametric inference procedure for detecting significant regions in 2D datasets. We simulate 101 replicates of fMRI data sets on a 2D-slice formed by  $64 \times 64$  coordinates. The number of scanned voxels  $\mathcal{V}_0 \cup \mathcal{V}_1$  is 578, and the number of truly active scanned voxels  $\mathcal{V}_1$  is 104, where  $\mathcal{V}_0$  and  $\mathcal{V}_1$  refer to the regions of background and activated voxels, respectively. Within each simulation, at each voxel  $v \in \mathcal{V}_0 \cup \mathcal{V}_1$ , the data are simulated via  $\mathbf{Y}^{(v)} = \mathbf{S}\mathbf{h}^{(v)} + \mathbf{d}^{(v)} + \boldsymbol{\varepsilon}^{(v)}$ , according to the set-up in Section 6.1, containing  $n = 300$  time points within each of 2 runs, with  $\ell = 1$  and  $m = 16$ . Here  $\mathbf{h}^{(v)} = \mathbf{0}$  at  $v \in \mathcal{V}_0$ , and  $\mathbf{h}^{(v)} \neq \mathbf{0}$  at  $v \in \mathcal{V}_1$ . For the non-zero  $\mathbf{h}^{(v)}$ , we adopt [14] to specify the HRF parameters as  $h_j(t_i) = \{g_1(1.5(i-1) - \tau_s)/g_1(a_1b_1) - cg_2(1.5(i-1) - \tau_s)/g_2(a_2b_2)\}I(1.5(i-1) - \tau_s > 0)$ , for  $i = 1, \dots, m$  and  $j = 1, \dots, \ell$ , where  $g_1(t) = t^{a_1}e^{-t/b_1}$  and  $g_2(t) = t^{a_2}e^{-t/b_2}$ , with  $a_1 = 5$ ,  $b_1 = 0.9$ ,  $a_2 = 12$ ,  $b_2 = 0.7$ ,  $c = 0.4$ , and  $\tau_s = 5.5$ . The region  $\mathcal{V}_1$  corresponds to two disks (as shown in red in the top left panel of Figure 7), centered at  $(47, 44)$  and  $(20, 27)$ , with radii 5 and 6, respectively. Recall from (1.1)

that stimuli in  $\mathbf{S}$  are common to all voxels  $v \in \mathcal{V}_0 \cup \mathcal{V}_1$  but vary with different simulations. Hence, the 2D data are spatially correlated. The signal to noise ratios (SNR) of simulated fMRI data at the activated voxels  $v \in \mathcal{V}_1$  are 1.31, 0.59, 1.36, 1.31, 0.59, and 1.35 under noise models 1–6, respectively, where  $\text{SNR} = \sum_{i=1}^n \text{var}(\mathbf{S}_i^T \mathbf{h}^{(v)}) / \sum_{i=1}^n \text{var}\{\varepsilon^{(v)}(t_i)\}$ .

To compare with the semi-parametric inference methods, we also include the parametric  $F$ -test combined with “parametric OLS” estimation, which assumes that the non-parametric component in (2.1) is fitted by a parametric quadratic model,  $d(t_i) = r_0 + r_1 t_i + r_2 t_i^2$  with unknown parameters  $(r_0, r_1, r_2)$ , and noise terms  $\{\varepsilon(t_i)\}$  are i.i.d. as  $N(0, \sigma^2)$ .

Two false discovery rate (FDR)-controlling procedures, the BH procedure [2] and the Storey *et al.* procedure, [36], are employed to simultaneously test the significance of  $\mathbf{h}^{(v)}$  in (6.1) at all 578 voxels, with the FDR control level set at 0.05. Table 1 compares the calculated false discovery proportions (FDP), averaged across 101 simulations. The proposed  $\mathbb{W}(\Sigma_n; b_{\text{oracle}})$  and  $\mathbb{W}(\widehat{\Sigma}_{\widehat{b}; \widehat{g}(\widehat{b})}; \widehat{b})$  both effectively control the FDR, whereas the “parametric OLS” method inflates the FDR due to misspecification of the non-parametric component.

TABLE 1  
Compare the calculated FDP averaged over 101 simulations.

noise model		OLS	$\mathbb{W}(\Sigma_n; b_{\text{oracle}})$	$\mathbb{W}(\widehat{\Sigma}_{\widehat{b}; \widehat{g}(\widehat{b})}; \widehat{b})$	
2D	1	BH	0.516	0.040	0.030
		Storey <i>et al.</i>	0.556	0.049	0.034
	2	BH	0.488	0.043	0.047
		Storey <i>et al.</i>	0.532	0.051	0.054
	3	BH	0.557	0.041	0.028
		Storey <i>et al.</i>	0.606	0.051	0.034
2D	4	BH	0.573	0.055	0.026
		Storey <i>et al.</i>	0.613	0.065	0.031
	5	BH	0.500	0.052	0.045
		Storey <i>et al.</i>	0.546	0.063	0.050
	6	BH	0.509	0.057	0.037
		Storey <i>et al.</i>	0.555	0.069	0.041

For graphical illustrations of the active regions detected by the parametric and semi-parametric methods, Figure 7 presents plots representing each of the three estimation methods (combined with the Storey *et al.* procedure). In this context, “representative” signifies the median performance, ranked according to the FDP, across 101 simulations. In Figure 7, the top left panel shows the true active regions (in red), while the top right, bottom left, and bottom right panels display the active regions (in purple) detected by the “parametric OLS” method, the proposed  $\mathbb{W}(\Sigma_n; b_{\text{oracle}})$ , and  $\mathbb{W}(\widehat{\Sigma}_{\widehat{b}; \widehat{g}(\widehat{b})}; \widehat{b})$ , respectively. The comparison reveals that the “semi-parametric data-driven” method outperforms the “parametric OLS” method in the presence of non-stationary non-Gaussian time series noise.

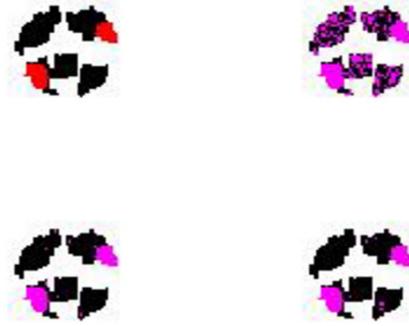


FIG 7. (*Noise model 5; 2D*) Top left: **truly** active regions; top right: **detected** active regions by the “parametric OLS”; bottom left: **detected** active regions by  $\mathbb{W}(\Sigma_n; b_{\text{oracle}})$ ; bottom right: **detected** active regions by  $\mathbb{W}(\hat{\Sigma}_{\hat{b}; \hat{g}(\hat{b})}; \hat{b})$ .

#### 6.4. Inference for 3D temporally and spatially correlated data

We simulate 101 sets of fMRI data on a 3D cube, with the number of 3D coordinates equal to  $64 \times 64 \times 8$ ; the number of scanned voxels is 4949, and the number of truly active scanned voxels is 773. The simulation setup at each voxel is similar to that for the 2D inference in Section 6.3. In Figure 8, the top left 8 slices show the true active regions (in red), while the top right, bottom left, and the bottom right 8 slices depict the active regions (in purple), which were identified by the “parametric OLS” method, the proposed  $\mathbb{W}(\Sigma_n; b_{\text{oracle}})$ , and  $\mathbb{W}(\hat{\Sigma}_{\hat{b}; \hat{g}(\hat{b})}; \hat{b})$ . The graphical assessment of detected regions in Figure 8, along with the numerical comparison of the FDPs summarized in Table 2, provides compelling evidence of the superior performance of the “semi-parametric data-driven” inference method over the parametric counterpart for data contaminated with non-stationary, non-Gaussian, temporally correlated noise.

TABLE 2  
Compare the calculated FDP averaged over 101 simulations.

noise model			OLS	$\mathbb{W}(\Sigma_n; b_{\text{oracle}})$	$\mathbb{W}(\hat{\Sigma}_{\hat{b}; \hat{g}(\hat{b})}; \hat{b})$
3D	4	BH	0.532	0.062	0.026
		Storey <i>et al.</i>	0.572	0.071	0.030
	5	BH	0.519	0.057	0.049
		Storey <i>et al.</i>	0.563	0.067	0.054
	6	BH	0.502	0.058	0.031
		Storey <i>et al.</i>	0.552	0.067	0.034

## 7. Real brain fMRI analysis

We analyze the StarPlus fMRI data originally collected at Carnegie Mellon University’s CCBI. Detailed information about the experiment design, data de-



FIG 8. (*Noise model 5*; 3D) Top left 8 slices: **truly** active regions; top right 8 slices: **detected** active regions by the “parametric OLS”; bottom left 8 slices: **detected** active regions by  $\mathbb{W}(\Sigma_n; b_{\text{oracle}})$ ; bottom right 8 slices: **detected** active regions by  $\mathbb{W}(\hat{\Sigma}_{\hat{b}; \hat{g}(\hat{b})}; \hat{b})$ .

scription, and subjects can be found at <http://www.cs.cmu.edu/afs/cs.cmu.edu/project/theo-81/www/>. Images were captured every 0.5 seconds for 6 subjects, with only a portion of each subject’s brain imaged. The data are marked with 25–30 anatomically defined regions referred to as “Regions of Interest” (ROIs). There are  $n = 2800$  images, featuring 3 types of stimuli, and  $m = 18$ . Please refer to Figure 9 for fMRI series plots of Subject 1 at 6 different voxels, where the non-stationarity pattern is visually evident. Figure 10 compares the active regions (in purple) identified by the “parametric OLS” method and the proposed “semi-parametric data-driven” method  $\mathbb{W}(\hat{\Sigma}_{\hat{b}; \hat{g}(\hat{b})}; \hat{b})$ . As clearly observed, the “semi-parametric data-driven” method outperforms the “parametric OLS” method by effectively capturing significant regions, including the ROIs. In contrast, the latter fails to detect a majority of the ROIs. Results from other subjects are similar but omitted due to space limitations.

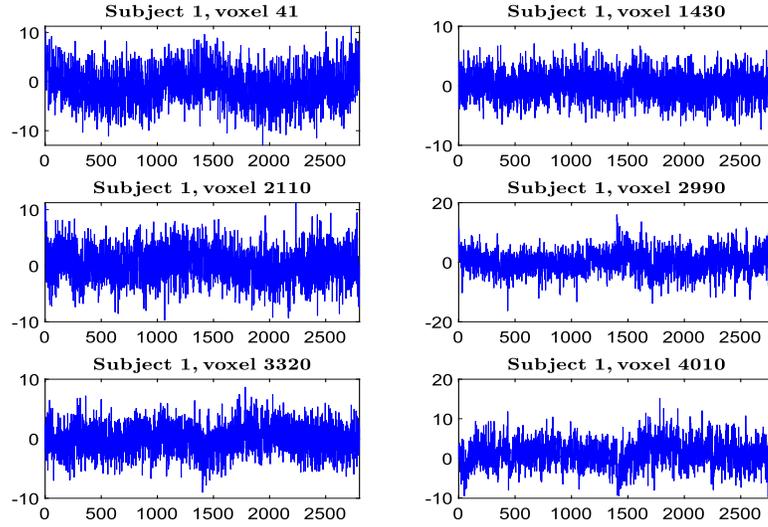


FIG 9. (*Real fMRI data*) Subject 1, the number of scanned voxels = 4949. Plots of fMRI time series at 6 voxels.

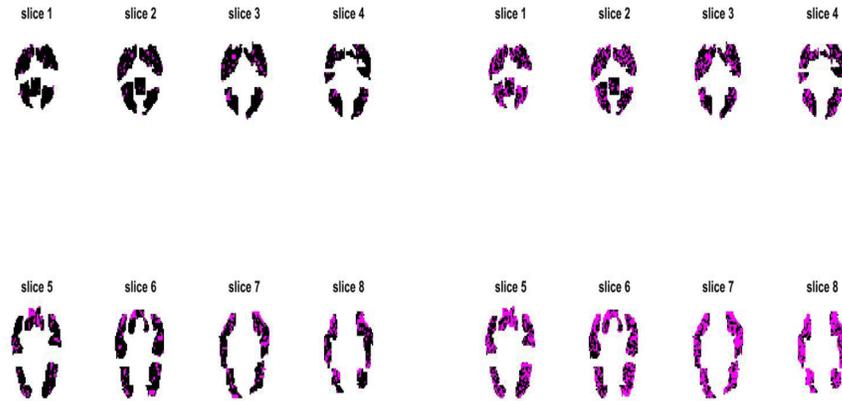


FIG 10. (*Real fMRI data*) Subject 1, the number of scanned voxels = 4949. Left 8 slices: *detected* active regions by the parametric OLS; right 8 slices: *detected* active regions by  $\mathbb{W}(\hat{\Sigma}_{\hat{b}, \hat{g}(\hat{b})}; \hat{b})$ .

## 8. Discussion

As high-throughput technology advances, statistical inference methods naturally emerge from the analysis of large-scale structured datasets. These datasets are characterized not only by high dimensionality but also by non-stationarity and temporal dependence in various forms. The development of new tools for stochastic modeling, computational algorithms, and statistical inference proce-

dures applied to these large-scale data is highly desirable. These tools enable scientists to efficiently analyze data with significantly increased flexibility and accuracy.

