Statistica Sinica Preprint No: SS-2022-0054	
Title	Activation Discovery With FDR Control: Application to FMRI Data
Manuscript ID	SS-2022-0054
URL	http://www.stat.sinica.edu.tw/statistica/
DOI	10.5705/ss.202022.0054
Complete List of Authors	Mengtao Wen, Guanghui Wang, Changliang Zou and Zhaojun Wang
Corresponding Authors	Guanghui Wang
E-mails	ghwang.nk@gmail.com

Statistica Sinica: Preprint
doi:10.5705/ss.202022.0054

Statistica Sinica

Activation discovery with FDR control: Application to fMRI data

Mengtao Wen[1], Guanghui Wang[2], Changliang Zou[1], Zhaojun Wang[1]

[1]*Nankai University,* [2]*East China Normal University*

Abstract: Time series from a large number of sources are ubiquitous, and may incur structural changes during data acquisition. For example, in fMRI analysis, brain regions associated with task-related stimuli or in a resting state become active. An activated time series can comprise readings from an activated region. Of interest is to control the *uncertainty* of discovering time series in activation (viz., activated regions in fMRI analysis) by using the false discovery rate (FDR) tool. We propose a simple, yet effective method that incorporates unknown asynchronous change patterns and spatial dependence. We justify the validity of our method in controlling the FDR using an asymptotic analysis. The results of our numerical experiments indicate that the proposed method is both accurate and powerful. An implementation is provided in the R package SLIP.

Key words and phrases: Change-point analysis, data splitting, false discovery rate, fMRI, regions of interest.

Statistica Sinica: Preprint
doi:10.5705/ss.202022.0054

1. Introduction

Time series from many sources are ubiquitous, and the underlying distribution of each time series may change during data acquisition, owing to external stimuli or internal evolution. A good example is functional magnetic resonance imaging, or functional MRI (fMRI), an image acquisition modality used to study the brain invivo. Research on fMRI focuses on changes in the blood oxygen level-dependent (BOLD) response (Ogawa et al., 1990), a surrogate measure of brain activity, typically caused by an externally controlled stimulus or task. Recently, researchers have begun paying greater attention to studying the BOLD response during rest, which reflects the brain's neuronal baseline activity; see, for example, Damoiseaux et al. (2006). During a task-related or resting-state fMRI experiment, the data comprise a series of magnetic resonance brain images: the BOLD responses over time from a large number of uniformly spaced volume elements (or voxels). A time series can be composed of readings from a voxel or a region of spatially contiguous voxels.

One fundamental goal of fMRI analysis is to discover regions or points of interest (ROIs or POIs), namely, regions or voxels activated by a task, or even in a baseline state. Excluding those consisting solely of background noise, not all time series are activated in a specific scene, and the time series

Statistica Sinica: Preprint
doi:10.5705/ss.202022.0054

in activation may react at different times during the experiment. Moreover, precise times of component-wise activations are usually unknown, owing to possible lags after a stimulus, or when the data are acquired in a resting state. A natural statistical approach is to use change-point or process control theory; see, for example, Lindquist et al. (2007) and Aston and Kirch (2012). These studies focus on modeling the fMRI data, voxelwise or treated as a whole. Few works focus on the inferential side, that is, the *uncertainty* of discovered voxels or regions in activation. This amounts to performing a component-wise hypothesis test of whether a change in the BOLD response occurs during an fMRI experiment, and provides a threshold for the resulting *activation map* of test statistics to meet for specific error rate control (Genovese et al., 2002; Nichols and Holmes, 2002).

An appealing statistical notion of the error rate is the false discovery rate (Benjamini and Hochberg, 1995, FDR), that is, the expected proportion of falsely rejected hypotheses. The authors also propose a procedure known as the Benjamini–Hochberg (BH) method, which controls the FDR for independent p-values corresponding to all null hypotheses. The BH method is widely used for neuroimaging data to determine the threshold of an activation map in a task-related fMRI with a known activation time (Genovese et al., 2002; Kriegeskorte et al., 2006), where each time series is

Statistica Sinica: Preprint
doi:10.5705/ss.202022.0054

associated with a statistic based on the BOLD responses before and after some stimulus (e.g., t-statistic). There are two problems with using this method in practical fMRI studies: (i) the precise time of activation is often unavailable for each region, especially for resting-state experiments, and need not be the same across regions; and (ii) component-wise comparisons are usually spatially correlated, which may cause the BH method to become conservative.

In this study, we borrow ideas from recent developments of FDR control methodologies, and propose a simple, yet effective procedure for discovering activated time series in an fMRI analysis with proper FDR control, while incorporating unknown asynchronous change patterns and spatial dependence. We call the proposed method SLIP, which comprises a sequence of steps: Splitting the data into two parts, Locating component-wise activation times based on one sample, Incorporating spatial dependence among time series, and Pooling the summary statistics from both samples. The SLIP method controls the FDR in the discovery of time series in activation.

Applications for the SLIP method are fairly widespread beyond fMRI data. For example, in studies on the effects of government policies (or unprecedented events, such as COVID-19) on the stock market, researchers often wish to discover which stocks or stock market sectors are affected from

Statistica Sinica: Preprint
doi:10.5705/ss.202022.0054

among a large number of candidates (Mazur et al., 2021). In addition, in modern manufacturing processes, transitions between multiple operating or environmental conditions may affect the product quality, and thus it would be helpful to find related quality characteristics that describe the underlying variations (Zou and Qiu, 2009; Capizzi, 2015). In both examples, we need to detect "activated" time series from among many candidates, where the reaction times of different time series corresponding to a policy release or some out-of-control state may differ. We can apply the proposed SLIP method to such scenarios with slight modifications, to control the FDR simultaneously. In this study, we focus on fMRI data analysis.

1.1 Problem formulation

Suppose p parallel time series are recorded at T time points, $\{Z_{ij}, i = 1, \ldots, T\}_{j=1}^{p}$. In fMRI, Z_{ij} can stand for the measured BOLD response at the jth voxel or region in the brain during the ith scan. We consider the mean-level change model

$$\begin{aligned} Z_{ij} &= \mu_{i,j} + \varepsilon_{ij},\ i = 1, \ldots, T,\ j = 1, \ldots, p, \\ \mu_{1,j} = \cdots = \mu_{\tau_j^*,j} &\neq \mu_{\tau_j^*+1,j} = \cdots = \mu_{T,j},\ j = 1, \ldots, p, \end{aligned} \tag{1.1}$$

where, for the jth time series (data sequence), the mean level of the BOLD signal remains unchanged if $\tau_j^* = T$, and experiences some change at

Statistica Sinica: Preprint
doi:10.5705/ss.202022.0054

τ_j^* if $1 \leq \tau_j^* < T$, due to a reaction to a stimulus or intrinsic evolution under a resting state, and ε_{ij} are random errors. We assume that $\boldsymbol{\varepsilon}_i := (\varepsilon_{i1}, \ldots, \varepsilon_{ip})^\top$, for $i = 1, \ldots, n$, are independent, with mean vector $\mathbf{0}$ and covariance matrix $\boldsymbol{\Sigma}$, where $\boldsymbol{\Sigma}$ reflects the spatial dependence structure among all time series. A discussion of temporal correlation is deferred to Section 4.5. We say a time series is *activated* or *in activation* if the associated $\tau_j^* < T$, such that a mean-level signal change occurs, where we borrow the terminology from fMRI applications. Our primary interest is to discover the set of activated time series, that is, $\mathcal{A} = \{1 \leq j \leq p : \tau_j^* < T\}$.

Any discovery procedure, say $\widehat{\mathcal{A}}$, can commit two kinds of errors, that is, it may include an inactivated time series, or it may exclude an activated one, referred to as a false positive and a false negative, respectively, in the terminology of *multiple testing*, where we conduct a sequence of hypothesis tests

$$H_{0j} : \tau_j^* = T \quad \text{versus} \quad H_{1j} : \tau_j^* < T, \ j = 1, \ldots, p. \tag{1.2}$$

Researchers tend to be reluctant to miss any true positives (e.g., activated regions), and so prefer a slight overestimation of $\mathcal{A}$, while guaranteeing the number (or rate) of false positives (i.e., inactivated regions in the selected set) at a prescribed level. Therefore, we use the FDR tool to control the expected proportion of false discoveries to handle the *uncertainty* in iden-

Statistica Sinica: Preprint
doi:10.5705/ss.202022.0054

tifying $\mathcal{A}$.

1.2 Connection with the literature on change-point detection

Recently, testing for the existence and estimating the times of activations (or change-points) for large-scale (or *high-dimensional*) time series have received much attention in the literature on change-point detection, where the locations of change-points, if they exist, are assumed to be shared across the time series; see, for example, Bai (2010), Cho and Fryzlewicz (2015), Wang et al. (2018), and Wang and Samworth (2018). These methods study either global testing for the existence of common change-points, or provide a consistent estimation of the number and locations of change-points. In contrast, we aim to identify time series that have encountered changes, rather than the change-points themselves. Moreover, we allow component-wise asynchronous change-points.

A more related work is that of Jirak (2015), who considers a consistent identification of the activation set $\mathcal{A}$, such that the probability of making even one false rejection, that is, the family-wise error rate (FWER), is controlled asymptotically. However, FWER-oriented procedures are known to be very conservative, motivating us to consider other uncertainty measures, such as the FDR.

Statistica Sinica: Preprint
doi:10.5705/ss.202022.0054

Prior studies have used the FDR tool to detect multiple change-points for a single data sequence, with varying definitions of false discoveries associated with the estimated change-points; see, for example, Hao et al. (2013), Li et al. (2016), and Wang et al. (2022). Specifically, Wang et al. (2022) determined the number of jumps in a regression curve, and quantified the uncertainty of the estimated change-points using the FDR.

Quantifying the uncertainty in multiple change-point detection has become an active research topic. Leveraging the approach of selective inference (Fithian et al., 2014; Lee et al., 2016), Hyun et al. (2018, 2021) and Jewell et al. (2022) developed valid tests for a change in the mean associated with estimated change-points for univariate data sequences, and Sugiyama et al. (2021) investigate multivariate scenarios. Furthermore, Fryzlewicz (2020) detects localized regions in a data sequence, with each region containing a change-point at a prescribed global significance level.

