ORIGINAL ARTICLE
Scandinavian Journal of Statistics

Efficient estimation via envelope chain in MRI-based studies

Lan Liu[1] | **Wei Li**[2*] | **Zhihua Su**[3] | **Dennis Cook**[1] | **Luca Vizioli**[4] | **Essa Yacoub**[4]

[1]School of Statistics, University of Minnesota at Twin Cities, Minneapolis, Minnesota, USA

[2]Center for Applied Statistics and School of Statistics, Renmin University of China, Beijing, China

[3]Department of Statistics, University of Florida, Gainesville, Florida, USA

[4]Department of Radiology, University of Minnesota at Twin Cities, Minneapolis, Minnesota, USA

Correspondence
Center for Applied Statistics and School of Statistics, Renmin University of China, Beijing, China
Email: weilistat@ruc.edu.cn

Funding information
NIH P41 EB015894; Grant-in-aid at the University of Minnesota at Twin Cities and NSF DMS 1916013; The Fundamental Research Funds for the Central Universities and the Research Funds of Renmin University of China.

Magnetic resonance imaging (MRI) is a technique that scans the anatomical structure of the brain, whereas functional magnetic resonance imaging (fMRI) use the same basic principles of atomic physics as MRI scans but image metabolic function. A major goal of MRI and fMRI study is to precisely delineate various types of tissues, anatomical structure, pathologies, and detect the brain regions that react to outer stimuli (e.g., viewing an image). As a key feature of these MRI-based neuroimaging data, voxels (cubic pixels of the brain volume) are highly correlated. However, the associations between voxels are often overlooked in the statistical analysis. We adapt a recently proposed dimension reduction method called the envelope method to analyze neuoimaging data taking into account correlation among voxels. We refer to the modified procedure the envelope chain procedure. Because the envelope chain procedure has not been employed before, we demonstrate in simulations the empirical performance of estimator, and examine its sensitivity when our assumptions are violated. We use the estimator to analyze the MRI data from ADHD-200 study. Data analyses demonstrate that leveraging the correlations among voxels can significantly increase the efficiency of the regression analysis, thus achieving higher detection power with small sample sizes.

KEYWORDS
Efficiency gain; Envelope method; Neuroimaging; Sufficient dimension reduction.

1 | INTRODUCTION

1.1 | Background

Magnetic resonance imaging (MRI) is a non-invasive imaging technology that produces three dimensional anatomical images. Functional magnetic resonance imaging (fMRI) is a technique that exploits the coupling between oxygenated blood flow and neuronal firing to infer cortical activity by measuring changes in the Blood Oxygen Level Dependent (BOLD) signal [1].

MRI image scans anatomical structure, whereas fMRI image metabolic function. A critical goal of MRI-based neuroimaging studies is to precisely detect the brain regions that react to diseases or outer stimuli. Such information helps us understand how the brain is organized, coordinated or connected and is useful to explore cortical functional organization and to examine neurological or mental disorders.

With rapid technological advances, MRI-based images can now generate about a million voxels (cubic pixels of brain volume). While such "big data" has revolutionized the prospects of brain imaging, a fundamental challenge of MRI and fMRI has been limited sample sizes due to prohibitive scanning costs and available magnet time. For example, even with a $30 million dollar budget, the human connectome project (HCP) data set was still limited to 1200 subjects. Although this is a unprecedented size for MRI-based studies, it is still limited given the high resolution of the data. Additionally, larger sample sizes are sometimes unrealistic for rare diseases, such as Gardner-Diamond syndrome. However, even if a large study is possible, maximizing the use and analytical quality of such "expensive" data is imperative.

In the neuroimaging literature, it is known that voxels in MRI-based scans typically display correlated structures. For example, Alexander et al. [2] reported healthy aging is associated with brain volume reductions in the frontal cortex, temporal, parietal, subcortical and cerebellar regions. Additionally, Biswal et al. [3] critically observed that, in BOLD recordings, anatomically connected brain systems show slow, spontaneous fluctuations, while other, unconnected regions demonstrate unrelated oscillations. Biswal et al. [3] and Van Dijk et al. [4] found a high correlation between left and right motor cortices and minimal correlation between motor cortex and visual cortex. Similarly, the correlated fluctuations have also been observed in other cortices such as the visual and auditory cortices [5]. This leads to the identification of a number of cortical networks, such as: the default network and the medial temporal lobe memory system [6], the language system [7], the dorsal attention system [8], and the frontoparietal control system [9]. Such spontaneous or unrelated fluctuations were not only observed in resting state, but also during task fMRI in various states of consciousness [10].

However, such features of BOLD images have not been fully utilized in the statistical analysis of voxel-level neuroimaging data. As we will show, leveraging on correlation across voxels can significantly reduce the dimensionality for voxels and increase analytical efficiency relatively to traditional voxel-wise approaches. This allows achieving higher statistical detection power with small sample sizes.

1.2 | Statistical methods for neuroimaging data and our contributions

A classic method in regression analysis for MRI-based data is to fit a univariate voxel-wise linear regression or mixed effect models [11, 12]. However, ignoring associations between voxels in the estimation process can lead to severe efficiency loss. Gaussian random field (GRF; [13]) method also depends on a voxel-wise linear model. Although GRF leverages the spacial correlation between voxels to increase power for signal detection, it has the same efficiency as the voxel-wise analysis. Independent component analysis (ICA; [14]) and principal component analysis (PCA; [15]) are alternative methods in the neuroimaging literature. However, these dimension reduction strategies are not suitable for regression problems because they do not incorporate any predictors when identifying the underlying components. Multivoxel pattern analyses (MVPA; [16]) adopt a classification technique such as support vector machine, where voxels serve as predictors to recognize a pattern of brain activity associated with one cognitive state. Similar to MVPA, the partial least squares (PLS) method also uses voxels as predictors to discriminate binary cognitive state [17, 18, 19]. In this paper, we are interested in investigating the difference of MRI between Attention Deficit Hyperactivity Disorder (ADHD) and normal individuals. The existing methods either suffer from efficiency loss or do not match our research goals. This motivates us to develop a new method to obtain efficient estimators in a multivariate regression problem.

Recently, Cook et al. [20] proposed a new sufficient dimension reduction method called the envelope method. The main assumption of the envelope method is the existence of linear combinations of the response variables that is irrelevant to the multivariate regression [20]. Given that such combinations of the responses do not depend on the predictors, nor are they associated with the relevant responses, they can be effectively discarded, thus reducing noise and improving efficiency of the regression analysis. Recent development of the envelope methods in different settings can be found in [21, 22, 23, 24, 25, 26]. However, the available envelope methods are applicable only for the cases where the dimension of response is smaller than the sample size [20, 23]. The number of parameters in the covariance matrix of the error terms can be on the scale of 10^{11}. Even with an unrealistically large sample size, the envelope method is not computationally tractable with any available statistical software.

Therefore, we adapt the envelope procedure and refer to the modified version the envelope chain procedure. Specifically, we assume that voxels can be partitioned into conditionally uncorrelated clusters and that within each cluster there exist combinations of voxels that are irrelevant to the regression. These assumptions are well grounded in neuroimage data. Under correct model specification and normality assumption on the error term, the envelope chain estimator corresponds to the maximum likelihood estimator. When the clustering is misspecified (for example, if all the voxels are correlated and we randomly partition them into clusters), the envelope chain estimator is still consistent for the parameters of interest. The envelope chain model is closely related to both voxel-wise regression and the envelope model. If there is only one cluster, then the envelope chain model reduces to the envelope model. If the number of clusters is the same as the sample size, then the envelope chain model reduces to voxel-wise regression. Thus, the envelope chain model can be seen as an intermediate model that represents an improvement upon a highly inefficient model, voxel-wise regression, as well as over a more efficient but not directly applicable model, envelope model.

1.3 | Outline

The remainder of this paper is organized as follows. A preliminary introduction about envelope procedure is given in Section 2. We adapt the envelope procedure and explore the properties of the modified version, i.e., the envelope chain procedure in Section 3. The envelope chain estimator is evaluated in the simulation study in Section 4. We use the estimator to analyze the MRI image from the ADHD-200 study in Section 5. We conclude the paper with a

discussion in Section 6. All proofs of propositions are given in the Appendix.