This paper proposed a new method for performing semi-parametric inference on large-scale imaging data, inspired by the analysis of fMRI data, where the underlying noise process exhibits non-stationary, non-Gaussian behavior, and temporal dependence. Our method departs from existing approaches that rely on the assumption of noise process stationarity. We propose a semi-parametric test statistic using a tapering-based estimator for the noise auto-covariance matrix and illustrate its advantages over current methods. On the theoretical side, our method relaxes the requirement for the noise covariance matrix estimator to be consistent. On the numerical side, our method avoids the direct inversion of the  $n \times n$  covariance matrix, preserving efficiency. It can be applied to both stationary and a broader range of non-stationary, non-Gaussian noise processes, making it particularly effective in addressing challenges posed by high-dimensional data.

In conclusion, we discuss potential directions for future research. First, the class of noise processes in Section 2 did not cover all possible cases, such as the stationary  $g_n$ -dependent process [15] and the long-range dependent  $1/f$ -like non-stationary noise process [26], which were not part of our assumptions. Investigating the feasibility of extensions on a case-by-case basis and addressing associated technical details present interesting avenues for further study. Second, our method for fitting the semi-parametric model estimates the parameter vector without imposing any prior structural knowledge. It would be valuable to explore how incorporating structural information, such as sparsity, grouping, and constraints, impacts parameter learning and related statistical inference. Third, conducting a comprehensive comparison between our proposed method and conventional approaches, including parametric and semi-parametric models from statistical, econometrics, and neuroimaging literature, would be beneficial for assessing its effectiveness on fMRI data.

## Appendix A: Proofs of main results

For Lemma 1 and Lemmas 2–6, which will be needed in the proofs of the main results, we present their proofs below.

*Proof of Lemma 1.* For any  $i = 1, \dots, n$ ,

$$\varepsilon(t_i) = \sum_{k=0}^{\infty} [\mathbb{E}\{\varepsilon(t_i) \mid e_{i-k}, \dots, e_{i+1}\} - \mathbb{E}\{\varepsilon(t_i) \mid e_{i-k+1}, \dots, e_{i+1}\}]. \quad (\text{A.1})$$

From (2.3),

$$\begin{aligned} \mathbb{E}\{\varepsilon(t_i) \mid e_{i-k+1}, \dots, e_{i+1}\} &= \mathbb{E}\{G(t_i; \mathcal{F}_i) \mid e_{i-k+1}, \dots, e_{i+1}\} \\ &= \mathbb{E}\{G(t_i; \mathcal{F}'_{i;i-k}) \mid e_{i-k+1}, \dots, e_{i+1}\}. \end{aligned}$$

Then, for any  $k = 0, 1, \dots$ ,

$$\begin{aligned}
 & \mathbb{E}[|\mathbb{E}\{\varepsilon(t_i) \mid e_{i-k}, \dots, e_{i+1}\} - \mathbb{E}\{\varepsilon(t_i) \mid e_{i-k+1}, \dots, e_{i+1}\}|^\omega] \\
 &= \mathbb{E}[|\mathbb{E}\{G(t_i; \mathcal{F}_i) \mid e_{i-k}, \dots, e_{i+1}\} - \mathbb{E}\{G(t_i; \mathcal{F}'_{i;i-k}) \mid e_{i-k+1}, \dots, e_{i+1}\}|^\omega] \\
 &= \mathbb{E}[|\mathbb{E}\{G(t_i; \mathcal{F}_i) \mid e_{i-k}, \dots, e_{i+1}\} - \mathbb{E}\{G(t_i; \mathcal{F}'_{i;i-k}) \mid e_{i-k}, \dots, e_{i+1}\}|^\omega] \\
 &= \mathbb{E}[|\mathbb{E}\{G(t_i; \mathcal{F}_i) - G(t_i; \mathcal{F}'_{i;i-k}) \mid e_{i-k}, \dots, e_{i+1}\}|^\omega] \\
 &\leq \mathbb{E}[\mathbb{E}\{|G(t_i; \mathcal{F}_i) - G(t_i; \mathcal{F}'_{i;i-k})|^\omega \mid e_{i-k}, \dots, e_{i+1}\}] \\
 &= \mathbb{E}\{|G(t_i; \mathcal{F}_i) - G(t_i; \mathcal{F}'_{i;i-k})|^\omega\} \leq \Delta_{k,\omega}^\omega.
 \end{aligned} \tag{A.2}$$

The first inequality in (A.2) is because of Jensen's inequality. From (A.1), Minkowski's inequality and (A.2), it follows that

$$\begin{aligned}
 \|\varepsilon(t_i)\|_\omega &= \left\| \sum_{k=0}^\infty [\mathbb{E}\{\varepsilon(t_i) \mid e_{i-k}, \dots, e_{i+1}\} - \mathbb{E}\{\varepsilon(t_i) \mid e_{i-k+1}, \dots, e_{i+1}\}] \right\|_\omega \\
 &\leq \sum_{k=0}^\infty \|\mathbb{E}\{\varepsilon(t_i) \mid e_{i-k}, \dots, e_{i+1}\} - \mathbb{E}\{\varepsilon(t_i) \mid e_{i-k+1}, \dots, e_{i+1}\}\|_\omega \\
 &\leq \sum_{k=0}^\infty \Delta_{k,\omega} = \Theta_{0,\omega}.
 \end{aligned}$$

Next, we will show  $\sup_{i=1, \dots, n-q} |\mathbb{E}\{\varepsilon(t_{i+q})\varepsilon(t_i)\}| \leq C\Theta_{q,\omega}$  for any  $q \in \{1, \dots, n-1\}$ . From (A.1),

$$\begin{aligned}
 \mathbb{E}\{\varepsilon(t_{i+q})\varepsilon(t_i)\} &= \mathbb{E}\left\{ \left( \sum_{k=0}^\infty [\mathbb{E}\{\varepsilon(t_{i+q}) \mid e_{i+q-k}, \dots, e_{i+q+1}\} - \mathbb{E}\{\varepsilon(t_{i+q}) \mid e_{i+q-k+1}, \dots, e_{i+q+1}\}] \right) \varepsilon(t_i) \right\} \\
 &= \mathbb{E}\left\{ \left( \sum_{k=0}^{q-1} [\mathbb{E}\{\varepsilon(t_{i+q}) \mid e_{i+q-k}, \dots, e_{i+q+1}\} - \mathbb{E}\{\varepsilon(t_{i+q}) \mid e_{i+q-k+1}, \dots, e_{i+q+1}\}] \right) \varepsilon(t_i) \right\} \\
 &\quad + \mathbb{E}\left\{ \left( \sum_{k=q}^\infty [\mathbb{E}\{\varepsilon(t_{i+q}) \mid e_{i+q-k}, \dots, e_{i+q+1}\} - \mathbb{E}\{\varepsilon(t_{i+q}) \mid e_{i+q-k+1}, \dots, e_{i+q+1}\}] \right) \varepsilon(t_i) \right\} \\
 &= \mathbb{E}\left\{ \left( \sum_{k=q}^\infty [\mathbb{E}\{\varepsilon(t_{i+q}) \mid e_{i+q-k}, \dots, e_{i+q+1}\} - \mathbb{E}\{\varepsilon(t_{i+q}) \mid e_{i+q-k+1}, \dots, e_{i+q+1}\}] \right) \varepsilon(t_i) \right\}
 \end{aligned}$$

which, together with the Cauchy-Schwartz inequality, Liapounov's inequality,

Minkowski's inequality, and (A.2), implies

$$\begin{aligned}
 |\mathbb{E}\{\varepsilon(t_{i+q})\varepsilon(t_i)\}| &\leq \left\| \sum_{k=q}^{\infty} [\mathbb{E}\{\varepsilon(t_{i+q}) \mid e_{i+q-k}, \dots, e_{i+q+1}\} \right. \\
 &\quad \left. - \mathbb{E}\{\varepsilon(t_{i+q}) \mid e_{i+q-k+1}, \dots, e_{i+q+1}\}] \right\|_2 \|\varepsilon(t_i)\|_2 \\
 &\leq \left\| \sum_{k=q}^{\infty} [\mathbb{E}\{\varepsilon(t_{i+q}) \mid e_{i+q-k}, \dots, e_{i+q+1}\} \right. \\
 &\quad \left. - \mathbb{E}\{\varepsilon(t_{i+q}) \mid e_{i+q-k+1}, \dots, e_{i+q+1}\}] \right\|_{\omega} \|\varepsilon(t_i)\|_{\omega} \\
 &\leq C \sum_{k=q}^{\infty} \|\mathbb{E}\{\varepsilon(t_{i+q}) \mid e_{i+q-k}, \dots, e_{i+q+1}\} \\
 &\quad - \mathbb{E}\{\varepsilon(t_{i+q}) \mid e_{i+q-k+1}, \dots, e_{i+q+1}\}\|_{\omega} \\
 &\leq C \sum_{k=q}^{\infty} \Delta_{k,\omega} = C\Theta_{q,\omega}.
 \end{aligned}$$

This completes the proof. □

**Lemma 2.** For  $\varepsilon(t_i)$  defined in (2.3) and  $S_b$  defined in (3.2), under Conditions A2, A3, A5 and A6, then

$$\sup_{i \in \{1, \dots, n\}} \mathbb{E}\left\{ \left| \sum_{j=1}^n \varepsilon(t_j) S_b(i, j) \right|^4 \right\} \leq C/(nb)^2.$$

*Proof.* By Lemma A1 of [45], for any  $i = 1, \dots, n$ ,

$$\mathbb{E}\left\{ \left| \sum_{j=1}^n \varepsilon(t_j) S_b(i, j) \right|^4 \right\} \leq C \left[ \sum_{j=1}^n \{S_b(i, j)\}^2 \right]^2.$$

Corollary A.1 of [46] indicates that  $nb|S_b(i, j)| \leq C$  for any  $i, j = 1, \dots, n$ ,  $\|S_b\|_1 \leq C$  and  $\|S_b\|_{\infty} \leq C$ . Then,

$$\left[ \sum_{j=1}^n \{S_b(i, j)\}^2 \right]^2 \leq \left\{ \frac{C}{nb} \sum_{j=1}^n |S_b(i, j)| \right\}^2 \leq C/(nb)^2.$$

The proof is finished. □

**Lemma 3.** Assume Conditions A2, A4, A5 and A6 hold. Denote  $\tilde{\mathbf{S}}_i = (\mathbf{I}_n - S_b)\mathbf{S}_i$  for  $i = 1, \dots, \ell$ . Let  $A = \{a_{i,j}\}_{i,j=1}^n$  be an  $n \times n$  matrix.

- (i) If  $A$  is random (or deterministic) and independent of  $\mathbf{S}$  and  $\max_{1 \leq i, j \leq n} \mathbb{E}(a_{i,j}^2) \leq C$ , then, for any  $j_1, j_2 = 1, \dots, \ell$  and  $u_1, u_2 = 1, \dots, m$ ,

$$\mathbb{E}\{(n^{-1} \mathbf{e}_{u_1; m}^T \tilde{\mathbf{S}}_{j_1}^T A \tilde{\mathbf{S}}_{j_2} \mathbf{e}_{u_2; m})^2\} = O(1);$$

additionally, if  $\max_{nbL \leq i \leq n-nbL, 1 \leq j \leq n} E(a_{i,j}^2) = o(1)$ , then

$$E\{(n^{-1} \mathbf{e}_{u_1;m}^T \tilde{\mathbf{S}}_{j_1}^T A \tilde{\mathbf{S}}_{j_2} \mathbf{e}_{u_2;m})^2\} = o(1).$$

(ii) If  $A$  is deterministic,  $\max_{1 \leq i,j \leq n} |a_{i,j}| \leq C$  and  $\|A\|_\infty = o(n)$ , then, for any  $j_1, j_2 = 1, \dots, \ell$  and  $u_1, u_2 = 1, \dots, m$ ,

$$\text{var}(n^{-1} \mathbf{e}_{u_1;m}^T \tilde{\mathbf{S}}_{j_1}^T A \tilde{\mathbf{S}}_{j_2} \mathbf{e}_{u_2;m}) = o(1).$$

*Proof.* Let  $\check{A} = \{\check{a}_{i,j}\}_{i,j=1}^n = (\mathbf{I}_n - S_b)^T A (\mathbf{I}_n - S_b)$ , and then

$$n^{-1} \mathbf{e}_{u_1;m}^T \tilde{\mathbf{S}}_{j_1}^T A \tilde{\mathbf{S}}_{j_2} \mathbf{e}_{u_2;m} = n^{-1} \mathbf{e}_{u_1;m}^T \mathbf{S}_{j_1}^T \check{A} \mathbf{S}_{j_2} \mathbf{e}_{u_2;m}.$$

For notational simplicity, denote  $Q = \{q_{i,j}\}_{i,j=1}^n = \mathbf{I}_n - S_b$ . Then,  $\check{a}_{i,j} = \sum_{v_1=1}^n \sum_{v_2=1}^n q_{v_1,i} a_{v_1,v_2} q_{v_2,j}$ . It follows that,

$$\begin{aligned} & n^{-1} \mathbf{e}_{u_1;m}^T \tilde{\mathbf{S}}_{j_1}^T A \tilde{\mathbf{S}}_{j_2} \mathbf{e}_{u_2;m} = n^{-1} \mathbf{e}_{u_1;m}^T \mathbf{S}_{j_1}^T \check{A} \mathbf{S}_{j_2} \mathbf{e}_{u_2;m} \\ &= n^{-1} \sum_{i=u_1}^n \sum_{k=u_2}^n s_{j_1}(t_{i-u_1}) s_{j_2}(t_{k-u_2}) \check{a}_{i,k} \\ &= n^{-1} \sum_{i=u_1}^n \sum_{k=u_2}^n \{s_{j_1}(t_{i-u_1}) - p_{j_1}\} \{s_{j_2}(t_{k-u_2}) - p_{j_2}\} \check{a}_{i,k} \\ &\quad + n^{-1} \sum_{i=u_1}^n \sum_{k=u_2}^n p_{j_1} p_{j_2} \check{a}_{i,k} \\ &\quad + n^{-1} \sum_{i=u_1}^n \sum_{k=u_2}^n p_{j_1} \{s_{j_2}(t_{k-u_2}) - p_{j_2}\} \check{a}_{i,k} \\ &\quad + n^{-1} \sum_{i=u_1}^n \sum_{k=u_2}^n \{s_{j_1}(t_{i-u_1}) - p_{j_1}\} p_{j_2} \check{a}_{i,k} \\ &= \text{I} + \text{II} + \text{III} + \text{IV}. \end{aligned} \tag{A.3}$$

For part (i), we will demonstrate  $E(\text{II}^2) = o(1)$ ,  $E(\text{III}^2) = o(1)$ ,  $E(\text{IV}^2) = o(1)$  and  $E(\text{I}^2) = O(1)$  under condition  $\max_{1 \leq i,j \leq n} E(a_{i,j}^2) \leq C$ .