1.3 Developments of FDR methodologies

Benjamini and Hochberg (1995) originally proposed the notion of FDR, and provided a procedure (called the BH method) that guarantees FDR control when p-values are independent. The effect of dependence on the FDR control of the BH method has been investigated widely in the literature;

Statistica Sinica: Preprint
doi:10.5705/ss.202022.0054

see, for example, Benjamini and Yekutieli (2001), Storey et al. (2004), and Clarke and Hall (2009). They suggest that the BH method could be valid under specific dependence, but may become very conservative. On the other hand, some recent studies have shown that the dependence structure can sometimes be informative, and could be used to improve the power; see Efron (2007) and Fan et al. (2012).

Component-wise p-values are needed to apply the BH method (or its variants) to current fMRI studies. For a time series in which a change may occur, we cannot calculate the p-value precisely. However, we can approximate it asymptotically using change-point theory (Csörgő and Horváth, 1997). Unfortunately, the convergence to the asymptotic null distribution of the component-wise test statistic is relatively slow, which may cause inaccurate approximations for finite-sample applications. Although Jirak (2015) provides simulation-based remedies, they rely heavily on the normality of the data distribution, and can be computationally heavy for large-scale multiple testing. In addition, it remains unknown how to incorporate spatial dependence among time series for better FDR control and power enhancement.

New FDR control procedures for simultaneously testing p null hypotheses have been proposed by abandoning direct usages of p-values; see, for

Statistica Sinica: Preprint
doi:10.5705/ss.202022.0054

example, Barber and Candès (2015) and Du et al. (2021). The idea is to construct a sequence of test statistics W_j, for $j = 1, \ldots, p$, that fulfills the *ranking* property, that is, $W_j > 0$, and tends to be large with a high probability if the index j corresponds to a true non-null hypothesis. Furthermore, the *symmetry* property, that is, the W_j corresponding to true null hypotheses are (asymptotically) symmetric about zero. Then, for a given threshold $L > 0$, the number of false discoveries can be approximated by $\#\{j : W_j < -L\}$. Consequently, by choosing the threshold

$$\widehat{L} = \widehat{L}(W_1, \ldots, W_p) = \inf_{L>0} \left\{ \frac{\#\{j : W_j < -L\}}{\#\{j : W_j > L\} \vee 1} \leq \alpha \right\}, \tag{1.3}$$

with the convention of $\inf \emptyset = +\infty$, for a prescribed FDR level α, the procedure rejects all null hypotheses corresponding to $W_j > \widehat{L}$, where $a \vee b$ stands for the maximum of a and b. This controls the FDR under some mild conditions. Du et al. (2021) propose using data-splitting strategies to construct W_j, providing a novel mechanism to enhance the detection power by exploiting the underlying dependence information, which directly motivates our method. However, our contributions are still nontrivial from methodological, theoretical, and practical aspects: (i) our null hypotheses incorporate temporal change patterns, and thus the approach of Du et al. (2021) is not directly applicable; (ii) adding change-point theory makes a theoretical analysis of the FDR validity much more complicated; and

Statistica Sinica: Preprint
doi:10.5705/ss.202022.0054

(iii) the current problem is relevant in real applications such as fMRI data analysis, and can be modified easily to meet different practical tasks; see Section 4 for further discussions.

1.4 Organization of the paper

The rest of this paper is organized as follows. In Section 2, we introduce the proposed SLIP method for discovering activated data sequences. Its asymptotic validity on the FDR control is presented in Section 3. Section 4 includes its variants for practical applications. Numerical studies are conducted in Section 5. Several concluding remarks are given in Section 6. Proofs of all theoretical conclusions and additional numerical results are deferred to the Supplementary Material.

1.5 Notation

For $a \in \mathbb{R}$, $\lfloor a \rfloor$ is the maximal integer less than or equal to a. Let $\mathbb{1}(\cdot)$ be the indicator function. The cardinality of a set $\mathcal{S}$ is denoted by $|\mathcal{S}|$. For $\mathbf{v} = (v_1, \ldots, v_p)^\top \in \mathbb{R}^p$, define $\mathrm{diag}(\mathbf{v}) = \mathrm{diag}(v_1, \ldots, v_p)$ as a diagonal matrix with diagonal elements $\mathbf{v}$. Denote $\|\mathbf{v}\|_2 = (\sum_{j=1}^p v_j^2)^{1/2}$ and $\|\mathbf{v}\|_\infty = \max_{1 \le j \le p} |v_j|$. Let $\mathbf{v}_\mathcal{S}$ be the sub-vector of $\mathbf{v}$ consisting of elements with indices that are in the set $\mathcal{S}$. For a square matrix $\mathbf{M} \in \mathbb{R}^{p \times p}$, denote by

Statistica Sinica: Preprint
doi:10.5705/ss.202022.0054

$\lambda_{\max}(\mathbf{M})$ and $\lambda_{\min}(\mathbf{M})$ the maximum and minimum of the eigenvalues of $\mathbf{M}$, respectively. For a matrix $\mathbf{M} \in \mathbb{R}^{n\times p}$, let $\|\mathbf{M}\|_2 = \lambda_{\max}^{1/2}(\mathbf{M}^\top\mathbf{M}) = \lambda_{\max}^{1/2}(\mathbf{M}\mathbf{M}^\top)$ and $\|\mathbf{M}\|_1 = \max_{1\le j\le p}\sum_{i=1}^{n}|M_{ij}|$, where M_{ij} is the (i,j)th element of $\mathbf{M}$. Let $\mathbf{M}_{\mathcal{S}}$ be the sub-matrix of $\mathbf{M}$ with columns with indices in $\mathcal{S}$. For $\mathbf{M}_1, \mathbf{M}_2 \in \mathbb{R}^{n\times p}$, $\mathbf{M}_1 \circ \mathbf{M}_2 \in \mathbb{R}^{n\times p}$ is the Hadamard product of $\mathbf{M}_1$ and $\mathbf{M}_2$.

2. Methodology

2.1 Our idea: independent setting

To fix the idea, we first consider a simplified setting in which all data sequences are independent of each other, and thus $\mathbf{\Sigma} = \mathrm{diag}(\sigma_{11}, \ldots, \sigma_{pp})$, with $\sigma_{jj} > 0$, for $j = 1, \ldots, p$, and σ_{jj} are known. A treatment for an unknown correlated $\mathbf{\Sigma}$ is deferred to Section 2.2. The key is to construct a sequence of activation statistics W_j for H_{0j}, $j = 1, \ldots, p$, in (1.2) that fulfills the symmetry and ranking properties (see Section 1.3).

Our construction is based on a specialized sample-splitting strategy, that is, the order-preserved splitting (OPS) proposed by Zou et al. (2020), which was originally used to estimate the number of change-points. Collect the BOLD responses during the ith scan as $\mathbf{Z}_i = (Z_{i1}, \ldots, Z_{ip})^\top$, for $i =$

Statistica Sinica: Preprint
doi:10.5705/ss.202022.0054

$1, \ldots, T$. The data are temporally split into two disjoint parts,

$$\mathcal{Z}_1 := \{\mathbf{Z}_i^{(1)} : i = 1, \ldots, T_1\} = \{\mathbf{Z}_i, i = 1, \ldots, T\} \backslash \mathcal{Z}_2 \text{ and}$$

$$\mathcal{Z}_2 := \{\mathbf{Z}_i^{(2)} : i = 1, \ldots, T_2\} = \{\mathbf{Z}_i : i = r, 2r, \ldots, \lfloor T/r \rfloor r\},$$

where $r > 1$ is an integer, discussed in Section 4.1, $T_2 = \lfloor T/r \rfloor$, and $T_1 = T - T_2$. Rather than splitting randomly, the OPS preserves the original change patterns during the time course of an fMRI experiment, as much as possible. Then, the sample $\mathcal{Z}_1$ is mainly used to locate component-wise activation times, and the sample $\mathcal{Z}_2$ plays a critical role in symmetrization.

First, a change-localizing algorithm to estimate possible activation times is applied to the data split $\mathcal{Z}_1$. To facilitate the presentation, we consider the traditional cumulative summation (CUSUM)-based procedure; other candidates that capture intrinsic change patterns are discussed in Section 4.2. Specifically, for $j = 1, \ldots, p$, if a change occurs to the jth data sequence of $\mathcal{Z}_1$, in other words, that data sequence is activated somewhere, the activation time is identified as

$$\widehat{\tau}_j^{(1)} = \arg\max_{\tau \in (\lfloor T_1 \varrho \rfloor, T_1 - \lfloor T_1 \varrho \rfloor]} \sqrt{\frac{\tau(T_1 - \tau)}{T_1}} \left| \bar{Z}_j^{(1)}(\tau, T_1) - \bar{Z}_j^{(1)}(0, \tau) \right|, \quad (2.1)$$

where $\bar{Z}_j^{(k)}(\ell_1, \ell_2) = (\ell_2 - \ell_1)^{-1} \sum_{\ell=\ell_1+1}^{\ell_2} Z_{\ell j}^{(k)}$, $Z_{\ell j}^{(k)}$ is the jth element of $\mathbf{Z}_\ell^{(k)}$, for $j = 1, \ldots, p$ and $k = 1, 2$, and $\varrho \in (0, 1/2)$ is a prespecified constant, discussed in Section 4.1. The potential component-wise activation times in the data split $\mathcal{Z}_2$ can then be approximated by $\widehat{\tau}_j^{(2)} = \lfloor T_2 \widehat{\tau}_j^{(1)} / T_1 \rfloor$, for

Statistica Sinica: Preprint
doi:10.5705/ss.202022.0054

$j = 1, \ldots, p$. If $\tau_j^* < T$ and is not near the boundary, and the activation signal is not too weak, then we expect that both $\widehat{\tau}_j^{(k)}$, for $k = 1, 2$, can recover τ_j^* well (up to a data-splitting ratio).