2 | PRELIMINARY

Consider the multivariate linear regression model

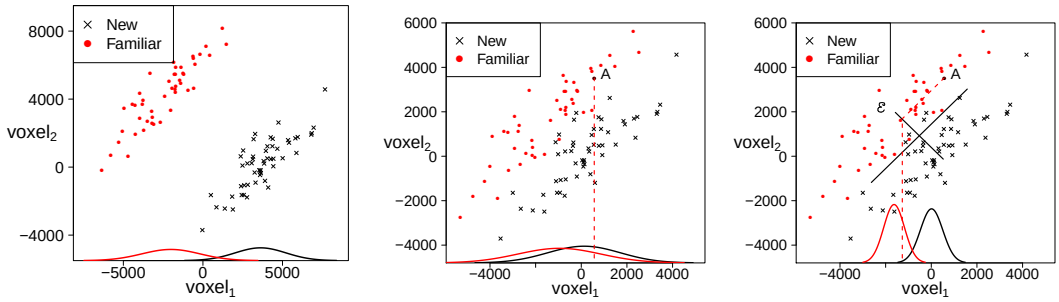
$$\mathbf{Y}_i = \boldsymbol{\alpha} + \boldsymbol{\beta}\mathbf{X}_i + \boldsymbol{\varepsilon}_i, \tag{1}$$

where $\mathbf{Y}_i = (Y_{i1}, \ldots, Y_{ir})^\top \in \mathbb{R}^{r\times 1}$ is the multivariate response containing the vectorized voxel intensities, $\mathbf{X}_i \in \mathbb{R}^{p\times 1}$ is the predictor vector, the error vector $\boldsymbol{\varepsilon}_i \in \mathbb{R}^{r\times 1}$ is normally distributed with mean $\mathbf{0}$ and covariance matrix $\boldsymbol{\Sigma} \in \mathbb{R}^{r\times r}$, $\boldsymbol{\alpha} \in \mathbb{R}^{r\times 1}$ is an unknown vector of intercepts and $\boldsymbol{\beta} \in \mathbb{R}^{r\times p}$ is an unknown matrix of regression coefficients for $i = 1, \ldots, n$.

Let $\tilde{\mathbf{X}} = (\mathbf{X}_1, \ldots, \mathbf{X}_n)$ and $\tilde{\mathbf{Y}} = (\mathbf{Y}_1, \ldots, \mathbf{Y}_n)$ denote the predictor and response matrices including all n individuals. Without loss of generality, we assume that the design matrix $\tilde{\mathbf{X}}$ has full row rank. A classic approach in regression analysis for the MRI-based data is to fit a univariate voxel-wise linear regression or a mixed effects model. The voxel-wise regression regresses one response at a time. Specifically, the voxel-wise estimator is $\hat{\boldsymbol{\beta}}^{\text{vol}} = \tilde{\mathbf{Y}}\tilde{\mathbf{X}}^\top(\tilde{\mathbf{X}}\tilde{\mathbf{X}}^\top)^{-1}$. This estimator does not utilize the covariance matrix $\boldsymbol{\Sigma}$, treating all voxels independently. However, by considering the correlations among brain voxel intensities, we can substantially reduce the dimension for the brain voxels and gain additional efficiency over voxel-wise estimator.

To illustrate this point, we consider simulated data. While our estimator effectively deals with thousands of voxels, to promote clarity, our examples will only portray the simplest 2 voxels scenarios. Suppose that the parameter of interest is the effect of a binary stimulus $\mathbf{X}$ (viewing an image of a familiar face versus a new face) on a voxel (voxel_1). We generated 2 scenarios: scenario 1, in which the responses of 2 voxels are highly distinguishable (i.e. the strong signal case, Figure 1a); and scenario 2, in which the responses of 2 voxels are largely overlapping (i.e. the weak signal case, Figure 1b). The curves at the bottom of Figures 1a and 1b are the density curves that voxel-wise regression used for inference. As shown in the density curves in Figure 1a, the exposed group $\mathbf{X} = 1$ is well distinguished from the unexposed group $\mathbf{X} = 0$ using the voxel-wise regression when the signal is so strong that the large variability does not matter. However, the voxel-wise estimator cannot distinguish the groups with weaker signals when variability is large (Figure 1b). This is because the voxel-wise regression utilizes only one voxel at a time, it suffers from efficiency loss, and thus it is only useful for detecting strong brain signals rather than moderate or weak ones. In other words, voxel-wise regression is more likely to lead to false negatives. Similarly, due to large variability, voxel-wise regression is also prone to false positive results in finite samples.

In contrast, our estimator utilizes multiple correlated voxels at the same time and therefore is more efficient. Figure 1c shows the use of a second voxel which is highly correlated with the first voxel. The key idea of our statistical procedure is to first identify highly correlated voxels, then to identify the direction that best distinguishes the differences between the groups, and to project data onto that direction to reduce noise (see Figure 1c). For example, while the voxel-wise regression procedure projects the data point A directly on the voxel_1-axis ignoring voxel_2 (Figure 1b), with our procedure, we first project the data point A to the direction $\mathcal{E}$, which best distinguishes the groups, and then project it down to the voxel_1-axis (Figure 1c). As shown in Figure 1c, the two groups have much smaller variability and thus are much easier to distinguish even when the effect of interest is not very pronounced due to low signal or high variability. Similarly, when the effects are null, our procedure is also less likely to have false positive findings.



(a) Voxel-wise estimate for strong signal. **(b)** Voxel-wise estimate for weak signal. **(c)** Envelope chain estimate for weak signal.

FIGURE 1 Graphical illustration of the voxel-wise and envelope chain estimators.

3 | ENVELOPE CHAIN

3.1 | Model and assumptions

We assume that the brain voxel intensities of the i^{th} individual, $\mathbf{Y}_i$, can be grouped into conditionally independent sub-vectors $\mathbf{Y}_{i(1)} \in \mathbb{R}^{r_1 \times 1}, \ldots, \mathbf{Y}_{i(g)} \in \mathbb{R}^{r_g \times 1}$, where g is the number of groups and $r_1 + \cdots + r_g = r$. That is,

Assumption 1 $\mathbf{Y}_i = (\mathbf{Y}_{i(1)}^T, \ldots, \mathbf{Y}_{i(g)}^T)^T$ *and* $\mathbf{Y}_{i(j)} \perp\!\!\!\perp \mathbf{Y}_{i(k)} | \mathbf{X}$ *for* $1 \le j \le g$, $1 \le k \le g$, $j \ne k$ *and* $i = 1, \ldots, n$.

Assumption 1 is a reasonable assumption for MRI-based data, since unconnected or different functional brain systems may have unrelated intensities [4]. As we will show below, the violation of Assumption 1 still leads to consistent estimator although not the most efficient. For now, we assume Assumption 1 holds. Since the groups are uncorrelated with each other, conditional on the predictors, the covariance matrix $\mathbf{\Sigma}$ of the error terms $\boldsymbol{\varepsilon}_i$ is a block-diagonal matrix with blocks $\mathbf{\Sigma}_{(1)}, \ldots, \mathbf{\Sigma}_{(g)}$. Thus, model (1) can be written as

$$\begin{pmatrix} \mathbf{Y}_{i(1)} \\ \vdots \\ \mathbf{Y}_{i(g)} \end{pmatrix} = \begin{pmatrix} \boldsymbol{\alpha}_{(1)} \\ \vdots \\ \boldsymbol{\alpha}_{(g)} \end{pmatrix} + \begin{pmatrix} \boldsymbol{\beta}_{(1)} \\ \vdots \\ \boldsymbol{\beta}_{(g)} \end{pmatrix} \mathbf{X}_i + \begin{pmatrix} \boldsymbol{\varepsilon}_{i(1)} \\ \vdots \\ \boldsymbol{\varepsilon}_{i(g)} \end{pmatrix},$$

where $\mathrm{Var}(\boldsymbol{\varepsilon}_i) = \mathbf{\Sigma} = \mathrm{bdiag}(\mathbf{\Sigma}_{(1)}, \ldots, \mathbf{\Sigma}_{(g)})$ and bdiag denotes the block-wise diagonal matrix. We denote the voxel-wise estimators of $\boldsymbol{\alpha}_{(j)}$ and $\boldsymbol{\beta}_{(j)}$ by $\hat{\boldsymbol{\alpha}}_{(j)}^{\mathrm{vol}}$ and $\hat{\boldsymbol{\beta}}_{(j)}^{\mathrm{vol}}$, and $\hat{\mathbf{\Sigma}}_{(j)}^{\mathrm{res}}$ is the sample estimate of $\mathbf{\Sigma}_{(j)}$, i.e., $\hat{\mathbf{\Sigma}}_{(j)}^{\mathrm{res}} = \sum_{i=1}^{n} (\mathbf{Y}_i - \hat{\boldsymbol{\alpha}}^{\mathrm{vol}} - \hat{\boldsymbol{\beta}}^{\mathrm{vol}} \mathbf{X}_i)(\mathbf{Y}_i - \hat{\boldsymbol{\alpha}}^{\mathrm{vol}} - \hat{\boldsymbol{\beta}}^{\mathrm{vol}} \mathbf{X}_i)^\top / n$, where $\hat{\boldsymbol{\alpha}}^{\mathrm{vol}} = ((\hat{\boldsymbol{\alpha}}_{(1)}^{\mathrm{vol}})^\top, \ldots, (\hat{\boldsymbol{\alpha}}_{(g)}^{\mathrm{vol}})^\top)^\top$ and $\hat{\boldsymbol{\beta}}^{\mathrm{vol}}$ is defined similarly. Under the clustering assumption, $\boldsymbol{\beta}_{(j)}$ can be estimated with only $\mathbf{Y}_{i(j)}$, ignoring the other components of $\mathbf{Y}_i$ for $i = 1, \ldots, n$.