Note that  $C$  is a finite constant which may vary from place to place. For term II, under condition  $\max_{1 \leq i,j \leq n} E(a_{i,j}^2) \leq C$ , with proper rearrangement of the embedded sums, we have,

$$\begin{aligned} & E(\text{II}^2) \\ &= n^{-2} p_{j_1}^2 p_{j_2}^2 \sum_{i=u_1}^n \sum_{k=u_2}^n \sum_{i'=u_1}^n \sum_{k'=u_2}^n E(\check{a}_{i,k} \check{a}_{i',k'}) \\ &= n^{-2} C \sum_{i=u_1}^n \sum_{k=u_2}^n \sum_{i'=u_1}^n \sum_{k'=u_2}^n \sum_{v_1=1}^n \sum_{v_2=1}^n \sum_{v_3=1}^n \sum_{v_4=1}^n \end{aligned}$$

$$\begin{aligned}
& \mathbb{E}(q_{v_1,i} a_{v_1,v_2} q_{v_2,k} q_{v_3,i'} a_{v_3,v_4} q_{v_4,k'}) \\
&= n^{-2} C \sum_{v_1=1}^n \sum_{v_2=1}^n \sum_{v_3=1}^n \sum_{v_4=1}^n \left\{ \left( \sum_{i=u_1}^n q_{v_1,i} \right) \left( \sum_{k=u_2}^n q_{v_2,k} \right) \left( \sum_{i'=u_1}^n q_{v_3,i'} \right) \right. \\
&\quad \left. \left( \sum_{k'=u_2}^n q_{v_4,k'} \right) \mathbb{E}(a_{v_1,v_2} a_{v_3,v_4}) \right\} \\
&\leq n^{-2} C \sum_{v_1=1}^n \sum_{v_2=1}^n \sum_{v_3=1}^n \sum_{v_4=1}^n \left( \left\| \sum_{i=u_1}^n q_{v_1,i} \right\| \left\| \sum_{k=u_2}^n q_{v_2,k} \right\| \right. \\
&\quad \left. \sum_{i'=u_1}^n q_{v_3,i'} \left\| \sum_{k'=u_2}^n q_{v_4,k'} \right\| \right) \max_{1 \leq i,j \leq n} \mathbb{E}(a_{i,j}^2) \\
&\leq n^{-2} C \left( \sum_{v_1=1}^n \left| \sum_{i=u_1}^n q_{v_1,i} \right| \right) \left( \sum_{v_2=1}^n \left| \sum_{k=u_2}^n q_{v_2,k} \right| \right) \\
&\quad \left( \sum_{v_3=1}^n \left| \sum_{i'=u_1}^n q_{v_3,i'} \right| \right) \left( \sum_{v_4=1}^n \left| \sum_{k'=u_2}^n q_{v_4,k'} \right| \right) \\
&= n^{-2} C \left\| (\mathbf{I}_n - S_b)(\mathbf{0}_{u_1-1}^T, \mathbf{1}_{n-u_1+1}^T)^T \right\|_1^2 \\
&\quad \left\| (\mathbf{I}_n - S_b)(\mathbf{0}_{u_2-1}^T, \mathbf{1}_{n-u_2+1}^T)^T \right\|_1^2. \tag{A.4}
\end{aligned}$$

The first inequality above is because of the triangular inequality and Cauchy-Schwartz inequality. The last step is due to  $(\sum_{i=u_1}^n q_{1,i}, \dots, \sum_{i=u_1}^n q_{n,i})^T = (\mathbf{I}_n - S_b)(\mathbf{0}_{u_1-1}^T, \mathbf{1}_{n-u_1+1}^T)^T$ , from the definition  $\{q_{j,i}\}_{j,i=1}^n = \mathbf{I}_n - S_b$ . From (3.2),

$$\begin{aligned}
\sum_{j=1}^n S_b(i, j) &= (1, 0) \{ \mathbf{X}(t_i)^T \mathbf{W}(t_i) \mathbf{X}(t_i) \}^{-1} \mathbf{X}(t_i)^T \mathbf{W}(t_i) \mathbf{1}_n \\
&= (1, 0) \{ \mathbf{X}(t_i)^T \mathbf{W}(t_i) \mathbf{X}(t_i) \}^{-1} \mathbf{X}(t_i)^T \mathbf{W}(t_i) \mathbf{X}(t_i) (1, 0)^T = 1.
\end{aligned}$$

Therefore,  $(\mathbf{I}_n - S_b) \mathbf{1}_n = \mathbf{0}_n$ . Since  $\|\mathbf{I}_n - S_b\|_1 \leq C$  (see Corollary A.1 of [46]),  $\|(\mathbf{I}_n - S_b)(\mathbf{0}_{u_1-1}^T, \mathbf{1}_{n-u_1+1}^T)^T\|_1 = \|(\mathbf{I}_n - S_b)(\mathbf{1}_{u_1-1}^T, \mathbf{0}_{n-u_1+1}^T)^T\|_1 \leq (u_1 - 1) \|\mathbf{I}_n - S_b\|_1 \leq C$ . Similarly,  $\|(\mathbf{I}_n - S_b)(\mathbf{0}_{u_2-1}^T, \mathbf{1}_{n-u_2+1}^T)^T\|_1 \leq C$ . From (A.4), we have  $\mathbb{E}(\text{II}^2) = O(n^{-2})$ .

For term III, noting that  $\mathbb{E}[\{s_{j_2}(t_{k-u_2}) - p_{j_2}\} \{s_{j_2}(t_{k'-u_2}) - p_{j_2}\}] \neq 0$  if and only if  $k = k'$ , we have

$$\begin{aligned}
& \mathbb{E}(\text{III}^2) \\
&= n^{-2} p_{j_1}^2 \sum_{i=u_1}^n \sum_{k=u_2}^n \sum_{i'=u_1}^n \sum_{k'=u_2}^n \mathbb{E}[\{s_{j_2}(t_{k-u_2}) - p_{j_2}\} \{s_{j_2}(t_{k'-u_2}) - p_{j_2}\}] \mathbb{E}(\check{a}_{i,k} \check{a}_{i',k'}) \\
&= n^{-2} p_{j_1}^2 \mathbb{E}[\{s_{j_2}(t_1) - p_{j_2}\}^2] \sum_{i=u_1}^n \sum_{k=u_2}^n \sum_{i'=u_1}^n \mathbb{E}(\check{a}_{i,k} \check{a}_{i',k}) \\
&= n^{-2} C \sum_{i=u_1}^n \sum_{k=u_2}^n \sum_{i'=u_1}^n \sum_{v_1=1}^n \sum_{v_2=1}^n \sum_{v_3=1}^n \sum_{v_4=1}^n \mathbb{E}(q_{v_1,i} a_{v_1,v_2} q_{v_2,k} q_{v_3,i'} a_{v_3,v_4} q_{v_4,k})
\end{aligned}$$

$$\begin{aligned}
&= n^{-2} C \sum_{k=u_2}^n \sum_{v_1=1}^n \sum_{v_3=1}^n \left[ \left( \sum_{i=u_1}^n q_{v_1,i} \right) \left( \sum_{i'=u_1}^n q_{v_3,i'} \right) \right. \\
&\quad \left. \left\{ \sum_{v_2=1}^n \sum_{v_4=1}^n q_{v_2,k} q_{v_4,k} \mathbf{E}(a_{v_1,v_2} a_{v_3,v_4}) \right\} \right] \\
&\leq n^{-2} C \sum_{k=u_2}^n \sum_{v_1=1}^n \sum_{v_3=1}^n \left( \left\| \sum_{i=u_1}^n q_{v_1,i} \right\| \left\| \sum_{i'=u_1}^n q_{v_3,i'} \right\| \sum_{v_2=1}^n |q_{v_2,k}| \sum_{v_4=1}^n |q_{v_4,k}| \right) \\
&\quad \cdot \max_{1 \leq i,j \leq n} \mathbf{E}(a_{i,j}^2) \\
&\leq n^{-2} C \sum_{k=u_2}^n \left\{ \left( \sum_{v_1=1}^n \left\| \sum_{i=u_1}^n q_{v_1,i} \right\| \right) \left( \sum_{v_3=1}^n \left\| \sum_{i'=u_1}^n q_{v_3,i'} \right\| \right) \right\} \\
&\leq n^{-2} C \sum_{k=u_2}^n \|(\mathbf{I}_n - S_b)(\mathbf{0}_{u_1-1}^T, \mathbf{1}_{n-u_1+1}^T)^T\|_1^2 = O(n^{-1}). \tag{A.5}
\end{aligned}$$

Similar arguments imply that  $\mathbf{E}(\mathbf{IV}^2) = O(n^{-1})$ .

Next, we will prove  $\mathbf{E}(\mathbf{I}^2) = O(1)$ . Note

$$\begin{aligned}
\mathbf{E}(\mathbf{I}^2) &= n^{-2} \sum_{i=u_1}^n \sum_{k=u_2}^n \sum_{i'=u_1}^n \sum_{k'=u_2}^n \mathbf{E}[\{s_{j_1}(t_{i-u_1}) - p_{j_1}\} \{s_{j_2}(t_{k-u_2}) - p_{j_2}\} \\
&\quad \{s_{j_1}(t_{i'-u_1}) - p_{j_1}\} \{s_{j_2}(t_{k'-u_2}) - p_{j_2}\}] \mathbf{E}(\check{a}_{i,k} \check{a}_{i',k'}).
\end{aligned}$$

Under Condition A4, the term expressed as  $\mathbf{E}[\{s_{j_1}(t_{i-u_1}) - p_{j_1}\} \{s_{j_2}(t_{k-u_2}) - p_{j_2}\} \{s_{j_1}(t_{i'-u_1}) - p_{j_1}\} \{s_{j_2}(t_{k'-u_2}) - p_{j_2}\}]$  is nonzero in the following four cases:

$$\begin{aligned}
\mathbf{I}_1 : & \quad i - u_1 = k - u_2 = i' - u_1 = k' - u_2; \\
\mathbf{I}_2 : & \quad \{i - u_1 = k - u_2\} \neq \{i' - u_1 = k' - u_2\}; \\
\mathbf{I}_3 : & \quad \{i - u_1 = i' - u_1\} \neq \{k - u_2 = k' - u_2\}; \\
\mathbf{I}_4 : & \quad \{i - u_1 = k' - u_2\} \neq \{k - u_2 = i' - u_1\}.
\end{aligned}$$

Hence,

$$\mathbf{E}(\mathbf{I}^2) = \mathbf{E}_{\mathbf{I}_1} + \mathbf{E}_{\mathbf{I}_2} + \mathbf{E}_{\mathbf{I}_3} + \mathbf{E}_{\mathbf{I}_4}, \tag{A.6}$$

where

$$\begin{aligned}
\mathbf{E}_{\mathbf{I}_1} &= n^{-2} \sum_{i=u_1}^n \mathbf{E}[\{s_{j_1}(t_{i-u_1}) - p_{j_1}\}^2 \{s_{j_2}(t_{i-u_1}) - p_{j_2}\}^2] \mathbf{E}(\check{a}_{i,i-u_1+u_2}^2), \\
\mathbf{E}_{\mathbf{I}_2} &= n^{-2} \sum_{i=u_1}^n \sum_{i'=u_1, i' \neq i}^n \mathbf{E}[\{s_{j_1}(t_{i-u_1}) - p_{j_1}\} \{s_{j_2}(t_{i-u_1}) - p_{j_2}\} \\
&\quad \{s_{j_1}(t_{i'-u_1}) - p_{j_1}\} \{s_{j_2}(t_{i'-u_1}) - p_{j_2}\}] \mathbf{E}(\check{a}_{i,i-u_1+u_2} \check{a}_{i',i'-u_1+u_2}), \\
\mathbf{E}_{\mathbf{I}_3} &= n^{-2} \sum_{i=u_1}^n \sum_{k=u_2, k \neq i-u_1+u_2}^n \mathbf{E}[\{s_{j_1}(t_{i-u_1}) - p_{j_1}\}^2 \{s_{j_2}(t_{k-u_2}) - p_{j_2}\}^2] \mathbf{E}(\check{a}_{i,k}^2),
\end{aligned}$$

$$E_{I_4} = n^{-2} \sum_{i=u_1}^n \sum_{k=u_2, k \neq i-u_1+u_2}^n E[\{s_{j_1}(t_{i-u_1}) - p_{j_1}\} \{s_{j_2}(t_{k-u_2}) - p_{j_2}\} \{s_{j_1}(t_{k-u_2}) - p_{j_1}\} \{s_{j_2}(t_{i-u_1}) - p_{j_2}\}] E(\check{a}_{i,k} \check{a}_{k-u_2+u_1, i-u_1+u_2}).$$