A simple, but important fact is that, conditional on the data split $\mathcal{Z}_1$, the CUSUM statistics based on $\mathcal{Z}_2$,

$$\xi_j^{(2)} = \sqrt{\frac{\widehat{\tau}_j^{(2)}\left(T_2 - \widehat{\tau}_j^{(2)}\right)}{T_2}} \left\{ \bar{Z}_j^{(2)}(\widehat{\tau}_j^{(2)}, T_2) - \bar{Z}_j^{(2)}(0, \widehat{\tau}_j^{(2)}) \right\}, \ j = 1, \ldots, p, \tag{2.2}$$

are *asymptotically* normally distributed. In addition, for inactivated data sequences, the corresponding $\xi_j^{(2)}$ are asymptotically symmetric about zero. This fact motivates the construction of our activation statistics

$$W_{j,\text{indep}} = \xi_j^{(1)} \xi_j^{(2)} / \sigma_{jj}, \ j = 1, \ldots, p, \tag{2.3}$$

where

$$\xi_j^{(1)} = \sqrt{\frac{\widehat{\tau}_j^{(1)}\left(T_1 - \widehat{\tau}_j^{(1)}\right)}{T_1}} \left\{ \bar{Z}_j^{(1)}(\widehat{\tau}_j^{(1)}, T_1) - \bar{Z}_j^{(1)}(0, \widehat{\tau}_j^{(1)}) \right\}, \ j = 1, \ldots, p$$

are constructed based on the data split $\mathcal{Z}_1$. Thus, we concluded that the $W_{j,\text{indep}}$ for $j \notin \mathcal{A}$ are conditionally (on $\mathcal{Z}_1$) asymptotically symmetric about zero. Moreover, for an activated data sequence, if both $\widehat{\tau}_j^{(k)}$, for $k = 1, 2$, are tracking τ_j^* well, then both $\xi_j^{(k)}$, for $k = 1, 2$, can recover an activation signal (including the sign and magnitude). Consequently, the products $W_{j,\text{indep}}$

Statistica Sinica: Preprint
doi:10.5705/ss.202022.0054

for $j \in \mathcal{A}$ are positive and large, regardless of the signs of the activation signals. Hence, the construction fulfills the ranking and symmetry properties simultaneously.

Let α be the nominal FDR level. A discovery is made if $W_{j,\text{indep}}$ passes a threshold $\widehat{L}_{\text{indep}} = \widehat{L}(W_{1,\text{indep}}, \ldots, W_{p,\text{indep}})$ (see Eq. (1.3)). Consequently, the estimate of the activation set is $\widehat{\mathcal{A}}_{\text{indep}} = \{1 \le j \le p : W_{j,\text{indep}} \ge \widehat{L}_{\text{indep}}\}$.

2.2 The SLIP method: the blessing of dependence

In real fMRI applications, data sequences are spatially dependent, which questions the validity of $\widehat{\mathcal{A}}_{\text{indep}}$. Strong correlations may break down the symmetry property of all inactivated $W_{j,\text{indep}}$ (i.e., $j \notin \mathcal{A}$), and thus distort the FDR control and make $\widehat{\mathcal{A}}_{\text{indep}}$ unreliable (see Figure 2(i)). Furthermore, ignoring component-wise dependence could make the procedure less powerful, even if the FDR is under control under a weak dependence structure (see Figure 2(ii)). In this section, we propose the SLIP method, which incorporates underlying spatial dependence to enhance the detection power, while still controlling the FDR at some prescribed level. The method comprises four steps: Splitting the data into two parts, Locating component-wise activation times based on one sample, Incorporating the spatial dependence among the data sequences, and Pooling the summary

Statistica Sinica: Preprint
doi:10.5705/ss.202022.0054

statistics from separate samples.

We observe that $\xi_j^{(2)}$ (see Eq. (2.2)) for inactivated data sequences has a zero mean, and we denote $\beta_j = \mathbb{E}(\xi_j^{(2)} \mid \mathcal{Z}_1)$ for activated data sequences (i.e., $j \in \mathcal{A}$). Let $\Xi_{jk} = \mathrm{Cov}(\xi_j^{(2)}, \xi_k^{(2)} \mid \mathcal{Z}_1)$, for $1 \le j, k \le p$. Note that the quantities β_j and Ξ_{jk} all depend on the data split $\mathcal{Z}_1$. Motivated by the work of Du et al. (2021), who consider the FDR control under general dependence by recasting the original mean testing problems into a regression framework, we introduce the following working model for $\boldsymbol{\xi}^{(2)} = (\xi_1^{(2)}, \ldots, \xi_p^{(2)})^\top$:

$$\boldsymbol{\xi}^{(2)} \approx \boldsymbol{\beta} + \boldsymbol{\epsilon},\ \mathbb{E}(\boldsymbol{\epsilon}) = \mathbf{0} \text{ and } \mathrm{Var}(\boldsymbol{\epsilon}) =: \boldsymbol{\Xi}, \tag{2.4}$$

where $\boldsymbol{\beta}$ is a p-dimensional vector with jth element β_j if $j \in \mathcal{A}$, and zero otherwise, and $\boldsymbol{\Xi}$ is a $p \times p$ matrix with (j,k)th element Ξ_{jk} for $1 \le j, k \le p$. Equivalently,

$$\mathbf{y}_2 \approx \mathbf{X}\boldsymbol{\beta} + \widetilde{\boldsymbol{\epsilon}},\ \mathbb{E}(\widetilde{\boldsymbol{\epsilon}}) = \mathbf{0} \text{ and } \mathrm{Var}(\widetilde{\boldsymbol{\epsilon}}) = \mathbf{I}, \tag{2.5}$$

where $\mathbf{X} = \boldsymbol{\Xi}^{-1/2}$, $\mathbf{y}_2 = \boldsymbol{\Xi}^{-1/2}\boldsymbol{\xi}^{(2)}$, and $\widetilde{\boldsymbol{\epsilon}} = \boldsymbol{\Xi}^{-1/2}\boldsymbol{\epsilon}$. The critical idea in Du et al. (2021) is to use least-squares (LS) estimates on a narrower subset of components, say $\widehat{\mathcal{S}}$, that captures nearly all nonzero signals, to replace the counterparts in the original control statistics. The LS estimates decorrelate the dependence and, in general, result in a higher signal-to-noise ratio (SNR). Hence, dependence becomes a blessing in large-scale multiple

Statistica Sinica: Preprint
doi:10.5705/ss.202022.0054

testing problems.

Tailored for our scenario, we construct $\widehat{\mathcal{S}}$ using a simple thresholding rule,

$$\widehat{\mathcal{S}} = \left\{1 \le j \le p : |\xi_j^{(1)}|/\widehat{\sigma}_{jj}^{1/2} \ge t_p\right\}, \tag{2.6}$$

to screen out some inactivated data sequences, with some threshold value t_p (specified in Section 4.1), so that activated data sequences with large signals tend to be selected (see Lemma S2.6 of the Supplementary Material), where $\widehat{\sigma}_{jj}$ are estimates of σ_{jj}, for $j = 1, \ldots, p$, specified in Section 4.3. Once $\widehat{\mathcal{S}}$ is selected, the LS solutions are obtained by minimizing $\|\mathbf{y}_2 - \mathbf{X}_{\widehat{\mathcal{S}}}\boldsymbol{\beta}_{\widehat{\mathcal{S}}}\|_2^2$, that is, $\widetilde{\beta}_j := (\mathbf{e}_j)_{\widehat{\mathcal{S}}}^\top (\mathbf{X}_{\widehat{\mathcal{S}}}^\top \mathbf{X}_{\widehat{\mathcal{S}}})^{-1} \mathbf{X}_{\widehat{\mathcal{S}}}^\top \mathbf{y}_2$, for $j \in \widehat{\mathcal{S}}$, where $\mathbf{e}_j$ is a p-tuple with all components being zero, except the jth, which is one. Let $\widetilde{\beta}_j = 0$, for $j \notin \widehat{\mathcal{S}}$. If the activation set $\mathcal{A}$ is covered in the selected set $\widehat{\mathcal{S}}$, then $\mathbb{E}(\widetilde{\beta}_j \mid \mathcal{Z}_1) = \mathbb{E}(\xi_j^{(2)} \mid \mathcal{Z}_1)$ and

$$\begin{aligned} \mathrm{Cov}(\widetilde{\beta}_j, \widetilde{\beta}_k \mid \mathcal{Z}_1) &= \mathrm{Cov}(\xi_j^{(2)}, \xi_k^{(2)} \mid \mathcal{Z}_1) - (\mathbf{e}_j)_{\widehat{\mathcal{S}}}^\top \boldsymbol{\Xi}_{\widehat{\mathcal{S}},\widehat{\mathcal{S}}^c} \boldsymbol{\Xi}_{\widehat{\mathcal{S}}^c,\widehat{\mathcal{S}}^c}^{-1} \boldsymbol{\Xi}_{\widehat{\mathcal{S}}^c,\widehat{\mathcal{S}}} (\mathbf{e}_k)_{\widehat{\mathcal{S}}} \\ &< \mathrm{Cov}(\xi_j^{(2)}, \xi_k^{(2)} \mid \mathcal{Z}_1), \end{aligned} \tag{2.7}$$

which motivates us to replace $\xi_j^{(2)}$ in the original activation statistics $W_{j,\mathrm{indep}}$ with $\widetilde{\beta}_j$ to increase the SNR. However, the LS solutions $\widetilde{\beta}_j$ cannot be used directly, because they depend on the unknown $\boldsymbol{\Sigma}$ via $\boldsymbol{\Xi}$, because $\boldsymbol{\Xi} = \mathbf{J} \circ \boldsymbol{\Sigma}$,