Below, we briefly introduce the concept of the envelope model for each group. More discussions of the envelope model can be found in Cook et al. [20] and Cook [26]. A matrix $\mathbf{O} \in \mathbb{R}^{t \times t}$ is orthonormal if and only if it satisfies $\mathbf{O}^\top \mathbf{O} = \mathbf{I}_t$, where $\mathbf{I}_t$ denotes the identity matrix with dimension t. Assume there exists an orthogonal matrix $(\mathbf{\Gamma}_{(j)}, \mathbf{\Gamma}_{0(j)}) \in \mathbb{R}^{r_j \times r_j}$ such that

Assumption 2 $\mathbf{\Gamma}_{0(j)}^\top \mathbf{Y}_{(j)} | \mathbf{X} \sim \mathbf{\Gamma}_{0(j)}^\top \mathbf{Y}_{(j)}$,

Assumption 3 $\mathbf{\Gamma}_{(j)}^\top \mathbf{Y}_{(j)} \perp\!\!\!\perp \mathbf{\Gamma}_{0(j)}^\top \mathbf{Y}_{(j)} | \mathbf{X}$,

where $\mathbf{\Gamma}_{(j)} \in \mathbb{R}^{r_j \times u_j}$, $\mathbf{\Gamma}_{0(j)} \in \mathbb{R}^{r_j \times (r_j - u_j)}$, $0 \leq u_j \leq r_j$ for $j = 1, \ldots, g$, and $\sim$ denotes identical distribution. There always exist $(\mathbf{\Gamma}_{(j)}, \mathbf{\Gamma}_{0(j)})$ that satisfies Assumptions 2 and 3. A trivial choice is $\mathbf{\Gamma}_{(j)} = \mathbf{I}_{r_j}$ and $\mathbf{\Gamma}_{0(j)}$ is null, i.e., $(\mathbf{\Gamma}_{(j)}, \mathbf{\Gamma}_{0(j)}) = \mathbf{\Gamma}_{(j)} = \mathbf{I}_{r_j}$. The subspace satisfying Assumptions 2 and 3 is not unique, but the j^{th} envelope is uniquely defined as the smallest subspace (in terms of u_j) satisfying these assumptions. The dimension u_j is the dimension of the j^{th} envelope.

Under Assumption 2, the marginal distribution of $\mathbf{\Gamma}_{0(j)}^{\top}\mathbf{Y}_{(j)}$ does not depend on $\mathbf{X}$. Under Assumption 3, $\mathbf{\Gamma}_{0(j)}^{\top}\mathbf{Y}_{(j)}$ also does not contribute to the estimation of $\boldsymbol{\beta}_{(j)}$ through its association with $\mathbf{\Gamma}_{(j)}^{\top}\mathbf{Y}_{(j)}$. Assumptions 2 and 3 together imply the redundancy of $\mathbf{\Gamma}_{0(j)}^{\top}\mathbf{Y}_{(j)}$ in the regression, i.e., $\mathbf{\Gamma}_{0(j)}^{\top}\mathbf{Y}_{(j)}$ does not contain any information on $\boldsymbol{\beta}_{(j)}$, thus it can be discarded without losing any regression information, and the subsequent analysis can be based only on $\mathbf{\Gamma}_{(j)}^{\top}\mathbf{Y}_{(j)}$. Similar to Cook et al. [20], $\mathbf{\Gamma}_{(j)}^{\top}\mathbf{Y}_{(j)}$ and $\mathbf{\Gamma}_{0(j)}^{\top}\mathbf{Y}_{(j)}$ are routinely named as the material part and immaterial part for group j, respectively. Assumptions 2 and 3 are also well grounded in MRI-based studies, because previous work has revealed that different voxels in connected brain systems may exhibit strong correlations or spontaneous fluctuations, indicating that the inclusion of all voxels in the regression is in fact redundant [3, 4].

As shown in Cook et al. [20], under Assumptions 2 and 3, we have $\boldsymbol{\beta}_{(j)} = \mathbf{\Gamma}_{(j)}\boldsymbol{\eta}_{(j)}$ and $\mathbf{\Sigma}_{(j)} = \mathbf{\Gamma}_{(j)}\mathbf{\Omega}_{(j)}\mathbf{\Gamma}_{(j)}^{T} + \mathbf{\Gamma}_{0(j)}\mathbf{\Omega}_{0(j)}\mathbf{\Gamma}_{0(j)}^{T}$, where $\boldsymbol{\eta}_{(j)} \in \mathbb{R}^{u_j \times p}$, $\mathbf{\Omega}_j = \mathbf{\Gamma}_{(j)}^{\top}\mathbf{\Sigma}_{(j)}\mathbf{\Gamma}_{(j)}$ and $\mathbf{\Omega}_{0(j)} = \mathbf{\Gamma}_{0(j)}^{\top}\mathbf{\Sigma}_{(j)}\mathbf{\Gamma}_{0(j)}$. Thus, under Assumptions 1–3, model (1) can be reparameterized as

$$\mathbf{Y}_{i(j)} = \boldsymbol{\alpha}_{(j)} + \mathbf{\Gamma}_{(j)}\boldsymbol{\eta}_{(j)}\mathbf{X}_i + \boldsymbol{\varepsilon}_{i(j)}, \tag{2}$$

for $j = 1, \ldots, g$, where $\mathrm{Var}(\boldsymbol{\varepsilon}_{i(j)}) = \mathbf{\Sigma}_{(j)} = \mathbf{\Gamma}_{(j)}\mathbf{\Omega}_{(j)}\mathbf{\Gamma}_{(j)}^{T} + \mathbf{\Gamma}_{0(j)}\mathbf{\Omega}_{0(j)}\mathbf{\Gamma}_{0(j)}^{T}$, and $\boldsymbol{\varepsilon}_{i(j)} \perp\!\!\!\perp \boldsymbol{\varepsilon}_{i(k)}$ for $j \neq k$, $i = 1, \ldots, n$. We call (2) the envelope chain model.

The envelope chain idea aligns with the general idea of handling high dimensional data by group partition and within group dimension reduction, which can be traced to Wold et al. [27]. They pointed out that *"There is a strong temptation to reduce the variables to a smaller, more manageable number. This reduction of variables, however, often removes information, makes the interpretation misleading and seriously increases the risk of spurious models. A better alternative is often to divide the variables into conceptually meaningful blocks and then apply hierarchical multi block PLS (or PC) models."* In our case, the dimension r of the response vectors is allowed to be larger than sample size n. To deal with the high-dimensional problem, we divide the response vectors into blocks and apply the envelope model to each block.

The envelope chain procedure is an intermediate model between and an improvement upon a highly inefficient model, voxel-wise regression, and an efficient but not directly applicable model, envelope model. If the block number $g = 1$, then the envelope chain model reduces to the envelope model. If $g = n$, then the envelope chain model reduces to the voxel-wise regression. With a moderate number of clusters, the dimension of each sub-vector $\mathbf{Y}_{(j)}$ is much lower, which makes the computation for each group feasible. Once the clustering has been decided, the computation for each group can be paralleled.

3.2 | Estimation

Although in some scenarios, prior information about clustering (such as that provided by anatomical connectivity) may be available, it is unknown for most applications. We use the hierarchical clustering algorithm [28] that assembles the multivariate responses that are highly correlated with each other in one group, and then implement the envelope chain model (2) based on such clustering. Specifically, we use $1 - abs\{cor(Y_{ij}, Y_{ik})\}$ as the metric to quantify the dissimilarity between the j^{th} and k^{th} voxels, where *cor* and *abs* denote the correlation and absolute value operations, respectively. The total number of clusters can be estimated using cross-validation by comparing the differences be-

tween the estimated voxel intensities versus those observed on the test data. Alternatively, the number of clusters can be evaluated by a sensitivity analysis. Once the total number of clusters is estimated, clustering membership can be obtained from the dedrogram produced by the hierarchical clustering algorithm. We write the estimated number of clusters as $\hat{g}$ and the estimated number of voxels in the j^{th} cluster as $\hat{r}_j$. We demonstrate such procedures in Section 4.

While the hierarchical clustering algorithm only groups highly correlated voxels into the same cluster, it is possible that voxels across clusters are also conditionally correlated. We will show that even if the cluster independence assumption does not hold, the envelope chain estimates are still consistent, although it is no longer the most efficient estimator. Additionally, as we demonstrate in the simulations, even when the clustering assumption is violated, the envelope chain still performs either better than or at least as well as voxel-wise estimator in finite samples.