For  $E_{I_2}$ ,

$$\begin{aligned} & |E_{I_2}| \\ & \leq Cn^{-2} \sum_{i=u_1}^n \sum_{i'=u_1}^n E(|\check{a}_{i, i-u_1+u_2} \check{a}_{i', i'-u_1+u_2}|) \\ & = Cn^{-2} \sum_{i=u_1}^n \sum_{i'=u_1}^n E\left(\left| \sum_{v_1=1}^n \sum_{v_2=1}^n \sum_{v_3=1}^n \sum_{v_4=1}^n q_{v_1, i} a_{v_1, v_2} q_{v_2, i-u_1+u_2} q_{v_3, i'} a_{v_3, v_4} q_{v_4, i'-u_1+u_2} \right|\right) \\ & \leq Cn^{-2} \sum_{i=u_1}^n \sum_{i'=u_1}^n \sum_{v_1=1}^n \sum_{v_2=1}^n \sum_{v_3=1}^n \sum_{v_4=1}^n |q_{v_1, i} q_{v_2, i-u_1+u_2} q_{v_3, i'} q_{v_4, i'-u_1+u_2}| \\ & \quad \cdot \max_{1 \leq i, j \leq n} E(a_{i, j}^2) \\ & \leq Cn^{-2} \sum_{i=u_1}^n \sum_{i'=u_1}^n \left\{ \left( \sum_{v_1=1}^n |q_{v_1, i}| \right) \left( \sum_{v_2=1}^n |q_{v_2, i-u_1+u_2}| \right) \right. \\ & \quad \left. \left( \sum_{v_3=1}^n |q_{v_3, i'}| \right) \left( \sum_{v_4=1}^n |q_{v_4, i'-u_1+u_2}| \right) \right\} \\ & \leq Cn^{-2} \sum_{i=u_1}^n \sum_{i'=u_1}^n C = O(1). \end{aligned}$$

Similarly, we can show that  $E_{I_1} = O(1)$ ,  $E_{I_3} = O(1)$ , and  $E_{I_4} = O(1)$ . Consequently,  $E(I^2) = O(1)$ . Hence, from (A.3), we conclude the proof for  $E\{(n^{-1} \mathbf{e}_{u_1; m}^T \tilde{\mathbf{S}}_{j_1}^T \tilde{\mathbf{A}} \tilde{\mathbf{S}}_{j_2}^T \mathbf{e}_{u_2; m})^2\} = O(1)$ .

Additionally, if  $\max_{nbL \leq i \leq n-nbL, 1 \leq j \leq n} E(a_{i, j}^2) = o(1)$ , then we will show that  $E(I^2) = o(1)$ . From (A.6),

$$\begin{aligned} & |E_{I_2}| \\ & \leq n^{-2} C \sum_{i=u_1}^n \sum_{i'=u_1}^n \sum_{v_1=1}^n \sum_{v_2=1}^n \sum_{v_3=1}^n \sum_{v_4=1}^n |q_{v_1, i} q_{v_2, i-u_1+u_2} q_{v_3, i'} q_{v_4, i'-u_1+u_2}| \\ & \quad |E(a_{v_1, v_2} a_{v_3, v_4})| \\ & \leq n^{-2} C \sum_{i=u_1}^n \sum_{i'=u_1}^n \sum_{v_1 \in M}^n \sum_{v_2=1}^n \sum_{v_3=1}^n \sum_{v_4=1}^n |q_{v_1, i} q_{v_2, i-u_1+u_2} q_{v_3, i'} q_{v_4, i'-u_1+u_2}| \\ & \quad \{E(a_{v_1, v_2}^2) E(a_{v_3, v_4}^2)\}^{1/2} \\ & \quad + n^{-2} C \sum_{i=u_1}^n \sum_{i'=u_1}^n \sum_{v_1 \in \bar{M}}^n \sum_{v_2=1}^n \sum_{v_3=1}^n \sum_{v_4=1}^n |q_{v_1, i} q_{v_2, i-u_1+u_2} q_{v_3, i'} q_{v_4, i'-u_1+u_2}| \end{aligned}$$

$$\begin{aligned}
& \{E(a_{v_1, v_2}^2)E(a_{v_3, v_4}^2)\}^{1/2} \\
& \leq n^{-2}C \max_{v_1 \in M, v_2 \in \{1, \dots, n\}} \{E(a_{v_1, v_2}^2)\}^{1/2} \sum_{i=u_1}^n \sum_{i'=u_1}^n C + n^{-2}C \sum_{i'=u_1}^n \sum_{v_1 \in \bar{M}} C \\
& \leq o(1) + Cb = o(1),
\end{aligned}$$

where  $M = [nbL, n - nbL]$  and  $\bar{M} = [1, nbL] \cup (n - nbL, n]$ . Similarly, we can show  $E_{I_1} = o(1)$ ,  $E_{I_3} = o(1)$  and  $E_{I_4} = o(1)$  and thus  $E(I^2) = o(1)$ , which together with (A.3) completes the proof of part (i).

For part (ii), from (A.3),

$$\begin{aligned}
E(I) &= n^{-1} \sum_{i=u_1}^n \sum_{k=u_2}^n E[\{s_{j_1}(t_{i-u_1}) - p_{j_1}\} \{s_{j_2}(t_{k-u_2}) - p_{j_2}\}] \check{\alpha}_{i,k} \\
&= n^{-1} \sum_{i=u_1}^n E[\{s_{j_1}(t_{i-u_1}) - p_{j_1}\} \{s_{j_2}(t_{i-u_1}) - p_{j_2}\}] \check{\alpha}_{i, i-u_1+u_2}.
\end{aligned}$$

Then,  $\{E(I)\}^2 = E_{I_2} + n^{-2} \sum_{i=u_1}^n (E[\{s_{j_1}(t_{i-u_1}) - p_{j_1}\} \{s_{j_2}(t_{i-u_1}) - p_{j_2}\}])^2 \times \check{\alpha}_{i, i-u_1+u_2}^2 = E_{I_2} + E_{\Delta}$ , and hence,  $\text{var}(I) = E_{I_1} + E_{I_3} + E_{I_4} - E_{\Delta}$  following (A.6). Next, we will show  $E_{I_1}$ ,  $E_{I_3}$ ,  $E_{I_4}$  and  $E_{\Delta}$  are all  $o(1)$ . For  $E_{I_1}$ ,

$$\begin{aligned}
& E_{I_1} \\
& \leq n^{-2}C \sum_{i=u_1}^n \sum_{v_1=1}^n \sum_{v_2=1}^n \sum_{v_3=1}^n \sum_{v_4=1}^n |q_{v_1, i} q_{v_2, i-u_1+u_2} q_{v_3, i} q_{v_4, i-u_1+u_2} a_{v_1, v_2} a_{v_3, v_4}| \\
& \leq n^{-2}C \sum_{i=u_1}^n \sum_{v_1=1}^n \sum_{v_2=1}^n \sum_{v_3=1}^n \sum_{v_4=1}^n |q_{v_1, i} q_{v_2, i-u_1+u_2} q_{v_3, i} q_{v_4, i-u_1+u_2}| \\
& \leq n^{-2}C \sum_{i=u_1}^n C = O(n^{-1}).
\end{aligned}$$

We can show  $E_{\Delta} = o(1)$  by similar arguments. For  $E_{I_3}$ , assumption  $\|A\|_{\infty} = o(n)$  implies that

$$\begin{aligned}
& E_{I_3} \\
& \leq n^{-2}C \sum_{i=u_1}^n \sum_{k=u_2}^n \check{\alpha}_{i,k}^2 \\
& \leq n^{-2}C \sum_{i=u_1}^n \sum_{k=u_2}^n \sum_{v_1=1}^n \sum_{v_2=1}^n \sum_{v_3=1}^n \sum_{v_4=1}^n |q_{v_1, i} q_{v_2, k} q_{v_3, i} q_{v_4, k} a_{v_1, v_2} a_{v_3, v_4}| \\
& \leq n^{-2}C \sum_{i=u_1}^n \sum_{v_1=1}^n |q_{v_1, i}| \sum_{v_3=1}^n |q_{v_3, i}| \sum_{v_2=1}^n |a_{v_1, v_2}| \sum_{k=u_2}^n |q_{v_2, k}| \sum_{v_4=1}^n |q_{v_4, k} a_{v_3, v_4}| \\
& \leq n^{-2}C \sum_{i=u_1}^n Co(n) = o(1).
\end{aligned}$$

Similar arguments imply that  $E_{I_4} = o(1)$ . Hence,  $\text{var}(I) = o(1)$ . Following (A.4) and (A.5), we can show that  $E(II^2)$ ,  $E(III^2)$  and  $E(IV^2)$  are  $o(1)$ . We finish the proof from (A.3). This completes the proof.  $\square$

**Lemma 4.** *Let  $A = \{a_{i,j}\}_{i,j=1}^n$  be an  $n \times n$  random matrix independent of  $\varepsilon$ . Assume  $B_n = \max(\max_{1 \leq i \leq n} \|\sum_{j=1}^n a_{i,j}^2\|_{\omega/2}^{1/2}, \max_{1 \leq j \leq n} \|\sum_{i=1}^n a_{i,j}^2\|_{\omega/2}^{1/2}) < \infty$ . Under Condition A3, it follows that*

$$\left\| \sum_{i=1}^n \sum_{j=1}^n [a_{i,j} \varepsilon(t_i) \varepsilon(t_j) - a_{i,j} E\{\varepsilon(t_i) \varepsilon(t_j)\}] \right\|_{\omega/2} \leq C \Theta_{0,\omega}^2 B_n \sqrt{n}.$$

*Proof.* Let  $L_n = \sum_{i=1}^n \sum_{j=1}^n a_{i,j} \varepsilon(t_i) \varepsilon(t_j)$  and  $Z_i = \sum_{j=1}^n a_{i,j} \varepsilon(t_j)$ . Then,  $L_n = \sum_{i=1}^n \varepsilon(t_i) Z_i$  and  $L_n - E(L_n | A) = \sum_{k=-\infty}^n \mathcal{P}_k(L_n)$ , where  $\mathcal{P}_k(\cdot) = E(\cdot | A, \mathcal{F}_k) - E(\cdot | A, \mathcal{F}_{k-1})$ . Let  $\varepsilon'(t_i; k) = G(t_i; \mathcal{F}'_{i;k})$  and  $Z'_{i;k} = \sum_{j=1}^n a_{i,j} \varepsilon'(t_j; k)$ . Noting that  $E\{\varepsilon(t_i) Z_i | A, \mathcal{F}_{k-1}\} = E\{\varepsilon'(t_i; k) Z'_{i;k} | A, \mathcal{F}_{k-1}\} = E\{\varepsilon'(t_i; k) Z'_{i;k} | A, \mathcal{F}_k\}$ , we have

$$\begin{aligned} & \|\mathcal{P}_k(L_n)\|_{\omega/2} \\ &= \left\| \sum_{i=1}^n [E\{\varepsilon(t_i) Z_i | A, \mathcal{F}_k\} - E\{\varepsilon(t_i) Z_i | A, \mathcal{F}_{k-1}\}] \right\|_{\omega/2} \\ &= \left\| \sum_{i=1}^n E\{\varepsilon(t_i) Z_i - \varepsilon'(t_i; k) Z'_{i;k} | A, \mathcal{F}_k\} \right\|_{\omega/2} \\ &= \left\| E \left[ \sum_{i=1}^n \{\varepsilon(t_i) Z_i - \varepsilon'(t_i; k) Z'_{i;k}\} \middle| A, \mathcal{F}_k \right] \right\|_{\omega/2} \\ &\leq \left\{ E \left( E \left[ \left| \sum_{i=1}^n \{\varepsilon(t_i) Z_i - \varepsilon'(t_i; k) Z'_{i;k}\} \right|^{\omega/2} \middle| A, \mathcal{F}_k \right] \right) \right\}^{2/\omega} \\ &= \left\| \sum_{i=1}^n \{\varepsilon(t_i) Z_i - \varepsilon'(t_i; k) Z'_{i;k}\} \right\|_{\omega/2} \\ &= \left\| \sum_{i=1}^n \{\varepsilon(t_i) Z_i - \varepsilon(t_i) Z'_{i;k} + \varepsilon(t_i) Z'_{i;k} - \varepsilon'(t_i; k) Z'_{i;k}\} \right\|_{\omega/2} \\ &\leq \left\| \sum_{i=1}^n \varepsilon(t_i) (Z_i - Z'_{i;k}) \right\|_{\omega/2} + \left\| \sum_{i=1}^n \{\varepsilon(t_i) - \varepsilon'(t_i; k)\} Z'_{i;k} \right\|_{\omega/2} \\ &= \left\| \sum_{j=1}^n \left[ \{\varepsilon(t_j) - \varepsilon'(t_j; k)\} \sum_{i=1}^n a_{i,j} \varepsilon(t_i) \right] \right\|_{\omega/2} \\ &\quad + \left\| \sum_{i=1}^n \{\varepsilon(t_i) - \varepsilon'(t_i; k)\} Z'_{i;k} \right\|_{\omega/2} \\ &\leq \sum_{j=1}^n \left\| \{\varepsilon(t_j) - \varepsilon'(t_j; k)\} \sum_{i=1}^n a_{i,j} \varepsilon(t_i) \right\|_{\omega/2} \end{aligned}$$