Statistica Sinica: Preprint
doi:10.5705/ss.202022.0054

where $\mathbf{J}$ is a $p \times p$ matrix with (j,k)th element

$$J_{jk} := \sqrt{\frac{(\widehat{\tau}_j^{(2)} \wedge \widehat{\tau}_k^{(2)})\{T_2 - \widehat{\tau}_j^{(2)} \vee \widehat{\tau}_k^{(2)}\}}{(\widehat{\tau}_j^{(2)} \vee \widehat{\tau}_k^{(2)})\{T_2 - \widehat{\tau}_j^{(2)} \wedge \widehat{\tau}_k^{(2)}\}}}, \tag{2.8}$$

for $1 \leq j,k \leq p$. Suppose we have a good estimate of $\mathbf{\Sigma}$, say $\widehat{\mathbf{\Sigma}}$, obtained based on the data split $\mathcal{Z}_1$ (see Section 4.3). A remedy is to use the plugged-in counterparts $\widehat{\beta}_j := (\mathbf{e}_j)_{\widehat{\mathcal{S}}}^\top (\widehat{\mathbf{X}}_{\widehat{\mathcal{S}}}^\top \widehat{\mathbf{X}}_{\widehat{\mathcal{S}}})^{-1} \widehat{\mathbf{X}}_{\widehat{\mathcal{S}}}^\top \widehat{\mathbf{y}}_2$, for $j \in \widehat{\mathcal{S}}$, where $\widehat{\mathbf{X}} = \widehat{\mathbf{\Xi}}^{-1/2}$, $\widehat{\mathbf{y}}_2 = \widehat{\mathbf{\Xi}}^{-1/2} \boldsymbol{\xi}^{(2)}$, and $\widehat{\mathbf{\Xi}} = \mathbf{J} \circ \widehat{\mathbf{\Sigma}}$. Hence, we propose the activation statistics as

$$W_{j,\mathrm{SLIP}} = \begin{cases} \left(\xi_j^{(1)} / \widehat{\sigma}_{jj}^{1/2}\right) \times \left(\widehat{\beta}_j / \widehat{V}_{jj}^{1/2}\right), & j \in \widehat{\mathcal{S}}, \\ 0, & j \notin \widehat{\mathcal{S}}, \end{cases} \tag{2.9}$$

where $\widehat{V}_{jj} = (\mathbf{e}_j)_{\widehat{\mathcal{S}}}^\top (\widehat{\mathbf{X}}_{\widehat{\mathcal{S}}}^\top \widehat{\mathbf{X}}_{\widehat{\mathcal{S}}})^{-1} (\mathbf{e}_j)_{\widehat{\mathcal{S}}}$ is the plugged-in estimate of $\mathrm{Var}(\widehat{\beta}_j \mid \mathcal{Z}_1)$, for $j \in \widehat{\mathcal{S}}$. The FDR threshold is thus determined according to the rule in Eq. (1.3), say $\widehat{L}_{\mathrm{SLIP}} = \widehat{L}(W_{1,\mathrm{SLIP}}, \ldots, W_{p,\mathrm{SLIP}})$. Consequently, the identified set of activated data sequences is given by $\widehat{\mathcal{A}}_{\mathrm{SLIP}} = \{j : W_{j,\mathrm{SLIP}} \geq \widehat{L}_{\mathrm{SLIP}}\}$.

3. Asymptotic properties

In this section, we investigate the validity of the proposed SLIP method in terms of FDR control in the asymptotic sense, that is, both T and p jointly diverge to infinity. Let $p_1 = |\mathcal{A}|$ and $p_0 = p - p_1$ be the numbers of activated and inactivated data sequences, respectively. Let $S = |\widehat{\mathcal{S}}|$ be the number of selected data sequences after screening, and $S_0 = |\widehat{\mathcal{S}} \backslash \mathcal{A}|$ be the number of

Statistica Sinica: Preprint
doi:10.5705/ss.202022.0054

selected data sequences that are still inactivated. Let $\delta_j^* = \mu_{\tau_j^*+1,j} - \mu_{\tau_j^*,j}$, for $j \in \mathcal{A}$, be the values of the activation signals (changes of mean levels) for activated data sequences, and set $\delta_j^* = 0$ for inactivated ones (i.e., $j \notin \mathcal{A}$). We focus on scenarios in which $\delta_j^* = O(1)$ uniformly, for all $j = 1, \ldots, p$. Let $\varepsilon_{ij}^{(k)} = Z_{ij}^{(k)} - \mathbb{E}(Z_{ij}^{(k)})$ be the random errors, for $k = 1, 2$, $j = 1, \ldots, p$, and $i = 1, \ldots, T$, and $\boldsymbol{\varepsilon}_i^{(k)} = (\varepsilon_{i1}^{(k)}, \ldots, \varepsilon_{ip}^{(k)})^\top$. Recall that $\mathbf{X} = \boldsymbol{\Xi}^{-1/2}$ and $\widehat{\mathbf{X}} = \widehat{\boldsymbol{\Xi}}^{-1/2}$, where $\boldsymbol{\Xi} = \mathbf{J} \circ \boldsymbol{\Sigma}$ and $\widehat{\boldsymbol{\Xi}} = \mathbf{J} \circ \widehat{\boldsymbol{\Sigma}}$. Denote $\mathbf{H} = (\widehat{\mathbf{X}}_{\widehat{\mathcal{S}}}^\top \widehat{\mathbf{X}}_{\widehat{\mathcal{S}}})^{-1} \widehat{\mathbf{X}}_{\widehat{\mathcal{S}}}^\top \widehat{\mathbf{X}}$. For $1 \le j \ne k \le |\widehat{\mathcal{S}}|$, let $V_{jk} = (\mathbf{e}_j)_{\widehat{\mathcal{S}}}^\top \mathrm{Var}(\widehat{\boldsymbol{\beta}}_{\widehat{\mathcal{S}}} \mid \mathcal{Z}_1)(\mathbf{e}_k)_{\widehat{\mathcal{S}}}$ and $R_{jk} = V_{jk}/\sqrt{V_{jj}V_{kk}}$.

Assumption 1 (Activation times)**.** There exists some constant $c_\tau \in (0, 1/2)$, such that $c_\tau \le \tau_j^*/T \le 1 - c_\tau$ uniformly, for $j \in \mathcal{A}$.

Assumption 2 (Activation signals)**.** There exists a partition of $\mathcal{A}$, that is, $\mathcal{A} = \mathcal{A}_* \cup (\mathcal{A} \backslash \mathcal{A}_*)$, such that, as $T, p \to \infty$, $\max_{j \in \mathcal{A} \backslash \mathcal{A}_*} T(\delta_j^*)^2 / \log T = O(1)$ and $\min_{j \in \mathcal{A}_*} T(\delta_j^*)^2 / \log T \to \infty$. Moreover, $p_{1*} := |\mathcal{A}_*| \to \infty$ as $T, p \to \infty$.

Assumption 3 (Thresholding)**.** With probability one, $S \le \bar{s}_p \asymp T^{c_1}$, for some nonrandom sequence $\bar{s}_p$ and constant $c_1 > 0$, $p_{1*} \le S_0/\alpha$, and $(p_1 - p_{1*})/S_0 \to 0$ as $T, p \to \infty$.

Assumption 4 (Random errors)**.** There exist two sequences $m_{p1} > 0$ and $m_{p2} > 0$, such that $\mathbb{E}\{\|\boldsymbol{\varepsilon}_1^{(1)}\|_\infty^\theta\} \le m_{p1}^\theta$ and $\mathbb{E}\{\|\mathbf{H}\boldsymbol{\varepsilon}_1^{(2)}\|_\infty^\theta \mid \mathbf{H}\} \le m_{p2}^\theta$, for some constant $\theta > 2$. In addition, as $T \to \infty$, $\sqrt{\log(Tp_{1*})} T^{-1/2+1/\theta+\epsilon_1} m_{p1} \to$

Statistica Sinica: Preprint
doi:10.5705/ss.202022.0054

0, and $\bar{s}_p^{c_2} T^{-1/2+1/\theta+\epsilon_1} m_{p2} \to 0$, for some sufficiently small constant $\epsilon_1 > 0$ and some constant $c_2 > 0$.

Assumption 5 (Covariance)**.** (i) There exist two constants $c_{\overline{\kappa}}$ and $c_{\underline{\kappa}}$, such that $0 < c_{\underline{\kappa}} \leq \lambda_{\min}(\mathbf{X}_{\widehat{\mathcal{S}}}^\top \mathbf{X}_{\widehat{\mathcal{S}}}) \leq \lambda_{\max}(\mathbf{X}_{\widehat{\mathcal{S}}}^\top \mathbf{X}_{\widehat{\mathcal{S}}}) \leq c_{\overline{\kappa}}$ holds with probability approaching one as $T, p \to \infty$; (ii) There exist two constants $c_{\underline{\sigma}}$ and $c_{\overline{\sigma}}$, such that $0 < c_{\underline{\sigma}} \leq \sigma_{jj} \leq c_{\overline{\sigma}}$ uniformly, for $j = 1, \ldots, p$; (iii) There exists a sequence $r_p > 0$, such that, for $j \in \widehat{\mathcal{S}} \backslash \mathcal{A}$, $|\mathcal{R}_j| \leq r_p$, where $\mathcal{R}_j = \left\{ k \in \widehat{\mathcal{S}} \backslash \mathcal{A} : |R_{jk}| \geq c_3 (\log T)^{-2-\epsilon_2} \right\}$, for some constant $c_3 > 0$ and sufficiently small constant $\epsilon_2 > 0$; in addition, $r_p / p_{1*} \to 0$ as $T, p \to \infty$.

Assumption 6 (Accuracy of covariance matrix estimation)**.** If $p_{1*} = p_1$, there exists a sequence $\omega_p > 0$, such that $\|\widehat{\mathbf{\Sigma}} - \mathbf{\Sigma}\|_2 < \omega_p$, where ω_p satisfies $\omega_p / \min\{1, \lambda_{\min}^2(\mathbf{\Sigma})\} \to 0$ as $T, p \to \infty$. If $p_{1*} < p_1$, assume $\|(\mathbf{X}_{\widehat{\mathcal{S}}}^\top \mathbf{X}_{\widehat{\mathcal{S}}})^{-1} \mathbf{X}_{\widehat{\mathcal{S}}}^\top \mathbf{X}_{\widehat{\mathcal{S}}^c} \boldsymbol{\beta}_{\widehat{\mathcal{S}}^c}\|_\infty = O_p(\sqrt{\log \bar{s}_p})$, and there exist two sequences $u_p > 0$ and $\omega_p > 0$, such that $\|\mathbf{X}_{\widehat{\mathcal{S}}}^\top \mathbf{X}_{\widehat{\mathcal{S}}^c}\|_1 \leq u_p$ and $\|\widehat{\mathbf{\Sigma}} - \mathbf{\Sigma}\|_2 < \omega_p$, with $(p_1 - p_{1*}) u_p \omega_p / \min\{1, \lambda_{\min}^2(\mathbf{\Sigma})\} \to 0$ as $T, p \to \infty$.