As shown in Cook et al. [20], an estimate of $\boldsymbol{\Gamma}_{(j)}$ can be obtained by minimizing $\log|\mathbf{G}^T\hat{\boldsymbol{\Sigma}}^{\text{res}}_{(j)}\mathbf{G}|+\log|\mathbf{G}^T\hat{\boldsymbol{\Sigma}}^{-1}_{\mathbf{Y}_{(j)}}\mathbf{G}|$, where $\hat{\boldsymbol{\Sigma}}_{\mathbf{Y}_{(j)}}$ is the sample variance of $\mathbf{Y}_{(j)}$. This is an optimization over a Grassmann manifold, which has been implemented in R and Matlab packages [29, 30]. Once we have $\hat{\boldsymbol{\Gamma}}_{(j)}$, the envelope chain estimator of $\boldsymbol{\beta}_{(j)}$ and $\boldsymbol{\Sigma}_{(j)}$ are $\hat{\boldsymbol{\beta}}^{\text{ec}}_{(j)} = \mathbf{P}_{\hat{\boldsymbol{\Gamma}}_{(j)}}\hat{\boldsymbol{\beta}}^{\text{vol}}_{(j)}$, where $\mathbf{P}_{\boldsymbol{s}} = \boldsymbol{s}(\boldsymbol{s}^{\top}\boldsymbol{s})^{-1}\boldsymbol{s}$ is the projection matrix onto the column space of $\boldsymbol{s}$. The dimension u_j can be selected by information criteria such as AIC, BIC, or likelihood ratio testing or cross validation as discussed in Cook et al. [20]. Eck and Cook [31] recommended using the BIC to decide the envelope dimension $\hat{u}_j$, because the AIC tends to overestimate the true dimensions and the likelihood ratio testing is inconsistent. Thus, throughout the paper, we use BIC to select the dimensions of the envelopes in envelope chain.

3.3 | Properties

Let $\boldsymbol{\Gamma} = \text{bdiag}(\boldsymbol{\Gamma}_{(1)}, \ldots, \boldsymbol{\Gamma}_{(g)})$, $\boldsymbol{\Gamma}_0 = \text{bdiag}(\boldsymbol{\Gamma}_{0(1)}, \ldots, \boldsymbol{\Gamma}_{0(g)})$, $\boldsymbol{\eta} = (\boldsymbol{\eta}^{\top}_{(1)}, \ldots, \boldsymbol{\eta}^{\top}_{(g)})^{\top}$, $\boldsymbol{\Omega} = \text{bdiag}(\boldsymbol{\Omega}_{(1)}, \ldots, \boldsymbol{\Omega}_{(g)})$, $\boldsymbol{\Omega}_0 = \text{bdiag}(\boldsymbol{\Omega}_{0(1)}, \ldots, \boldsymbol{\Omega}_{0(g)})$, and let $u = u_1 + \cdots + u_g$. The following proposition reveals that an envelope chain model is also an envelope model.

Proposition 1 *Under model* (2)*, we have* $\mathbf{Y}_i = \boldsymbol{\alpha} + \boldsymbol{\Gamma}\boldsymbol{\eta}\mathbf{X}_i + \boldsymbol{\varepsilon}_i$*, where* $\text{Var}(\boldsymbol{\varepsilon}_i) = \boldsymbol{\Sigma} = \boldsymbol{\Gamma}\boldsymbol{\Omega}\boldsymbol{\Gamma}^T + \boldsymbol{\Gamma}_0\boldsymbol{\Omega}_0\boldsymbol{\Gamma}_0^T$ *for* $i = 1, \ldots, n$*. That is,* $\text{span}(\boldsymbol{\Gamma})$ *is the envelope for* $\boldsymbol{\beta}$ *with dimension* u*.*

Similar to the envelope model, the envelope chain model also reduces the number of parameters. The number of free parameters under the voxel-wise model is $N^{\text{vol}} = r + rp + \sum_{j=1}^{g} r_j(r_j+1)/2$, and the number of free parameters under the envelope chain model is $N^{\text{ec}} = r + up + \sum_{j=1}^{g} r_j(r_j+1)/2$. It is easy to see that $N^{\text{vol}} \geq N^{\text{ec}}$. When u is small, the difference between the numbers of the parameters in the two models is substantial. Following Cook et al. [20], the envelope chain estimator achieves an efficiency gain over the standard voxel-wise estimator.

Proposition 2 *Assume model* (2) *holds. If the clustering structure and the envelope dimension for each cluster are correctly specified, then* $\sqrt{n}\{\text{vec}(\hat{\boldsymbol{\beta}}^{\text{vol}}) - \text{vec}(\boldsymbol{\beta})\} \xrightarrow{d} N(\mathbf{0}, \mathbf{V}^{\text{vol}})$ *and* $\sqrt{n}\{\text{vec}(\hat{\boldsymbol{\beta}}^{\text{ec}}) - \text{vec}(\boldsymbol{\beta})\} \xrightarrow{d} N(\mathbf{0}, \mathbf{V}^{\text{ec}})$*, where* $\mathbf{V}^{\text{vol}} = bdiag(\mathbf{V}^{\text{vol}}_{(1)}, \ldots, \mathbf{V}^{\text{vol}}_{(g)})$, $\mathbf{V}^{\text{ec}} = bdiag(\mathbf{V}^{\text{ec}}_{(1)}, \ldots, \mathbf{V}^{\text{ec}}_{(g)})$, $\mathbf{V}^{\text{vol}}_{(j)} = \boldsymbol{\Sigma}_{\mathbf{X}}^{-1} \otimes \boldsymbol{\Sigma}_{(j)}$*, and* $\mathbf{V}^{\text{ec}}_{(j)} = \boldsymbol{\Sigma}_{\mathbf{X}}^{-1} \otimes \boldsymbol{\Gamma}_{(j)}\boldsymbol{\Omega}_{(j)}\boldsymbol{\Gamma}^T_{(j)} + (\boldsymbol{\eta}^T_{(j)} \otimes \boldsymbol{\Gamma}_{0(j)})(\boldsymbol{\eta}_{(j)}\boldsymbol{\Sigma}_{\mathbf{X}}\boldsymbol{\eta}^T_{(j)} \otimes \boldsymbol{\Omega}^{-1}_{0(j)} + \boldsymbol{\Omega}_{(j)} \otimes \boldsymbol{\Omega}^{-1}_{0(j)} + \boldsymbol{\Omega}^{-1}_{(j)} \otimes \boldsymbol{\Omega}_{0(j)} - 2\mathbf{I}_{u_j} \otimes \mathbf{I}_{r_j-u_j})^{\dagger}(\boldsymbol{\eta}_{(j)} \otimes \boldsymbol{\Gamma}^T_{0(j)})$ *for* $j = 1, \ldots, g$*, where* $\dagger$ *is the Moore-Penrose inverse. Moreover,* $\mathbf{V}^{\text{ec}} \leq \mathbf{V}^{\text{vol}}$*.*

The efficiency gain of the envelope chain estimator over the voxel-wise estimator can be obtained by noting $\mathbf{V}^{\text{ec}} \leq \mathbf{V}^{\text{vol}}$ when the clustering structure is correctly specified. A nice feature of the proposed procedure is that although we impose the clustering structure, i.e., Assumption 1 on the brain voxels, this assumption does not need to hold for our estimator to be $\sqrt{n}$-consistent. That is, we have

Proposition 3 *Under model* (1)*, for a given clustering structure that possibly violates Assumption 1, if the envelope dimension is estimated using BIC, then the envelope chain estimator* $\hat{\boldsymbol{\beta}}^{\text{ec}}$ *is a* $\sqrt{n}-$*consistent estimator for* $\boldsymbol{\beta}$.

The proof of Proposition 3 follows from the fact that, in each cluster BIC will select the correct envelope dimension with probability tending to 1 as $n \to \infty$ [26, 31, 32], and the envelope estimator is consistent if the envelope dimension is correct. It is a challenging task to account for clustering variability theoretically, and hence, the theoretical results in Propositions 2 and 3 are established with a specified clustering structure. Given such a clustering structure, the envelope chain method partitions the high dimensional voxels into lower dimensional groups so that the theories of existing envelope methods are still applicable. Interestingly, Proposition 3 implies that even if the clustering structure is incorrectly specified, the envelope chain estimator is still $\sqrt{n}-$consistent. Thus, we can use any off-the-shelf methods, e.g., the hierarchical clustering algorithm, to conduct clustering analysis in practice without negatively impacting the consistency of the resulting envelope chain estimator. Besides, we suggest using the bootstrap procedure to account for clustering variability. We need to acknowledge that bootstrapping is used here without theoretical justification, similar to some other works of high-dimensional studies [33, 34].

4 | SIMULATIONS

In this section, we demonstrate finite sample performance of the envelope chain estimator in simulations, and compare with the voxel-wise and principal component regression (PCR) estimators. As the classic envelope estimator has been thoroughly investigated in Cook et al. [20], we only examine the performances of the envelope chain estimator under 3 scenarios: 1) when the clustering assumption, i.e., Assumption 1, holds (Section 4.1); 2) when it is moderately violated (Section 4.2); and 3) when it is seriously violated (Section 4.2). In each scenario, we keep the results of the voxel-wise estimators as a reference. The comparison results between the envelope chain and voxel-wise estimator are shown in figure for all the above 3 scenarios. However, for ease of presentation, the comparison results between PCR and voxel-wise estimators are shown in figure for only scenario 3), and the corresponding results for the other 2 scenarios are stated in text.