$$\begin{aligned}
& + \sum_{i=1}^n \|\{\varepsilon(t_i) - \varepsilon'(t_i; k)\} Z'_{i;k}\|_{\omega/2} \\
& \leq \sum_{j=1}^n \|\varepsilon(t_j) - \varepsilon'(t_j; k)\|_{\omega} \left\| \sum_{i=1}^n a_{i,j} \varepsilon(t_i) \right\|_{\omega} \\
& \quad + \sum_{i=1}^n \|\varepsilon(t_i) - \varepsilon'(t_i; k)\|_{\omega} \|Z'_{i;k}\|_{\omega},
\end{aligned} \tag{A.7}$$

where the first inequality comes from Jensen's inequality. Lemma A1 in [45] implies that  $\|Z_i\|_{\omega} = [\mathbb{E}\{\mathbb{E}(|Z_i|^{\omega} | A)\}]^{1/\omega} \leq C \|\sum_{j=1}^n a_{i,j}^2\|_{\omega/2}^{1/2} \Theta_{0,\omega} \leq C \Theta_{0,\omega} B_n$ . Similar arguments imply that  $\|\sum_{i=1}^n a_{i,j} \varepsilon(t_i)\|_{\omega} \leq C \Theta_{0,\omega} B_n$  and  $\|Z'_{i;k}\|_{\omega} \leq C \Theta_{0,\omega} B_n$ . Hence,

$$\begin{aligned}
& \|L_n - \mathbb{E}(L_n | A)\|_{\omega/2}^2 \\
& = \left\| \sum_{k=-\infty}^n \mathcal{P}_k(L_n) \right\|_{\omega/2}^2 = \left[ \mathbb{E} \left\{ \lim_{h \rightarrow -\infty} \left| \sum_{k=h}^n \mathcal{P}_k(L_n) \right|^{\omega/2} \right\} \right]^{4/\omega} \\
& \leq \lim_{h \rightarrow -\infty} \left\| \sum_{k=h}^n \mathcal{P}_k(L_n) \right\|_{\omega/2}^2 \\
& \leq C \lim_{h \rightarrow -\infty} \sum_{k=h}^n \|\mathcal{P}_k(L_n)\|_{\omega/2}^2 = C \sum_{k=-\infty}^n \|\mathcal{P}_k(L_n)\|_{\omega/2}^2 \\
& \leq C \sum_{k=-\infty}^n \left\{ \sum_{j=1}^n \|\varepsilon(t_j) - \varepsilon'(t_j; k)\|_{\omega} \left\| \sum_{i=1}^n a_{i,j} \varepsilon(t_i) \right\|_{\omega} \right\}^2 \\
& \quad + C \sum_{k=-\infty}^n \left\{ \sum_{i=1}^n \|\varepsilon(t_i) - \varepsilon'(t_i; k)\|_{\omega} \|Z'_{i;k}\|_{\omega} \right\}^2 \\
& \leq C \Theta_{0,\omega}^2 B_n^2 \sum_{k=-\infty}^n \left( \sum_{j=\max(1,k)}^n \Delta_{j-k,\omega} \right)^2 \\
& \leq C \Theta_{0,\omega}^3 B_n^2 \sum_{k=-\infty}^n \sum_{j=\max(1,k)}^n \Delta_{j-k,\omega} \\
& = C \Theta_{0,\omega}^3 B_n^2 \sum_{j=1}^n \sum_{k=-\infty}^j \Delta_{j-k,\omega} \leq C \Theta_{0,\omega}^4 B_n^2 n,
\end{aligned}$$

where the first inequality comes from Fatou's lemma, the second inequality is due to Lemma 7 of [43] and the third inequality is from (A.7). This completes the proof.  $\square$

**Lemma 5.** Under Conditions A1–A6 and A8, if  $n^{-1} \tilde{\mathbf{S}}^T \tilde{\mathbf{S}} \xrightarrow{\mathbb{P}} M_1 \succ 0$  and  $n^{-1} \tilde{\mathbf{S}}^T \Sigma_n \tilde{\mathbf{S}} \xrightarrow{\mathbb{P}} M_2 \succ 0$ , then  $n^{1/2}(\hat{\mathbf{h}} - \mathbf{h}) \xrightarrow{\mathcal{D}} N(\mathbf{0}, M_1^{-1} M_2 M_1^{-1})$ .

*Proof.* Using expressions  $\hat{\mathbf{h}} = (\tilde{\mathbf{S}}^T \tilde{\mathbf{S}})^{-1} \tilde{\mathbf{S}}^T \tilde{\mathbf{Y}} = (\tilde{\mathbf{S}}^T \tilde{\mathbf{S}})^{-1} \tilde{\mathbf{S}}^T (\tilde{\mathbf{S}} \mathbf{h} + \tilde{\mathbf{d}} + \tilde{\varepsilon}) = \mathbf{h} + (\tilde{\mathbf{S}}^T \tilde{\mathbf{S}})^{-1} \tilde{\mathbf{S}}^T (\tilde{\mathbf{d}} + \tilde{\varepsilon})$ , we have

$$\begin{aligned} n^{1/2}(\hat{\mathbf{h}} - \mathbf{h}) &= (n^{-1} \tilde{\mathbf{S}}^T \tilde{\mathbf{S}})^{-1} \{n^{-1/2} \tilde{\mathbf{S}}^T (\tilde{\mathbf{d}} + \tilde{\varepsilon})\} \\ &= (n^{-1} \tilde{\mathbf{S}}^T \tilde{\mathbf{S}})^{-1} (n^{-1/2} \tilde{\mathbf{S}}^T \tilde{\mathbf{d}} + n^{-1/2} \tilde{\mathbf{S}}^T \varepsilon - n^{-1/2} \tilde{\mathbf{S}}^T S_b \varepsilon). \end{aligned}$$

To complete the proof, we will show that

$$n^{-1/2} \tilde{\mathbf{S}}^T \tilde{\mathbf{d}} = o_P(1), \quad n^{-1/2} \tilde{\mathbf{S}}^T \varepsilon \xrightarrow{D} N(\mathbf{0}_{\ell m}, M_2), \quad \text{and} \quad n^{-1/2} \tilde{\mathbf{S}}^T S_b \varepsilon = o_P(1).$$

By part (i) of Lemma 3, we can show  $n^{-1/2} \tilde{\mathbf{S}}^T \tilde{\mathbf{d}} = o_P(1)$  and  $n^{-1/2} \tilde{\mathbf{S}}^T S_b \varepsilon = o_P(1)$  using (A.15) of [46] and Lemma 2, respectively. Next, we only need to prove  $n^{-1/2} \tilde{\mathbf{S}}^T \varepsilon \xrightarrow{D} N(\mathbf{0}_{\ell m}, M_2)$ .

Let  $l_n = \lfloor \log(n) \rfloor$ ,  $a_n = \lfloor n^{1/2-\alpha} \rfloor$ , and  $w_n = \lfloor n/l_n - a_n \rfloor$ , where  $\alpha \in (0, 1/2)$  is a constant. Define  $\varepsilon^* = (\varepsilon^*(t_1), \dots, \varepsilon^*(t_n))^T$  with  $\varepsilon^*(t_i) = E\{\varepsilon(t_i) \mid e_{i-a_n}, \dots, e_i\}$ . Then, to show  $n^{-1/2} \tilde{\mathbf{S}}^T \varepsilon \xrightarrow{D} N(\mathbf{0}_{\ell m}, M_2)$  is equivalent to showing  $n^{-1/2} \tilde{\mathbf{S}}^T \varepsilon^* \xrightarrow{D} N(\mathbf{0}_{\ell m}, M_2)$  because

$$\text{var}\{n^{-1/2} e_{i;\ell m}^T \tilde{\mathbf{S}}^T (\varepsilon - \varepsilon^*)\} \leq CE(\|n^{-1/2} e_{i;\ell m}^T \tilde{\mathbf{S}}^T\|_2^2) \Theta_{a_n+1,2}^2 \leq C \Theta_{a_n+1,2}^2 = o(1)$$

for any  $i = 1, \dots, \ell m$ , by Lemma A1 in [45].

To demonstrate  $n^{-1/2} \tilde{\mathbf{S}}^T \varepsilon^* \xrightarrow{D} N(\mathbf{0}_{\ell m}, M_2)$ , it suffices to prove the result as follows:  $n^{-1/2} (n^{-1} \tilde{\mathbf{S}}^T \Sigma_n \tilde{\mathbf{S}})^{-1/2} \tilde{\mathbf{S}}^T \varepsilon^* \xrightarrow{D} N(\mathbf{0}_{\ell m}, \mathbf{I}_{\ell m})$ , which is equivalent to showing  $n^{-1/2} \mathbf{u}^T (n^{-1} \tilde{\mathbf{S}}^T \Sigma_n \tilde{\mathbf{S}})^{-1/2} \tilde{\mathbf{S}}^T \varepsilon^* \xrightarrow{D} N(0, 1)$  for any  $\mathbf{u} \in \mathbb{R}^{\ell m}$  with  $\|\mathbf{u}\|_2 = 1$ .

First, we will show that, given  $\mathbf{S} = \mathbf{s}$ , if  $\|(n^{-1} \tilde{\mathbf{s}}^T \Sigma_n \tilde{\mathbf{s}})^{-1/2}\|_2 \leq C < \infty$  for  $n$  large enough, where  $\tilde{\mathbf{s}} = (\mathbf{I}_n - S_b)\mathbf{s}$ , then

$$P\{n^{-1/2} \mathbf{u}^T (n^{-1} \tilde{\mathbf{s}}^T \Sigma_n \tilde{\mathbf{s}})^{-1/2} \tilde{\mathbf{S}}^T \varepsilon^* \leq x\} \rightarrow \Phi(x),$$

where  $\Phi(x)$  is the C.D.F. of  $N(0, 1)$ . Denote by  $\mathbf{v} = (v_1, \dots, v_n) = n^{-1/2} \mathbf{u}^T (n^{-1} \tilde{\mathbf{s}}^T \Sigma_n \tilde{\mathbf{s}})^{-1/2} \tilde{\mathbf{S}}^T$ . Define

$$\begin{aligned} \Xi_i &= \sum_{j=(i-1)(w_n+a_n)+1}^{(i-1)(w_n+a_n)+w_n} v_j \varepsilon^*(t_j), \quad i = 1, \dots, l_n, \\ \Xi'_i &= \begin{cases} \sum_{j=(i-1)(w_n+a_n)+w_n+1}^{i(w_n+a_n)} v_j \varepsilon^*(t_j), & i = 1, \dots, l_n - 1, \\ \sum_{j=(i-1)(w_n+a_n)+w_n+1}^n v_j \varepsilon^*(t_j), & i = l_n. \end{cases} \end{aligned}$$

Therefore,  $\sum_{i=1}^{l_n} \Xi_i + \sum_{i=1}^{l_n} \Xi'_i = \mathbf{v} \varepsilon^*$ . For  $n$  large enough,  $\{\Xi_i\}_{i=1}^{l_n}$  are independent and  $\{\Xi'_i\}_{i=1}^{l_n}$  are independent, since  $\varepsilon^*$  is  $a_n$ -dependent.