Assumption 1 requires that coordinate-wise activation times are not at the boundaries, which is frequently considered in the literature on change-point detection (Csörgő and Horváth, 1997; Bai, 2010). Assumption 2 separates the set of activated signals into two parts, $\mathcal{A}_*$ and $\mathcal{A} \backslash \mathcal{A}_*$, according to the change magnitudes. Jirak (2015) discusses the consistent selection of the

Statistica Sinica: Preprint
doi:10.5705/ss.202022.0054

set of activated data sequences when $p_1 = p_{1*}$, that is, $T(\delta_j^*)^2/\log T \to \infty$, for all $\delta_j \neq 0$. Valid FDR control can be achieved, even allowing for some weaker signals (i.e., $p_{1*} < p_1$). Assumption 3 restricts the size of the selected set $\mathcal{S}$, which can be guaranteed by selecting a large $\widehat{\mathcal{S}}$. For example, we can choose the components corresponding to the first $d = \lfloor T/3 \rfloor$ largest $|\xi_j^{(1)}|/\widehat{\sigma}_{jj}^{1/2}$ in practice, motivated by the variable-screening procedures (Fan and Lv, 2008). Assumption 4 places moment constraints on the random noise, which can be verified for sub-Gaussian noise. Assumption 5 imposes restrictions on the dependence structures across different data sequences. In particular, in Assumption 5–(iii), R_{jk} measures the dependence among the de-correlated (by LS estimates) data sequences, in which there should not be many data sequences with strong correlations (Du et al., 2021; Xia et al., 2020). Assumption 6 places restrictions on the estimated covariance matrix. To better illustrate Assumptions 5–6, we consider a special scenario in which all changes can happen only at a common change-point, which is widely investigated in the literature on high-dimensional change-point detection (Bai, 2010; Cho and Fryzlewicz, 2015; Wang and Samworth, 2018); correspondingly, we obtain a certain common change-point estimator.

Lemma 1. *Assume that all changes can happen only at a common change-point that is not at the boundary, and that the signal magnitude satisfies*

Statistica Sinica: Preprint
doi:10.5705/ss.202022.0054

$T(\delta_j^*)^2/\log T \to \infty$, *for* $\delta_j^* \neq 0$. *Assumptions 5–6 are satisfied, provided that* $\mathbf{\Sigma}$ *has uniformly bounded eigenvalues,* $|\mathcal{R}_j| \leq r_p$, *for* $j \in \widehat{\mathcal{S}} \setminus \mathcal{A}$ *satisfying* $r_p/p_1 \to 0$, *and* $\|\widehat{\mathbf{\Sigma}} - \mathbf{\Sigma}\|_2 \to 0$ *as* $T, p \to \infty$.

Remark 1. The condition involved requires a consistent estimation of $\mathbf{\Sigma}$. If the change-points can be precisely recovered, we can use the centered data to estimate $\mathbf{\Sigma}$ by using state-of-the-art high-dimensional covariance estimation procedures (Bickel and Levina, 2008b; Fan et al., 2013; Cai and Liu, 2011). If the change-points are consistently estimated with a fine precision, we conjecture that this conclusion also holds under some conditions. In fact, high-dimensional covariance matrix recovery in the presence of change-points is challenging even for synchronous change patterns, which deserves further research; see Section 4.3.

Remark 2. If we further permit the existence of weaker signals (i.e., $p_{1*} < p_1$), we discuss how Assumption 6 can be verified by considering two specific covariance structures. Let $U_p := \|\mathbf{X}_{\widehat{\mathcal{S}}}^\top \mathbf{X}_{\widehat{\mathcal{S}}^c}\|_1$ and $V_p := \|(\mathbf{X}_{\widehat{\mathcal{S}}}^\top \mathbf{X}_{\widehat{\mathcal{S}}})^{-1} \mathbf{X}_{\widehat{\mathcal{S}}}^\top \mathbf{X}_{\widehat{\mathcal{S}}^c} \boldsymbol{\beta}_{\widehat{\mathcal{S}}^c}\|_\infty$. It can be shown that (i) $U_p = V_p = 0$ and $\lambda_{\min}(\mathbf{\Sigma}) = 1$ if $\mathbf{\Sigma} = \mathbf{I}_p$, and (ii) $U_p \leq 1$, $V_p \leq (p_1 - p_{1*})(p - \bar{s}_p)^{-1}\sqrt{\log \bar{s}_p}$, and $\lambda_{\min}(\mathbf{\Sigma}) = 1-\rho$ if $\mathbf{\Sigma} = \rho \mathbf{1}\mathbf{1}^\top + (1-\rho)\mathbf{I}$. Consequently, in either case, by choosing $u_p = 1$, Assumption 6 holds, provided that $(p_1 - p_{1*})\|\widehat{\mathbf{\Sigma}} - \mathbf{\Sigma}\|_2 \to 0$ and $(p_1 - p_{1*})/(p - \bar{s}_p) \to c \in [0, +\infty)$ as $T, p \to \infty$.

Statistica Sinica: Preprint
doi:10.5705/ss.202022.0054

Theorem 1. *Under Assumptions 1–6, for any prespecified* $\alpha \in (0, 1)$,

$$\text{FDP} := \frac{\#\{j : W_{j,\text{SLIP}} \geq \widehat{L}_{\text{SLIP}},\ j \notin \mathcal{A}\}}{\#\{j : W_{j,\text{SLIP}} \geq \widehat{L}_{\text{SLIP}}\} \vee 1} \leq \alpha + o_p(1),$$

and consequently $\limsup_{T,p\to\infty} \mathbb{E}(\text{FDP}) \leq \alpha$.

Proposition 1. *Under Assumptions 1–6,* $\Pr(\mathcal{A}_* \subseteq \widehat{\mathcal{A}}_{\text{SLIP}}) \to 1$ *as* $T, p \to \infty$.

Theorem 1 shows that the SLIP method asymptotically controls the false discovery proportion (FDP), and thus the FDR, at the nominal FDR level α. Note that the procedure developed under the independent setting (see Eq. (2.3)) can control the FDR asymptotically under some correlation restrictions (see Assumption 5–(iii)). The current dependence-assisted scheme helps raise the SNR in the sense of Eq. (2.7). Proposition 1 shows that all strong signals can be detected with probability tending to one.

We defer the proofs of Theorem 1 and Proposition 1 to Appendix S2 of the Supplementary Material. Remarkably, the working model in Eq. (2.4) and (2.5) is presented only for a convenient introduction to the SLIP procedure. The approximation error induced by the uncertainty of the change-localizing algorithm is handled delicately in the proof, which makes the theoretical development much more involved than that of Du et al. (2021). In addition, we consider the estimation uncertainty of $\widehat{\mathbf{\Sigma}}$ in the proof.

Statistica Sinica: Preprint
doi:10.5705/ss.202022.0054

4. Practical guidelines

4.1 Tuning parameters

For the data split ratio r, we recommend using $r = 3$, because the data split $\mathcal{Z}_1$ is responsible for identifying the activation times, which has a direct impact on the detection ability. The boundary parameter $\varrho \in (0, 1/2)$ is frequently considered in the literature (Csörgő and Horváth, 1997; Yu and Chen, 2021) to avoid certain boundary problems, and can be set as $\varrho = 0.1$ for practical purposes. The threshold parameter t_p (see Eq. (2.6)) can be chosen as $t_p = \sqrt{C \log T_1}$, with $C = 1.5$, which yields satisfactory performance for both FDR control and power enhancement, as discussed in Section 5.

4.2 Other change patterns

In some task-related fMRI experiments, researchers give multiple stimuli to subjects in sequence, and thus multiple change-points may exist in each data sequence. The proposed method can be extended to multiple change-point scenarios, because it can be thought of as the first step of a binary segmentation change-detection algorithm. It can be shown that the highly "significant" change-point can be recovered consistently in this step under

Statistica Sinica: Preprint
doi:10.5705/ss.202022.0054

some conditions (Fryzlewicz, 2014). Furthermore, variants of the binary segmentation algorithm (Fryzlewicz, 2014; Eichinger and Kirch, 2018) are also feasible. On the other hand, if we have some priors on a specific data distribution, the log-likelihood ratio statistic (Csörgő and Horváth, 1997) can be employed instead of the CUSUM statistic, which yields better change-point estimators.

In experiments in which some brain regions are activated simultaneously, changes may happen in a small part of data sequences simultaneously. If we have such prior knowledge, we can use high-dimensional change-point methodologies (Jirak, 2015; Wang and Samworth, 2018) to identify the common change-point as a single $\widehat{\tau} = \widehat{\tau}_j^{(1)}$, for $j = 1, \ldots, p$, to which we can still apply the SLIP method.

4.3 Estimation routines for the covariance matrix

High-dimensional covariance matrix estimation is a fundamental problem. To achieve consistent estimations, we typically need additional structural assumptions on $\mathbf{\Sigma}$, including banded or sparse assumptions or certain low-dimensional representations, such as factor models; see, for example, Bickel and Levina (2008a), Bickel and Levina (2008b), Friedman et al. (2007), and Fan et al. (2013). However, these methods cannot be applied directly in

Statistica Sinica: Preprint
doi:10.5705/ss.202022.0054

the presence of mean changes, as in our model (1.1).