4.1 | Efficiency gain

When the variance $\boldsymbol{\Sigma}$ has a blockwise structure, the envelope chain estimator is asymptotically unbiased. The variability of the envelope chain estimator is composed of three parts: clustering, the estimation of the envelope dimension and the envelope estimation. In order to have a relative comparison among them, we consider envelope chain estimator when: (i) both clustering membership and u_j are known; (ii) only clustering membership is known; (iii) neither clustering membership nor u_j is known. That is, if the clustering membership and/or the envelope dimensions are known, we use their true values, otherwise, we use hierarchical clustering (with the total number of clusters set as the true value), and BIC to estimate them. We carry out the simulations as follows.

Step 1: Set $p = 5$ and independently generate a vector of covariates $\mathbf{X}_i$ from a multivariate standard normal with $\mathbf{I}_5$ as covariance matrix for $i = 1, \dots, 200$.

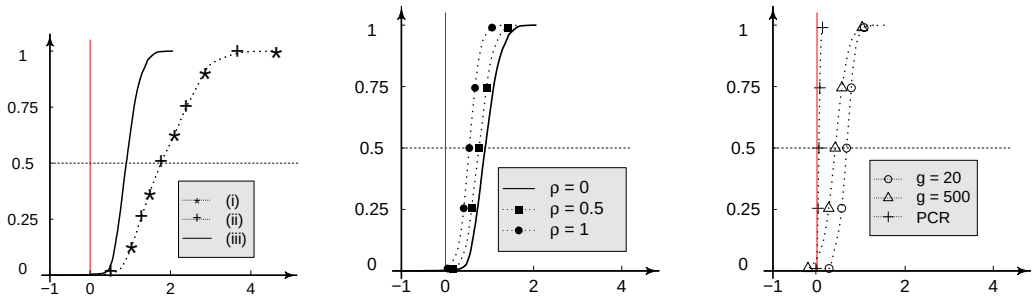
Step 2: Set $g = 20$, and r_j is drawn from uniform distribution $U(30, 50)$ for $j = 1, \dots, g$. As a result, we have $r = 799$. Set $u_j = 1$ for $j = 1, \dots, g$. Generate $\tilde{\boldsymbol{\Gamma}}_{(j)} \in \mathbb{R}^{r_j \times u_j}$, where each element of $\tilde{\boldsymbol{\Gamma}}_{(j)}$ is generated from $N(0, 1)$. Set $\boldsymbol{\Gamma}_{(j)} = \tilde{\boldsymbol{\Gamma}}_{(j)} (\tilde{\boldsymbol{\Gamma}}_{(j)}^{\top} \tilde{\boldsymbol{\Gamma}}_{(j)})^{-1/2}$, and therefore $\boldsymbol{\Gamma}_{(j)}^{\top} \boldsymbol{\Gamma}_{(j)} = \mathbf{I}_{u_j}$. Generate $\boldsymbol{\Gamma}_{0(j)} \in \mathbb{R}^{r_j \times (r_j - u_j)}$ such that $(\boldsymbol{\Gamma}_{(j)}, \boldsymbol{\Gamma}_{0(j)})$ is orthonormal. Generate $\boldsymbol{\Omega}_{(j)} \in \mathbb{R}^{u_j \times u_j}$, where the element is firstly generated from a χ_1^2-distribution and then multiplied by

2. Generate $\mathbf{M}_j \in \mathbb{R}^{(r_j-u_j)\times(r_j-u_j)}$, where each element of $\mathbf{M}_j$ is generated from $U(0, 1)$ and set $\boldsymbol{\Omega}_{0(j)} = \mathbf{M}_j^\top \mathbf{M}_j$. Generate $\boldsymbol{\eta}_{(j)} \in \mathbb{R}^{u_j \times p}$, where each element of $\boldsymbol{\eta}_{(j)}$ is firstly generated from $N(0, 1)$, then multiplied by 20, and finally we set $\boldsymbol{\beta}_{(j)} = \boldsymbol{\Gamma}_{(j)}\boldsymbol{\eta}_{(j)}$.

Step 3: Generate $\mathbf{Y}$ according to model (2) using the generated parameters in **Step 2**. We randomly divide the data set into training and test parts according to the proportion with 4 : 1. Using the training data, we calculate the envelope chain in (i)–(iii), PCR, and voxel-wise estimators, and then we compare their efficiency and MSE. We also calculate prediction bias in the test set.

Step 4: Repeat **Step 3** for 100 times.

We report the empirical distribution of the log ratio of Monte Carlo mean squared errors (MSEs) of the voxel-wise and the envelope chain estimators in (i)–(iii) in Figure 2a. A similar trend was observed for the Monte Carlo variances. Due to space constraints, we only graph the results for the MSEs and describe those for the variances in text. In Figure 2a, the x-axis represents the log ratio of the MSEs, while the y-axis represents its empirical distribution across elements of $\boldsymbol{\beta}$. Thus, the red vertical line at 0 shows where the voxel-wise and the envelope chain estimators have the same MSE. The right and left hand sides of the red line represent, respectively, the cases where the envelope chain is more accurate in terms of MSE or where the voxel-wise is more accurate. The solid curve, dotted curve with crosses, and dotted curve with stars denote the corresponding results for cases (i), (ii) and (iii), respectively.



(a) Blockwise structure holds for variance. **(b)** Blockwise structure is moderately violated. **(c)** Blockwise structure is seriously violated.

FIGURE 2 Empirical cumulative distribution functions of the log ratio of Monte Carlo MSEs between the voxel-wise and the envelope chain estimators when (a) the envelope structure holds: (i) both clustering membership and u_j are known; (ii) only clustering membership is known; (iii) neither clustering membership or u_j is known; (b) the envelope structure is moderately violated: sensitivity parameter ρ is chosen as 0.5 and 1; and (c) the envelope structure is seriously violated: the specified number of clusters are 20 and 500. PCR results are also included in (c).

When the envelope dimension is estimated by $\hat{u}_j$, it is mostly ≥ 1 for $j = 1, \ldots, g$. An over estimation of u_j results in an asymptotically unbiased estimator $\hat{\boldsymbol{\beta}}_{(j)}$, but it is less efficient compared to a correctly estimated u_j. The median log ratio of the MSEs of the voxel-wise versus the envelope chain estimator for the three curves are: 1.77 (case (i)), 1.76 (case (ii)) and 0.90 (case (iii)). The envelope chain estimators in these three cases all perform better than the voxel-wise estimators in terms of MSE. When both clustering membership and u_j are known, as portrayed in case (i), the envelope chain estimator has the smallest MSEs compared to the two other cases. The estimation in case (ii)

is similar to that in case (i), which indicates good performance of BIC in choosing the envelope dimension u_j. The estimators in these two cases are both better than that in case (iii). This is expected since the estimators are less variable when more information is available. In comparison, the median log ratio of MSEs of the voxel-wise versus the PCR estimator is -0.01. This result indicates that the performance of PCR estimator is comparable to that of the voxel-wise estimator, and it is worse than that of the proposed envelope chain estimator even when no clustering membership or envelope dimension is unknown.

We observe that all 3995 elements in the envelope chain estimator for $\boldsymbol{\beta} \in \mathbb{R}^{799\times 5}$ are more efficient compared to the voxel-wise estimator in cases (i)–(ii), and there are only 11 elements of the envelope chain estimator that are less efficient in case (iii). This demonstrates the dominating efficiency gain of the envelope chain estimator even in a finite sample. The median log ratio of the Monte Carlo variances of the voxel-wise versus the envelope chain estimator for the three curves are: 3.52 (case (i)), 3.51 (case (ii)), and 1.79 (case (iii)). This shows that to achieve the same power for the median performance among all the elements in voxel-wise regression, we only need about one fifth (e.g., for case (iii), $1/\exp(1.79) = 16.7\%$) of the original sample size when using the envelope chain estimators! Such an efficiency gain can be even more drastic if the eigenvalues of the immaterial part are larger compared to those of the material part. Here, we only show the simulation of a conservative setting that is more comparable with the real data. In comparison, the median log ratio of the Monte Carlo variances of the voxel-wise versus the PCR estimator is -0.04. This shows that the PCR estimator has a similar efficiency as the voxel-wise estimator in this setting.

Finally, we report the empirical distribution of the log ratio of squared Monte Carlo prediction biases of the voxel-wise and the envelope chain estimators on test data in (i)–(iii) in Figure 9a, which is relegated to the Appendix. The results are similar to those in Figure 2a, which again demonstrates the superiority of the envelope chain estimator over the voxel-wise estimator in terms of prediction for all cases. The median log ratio of squared prediction biases of the voxel-wise versus the envelope chain estimator for the three curves are: 0.90 (case (i)), 0.89 (case (ii)) and 0.70 (case (iii)). In comparison, the median log ratio of squared prediction biases of the voxel-wise versus the PCR estimator is -0.01. This result indicates that the PCR estimator exhibits similar prediction performance with that of the voxel-wise estimator.