In the following, we will show  $\sum_{i=1}^{l_n} \Xi_i \xrightarrow{D} N(0, 1)$  and  $\sum_{i=1}^{l_n} \Xi'_i = o_P(1)$ . Since  $\|\mathbf{I}_n - S_b\|_1 \leq C$ ,  $\|\mathbf{I}_n - S_b\|_\infty \leq C$ ,  $\|(n^{-1}\tilde{\mathbf{s}}^T \Sigma_n \tilde{\mathbf{s}})^{-1/2}\|_2 \leq C$ , and  $\max_{1 \leq i \leq n, 1 \leq j \leq \ell m} |\mathbf{s}(i, j)| \leq C$ , we have  $\max_{j \in \{1, \dots, n\}} |v_j| = O(n^{-1/2})$ . Therefore,  $\text{var}(\sum_{i=1}^{l_n} \Xi'_i) = \sum_{i=1}^{l_n} \text{var}(\Xi'_i) = O(l_n a_n^2/n) = o(1)$ , which, together with  $E(\sum_{i=1}^{l_n} \Xi'_i) = 0$  implies  $\sum_{i=1}^{l_n} \Xi'_i = o_P(1)$ . By Lemma A1 in [45],  $\text{var}\{\mathbf{v}(\boldsymbol{\varepsilon} - \boldsymbol{\varepsilon}^*)\} \leq C\|\mathbf{v}\|_2^2 \Theta_{a_n+1,2}^2 \leq C\Theta_{a_n+1,2}^2 = o(1)$ . Then,  $\lim_{n \rightarrow \infty} \text{var}(\sum_{i=1}^{l_n} \Xi_i + \sum_{i=1}^{l_n} \Xi'_i) = \lim_{n \rightarrow \infty} \text{var}(\mathbf{v}\boldsymbol{\varepsilon}^*) = \text{var}(\mathbf{v}\boldsymbol{\varepsilon}) = 1$ , implying that  $\lim_{n \rightarrow \infty} \text{var}(\sum_{i=1}^{l_n} \Xi_i) = 1$ , since  $\text{var}(\sum_{i=1}^{l_n} \Xi'_i) = o(1)$ . From Lemma A1 of [45], the Liapounov condition follows

$$\begin{aligned} \sum_{i=1}^{l_n} E(|\Xi_i|^{2+c}) &\leq C \sum_{i=1}^{l_n} \left\{ \sum_{j=(i-1)(w_n+a_n)+1}^{(i-1)(w_n+a_n)+w_n} v_j^2 \right\}^{1+c/2} \Theta_{0,2+c}^{2+c} \\ &= O(l_n w_n^{1+c/2}/n^{1+c/2}) = o(1) \end{aligned}$$

for a small constant  $c > 0$ . Hence, by the central limit theorem,

$$\begin{aligned} &P\{n^{-1/2} \mathbf{u}^T (n^{-1} \tilde{\mathbf{S}}^T \Sigma_n \tilde{\mathbf{S}})^{-1/2} \tilde{\mathbf{S}}^T \boldsymbol{\varepsilon}^* \leq x \mid \mathbf{S} = \mathbf{s}\} \\ &= P\{n^{-1/2} \mathbf{u}^T (n^{-1} \tilde{\mathbf{s}}^T \Sigma_n \tilde{\mathbf{s}})^{-1/2} \tilde{\mathbf{s}}^T \boldsymbol{\varepsilon}^* \leq x\} \rightarrow \Phi(x), \end{aligned}$$

when  $\|(n^{-1} \tilde{\mathbf{s}}^T \Sigma_n \tilde{\mathbf{s}})^{-1/2}\|_2 \leq C$  for  $n$  large enough.

Since  $n^{-1} \tilde{\mathbf{S}}^T \Sigma_n \tilde{\mathbf{S}} \xrightarrow{P} M_2 \succ 0$ , then for any subsequence  $\{n_l\}_{l=1}^\infty$ , there exists a further subsequence  $\{n_{l_j}\}_{j=1}^\infty$  such that  $n_{l_j}^{-1} \tilde{\mathbf{S}}^T \Sigma_{n_{l_j}} \tilde{\mathbf{S}} \xrightarrow{\text{a.s.}} M_2$  as  $j \rightarrow \infty$ . Therefore, for almost sure  $\mathbf{S} = \mathbf{s}$ , there exists a constant  $C > 0$  such that  $\|(n_{l_j}^{-1} \tilde{\mathbf{s}}^T \Sigma_{n_{l_j}} \tilde{\mathbf{s}})^{-1/2}\|_2 \leq C$  for  $n_{l_j}$  large enough. Then, as  $j \rightarrow \infty$ ,

$$\begin{aligned} &P\{n_{l_j}^{-1/2} \mathbf{u}^T (n_{l_j}^{-1} \tilde{\mathbf{S}}^T \Sigma_{n_{l_j}} \tilde{\mathbf{S}})^{-1/2} \tilde{\mathbf{S}}^T \boldsymbol{\varepsilon}^* \leq x \mid \mathbf{S} = \mathbf{s}\} \\ &= P\{n_{l_j}^{-1/2} \mathbf{u}^T (n_{l_j}^{-1} \tilde{\mathbf{s}}^T \Sigma_{n_{l_j}} \tilde{\mathbf{s}})^{-1/2} \tilde{\mathbf{s}}^T \boldsymbol{\varepsilon}^* \leq x\} \rightarrow \Phi(x), \end{aligned}$$

which implies that

$$\begin{aligned} &P\{n_{l_j}^{-1/2} \mathbf{u}^T (n_{l_j}^{-1} \tilde{\mathbf{S}}^T \Sigma_{n_{l_j}} \tilde{\mathbf{S}})^{-1/2} \tilde{\mathbf{S}}^T \boldsymbol{\varepsilon}^* \leq x\} \\ &= E[P\{n_{l_j}^{-1/2} \mathbf{u}^T (n_{l_j}^{-1} \tilde{\mathbf{S}}^T \Sigma_{n_{l_j}} \tilde{\mathbf{S}})^{-1/2} \tilde{\mathbf{S}}^T \boldsymbol{\varepsilon}^* \leq x \mid \mathbf{S}\}] \rightarrow \Phi(x) \end{aligned}$$

by dominated convergence theorem.

Then, we conclude that

$$P\{n^{-1/2} \mathbf{u}^T (n^{-1} \tilde{\mathbf{S}}^T \Sigma_n \tilde{\mathbf{S}})^{-1/2} \tilde{\mathbf{S}}^T \boldsymbol{\varepsilon}^* \leq x\} \rightarrow \Phi(x) \quad \text{as } n \rightarrow \infty.$$

This completes the proof. □

**Lemma 6.** Under Condition A, if  $n^{-1} \tilde{\mathbf{S}}^T \tilde{\mathbf{S}} \xrightarrow{P} M_1$  and  $n^{-1} \tilde{\mathbf{S}}^T \Sigma_n \tilde{\mathbf{S}} \xrightarrow{P} M_2$ , where  $M_1$  and  $M_2$  are  $(\ell m) \times (\ell m)$  positive-definite matrices, then  $n^{-1} \tilde{\mathbf{S}}^T \tilde{\Sigma} \tilde{\mathbf{S}} - n^{-1} \tilde{\mathbf{S}}^T \Sigma_n \tilde{\mathbf{S}} = o_P(1)$ .

*Proof.* Let

$$\begin{aligned} & n^{-1}\tilde{\mathbf{S}}^T\widehat{\Sigma}\tilde{\mathbf{S}} - n^{-1}\tilde{\mathbf{S}}^T\Sigma_n\tilde{\mathbf{S}} \\ &= (n^{-1}\tilde{\mathbf{S}}^T\widehat{\Sigma}\tilde{\mathbf{S}} - n^{-1}\tilde{\mathbf{S}}^T\widehat{\Sigma}^*\tilde{\mathbf{S}}) + \{n^{-1}\tilde{\mathbf{S}}^T\widehat{\Sigma}^*\tilde{\mathbf{S}} - n^{-1}\tilde{\mathbf{S}}^T(\Sigma_n \circ \mathbf{T})\tilde{\mathbf{S}}\} \\ & \quad + \{n^{-1}\tilde{\mathbf{S}}^T(\Sigma_n \circ \mathbf{T})\tilde{\mathbf{S}} - n^{-1}\tilde{\mathbf{S}}^T\Sigma_n\tilde{\mathbf{S}}\} \\ &= \text{I} + \text{II} + \text{III}, \end{aligned}$$

where  $\widehat{\Sigma}^* = \{(\tilde{\mathbf{Y}} - \tilde{\mathbf{S}}\mathbf{h})(\tilde{\mathbf{Y}} - \tilde{\mathbf{S}}\mathbf{h})^T\} \circ \mathbf{T}$ . It suffices to show that  $\text{I} = o_{\mathbb{P}}(1)$ ,  $\text{II} = o_{\mathbb{P}}(1)$  and  $\text{III} = o_{\mathbb{P}}(1)$ .

First, for term I,

$$\begin{aligned} & \widehat{\Sigma} - \widehat{\Sigma}^* - \kappa_n \mathbf{I}_n = \{(\tilde{\mathbf{Y}} - \tilde{\mathbf{S}}\widehat{\mathbf{h}})(\tilde{\mathbf{Y}} - \tilde{\mathbf{S}}\widehat{\mathbf{h}})^T - (\tilde{\mathbf{Y}} - \tilde{\mathbf{S}}\mathbf{h})(\tilde{\mathbf{Y}} - \tilde{\mathbf{S}}\mathbf{h})^T\} \circ \mathbf{T} \\ &= [\{\tilde{\mathbf{S}}(\mathbf{h} - \widehat{\mathbf{h}}) + \tilde{\mathbf{d}} + \tilde{\varepsilon}\}\{\tilde{\mathbf{S}}(\mathbf{h} - \widehat{\mathbf{h}}) + \tilde{\mathbf{d}} + \tilde{\varepsilon}\}^T - (\tilde{\mathbf{d}} + \tilde{\varepsilon})(\tilde{\mathbf{d}} + \tilde{\varepsilon})^T] \circ \mathbf{T} \\ &= \{\tilde{\mathbf{S}}(\mathbf{h} - \widehat{\mathbf{h}})(\mathbf{h} - \widehat{\mathbf{h}})^T\tilde{\mathbf{S}}^T + \tilde{\mathbf{S}}(\mathbf{h} - \widehat{\mathbf{h}})(\tilde{\mathbf{d}} + \tilde{\varepsilon})^T + (\tilde{\mathbf{d}} + \tilde{\varepsilon})(\mathbf{h} - \widehat{\mathbf{h}})^T\tilde{\mathbf{S}}^T\} \circ \mathbf{T}, \end{aligned}$$

which implies

$$\begin{aligned} \text{I} - \kappa_n/n\tilde{\mathbf{S}}^T\tilde{\mathbf{S}} &= n^{-1}\tilde{\mathbf{S}}^T(\widehat{\Sigma} - \widehat{\Sigma}^* - \kappa_n \mathbf{I}_n)\tilde{\mathbf{S}} \\ &= n^{-1}\sum_{i=1}^n\sum_{j=1}^n(e_{i;n}^T\tilde{\mathbf{S}})^T(e_{j;n}^T\tilde{\mathbf{S}})\{e_{i;n}^T\tilde{\mathbf{S}}(\mathbf{h} - \widehat{\mathbf{h}})(\mathbf{h} - \widehat{\mathbf{h}})^T\tilde{\mathbf{S}}^T e_{j;n} \\ & \quad + e_{i;n}^T\tilde{\mathbf{S}}(\mathbf{h} - \widehat{\mathbf{h}})(\tilde{\mathbf{d}} + \tilde{\varepsilon})^T e_{j;n} + e_{i;n}^T(\tilde{\mathbf{d}} + \tilde{\varepsilon})(\mathbf{h} - \widehat{\mathbf{h}})^T\tilde{\mathbf{S}}^T e_{j;n}\}\mathbf{T}(i, j) \\ &= n^{-1}\sum_{i=1}^n\sum_{j=1}^n(e_{i;n}^T\tilde{\mathbf{S}})^T(e_{j;n}^T\tilde{\mathbf{S}})\{(\mathbf{h} - \widehat{\mathbf{h}})^T\tilde{\mathbf{S}}^T e_{j;n} e_{i;n}^T\tilde{\mathbf{S}} \\ & \quad + (\tilde{\mathbf{d}} + \tilde{\varepsilon})^T e_{j;n} e_{i;n}^T\tilde{\mathbf{S}} + (\tilde{\mathbf{d}} + \tilde{\varepsilon})^T e_{i;n} e_{j;n}^T\tilde{\mathbf{S}}\}\mathbf{T}(i, j)(\mathbf{h} - \widehat{\mathbf{h}}). \end{aligned}$$

Since  $n^{-1}\tilde{\mathbf{S}}^T\tilde{\mathbf{S}} \xrightarrow{\mathbb{P}} M_1$  and  $\kappa_n = o(1)$ , we have  $\kappa_n/n\tilde{\mathbf{S}}^T\tilde{\mathbf{S}} = o_{\mathbb{P}}(1)$ . From Lemma 5, we have  $\mathbf{h} - \widehat{\mathbf{h}} = O_{\mathbb{P}}(n^{-1/2})$ . Since  $\|e_{i;n}^T\tilde{\mathbf{S}}\|_2 \leq C$  almost surely, and  $\mathbb{E}\{[(\tilde{\mathbf{d}} + \tilde{\varepsilon})^T e_{i;n}]\} \leq C$ , we have  $\mathbb{E}\{[(\mathbf{h} - \widehat{\mathbf{h}})^T\tilde{\mathbf{S}}^T e_{j;n} e_{i;n}^T\tilde{\mathbf{S}} + (\tilde{\mathbf{d}} + \tilde{\varepsilon})^T e_{j;n} e_{i;n}^T\tilde{\mathbf{S}} + (\tilde{\mathbf{d}} + \tilde{\varepsilon})^T e_{i;n} e_{j;n}^T\tilde{\mathbf{S}}]\|_2\} \leq C$  for any  $i, j = 1, \dots, n$ . Hence,  $\text{I} = O_{\mathbb{P}}(r_n/n^{1/2})$ , where  $r_n = n^{-1}\sum_{i=1}^n\sum_{j=1}^n|\mathbf{T}(i, j)| \leq \|\mathbf{T}\|_1 = o(n^{1/2})$  from Condition A9.