The difference-based covariance estimators are good choices in the presence of change-points. Consistency can be reached under low-dimensional scenarios if the number of change-points is not too high and the change magnitudes are not too large (Rice, 1984; Chan, 2022). However, for an asynchronous change pattern (1.1), where τ_j^* are not required to be the same, difference-based estimators may bring unnecessary bias accumulation, and thus may be not a good candidate, even under low-dimensional scenarios.

Hence, we turn to a simple change-removing strategy to alleviate the bias induced by change-points. In particular, we centralize each data point in $\mathcal{Z}_1$ by subtracting the estimated mean, that is,

$$\widetilde{Z}_{ij}^{(1)} = Z_{ij}^{(1)} - \mathbb{1}(i \leq \widehat{\tau}_j^{(1)})\bar{Z}_j^{(1)}(0, \widehat{\tau}_j^{(1)}) - \mathbb{1}(i > \widehat{\tau}_j^{(1)})\bar{Z}_j^{(1)}(\widehat{\tau}_j^{(1)}, T_1).$$

Then, we can apply state-of-the-art high-dimensional covariance estimation procedures to the centered data $\{\widetilde{Z}_{ij},\, i = 1, \ldots, T_1,\, j = 1, \ldots, p\}$, which yields satisfactory performance in our numerical studies. Note that the covariance may be underestimated, because some data sequences do not contain a change. We conjecture that a refitted cross-validation strategy (Fan et al., 2012) may alleviate this phenomenon. In fact, high-dimensional covariance estimation in the presence of change-points is challenging, espe-

Statistica Sinica: Preprint
doi:10.5705/ss.202022.0054

cially for asynchronous change patterns, which warrants future research.

4.4 A LASSO-based screening strategy

In Section 2, we provided a thresholding rule t_p (see Eq. (2.6)) to obtain a narrower subset of coordinates, $\widehat{\mathcal{S}}$. Here, we offer another candidate strategy for screening redundant data sequences, upon which the LS strategy can be applied.

We revisit the working model (2.5) for $\boldsymbol{\xi}^{(2)}$. Note that the nonzero components of $\boldsymbol{\beta}$ correspond to activated data sequences. If the number of activated data sequences is not very large, intuitively, we can use a sparse estimate of $\boldsymbol{\beta}$ by using, for example, the LASSO method (Tibshirani, 1996). Specifically, we achieve this based on the data split $\mathcal{Z}_1$, that is,

$$\check{\boldsymbol{\beta}} := \arg\min_{\boldsymbol{\gamma}} \left\{ \frac{1}{2} \|\widehat{\mathbf{y}}_1 - \widehat{\mathbf{X}}\boldsymbol{\gamma}\|_2^2 + \lambda \|\boldsymbol{\gamma}\|_1 \right\},$$

where $\widehat{\mathbf{y}}_1 = \widehat{\boldsymbol{\Xi}}^{-1/2} \boldsymbol{\xi}^{(1)}$ and λ is a tuning parameter that can be determined using cross-validation or some information criteria, such as the AIC. Then, we select the set of nonzero components of $\check{\boldsymbol{\beta}}$ as $\widehat{\mathcal{S}}$. Finally, the activation statistics can be constructed as

$$W_{j,\mathrm{LASSO}} = \begin{cases} \left(\mathbf{e}_j^\top \check{\boldsymbol{\beta}} \times \widehat{\beta}_j \right) / \widehat{V}_{jj} & j \in \widehat{\mathcal{S}}, \\ 0, & j \notin \widehat{\mathcal{S}}. \end{cases} \tag{4.1}$$

Statistica Sinica: Preprint
doi:10.5705/ss.202022.0054

In contrast to the thresholding rule that solely compares components of $\boldsymbol{\xi}^{(1)}$, the LASSO-based screening strategy incorporates the dependence information among all components, to a certain degree. Hence, it should be more efficient when the activation signal is relatively weak. Our numerical results also show that the SLIP method combined with the LASSO strategy performs satisfactorily. A theoretical investigation certainly warrants future research.

4.5 Temporal dependence

In some fMRI data analyses, coordinate-wise temporal correlations exist, especially in contiguous sampling periods. Although the asymptotic FDR validity is established by assuming that the $\mathbf{Z}_i$ are independent, we may expect the proposed SLIP method to be applicable, with slight modifications, for temporally dependent cases. Motivated by the idea of a moving block bootstrap for stationary series (Kunsch, 1989), we suggest a local averaging-based procedure capable of alleviating the effect of auto-correlations, to a certain degree. For each data sequence, we first partition the time-ordered observations into a sequence of blocks of roughly the same length, and then take the average of the observations in each block as a new pseudo observation. These pseudo observations then constitute a new data sequence, in

Statistica Sinica: Preprint
doi:10.5705/ss.202022.0054

which the temporal correlation tends to be weakened. Then, we can apply the proposed SLIP method to the newly constructed data.

4.6 R package

To facilitate the implementation of the proposed SLIP method, we have developed an R package called SLIP, which is available at `https://github.com/MengtaoWen/SLIP`.

5. Experiments

5.1 Synthetic data

To evaluate the performance of the proposed SLIP method in terms of discovering activated data sequences, we first introduce two benchmark procedures by applying the BH method. To test each H_{0j} (see Eq. (1.2)), for $j = 1, \ldots, p$, we adopt the CUSUM statistics based on the entire data, that is,

$$\frac{1}{\widehat{\sigma}_{jj}^{1/2}} \max_{1 \le \tau < n} \sqrt{\frac{\tau(T-\tau)}{T}} \left| \frac{1}{\tau} \sum_{i=1}^{\tau} Z_{ij} - \frac{1}{T-\tau} \sum_{i=\tau+1}^{T} Z_{ij} \right|, \tag{5.1}$$

which converges in distribution to an extreme value distribution under H_{0j} (Csörgő and Horváth, 1997). Hence, we can calculate the associated p-values using the asymptotic null distribution. We refer to this benchmark

Statistica Sinica: Preprint
doi:10.5705/ss.202022.0054

procedure as *BH-asymp*. Jirak (2015) proposes a simulation-based approach that mimics the null distribution by generating the data $Z_{ij} \overset{\text{iid}}{\sim} N(0,1)$, and recalculates the CUSUM statistics in Eq. (5.1) repeatedly. We refer to this procedure as *BH-simul*. For our method, we consider three variants: the SLIP method with activation statistics $W_{j,\text{indep}}$ in Eq. (2.3), ignoring the dependence information (see Section 2.1); the SLIP method with activation statistics $W_{j,\text{SLIP}}$, together with the thresholding rule (see Eq. (2.9)); and the SLIP method with activation statistics $W_{j,\text{LASSO}}$, combined with the LASSO screening strategy (see Section 4.4). These three procedures are named *SLIP-indep*, *SLIP-thresh*, and *SLIP-lasso*, respectively, for notational convenience. The estimation of the unknown covariance matrix follows the guidelines in Section 4.3, together with some state-of-the-art covariance matrix estimation routines, specified later. We implement these procedures in the R package SLIP to facilitate practical use. We set the nominal FDR level $\alpha = 20\%$. We conduct 500 replications to estimate the FDR and the power of each procedure, where the latter is defined as the proportion of discovered activated data sequences to all activated data sequences.

It is well known that the validity of the BH method is guaranteed if the p-values are independent. Hence, we first conduct a simulation study by

Statistica Sinica: Preprint
doi:10.5705/ss.202022.0054

setting $\mathbf{\Sigma} = \mathbf{I}$ to investigate the performance of the considered procedures. For now, we assume $\mathbf{\Sigma}$ is known. The random errors are generated following (i) a normal distribution, and (ii) a t-distribution with degrees of freedom five. We set the proportion of activated data sequences $p_1 = \lfloor 0.15p \rfloor$, the indices of which are chosen randomly from $\{1, \ldots, p\}$. For each activated data sequence, the activation time τ_j^* is sampled randomly from $\{\lfloor T\varrho \rfloor + 1, \ldots, T - 1 - \lfloor T\varrho \rfloor\}$, with $\varrho = 0.05$, and the change magnitude δ_j^* is first sampled uniformly from the interval $[\delta - 0.1, \delta + 0.1]$ with $\delta > 0.1$, and then its sign is flipped with probability 0.5, where δ is a parameter controlling the signal strength.

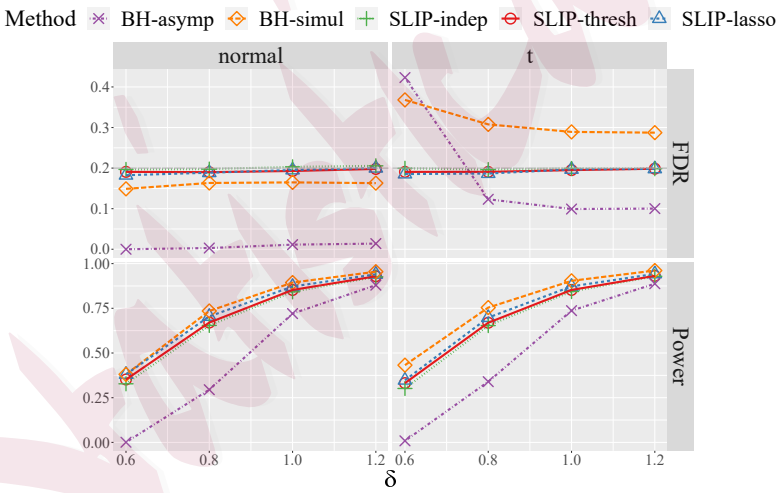

Figure 1: Empirical FDR and power of the SLIP and BH methods when $(T, p) = (120, 800)$ under the independent scenario with $\mathbf{\Sigma} = \mathbf{I}$.

Statistica Sinica: Preprint
doi:10.5705/ss.202022.0054

Figure 1 depicts the empirical FDR and the power of the SLIP and BH methods when $(T, p) = (120, 800)$ under the independent scenario, where δ takes values from $\{0.6, 0.8, 1.0, 1.2\}$. When the data are normally distributed, we observe that all procedures maintain the nominal FDR level and have comparable power, but that BH-asymp exhibits conservative performance (it makes fewer discoveries than the others do). When the data are from a t-distribution, BH-asymp is still conservative, and fails to control the FDR, even for low signal scenarios, which may be due to the quite slow convergence to the asymptotic null distribution. BH-simul has inflated FDR levels, and thus unnecessarily higher power, which is expected, because it uses normally distributed samples to approximate the null distribution, and thus is sensitive to the normality of the data distribution. In contrast, the proposed SLIP procedures perform very well in terms of both FDR and power, as in the normal scenario. Moreover, the performance difference between SLIP-indep, SLIP-thresh, and SLIP-lasso is negligible.