4.2 | Sensitivity analysis

Next, we perform several sensitivity analyses to evaluate the robustness of our proposed procedure against possible violations of the blockwise structure in the covariance matrix of error terms. We first carry out a simulation where the blockwise structure is moderately violated. The simulation procedure is similar to that in Section 4.1, except that the variance of $\boldsymbol{\varepsilon}_i$ in **Step 2** is set to $\mathrm{Var}(\boldsymbol{\varepsilon}_i) = \tilde{\boldsymbol{\Sigma}} = \boldsymbol{\Sigma} + \rho\boldsymbol{\Sigma}_0$, where $\boldsymbol{\Sigma}$ is the blockwise covariance matrix specified above, $\boldsymbol{\Sigma}_0 = \mathbf{U}\boldsymbol{\Lambda}\mathbf{U}^{\mathsf{T}}$, $\mathbf{U}$ is an orthonormal matrix and $\boldsymbol{\Lambda}$ is a diagonal matrix with each elements generated from $U(0, 10)$ and ρ is a sensitivity parameter. The number of clusters for $\tilde{\boldsymbol{\Sigma}}$ is misspecified as 20, which is the same as the number of clusters in $\boldsymbol{\Sigma}$. The log ratio of the Monte Carlo MSEs of the voxel-wise estimator versus the envelope chain estimator is given in Figure 2b. Note $\rho = 0$ corresponds to the case where the blockwise structure still holds. Thus, the solid line in Figure 2b is the same as that in Figure 2a. We keep it here as a reference. As expected, when ρ deviates from 0, the log ratio of the MSEs for both estimators decreases. For example, the median of the log ratio of MSEs are 0.76 and 0.54 when $\rho = 0.5$ and $\rho = 1$, respectively. Although they are both smaller than that when $\rho = 0$ (0.92), they are still greater than 0, demonstrating the superior performance of the envelope chain estimator over the voxel-wise regression when violations of the block structure are moderate. The median of the log ratio of Monte Carlo variances are 0.76 and 0.54 when $\rho = 0.5$ and $\rho = 1$. This indicates that only about a half ($1/\exp(0.76) = 0.47$) of the original sample is needed to achieve the same median performance among all the coefficients when using the envelope chain

estimator compared to voxel-wise regression. Out of the 3995 elements in $\boldsymbol{\beta} \in \mathbb{R}^{779\times 5}$, only 17 and 23 elements in the envelope chain estimator are less efficient compared to those in the voxel-wise estimator for $\rho = 0.5$ and 1, respectively. In comparison, the median of the log ratio of MSEs between the voxel-wise estimator and the PCR estimator are -0.004 and -0.001 when $\rho = 0.5$ and $\rho = 1$, respectively. These small negative values imply that the PCR estimator performs slightly worse than the voxel-wise estimator in these three cases, although such a difference may not be significant. Similar trends are observed for Monte Carlo variances of these two estimators. We also report the log ratio of squared prediction biases of the voxel-wise estimator versus the envelope chain estimator in Figure 9b. The results are similar to those in Figure 2b. As ρ increases, the log ratios of squared prediction biases are expected to decrease. The median of the log ratio of squared prediction biases are 0.62 and 0.48 when $\rho = 0.5$ and $\rho = 1$, respectively. In comparison, the median of the log ratio of squared prediction biases between the voxel-wise estimator and PCR estimator are -0.004 and -0.001 when $\rho = 0.5$ and $\rho = 1$, respectively. These results again demonstrate that the envelope chain estimators outperform both PCR and voxel-wise estimators in terms of prediction even when the blockwise structure of the covariance matrix is moderately violated.

Now we consider a case where the block-wise structure is seriously violated. The simulation setting is similar to that in Section 4.1, except that **Step 2** is replaced with

Step 2*: Set $r = 799$ and $u = 1$. Generate $\tilde{\boldsymbol{\Gamma}} \in \mathbb{R}^{r\times u}$, where each element of $\tilde{\boldsymbol{\Gamma}}$ is generated from $N(0, 1)$. Set $\boldsymbol{\Gamma} = \tilde{\boldsymbol{\Gamma}}(\tilde{\boldsymbol{\Gamma}}^\top\tilde{\boldsymbol{\Gamma}})^{-1/2}$, and therefore $\boldsymbol{\Gamma}^\top\boldsymbol{\Gamma} = \mathbf{I}_u$. Generate $\boldsymbol{\Gamma}_0 \in \mathbb{R}^{r\times(r-u)}$ such that $(\boldsymbol{\Gamma}, \boldsymbol{\Gamma}_0)$ is an orthonormal matrix. Generate $\boldsymbol{\Omega} \in \mathbb{R}^{u\times u}$, where the element is firstly generated from a χ^2_1-distribution, and then multiplied by 2. Generate $\mathbf{M} \in \mathbb{R}^{(r-u)\times(r-u)}$, where each element of $\mathbf{M}$ is generated from $U(0, 1)$ and set $\boldsymbol{\Omega}_0 = \mathbf{M}^\top\mathbf{M}$. Generate $\boldsymbol{\eta} \in \mathbb{R}^{u\times p}$, where each element of $\boldsymbol{\eta}$ is firstly generated from $N(0, 1)$, then multiplied by 20, and finally we set $\boldsymbol{\beta} = \boldsymbol{\Gamma}\boldsymbol{\eta}$.

Additionally, in **Step 3**, we now generate $\mathbf{Y}$ from model (1) with the $\boldsymbol{\beta}$ given above and $\boldsymbol{\Sigma} = \boldsymbol{\Gamma}\boldsymbol{\Omega}\boldsymbol{\Gamma}^\top + \boldsymbol{\Gamma}_0\boldsymbol{\Omega}_0\boldsymbol{\Gamma}_0^\top$. While the envelope assumptions 2–3 still hold under the choices of parameters in **Step 2***, the clustering assumption 1 no longer holds.

The empirical distribution of the log ratios of Monte Carlo MSEs for the envelope chain and the PCR estimators versus the voxel-wise estimator are given in Figure 2c. The dotted curves with hollow circles and triangles show the results obtained when the number of clusters in the envelope chain estimator is specified as 20 or 500. The dotted curve with crosses shows the results obtained from PCR. Although the true data generating mechanism does not have any clustering structure, when we artificially specified the number of clusters to be 20, the median log ratio of Monte Carlo MSEs among all elements is 0.67. This is smaller than those reported in Figure 2b, where the blockwise structure of the covariance matrix is moderately violated. As the number of clusters increases to be close to the number of responses, say 500 in Figure 2c, the advantage of the envelope chain decreases (median log ratio of MSEs is 0.41 and 219 elements have larger MSEs). This decrease is also expected, as when the number of clusters is the same as the number of responses, the envelope chain estimator reduces to the voxel-wise estimator. The median log ratio of MSEs for the PCR estimator versus the voxel-wise estimator is 0.04. These results indicate the superiority of the envelope chain estimator over the PCR and voxel-wise estimators in terms of MSE. The median log ratios of Monte Carlo variances for the envelope chain estimator with 20 and 500 clusters, and the PCR estimator versus the voxel-wise estimator are 0.001, -0.001, 0.08, respectively. We observe that the envelope chain estimator is comparable to the voxel-wise estimator, but seems slightly worse than the PCR estimator in terms of variance. This may be specific to this simulation setting. Finally, we report the log ratio of squared prediction bias of the voxel-wise estimator versus the envelope chain estimator in Figure 9c. The results in this figure are similar to the MSE results in

Figure 2c, which again demonstrates the advantage of the envelope chain estimator over existing estimators in terms of prediction.

These sensitivity analyses reveal that the envelope chain estimator can achieve finite sample efficiency gain over the voxel-wise estimator when the deviation of the covariance matrix of the errors is moderate. Such a gain decreases when the covariance matrix further deviates from the blockwise structure and when the number of the clusters specified in the envelope chain estimator is large.

5 | STATISTICAL ANALYSIS

Attention Deficit Hyperactivity Disorder (ADHD) is a brain disorder with a number of symptoms, including inattention and hyperactivity. In the US, ADHD is associated with substantial lifelong impairment, with annual direct costs exceeding $36 billion/year. It is thus important to understand its pathophysiology, and to be able to perform early diagnosis.

The ADHD-200 study [35] was carried out among 776 children and adolescents across 8 independent imaging sites. A total of 15 individuals are removed because of low quality data. The study sample is composed of 491 typically developing individuals and 285 children and adolescents with ADHD between 7 and 21 years old. Phenotypic information including diagnostic status, dimensional ADHD symptom measures, age, sex, intelligence quotient (IQ) and lifetime medication status were collected. The data are accessible at `http://neurobureau.projects.nitrc.org/ADHD200/Data.html`.