Second, for the term II,

$$\begin{aligned} \text{II} &= n^{-1}\tilde{\mathbf{S}}^T[\{(\tilde{\mathbf{d}} + \tilde{\varepsilon})(\tilde{\mathbf{d}} + \tilde{\varepsilon})^T\} \circ \mathbf{T}]\tilde{\mathbf{S}} - n^{-1}\tilde{\mathbf{S}}^T(\Sigma_n \circ \mathbf{T})\tilde{\mathbf{S}} \\ &= n^{-1}\tilde{\mathbf{S}}^T[\{(\tilde{\mathbf{d}} - S_b\varepsilon)(\tilde{\mathbf{d}} - S_b\varepsilon)^T + (\tilde{\mathbf{d}} - S_b\varepsilon)\varepsilon^T + \varepsilon(\tilde{\mathbf{d}} - S_b\varepsilon)^T\} \circ \mathbf{T}]\tilde{\mathbf{S}} \\ & \quad + n^{-1}\tilde{\mathbf{S}}^T[\{\varepsilon\varepsilon^T - \mathbb{E}(\varepsilon\varepsilon^T)\} \circ \mathbf{T}]\tilde{\mathbf{S}} \\ &= n^{-1}\tilde{\mathbf{S}}^T[\{(\tilde{\mathbf{d}} - S_b\varepsilon)(\tilde{\mathbf{d}} - S_b\varepsilon)^T\} \circ \mathbf{T}]\tilde{\mathbf{S}} + n^{-1}\tilde{\mathbf{S}}^T[\{(\tilde{\mathbf{d}} - S_b\varepsilon)\varepsilon^T\} \circ \mathbf{T}]\tilde{\mathbf{S}} \\ & \quad + n^{-1}\tilde{\mathbf{S}}^T[\{\varepsilon(\tilde{\mathbf{d}} - S_b\varepsilon)^T\} \circ \mathbf{T}]\tilde{\mathbf{S}} \\ & \quad + n^{-1}\sum_{i=1}^n\sum_{j=1}^n(e_{i;n}^T\tilde{\mathbf{S}})^T(e_{j;n}^T\tilde{\mathbf{S}})[\varepsilon(t_i)\varepsilon(t_j) - \mathbb{E}\{\varepsilon(t_i)\varepsilon(t_j)\}]\mathbf{T}(i, j) \end{aligned}$$

$$= \text{II}_1 + \text{II}_2 + \text{II}_3 + \text{II}_4.$$

From (A.15) of [46], we know that  $\sup_{nbL \leq i \leq n-nbL} |\mathbf{e}_{i;n}^T \tilde{\mathbf{d}}| = o(1)$  and  $\sup_{1 \leq i \leq n} |\mathbf{e}_{i;n}^T \tilde{\mathbf{d}}| \leq C$ , which, together with  $\sup_{1 \leq i \leq n} \mathbb{E}\{(\mathbf{e}_{i;n}^T S_b \boldsymbol{\varepsilon})^4\} = o(1)$  (by Lemma 2), implies

$$\begin{aligned} \sup_{nbL \leq i \leq n-nbL, 1 \leq j \leq n} \mathbb{E}\{[\mathbf{e}_{i;n}^T (\tilde{\mathbf{d}} - S_b \boldsymbol{\varepsilon})(\tilde{\mathbf{d}} - S_b \boldsymbol{\varepsilon})^T \mathbf{e}_{j;n} \mathbf{T}(i, j)]^2\} &= o(1), \\ \sup_{nbL \leq i \leq n-nbL, 1 \leq j \leq n} \mathbb{E}\{[\mathbf{e}_{i;n}^T (\tilde{\mathbf{d}} - S_b \boldsymbol{\varepsilon}) \boldsymbol{\varepsilon}^T \mathbf{e}_{j;n} \mathbf{T}(i, j)]^2\} &= o(1). \end{aligned}$$

Part (i) of Lemma 3 implies  $\text{II}_1, \text{II}_2 = o_P(1)$ . Since  $\text{II}_3 = \text{II}_2^T$ , we have  $\text{II}_3 = o_P(1)$ . From Lemma 4 and the fact that  $\|\mathbf{e}_{i;n}^T \tilde{\mathbf{S}}\|_2 \leq C$  almost surely, we know  $\text{II}_4 = O_P(B_n/n^{1/2})$ , where  $B_n = \max(\max_{1 \leq j \leq n} [\sum_{i=1}^n \{\mathbf{T}(i, j)\}^2]^{1/2}, \max_{1 \leq i \leq n} [\sum_{j=1}^n \{\mathbf{T}(i, j)\}^2]^{1/2})$ . From Condition A9,  $\sum_{i=1}^n \{\mathbf{T}(i, j)\}^2 \leq \sum_{i=1}^n |\mathbf{T}(i, j)| \leq \|\mathbf{T}\|_1 = o(n^{1/2})$  for any  $j = 1, \dots, n$ . Since  $\mathbf{T}$  is symmetric,  $B_n = o(n^{1/4})$ . Hence,  $\text{II}_4 = o_P(1)$ .

Finally,  $\text{III} = o(1)$  due to part (i) of Lemma 3 and the fact that  $\max_{1 \leq i, j \leq n} |\mathbf{e}_{i;n}^T (\Sigma_n \circ \mathbf{T} - \Sigma_n) \mathbf{e}_{j;n}| = o(1)$  following Condition A9 and Lemma 1. This completes the proof.  $\square$

*Proof of Theorem 1.* From Condition A8, we have  $\|\Sigma_n\|_1 = \|\Sigma_n\|_\infty \leq n^{1/2} \|\Sigma_n\|_2 = n^{1/2} \lambda_{\max}(\Sigma_n) = O(n^{1/2}) = o(n)$ . Lemma 3 implies  $n^{-1} \tilde{\mathbf{S}}^T \Sigma_n \tilde{\mathbf{S}} - \mathbb{E}(n^{-1} \tilde{\mathbf{S}}^T \Sigma_n \tilde{\mathbf{S}}) = o_P(1)$ ,  $n^{-1} \tilde{\mathbf{S}}^T \tilde{\mathbf{S}} - \mathbb{E}(n^{-1} \tilde{\mathbf{S}}^T \tilde{\mathbf{S}}) = o_P(1)$ ,  $\|\mathbb{E}(n^{-1} \tilde{\mathbf{S}}^T \Sigma_n \tilde{\mathbf{S}})\|_2 \leq C$ , and  $\|\mathbb{E}(n^{-1} \tilde{\mathbf{S}}^T \tilde{\mathbf{S}})\|_2 \leq C$ . Hence, for any subsequence  $\{n_l\}_{l=1}^\infty$ , there is a further subsequence  $\{n_{l_j}\}_{j=1}^\infty$  such that  $n_{l_j}^{-1} \tilde{\mathbf{S}}^T \tilde{\mathbf{S}} \xrightarrow{P} M_1^*$  and  $n_{l_j}^{-1} \tilde{\mathbf{S}}^T \Sigma_{n_{l_j}} \tilde{\mathbf{S}} \xrightarrow{P} M_2^*$  as  $j \rightarrow \infty$ . From the proof of Lemma A.6 of [46] and Condition A10, we can show that  $M_1^*$  and  $M_2^*$  are positive definite. From Lemma 5,  $n_{l_j}^{1/2} (\hat{\mathbf{h}} - \mathbf{h}) \xrightarrow{D} N(\mathbf{0}, M_1^{*-1} M_2^* M_1^{*-1})$ , which implies  $\{A(\tilde{\mathbf{S}}^T \tilde{\mathbf{S}})^{-1} \tilde{\mathbf{S}}^T \Sigma_{n_{l_j}} \tilde{\mathbf{S}} (\tilde{\mathbf{S}}^T \tilde{\mathbf{S}})^{-1} A^T\}^{-1/2} (A \hat{\mathbf{h}} - A \mathbf{h}) \xrightarrow{D} N(\mathbf{0}_k, \mathbf{I}_k)$  as  $j \rightarrow \infty$ . Hence, as  $n \rightarrow \infty$ :

$$\begin{aligned} \{A(\tilde{\mathbf{S}}^T \tilde{\mathbf{S}})^{-1} \tilde{\mathbf{S}}^T \Sigma_n \tilde{\mathbf{S}} (\tilde{\mathbf{S}}^T \tilde{\mathbf{S}})^{-1} A^T\}^{-1/2} (A \hat{\mathbf{h}} - A \mathbf{h}) &\xrightarrow{D} N(\mathbf{0}_k, \mathbf{I}_k) \\ \text{and } \mathbb{W}(\Sigma_n; b) &\xrightarrow{D} \chi_k^2. \end{aligned}$$

From Lemma 6,  $n_{l_j}^{-1} \tilde{\mathbf{S}}^T \hat{\Sigma} \tilde{\mathbf{S}} - n_{l_j}^{-1} \tilde{\mathbf{S}}^T \Sigma_{n_{l_j}} \tilde{\mathbf{S}} = o_P(1)$ . Then,  $n_{l_j}^{-1} \tilde{\mathbf{S}}^T \hat{\Sigma} \tilde{\mathbf{S}} \xrightarrow{P} M_2^*$  and thus,  $\{A(\tilde{\mathbf{S}}^T \tilde{\mathbf{S}})^{-1} \tilde{\mathbf{S}}^T \hat{\Sigma} \tilde{\mathbf{S}} (\tilde{\mathbf{S}}^T \tilde{\mathbf{S}})^{-1} A^T\}^{-1/2} (A \hat{\mathbf{h}} - A \mathbf{h}) \xrightarrow{D} N(\mathbf{0}_k, \mathbf{I}_k)$  as  $j \rightarrow \infty$ , which implies that as  $n \rightarrow \infty$ ,

$$\{A(\tilde{\mathbf{S}}^T \tilde{\mathbf{S}})^{-1} \tilde{\mathbf{S}}^T \hat{\Sigma} \tilde{\mathbf{S}} (\tilde{\mathbf{S}}^T \tilde{\mathbf{S}})^{-1} A^T\}^{-1/2} (A \hat{\mathbf{h}} - A \mathbf{h}) \xrightarrow{D} N(\mathbf{0}_k, \mathbf{I}_k) \text{ and } \mathbb{W}(\hat{\Sigma}; b) \xrightarrow{D} \chi_k^2.$$

This completes the proof.  $\square$

*Proof of Theorem 2.* Note that

$$\mathbb{W}(\Sigma_n; b) = (A \hat{\mathbf{h}} - A \mathbf{h})^T \{A(\tilde{\mathbf{S}}^T \tilde{\mathbf{S}})^{-1} \tilde{\mathbf{S}}^T \Sigma_n \tilde{\mathbf{S}} (\tilde{\mathbf{S}}^T \tilde{\mathbf{S}})^{-1} A^T\}^{-1} (A \hat{\mathbf{h}} - A \mathbf{h})$$

$$\begin{aligned}
& + 2(\mathbf{A}\mathbf{h})^T \{ \mathbf{A}(\tilde{\mathbf{S}}^T \tilde{\mathbf{S}})^{-1} \tilde{\mathbf{S}}^T \Sigma_n \tilde{\mathbf{S}} (\tilde{\mathbf{S}}^T \tilde{\mathbf{S}})^{-1} \mathbf{A}^T \}^{-1} (\mathbf{A}\hat{\mathbf{h}} - \mathbf{A}\mathbf{h}) \\
& + (\mathbf{A}\mathbf{h})^T \{ \mathbf{A}(\tilde{\mathbf{S}}^T \tilde{\mathbf{S}})^{-1} \tilde{\mathbf{S}}^T \Sigma_n \tilde{\mathbf{S}} (\tilde{\mathbf{S}}^T \tilde{\mathbf{S}})^{-1} \mathbf{A}^T \}^{-1} (\mathbf{A}\mathbf{h}) \\
& = \text{I} + \text{II} + \text{III}.
\end{aligned}$$

From the proof of Theorem 1 and assumptions  $n^{-1}\tilde{\mathbf{S}}^T \tilde{\mathbf{S}} \xrightarrow{P} M_1$  and  $n^{-1}\tilde{\mathbf{S}}^T \Sigma_n \tilde{\mathbf{S}} \xrightarrow{P} M_2$ , it follows that  $n^{-1}\text{I} = O_P(n^{-1})$ ,  $n^{-1}\text{II} = O_P(n^{-1/2})$ , and  $n^{-1}\text{III} \xrightarrow{P} (\mathbf{A}\mathbf{h})^T (\mathbf{A}M_1^{-1}M_2M_1^{-1}\mathbf{A}^T)^{-1}(\mathbf{A}\mathbf{h})$ . Similar arguments and Lemma 6 imply  $n^{-1}\mathbb{W}(\hat{\Sigma}; b) \xrightarrow{P} (\mathbf{A}\mathbf{h})^T (\mathbf{A}M_1^{-1}M_2M_1^{-1}\mathbf{A}^T)^{-1}(\mathbf{A}\mathbf{h})$ . This completes the proof.  $\square$

*Proof of Corollary 1.* As in the proof of Theorem 1, for any subsequence  $\{n_l\}_{l=1}^\infty$ , there is a further subsequence  $\{n_{l_j}\}_{j=1}^\infty$  such that  $n_{l_j}^{-1}\tilde{\mathbf{S}}^T \tilde{\mathbf{S}} \xrightarrow{P} M_1^*$ , and  $n_{l_j}^{-1}\tilde{\mathbf{S}}^T \Sigma_{n_{l_j}} \tilde{\mathbf{S}} \xrightarrow{P} M_2^*$  as  $j \rightarrow \infty$ , where  $M_1^*$  and  $M_2^*$  are positive definite matrices. From the proof of Theorem 2,  $n_{l_j}^{-1}\mathbb{W}(\Sigma_{n_{l_j}}; b) \xrightarrow{P} (\mathbf{A}\mathbf{h})^T (\mathbf{A}M_1^{*-1}M_2^*M_1^{*-1}\mathbf{A}^T)^{-1}(\mathbf{A}\mathbf{h})$ . Therefore, for any subsequence  $\{n_l\}_{l=1}^\infty$ , there is a further subsequence  $\{n_{l_j}\}_{j=1}^\infty$  such that  $\mathbb{W}(\Sigma_{n_{l_j}}; b) \xrightarrow{P} \infty$  as  $j \rightarrow \infty$ . Hence,  $\mathbb{W}(\Sigma_n; b) \xrightarrow{P} \infty$  as  $n \rightarrow \infty$ . Similarly, we can show  $\mathbb{W}(\hat{\Sigma}; b) \xrightarrow{P} \infty$  as  $n \rightarrow \infty$  by Lemma 6. This completes the proof.  $\square$