To investigate the impact of the dependence structure, we conduct simulation experiments with different patterns of covariance matrices: Scenario-(i) is a compound symmetric matrix, the elements of which are all ρ, except the diagonal elements, which are equal to one; and Scenario-(ii) is a matrix with a first-order autoregressive structure, in which the (i, j)th element is

Statistica Sinica: Preprint
doi:10.5705/ss.202022.0054

$\rho^{|i-j|}$, where ρ in both cases controls the degree of the correlations. The data-generating process is the same as earlier, with the random error distributed as a t-distribution with degrees of freedom five. For Scenario-(i), the covariance matrix estimation routine is chosen as the POET proposed by Fan et al. (2013), and for Scenario-(ii), it is chosen as the thresholding rule in Lee and Lee (2021). The left panel of Figure 2 shows the empirical

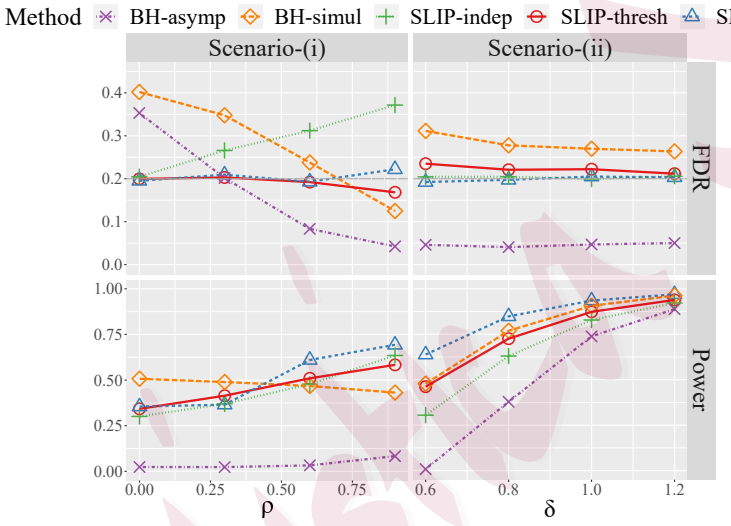

Figure 2: Empirical FDR and power of the SLIP and BH methods when $(T, p) = (120, 800)$ under the dependent scenario, where $\mathbf{\Sigma}$ is specified as either a compound symmetric (i) or autoregressive (ii) structure.

FDR and power of the SLIP and BH methods when $(T, p) = (120, 800)$, $\delta = 0.6$, and $\mathbf{\Sigma}$ is chosen as in Scenario-(i), with ρ ranging over the val-

Statistica Sinica: Preprint
doi:10.5705/ss.202022.0054

ues $\{0, 0.3, 0.6, 0.9\}$. The right panel of Figure 2 shows the results when $(T, p) = (120, 800)$, $\mathbf{\Sigma}$ is chosen as in Scenario-(ii), with $\rho = 0.8$, and δ is selected from the values $\{0.6, 0.8, 1.0, 1.2\}$. Again, BH-asymp is quite conservative, and BH-simul has inflated FDR levels, in most cases. SLIP-indep also has inflated FDR levels when the correlations among the data sequences are strong, for example, in (i). In contrast, SLIP-thresh and SLIP-lasso perform well, with good power, while maintaining the nominal FDR level. In addition, SLIP-lasso sometimes has slightly better power than that of SLIP-thresh, because it uses dependence information to construct the narrow set containing activated data sequences.

5.2 An fMRI data analysis

As an illustrative example, we apply the SLIP method to analyze the task-related fMRI data in Mitchell et al. (2008), who study the brain activation associated with thinking about concrete nouns to examine how the human brain represents and organizes conceptual knowledge. The stimuli are line drawings and noun labels of some concrete objects. Each stimulus item is presented, and the participant is instructed to think about the object's properties, followed by a rest period. A sequence of 360 images of the participant was collected and processed, with each brain image containing

Statistica Sinica: Preprint
doi:10.5705/ss.202022.0054

about 21,000 voxels. These voxels were divided into 90 ROIs (Tzourio-Mazoyer et al., 2002), generally believed to be anatomically and functionally distinct. The ROIs vary greatly in size. Following the strategy of Wehbe et al. (2015), to achieve size uniformity, we further divide the ROIs into regions with 100 voxels or fewer, yielding $p = 264$ ROIs. We also discard five images from the data sequence to ensure there is no signal leakage due to the slow decay of the hemodynamic responses (Wehbe et al., 2015). Then, we use the average of every 10 consecutive images as our data for analysis to mitigate potential temporal correlations (see Section 4.5), and thus $T = 35$. The processed data are available in the SLIP package. We estimate the covariance matrix using the POET routine (Fan et al., 2013), and set the nominal FDR level at $\alpha = 20\%$ to implement the SLIP method.

Figure 3 presents a sliced brain map of the estimated component-wise change magnitude, scaled by the estimated component-wise variance, discovered by SLIP-thresh and SLIP-lasso, respectively. A large absolute value in the map indicates a very likely change. The two procedures discover 10 and 16 ROIs, respectively, and have six in common. The boundaries of the discovered ROIs are marked in red, and the names of the discovered ROIs are listed in Appendix S4 of the Supplementary Material (Table S1). We observe that the detected regions correspond to those with larger absolute

Statistica Sinica: Preprint
doi:10.5705/ss.202022.0054

values, especially for the commonly discovered ones. The two procedures differ in a few regions, with relatively more minor absolute values. These may be false positives, and need further examination.

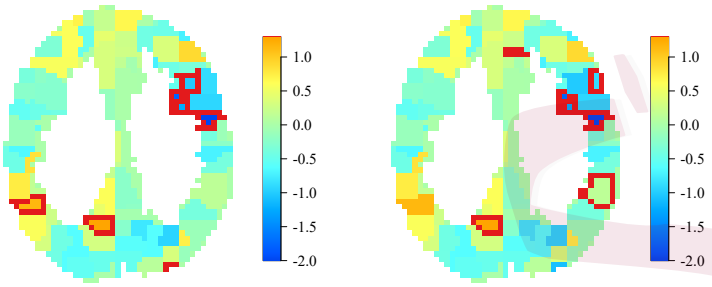

Figure 3: A sliced map of the estimated component-wise change magnitude scaled by the estimated component-wise variance, discovered by SLIP-thresh (*left*) and SLIP-lasso (*right*). The boundaries of the discovered ROIs are marked in red.

In addition, an interesting, yet reasonable phenomenon is that contiguous regions tend to be activated at the same time. Other slices of the brain maps are collected in Appendix S4 of the Supplementary Material (Figures S3–S4), revealing a similar pattern. Whether such "clustering" information can be used to enhance the detection ability, while still guaranteeing some notion of the FDR, warrants further research.

Statistica Sinica: Preprint
doi:10.5705/ss.202022.0054

6. Conclusion

We have investigated the uncertainty in selecting activated data sequences that encounter asynchronous changes by leveraging the FDR tool. We propose the SLIP method to guarantee the FDR being controlled at a prescribed level, while also incorporating potential dependence structures. There is still some room for improvement. For instance, empirical numerical results reveal that the identified regions are usually spatially contiguous. Such information of spatially structured data sequences may improve the power, for example, by using some local aggregation strategy (Zhang et al., 2011). In addition, it would be interesting to extend the proposed method beyond mean-level changes. For example, a correlation-level change model is more suitable for studying the connectivity between distinct brain regions (Xia and Li, 2017). Another interesting direction is to estimate a high-dimensional covariance matrix in the presence of change-points, especially for the current asynchronous change patterns. Tailoring the difference- and change-removing-based procedures to achieve this purpose warrants future research.

Statistica Sinica: Preprint
doi:10.5705/ss.202022.0054

Supplementary Material

The online Supplementary Material includes proofs of all our theoretical results, as well as some additional numerical experiments.

Acknowledgments

The authors contributed equally to this work. The authors would like to acknowledge the editor, associate editor, and three referees for their constructive comments and suggestions. Wang, G. was supported by the National Key R&D Program of China (No. 2021YFA1000100, 2021YFA1000101) and NNSF of China Grant (No. 11901314). Wen, M., Zou, C., and Wang, Z. were supported by NNSF of China Grants (Nos. 11925106, 12231011, 11931001, and 11971247), the China National Key R&D Program (Grant Nos. 2019YFC1908502, 2022YFA1003703, 2022YFA1003802, 2022YFA1003803), and NSF of Tianjin Grant (No. 18JCJQJC46000).

References

Aston, J. A. D. and C. Kirch (2012). Evaluating stationarity via change-point alternatives with applications to fMRI data. *The Annals of Applied Statistics 6*(4), 1906–1948.

Bai, J. (2010). Common breaks in means and variances for panel data. *Journal of Econometrics 157*(1), 78–92.

Statistica Sinica: Preprint
doi:10.5705/ss.202022.0054

REFERENCES

Barber, R. F. and E. J. Candès (2015). Controlling the false discovery rate via knockoffs. *The Annals of Statistics 43*(5), 2055–2085.

Benjamini, Y. and Y. Hochberg (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society: Series B (Methodological) 57*(1), 289–300.

Benjamini, Y. and D. Yekutieli (2001). The control of the false discovery rate in multiple testing under dependency. *The Annals of Statistics 29*(4), 1165–1188.

Bickel, P. J. and E. Levina (2008a, 12). Covariance regularization by thresholding. *The Annals of Statistics 36*(6), 2577–2604.

Bickel, P. J. and E. Levina (2008b). Regularized estimation of large covariance matrices. *The Annals of Statistics 36*(1), 199–227.