We consider a dataset comprising 3D MRI images (with dimensions 30 × 36 × 30). The MRI dataset was preprocessed by standard steps including AC (anterior commissure) and PC (posterior commissure) correction, N2 bias field correction, skull-stripping, intensity inhomogeneity correction, cerebellum removal, segmentation, and registration (details given in [36]). The local volumetric group differences were quantified by generating RAVENS maps for the whole brain as well as each of the segmented tissue type (gray matter, white matter, ventricle and cerebrospinal fluid), respectively, using the deformation field obtained during registration. The 3D RAVENS map has dimensions 30 × 36 × 30 and the data contains a number of covariates, including ADHD status (binary), age, gender and handiness. After excluding all voxels with values of 0, located outside the brain volume, we count 11,607 voxels, which we use as responses in our models. The dimension of voxels is much larger than the sample size, thus, the traditional envelope method does not apply directly. Below, we modify the envelope method into an envelope chain method to address the high dimentionality and leveraging correlation structure simultaneously.

In order to estimate the number of clusters, we partition that data into training and test. We randomly allocate 600 observations as training data and 162 observations to the test data. We carry out hierarchical clustering to identify highly correlated voxels in the training data. We use the median square error loss among all the elements of $\mathbf{Y}$ as the loss function. The number of clusters that produce the smallest median squared error loss is 500 (see Figure 3), although the difference among the MSEs is not large. We therefore partition the voxels into 500 clusters. The sizes of the clusters range from 2 to 280 with a median of 11. In Figure 4, we present a colored map of the clustering membership for the entire brain volume, when the membership is plotted as a continuous variable. Here, we observe a high degree of cluster symmetry across the 2 hemispheres, with contra-lateral cortical regions devoted to the same neural processes typically clustering together.

As shown in Figure 5, the estimated envelope sizes $\hat{u}_j$ range from 0 to 23 with a median of 1. These are substantially smaller than those of the $\hat{r}_j$, indicating a potential efficiency gain. To further illustrate our estimator, we generated Figure 6 using 2 voxels from the ADHD study. The figure shows that our estimator is more efficient than

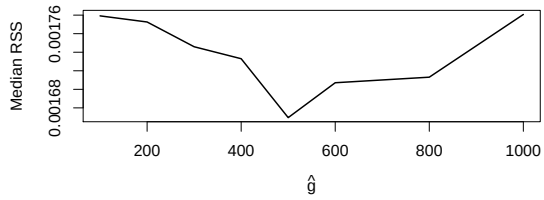
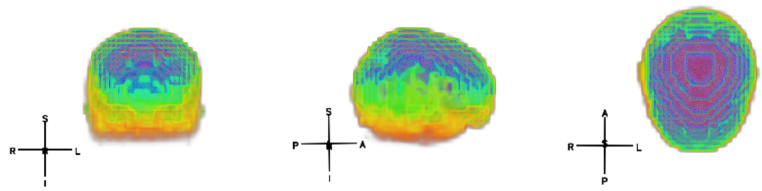


FIGURE 3 Median square error loss with different number of clusters $\hat{g}$ specified for the envelope chain estimator.

the voxel-wise estimator: while the voxel-wise regression estimator cannot distinguish the two groups (Figure 6a; p-value is 0.5), our estimator does so successfully (Figure 6b; p-value < 0.001).



(a) Coronal view **(b)** Sagittal view **(c)** Transverse view

FIGURE 4 Display of cluster membership as a continuous variable for ADHD data.

Figure 7 compares the variances between the voxel-wise and envelope chain estimators using either the closed form variance given in Proposition 2 (Figure 7a) or using bootstrap (Figure 7b). As indicated by Proposition 2, the variance of the voxel-wise estimator is guaranteed to be larger than that of the envelope chain estimator, conditional on the envelope dimension and cluster membership. This is revealed in Figure 7a, as the entire cumulative density function of the empirical distribution of the log ratio of two variances is on the right hand side of 0, with a median of 1.89. The nonparametric bootstrap (sampling individuals with replacement) variance takes into account the uncertainty in estimating u_j and clustering membership as well as the variability in the predictor distribution. Despite the additional variability, about 70.8% of voxels show an efficiency gain when using the envelope chain estimator. The median and mean efficiency gain (ratio of two bootstrap variance estimators) are 1.2 and 1.9, respectively, indicating about 16.7% and 47.4% of data can be saved using envelope chain estimator to achieve the same power for the median and average performance in estimating all the regression coefficients.

To identify the brain regions that are significantly different among ADHD and normal individuals, we calculate the Z value for the voxel-wise regression and the envelope chain estimator. We use the bootstrap variance as it accommodates the variabilities due to clustering and the selection of the envelope dimensions. As adjusting for multiple comparison is not straightforward due to the correlations among voxels [24, 37], we use Z value ≥ 3 as the threshold for significance. The results are shown in Figure 8. The areas that envelope chain estimator finds to be significantly related with ADHD include cerebellum, hippocampus, and caudate-nucleus. These areas are well

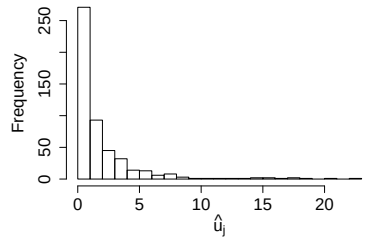


FIGURE 5 Estimated envelope dimension for each envelope chain across clusters in different bootstrap samples.

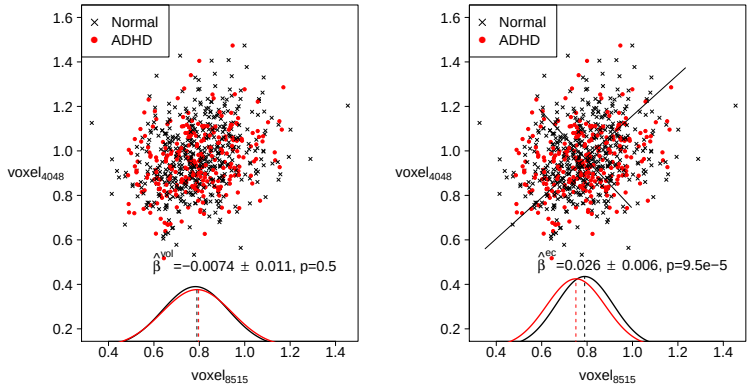


FIGURE 6 The voxel-wise estimator and the envelope chain estimator of the effect of having ADHD on two voxels in ADHD-200 study.

documented in the neuroscience literature to be affected by ADHD [38, 39, 40]. As the signal of the effect of ADHD on the brain activities in this study is quite strong, both the standard estimator and the envelope chain lead to comparable results. However, the standard estimator produces noisier results, with a number of significant regions scattered across the brain (e.g., the region with red circle in Figure 8). The regions such as inferior temporal cortex, inferior frontal gyrus opercular part, postcentral gyrus are detected to be significantly different among ADHD and normal individuals by voxel-wise estimator but not envelope chain estimator, some of which were also noted in the MRI literature [41, 42].

We also implement the PCR estimator. We choose the smallest number of principal components that can explain 90% of variance. The reduced outcome is then used in a linear regression to obtain the regression coefficient. The variance of the PCR are evaluated using bootstrap. The *Z*-values of the PCR corresponds to the coefficient for ADHD status are shown in Figure 10. The PCR estimator also reveals that cerebellum, hippocampus are significantly related to ADHD status. However, the PCR estimator finds more significant but smaller regions than the voxel-wise and envelope estimators (additional regions include rectus gyrus, vermis and some scattered voxels). We also implement the PCR estimator with 70% of variance explained. The result is similar to that of 90% with slightly fewer regions detected. Because the dimension reduction step in PCR estimator does not use the predictors, the results may be

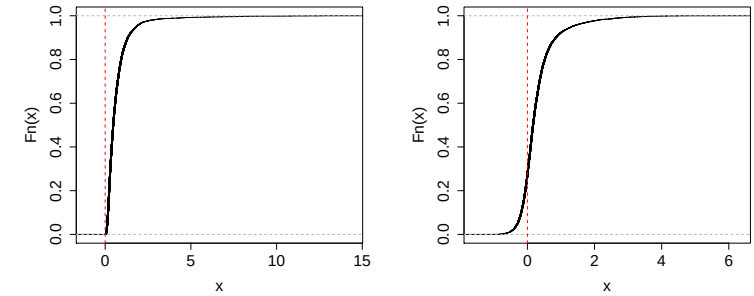


FIGURE 7 Empirical cumulative distribution of the log ratio of the Monte Carlo variances of the voxel-wise estimator versus envelope chain estimator.