*Proof of Theorem 3.* Following the proof of Theorem 1, we have

$$\{ \mathbf{A}(\tilde{\mathbf{S}}^T \tilde{\mathbf{S}})^{-1} \tilde{\mathbf{S}}^T \Sigma_n \tilde{\mathbf{S}} (\tilde{\mathbf{S}}^T \tilde{\mathbf{S}})^{-1} \mathbf{A}^T \}^{-1/2} (\mathbf{A}\hat{\mathbf{h}} - \mathbf{A}\mathbf{h}) \xrightarrow{D} N(\mathbf{0}_k, \mathbf{I}_k), \text{ as } n \rightarrow \infty.$$

Since  $n^{-1}\tilde{\mathbf{S}}^T \tilde{\mathbf{S}} \xrightarrow{P} M_1$  and  $n^{-1}\tilde{\mathbf{S}}^T \Sigma_n \tilde{\mathbf{S}} \xrightarrow{P} M_2$ , it follows that

$$\{ \mathbf{A}(\tilde{\mathbf{S}}^T \tilde{\mathbf{S}})^{-1} \tilde{\mathbf{S}}^T \Sigma_n \tilde{\mathbf{S}} (\tilde{\mathbf{S}}^T \tilde{\mathbf{S}})^{-1} \mathbf{A}^T \}^{-1/2} (\mathbf{A}\mathbf{h}) \xrightarrow{P} (\mathbf{A}M_1^{-1}M_2M_1^{-1}\mathbf{A}^T)^{-1/2} \mathbf{c}.$$

Therefore,

$$\{ \mathbf{A}(\tilde{\mathbf{S}}^T \tilde{\mathbf{S}})^{-1} \tilde{\mathbf{S}}^T \Sigma_n \tilde{\mathbf{S}} (\tilde{\mathbf{S}}^T \tilde{\mathbf{S}})^{-1} \mathbf{A}^T \}^{-1/2} (\mathbf{A}\hat{\mathbf{h}}) \xrightarrow{D} N((\mathbf{A}M_1^{-1}M_2M_1^{-1}\mathbf{A}^T)^{-1/2} \mathbf{c}, \mathbf{I}_k),$$

which implies that  $\mathbb{W}(\Sigma_n; b) \xrightarrow{D} \chi_k^2(\tau^2)$ . The result for  $\mathbb{W}(\hat{\Sigma}; b)$  follows from Lemma 6. This completes the proof.  $\square$

*Proof of Proposition 1.* From (5.4),  $\mathbf{T}_{g_n}$  is symmetric, Toeplitz and satisfies  $1 = \mathbf{T}_{g_n}(1, 1) \geq \mathbf{T}_{g_n}(2, 1) \geq \dots \geq \mathbf{T}_{g_n}(n, 1) \geq 0$ . Since  $g_n = o(n^{1/2})$ , we have

$$\|\mathbf{T}_{g_n}\|_1 \leq 2 \sum_{i=1}^{g_n} \{1 - (i-1)/g_n\} = O(g_n) = o(n^{1/2}).$$

Let  $x_n = \lfloor g_n^{1/2} \rfloor$ . Since  $\lim_{n \rightarrow \infty} g_n = \infty$ , it follows that  $\lim_{n \rightarrow \infty} x_n = \infty$  and  $\mathbf{T}_{g_n}(x_n, 1) = 1 - (x_n - 1)/g_n = 1 - (\lfloor g_n^{1/2} \rfloor - 1)/g_n \rightarrow 1$  as  $n \rightarrow \infty$ .

It remains to show that  $\mathbf{T}_{g_n}$  is positive semidefinite. Consider an MA( $g_n - 1$ ) process  $\{\zeta_i\}_{i=1}^{g_n}$  defined as  $\zeta_i = z_i + z_{i-1} + \dots + z_{i-(g_n-1)}$ , where  $\{z_i\}$  are i.i.d. white noise terms with  $E(z_i) = 0$  and  $\text{var}(z_i) = 1$ . The auto-covariance matrix of  $\{\zeta_i\}_{i=1}^{g_n}$  is  $g_n \mathbf{T}_{g_n}$  which indicates that  $\mathbf{T}_{g_n} \succ 0$ . This completes the proof.  $\square$

## Acknowledgments

The authors thank the Editor, Associate Editor, and three reviewers for their insightful comments.

## Funding

C. Zhang's work was supported by U.S. National Science Foundation grants DMS-2013486 and DMS-1712418, and provided by the University of Wisconsin-Madison Office of the Vice Chancellor for Research and Graduate Education with funding from the Wisconsin Alumni Research Foundation. M. Chen's research was supported by the National Natural Science Foundation of China, grants 11690014 and 11731051. X. Guo's research is supported by the National Natural Science Foundation of China, grant 12071452.

## References

- [1] Ali, S. T., and Feigl, K. L. (2012). A new strategy for estimating geophysical parameters from InSAR data: Application to the Krafla central volcano in Iceland. *Geochem. Geophys. Geosyst.*, **13**, Q06005.
- [2] Benjamini, Y. and Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. R. Statist. Soc. B*, **57**, 289–300. [MR1325392](#)
- [3] Bickel, P. J. and Levina, E. (2008). Regularized estimation of large covariance matrices. *Ann. Statist.*, **36**, 199–227. [MR2387969](#)
- [4] 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.
- [5] Cai, T. T. and Liu, W. (2011). Adaptive thresholding for sparse covariance matrix estimation. *J. Amer. Statist. Assoc.*, **106**, 672–684. [MR2847949](#)
- [6] Cai, T. T., Zhang, C.-H. and Zhou, H. H. (2010). Optimal rates of convergence for covariance matrix estimation. *Ann. Statist.*, **38**, 2118–2144. [MR2676885](#)
- [7] Cai, T. T. and Zhou, H. H. (2012). Minimax estimation of large covariance matrices under  $\ell_1$ -norm. *Statist. Sinica*, **22**, 1319–1378. [MR3027084](#)
- [8] Chang, L. and Hanssen, R. F. (2016). A probabilistic approach for InSAR time-series postprocessing. *IEEE Transactions on Geoscience and Remote Sensing*, **54**, 421–430.
- [9] Diedrichsen, J. and Shadmehr, R. (2005). Detecting and adjusting for artifacts in fMRI time series data. *NeuroImage*, **27**, 624–634.
- [10] Dong, C., Gao, J. and Tjøstheim, D. (2016). Estimation for single-index and partially linear single-index integrated models. *Ann. Statist.*, **44**, 425–453. [MR3449774](#)

- [11] U.S. EPA, Fact Sheet: Revised air quality standards for particle pollution and updates to the air quality index (AQI), (2012) Available at: <http://www.epa.gov/airquality/particlepollution/2012/decfsstandards.pdf>.
- [12] Fan, J. and Gijbels, I. (1996). *Local Polynomial Modeling and Its Applications*. Chapman and Hall, London. [MR1383587](#)
- [13] Gao, J. and Phillips, P. C. B. (2013). Semiparametric estimation in triangular system equations with nonstationarity. *Journal of Econometrics*, **176**, 59–79. [MR3067025](#)
- [14] Glover, G. H. (1999). Deconvolution of impulse response in event-related BOLD fMRI. *NeuroImage*, **9**, 416–429.
- [15] Guo, X. and Zhang, C. M. (2015). Estimation of the noise autocorrelation matrix in semi-parametric model for fMRI data. *Statist. Sinica*, **25**, 475–498. [MR3379083](#)
- [16] Huang, L., Jiang, H. and Wang, H. (2019). A novel partial-linear single-index model for time series data. *Computational Statistics and Data Analysis*, **134**, 110–122. [MR3926489](#)
- [17] Jezzard, P. (1999). Physiological noise: Strategies for correction, in *Functional MRI*, C. T. W. Moonen and P. A. Bandettini, Eds. New York: Springer-Verlag, pp. 173–182.
- [18] El Karoui, N. (2008). Operator norm consistent estimation of large-dimensional sparse covariance matrices. *Ann. Statist.*, **36**, 2717–2756. [MR2485011](#)
- [19] Lange, N. (1996). Tutorial in biostatistics: Statistical approaches to human brain mapping by functional magnetic resonance imaging. *Statist. Med.*, **15**, 389–428.
- [20] 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.
- [21] Li, Q. (1996). On the root- $N$ -consistent semiparametric estimation of partially linear models. *Economics Letters*, **51**, 277–285. [MR1403469](#)
- [22] Li, D., Chen, J. and Lin, Z. (2011). Statistical inference in partially time-varying coefficient models. *Journal of Statistical Planning and Inference*, **141**, 995–1013. [MR2732967](#)
- [23] Li, J., Chen, L., Kim, K. H. and Zhou, T. (2022). Simultaneous inference of trend in partially linear time series. [arXiv:2212.10359v1](#).
- [24] Long, C. J., Brown, E. N., Triantafyllou, C., Aharon, I., Wald, L. L. and Solo, V. (2005). Nonstationary noise estimation in functional MRI. *NeuroImage*, **28**, 890–903.
- [25] Lund, R.B., Seymour, P.L. and Kafadar, K. (2001). Temperature trends in the United States. *Environmetrics*, **12**, 673–690.
- [26] Luo, H. and Puthusserypady, S. (2007). fMRI data analysis with nonstationary noise models: a bayesian approach. *IEEE Trans. Biomed. Eng.*, **54**, 1621–1630.
- [27] Meyer, Francois G. (2003). Wavelet based estimation of a semi parametric generalized linear model of fMRI time-series. *IEEE Transactions on Medical Imaging*, **22**, 315–322.

- [28] 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. [MR2732601](#)
- [29] Oikonomou, V. P., Tripoliti, E. E. and Fotiadis, D. I. (2010). Bayesian methods for fMRI time-series analysis using a nonstationary model for the noise. *IEEE Trans. Inf. Technol. Biomed.*, **14**, 664–674.
- [30] Politis, D. N., Romano, J. P. and Wolf, M. (1999). *Subsampling*. New York: Springer-Verlag. [MR1707286](#)
- [31] 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.
- [32] Robinson, P. M. (1988). Root- $N$ -consistent semiparametric regression. *Econometrica*, **56**, 931–954. [MR0951762](#)
- [33] Schur, J. (1911). Bemerkungen zur theorie der beschränkten bilinearformen mit unendlich vielen veränderlichen. *J. Reine Math.*, **140**, 1–28. [MR1580823](#)
- [34] Silverman, B. W. (1986). *Density Estimation for Statistics and Data Analysis*. London: Chapman and Hall. [MR0848134](#)
- [35] Song, S., Härdle, W. K. and Ritov, Y. (2014). Generalized dynamic semi-parametric factor models for high-dimensional non-stationary time series. *Econometrics Journal*, **17**, S101–S131. [MR3219152](#)
- [36] Storey, J.D., Taylor, J.E. and Siegmund, D. (2004). Strong control, conservative point estimation and simultaneous conservative consistency of false discovery rates: a unified approach. *J. R. Statist. Soc. B*, **66**, 187–205. [MR2035766](#)
- [37] Thomas, B. R., Kent, E.C. & Swail, V. R. (2005). Methods to homogenize wind speeds from ships and buoys. *International Journal of Climatology*, **25**, 979–995.
- [38] Wang, X. and Chen, S. (2020). Semiparametric estimation of generalized transformation panel data models with nonstationary error. *Econometrics Journal*, **23**, 386–402. [MR4165464](#)
- [39] 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.
- [40] Wu, W. B. and Pourahmadi, M. (2009). Banding sample autocovariance matrices of stationary processes. *Statist. Sinica*, **19**, 1755–1768. [MR2589209](#)
- [41] Wu, W. B. and Xiao, H. (2011). Covariance matrix estimation in time series. *Handbook of Statistics*, **30**, 187–209.
- [42] Wu, W. B. and Zhao, Z. (2007). Inference of trends in time series. *J. R. Statist. Soc. B*, **69**, 391–410. [MR2323759](#)
- [43] Xiao, H. and Wu, W. B. (2012). Covariance matrix estimation for stationary time series. *Ann. Statist.*, **40**, 466–493. [MR3014314](#)
- [44] Xue, L., Ma, S. and Zou, H. (2012). Positive-definite  $\ell_1$ -penalized estimation of large covariance matrices. *J. Amer. Statist. Assoc.*, **107**, 1480–1491. [MR3036409](#)
- [45] Zhang, T. and Wu, W. B. (2011). Testing parametric assumptions of trends

- of nonstationary time series. *Biometrika*, **98**, 599–614. [MR2836409](#)
- [46] Zhang, C. M. and Yu, T. (2008). Semiparametric detection of significant activation for brain fMRI. *Ann. Statist.*, **36**, 1693–1725. [MR2435453](#)
- [47] Zhou, Z. and Wu, W. B. (2009). Local linear quantile estimation for non-stationary time series. *Ann. Statist.*, **37**, 2696–2729. [MR2541444](#)