Cai, T. and W. Liu (2011). Adaptive thresholding for sparse covariance matrix estimation. *Journal of the American Statistical Association 106*(494), 672–684.

Capizzi, G. (2015). Recent advances in process monitoring: Nonparametric and variable-selection methods for phase i and phase ii. *Quality Engineering 27*(1), 44–67.

Chan, K. W. (2022). Optimal difference-based variance estimators in time series: A general framework. *The Annals of Statistics 50*(3), 1376 – 1400.

Cho, H. and P. Fryzlewicz (2015). Multiple-change-point detection for high dimensional time series via sparsified binary segmentation. *Journal of the Royal Statistical Society: Series B (Statistical Methodology) 77*(2), 475–507.

Statistica Sinica: Preprint
doi:10.5705/ss.202022.0054

REFERENCES

Clarke, S. and P. Hall (2009). Robustness of multiple testing procedures against dependence. *The Annals of Statistics 37*(1), 332–358.

Csörgő, M. and L. Horváth (1997). *Limit theorems in change-point analysis*. Wiley Series in Probability and Statistics. John Wiley & Sons, Ltd., Chichester.

Damoiseaux, J. S., S. A. R. B. Rombouts, F. Barkhof, P. Scheltens, C. J. Stam, S. M. Smith, et al. (2006). Consistent resting-state networks across healthy subjects. *Proceedings of the National Academy of Sciences 103*(37), 13848–13853.

Du, L., X. Guo, W. Sun, and C. Zou (2021). False discovery rate control under general dependence by symmetrized data aggregation. *Journal of the American Statistical Association*, 1–15. doi:10.1080/01621459.2021.1945459.

Efron, B. (2007). Correlation and large-scale simultaneous significance testing. *Journal of the American Statistical Association 102*(477), 93–103.

Eichinger, B. and C. Kirch (2018). A MOSUM procedure for the estimation of multiple random change points. *Bernoulli 24*(1), 526 – 564.

Fan, J., S. Guo, and N. Hao (2012). Variance estimation using refitted cross-validation in ultrahigh dimensional regression. *Journal of the Royal Statistical Society: Series B (Statistical Methodology) 74*(1), 37–65.

Fan, J., X. Han, and W. Gu (2012). Estimating false discovery proportion under arbitrary covariance dependence. *Journal of the American Statistical Association 107*(499), 1019–1035.

Statistica Sinica: Preprint
doi:10.5705/ss.202022.0054

REFERENCES

Fan, J., Y. Liao, and M. Mincheva (2013). Large covariance estimation by thresholding principal orthogonal complements. *Journal of the Royal Statistical Society: Series B (Statistical Methodology) 75*(4), 603–680.

Fan, J. and J. Lv (2008). Sure independence screening for ultrahigh dimensional feature space. *Journal of the Royal Statistical Society: Series B (Statistical Methodology) 70*(5), 849–911.

Fithian, W., D. Sun, and J. Taylor (2014). Optimal inference after model selection.

Friedman, J., T. Hastie, and R. Tibshirani (2007, 12). Sparse inverse covariance estimation with the graphical lasso. *Biostatistics 9*(3), 432–441.

Fryzlewicz, P. (2014). Wild binary segmentation for multiple change-point detection. *The Annals of Statistics 42*(6), 2243 – 2281.

Fryzlewicz, P. (2020). Narrowest significance pursuit: inference for multiple change-points in linear models. *arXiv preprint.* arXiv:2009.05431.

Genovese, C. R., N. A. Lazar, and T. Nichols (2002). Thresholding of statistical maps in functional neuroimaging using the false discovery rate. *NeuroImage 15*(4), 870–878.

Hao, N., Y. S. Niu, and H. Zhang (2013). Multiple change-point detection via a screening and ranking algorithm. *Statistica Sinica 23*(4), 1553–1572.

Hyun, S., M. G'Sell, and R. J. Tibshirani (2018). Exact post-selection inference for the generalized lasso path. *Electronic Journal of Statistics 12*(1), 1053 – 1097.

Hyun, S., K. Z. Lin, M. G'Sell, and R. J. Tibshirani (2021). Post-selection inference for changepoint detection algorithms with application to copy number variation data. *Bio-*

Statistica Sinica: Preprint
doi:10.5705/ss.202022.0054

metrics 77(3), 1037–1049.

Jewell, S., P. Fearnhead, and D. Witten (2022). Testing for a change in mean after changepoint detection. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 1 – 23. doi:10.1111/rssb.12501.

Jirak, M. (2015). Uniform change point tests in high dimension. *The Annals of Statistics 43*(6), 2451–2483.

Kriegeskorte, N., R. Goebel, and P. Bandettini (2006). Information-based functional brain mapping. *Proceedings of the National Academy of Sciences 103*(10), 3863–3868.

Kunsch, H. R. (1989). The jackknife and the bootstrap for general stationary observations. *The Annals of Statistics 17*(3), 1217 – 1241.

Lee, J. D., D. L. Sun, Y. Sun, and J. E. Taylor (2016). Exact post-selection inference, with application to the lasso. *The Annals of Statistics 44*(3), 907 – 927.

Lee, K. and J. Lee (2021). Estimating large precision matrices via modified Cholesky decomposition. *Statistica Sinica 31*, 173–196.

Li, H., A. Munk, and H. Sieling (2016). FDR-control in multiscale change-point segmentation. *Electronic Journal of Statistics 10*(1), 918–959.

Lindquist, M. A., C. Waugh, and T. D. Wager (2007). Modeling state-related fMRI activity using change-point theory. *NeuroImage 35*(3), 1125–1141.

Mazur, M., M. Dang, and M. Vega (2021). Covid-19 and the march 2020 stock market crash: Evidence from s&p500. *Finance Research Letters 38*, 101690.

Statistica Sinica: Preprint
doi:10.5705/ss.202022.0054

REFERENCES

Mitchell, T. M., S. V. Shinkareva, A. Carlson, K.-M. Chang, V. L. Malave, R. A. Mason, et al. (2008). Predicting human brain activity associated with the meanings of nouns. *Science 320*(5880), 1191–1195.

Nichols, T. E. and A. P. Holmes (2002). Nonparametric permutation tests for functional neuroimaging: A primer with examples. *Human Brain Mapping 15*(1), 1–25.

Ogawa, S., T.-M. Lee, A. S. Nayak, and P. Glynn (1990). Oxygenation-sensitive contrast in magnetic resonance image of rodent brain at high magnetic fields. *Magnetic resonance in medicine 14*(1), 68–78.

Rice, J. (1984). Bandwidth Choice for Nonparametric Regression. *The Annals of Statistics 12*(4), 1215 – 1230.

Storey, J. D., J. E. Taylor, and D. Siegmund (2004). Strong control, conservative point estimation and simultaneous conservative consistency of false discovery rates: a unified approach. *Journal of the Royal Statistical Society: Series B (Statistical Methodology) 66*(1), 187–205.

Sugiyama, R., H. Toda, V. N. L. Duy, Y. Inatsu, and I. Takeuchi (2021). Valid and exact statistical inference for multi-dimensional multiple change-points by selective inference. *arXiv preprint*. arXiv:2110.08989.

Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society: Series B (Methodological) 58*(1), 267–288.

Tzourio-Mazoyer, N., B. Landeau, D. Papathanassiou, F. Crivello, O. Etard, N. Delcroix, B. Mazoyer, and M. Joliot (2002). Automated anatomical labeling of activations in spm using a

Statistica Sinica: Preprint
doi:10.5705/ss.202022.0054

macroscopic anatomical parcellation of the mni mri single-subject brain. *NeuroImage 15*(1), 273–289.

Wang, G., C. Zou, and P. Qiu (2022). Data-driven determination of the number of jumps in regression curves. *Technometrics 64*(3), 312–322.

Wang, G., C. Zou, and G. Yin (2018). Change-point detection in multinomial data with a large number of categories. *The Annals of Statistics 46*(5), 2020–2044.

Wang, T. and R. J. Samworth (2018). High dimensional change point estimation via sparse projection. *Journal of the Royal Statistical Society: Series B (Statistical Methodology) 80*(1), 57–83.

Wehbe, L., A. Ramdas, R. C. Steorts, and C. R. Shalizi (2015). Regularized brain reading with shrinkage and smoothing. *The Annals of Applied Statistics 9*(4), 1997 – 2022.

Xia, Y., T. T. Cai, and W. Sun (2020). Gap: A general framework for information pooling in two-sample sparse inference. *Journal of the American Statistical Association 115*(531), 1236–1250.

Xia, Y. and L. Li (2017). Hypothesis testing of matrix graph model with application to brain connectivity analysis. *Biometrics 73*(3), 780–791.

Yu, M. and X. Chen (2021). Finite sample change point inference and identification for high-dimensional mean vectors. *Journal of the Royal Statistical Society: Series B (Statistical Methodology) 83*(2), 247–270.

Zhang, C., J. Fan, and T. Yu (2011). Multiple testing via FDRL for large-scale imaging data.

Statistica Sinica: Preprint
doi:10.5705/ss.202022.0054

The Annals of Statistics 39(1), 613–642.

Zou, C. and P. Qiu (2009). Multivariate Statistical Process Control Using LASSO. *Journal of the American Statistical Association 104*(488), 1586–1596.

Zou, C., G. Wang, and R. Li (2020). Consistent selection of the number of change-points via sample-splitting. *The Annals of Statistics 48*(1), 413–439.

Mengtao Wen, School of Statistics and Data Science, LPMC, KLMDASR and LEBPS, Nankai University

E-mail: mtwen97@gmail.com

Guanghui Wang, School of Statistics, Academy of Statistics and Interdisciplinary Sciences, and KLATASDS-MOE, East China Normal University

E-mail: ghwang.nk@gmail.com

Changliang Zou, School of Statistics and Data Science, LPMC, KLMDASR and LEBPS, Nankai University

E-mail: nk.chlzou@gmail.com

Zhaojun Wang, School of Statistics and Data Science, LPMC, KLMDASR and LEBPS, Nankai University

E-mail: zjwangnk@126.com