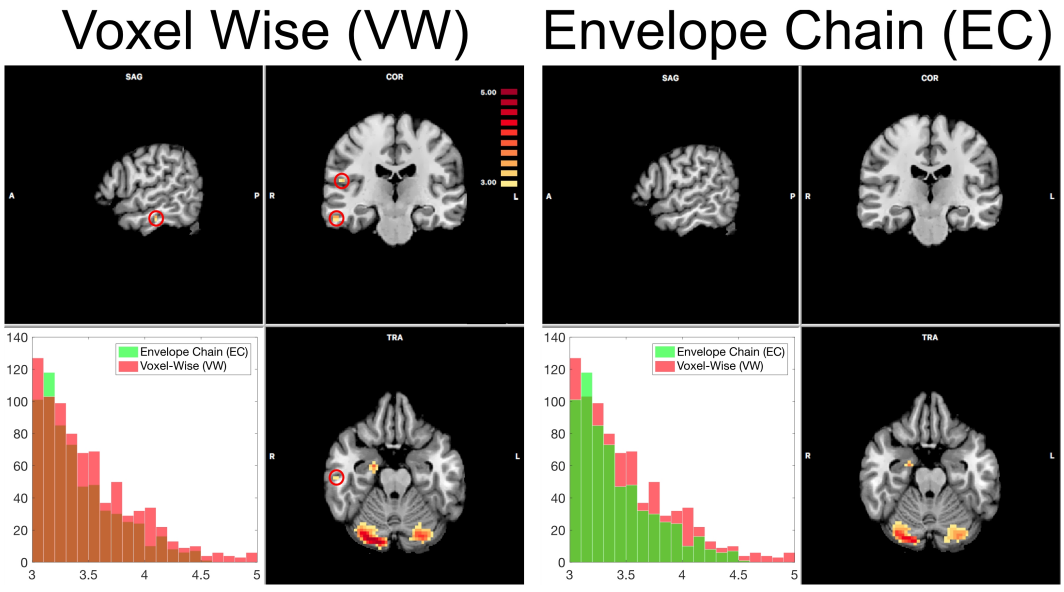


FIGURE 8 Brain areas that are significantly different in ADHD versus normal individuals using the voxel-wise (VW) and envelope chain (EC) estimators. The subfigure at the lower lefthand side is the distribution of Z scores for each estimator. The two histograms are identical showing the Z value from the two estimators with different emphasizes of the contrast (red on top vs green on top).

subject to information loss and thus may not be reliable.

6 | DISCUSSION

In this paper, we adapt the envelope procedure and apply the envelope chain procedure as a modified tool to carry out sufficient dimension reduction in high dimensional brain imaging data. The resulting estimator leads to a significant efficiency gain over the traditional voxel-wise estimator by leveraging the correlations between voxels. This framework is applicable to both MRI and fMRI. Due to space constraint, we only illustrate our method in an MRI study where redundant information is present in the spatial correlation. Similar approaches can be extended where temporal and spatial intensities are stacked in the multivariate response without distinction. Alternatively, the temporal information can be treated differently from the spatial information by considering a mixed effects model, in which the temporal correlation is modeled with additional assumptions. We report the results and comparisons between different approaches elsewhere.

As an early foray into the use of correlation to gain efficiency in neuroimaging studies, this study leaves gaps to be filled with regard to both theory and computation. However, the empirical results presented here are rather compelling as an improvement over the traditional estimators. The adapted procedure can also be applied to other high dimensional data problems, such as genetic data, as long as the problem can be framed as a multivariate linear regression.

We used the hierarchical clustering algorithm to divide the voxels into smaller groups. As long as there are no dominating clusters, an increase in the number of clusters leads to a decrease in the size of each cluster. Other clustering methods could be used as well. It would be interesting to see how the method would work using alternative clustering algorithms.

ACKNOWLEDGMENT

We sincerely thank the editor, associate editor, and reviewers for their helpful and insightful comments. Professor Essa Yacoub is supported by NIH P41 EB015894. Lan Liu is supported by Grant-in-aid at the University of Minnesota at Twin Cities and NSF DMS 1916013. Wei Li is supported by the Fundamental Research Funds for the Central Universities and the Research Funds of Renmin University of China. The authors want to thank the audience at the Department of Biostatistics at the University of Pennsylvania, the Department of Statistics at Peking University for insightful comments and suggestions. The authors also express special thanks to Prof. Lexin Li and Prof. Joerg Polzehl for their tremendous help, patience and suggestions in the data processing procedure. The content is solely the responsibility of the authors.

APPENDIX

| Proof of Proposition 1

Proof According to model (2), we have

$$\mathbf{Y}_{i(j)} = \boldsymbol{\alpha}_{(j)} + \Gamma_{(j)}\boldsymbol{\eta}_{(j)}\mathbf{X}_i + \boldsymbol{\varepsilon}_{i(j)}$$

for each $j = 1, \ldots, g$. By stacking these equations, we have

$$\begin{pmatrix} \mathbf{Y}_{i(1)} \\ \vdots \\ \mathbf{Y}_{i(g)} \end{pmatrix} = \begin{pmatrix} \boldsymbol{\alpha}_{(1)} \\ \vdots \\ \boldsymbol{\alpha}_{(g)} \end{pmatrix} + \begin{pmatrix} \boldsymbol{\Gamma}_{(1)}\boldsymbol{\eta}_{(1)} \\ \vdots \\ \boldsymbol{\Gamma}_{(g)}\boldsymbol{\eta}_{(g)} \end{pmatrix} \mathbf{X}_i + \begin{pmatrix} \boldsymbol{\varepsilon}_{i(1)} \\ \vdots \\ \boldsymbol{\varepsilon}_{i(g)} \end{pmatrix}. \tag{3}$$

Note that $\mathbf{Y}_i = (\mathbf{Y}_{i(1)}^T, \ldots, \mathbf{Y}_{i(g)}^T)^T$, $\boldsymbol{\alpha} = (\boldsymbol{\alpha}_{i(1)}^T, \ldots, \boldsymbol{\alpha}_{i(g)}^T)^T$, $\boldsymbol{\Gamma} = \text{bdiag}(\boldsymbol{\Gamma}_{(1)}, \ldots, \boldsymbol{\Gamma}_{(g)})$, $\boldsymbol{\eta} = (\boldsymbol{\eta}_{(1)}^\top, \ldots, \boldsymbol{\eta}_{(g)}^\top)^\top$, and $\boldsymbol{\varepsilon}_i = (\boldsymbol{\varepsilon}_{i(1)}^T, \ldots, \boldsymbol{\varepsilon}_{i(g)}^T)^T$. We can rewrite (3) as follows:

$$\mathbf{Y}_i = \boldsymbol{\alpha} + \boldsymbol{\Gamma}\boldsymbol{\eta}\mathbf{X}_i + \boldsymbol{\varepsilon}_i.$$

Under Assumption 1, we have $\text{Var}(\boldsymbol{\varepsilon}_i \mid \mathbf{X}_i) = \text{bdiag}(\boldsymbol{\Sigma}_{(1)}, \ldots, \boldsymbol{\Sigma}_{(g)})$. By the expressions of $\boldsymbol{\Sigma}_{(j)}$ below (2) for each j and that $\boldsymbol{\Gamma}_0 = \text{bdiag}(\boldsymbol{\Gamma}_{0(1)}, \ldots, \boldsymbol{\Gamma}_{0(g)})$, $\boldsymbol{\Omega} = \text{bdiag}(\boldsymbol{\Omega}_{(1)}, \ldots, \boldsymbol{\Omega}_{(g)})$, $\boldsymbol{\Omega}_0 = \text{bdiag}(\boldsymbol{\Omega}_{0(1)}, \ldots, \boldsymbol{\Omega}_{0(g)})$, we have $\boldsymbol{\Sigma} = \boldsymbol{\Gamma}\boldsymbol{\Omega}\boldsymbol{\Gamma}^T + \boldsymbol{\Gamma}_0\boldsymbol{\Omega}_0\boldsymbol{\Gamma}_0^T$.

Note that $\boldsymbol{\Gamma}\boldsymbol{\Omega}\boldsymbol{\Gamma}^T$ and $\boldsymbol{\Gamma}_0\boldsymbol{\Omega}_0\boldsymbol{\Gamma}_0^T$ are symmetric positive semi-definite matrics with $\boldsymbol{\Gamma}\boldsymbol{\Omega}\boldsymbol{\Gamma}^T\boldsymbol{\Gamma}_0\boldsymbol{\Omega}_0\boldsymbol{\Gamma}_0^T = 0$. Then according to Corollary 3.1 in Cook et al. [20], the envelope for $\boldsymbol{\beta}$ is $\text{span}(\boldsymbol{\Gamma}\boldsymbol{\Omega}\boldsymbol{\Gamma}^T)$. Because $\boldsymbol{\Gamma}$ is orthogonal and $\boldsymbol{\Omega}$ is non-singular, we have $\text{span}(\boldsymbol{\Gamma}\boldsymbol{\Omega}\boldsymbol{\Gamma}^T) = \text{span}(\boldsymbol{\Gamma})$ with dimension $u = u_1 + \cdots + u_g$. □

| Proof of Proposition 2

Proof Suppose that model (2) holds and the errors $\boldsymbol{\varepsilon}_i$ are normally distributed. Then according to Theorem 5.1 in Cook et al. [20], for each cluster $j = 1, \ldots, p$, we have

$$\sqrt{n}\{\text{vec}(\hat{\boldsymbol{\beta}}_{(j)}^{\text{vol}}) - \text{vec}(\boldsymbol{\beta}_{(j)})\} \xrightarrow{d} N(\mathbf{0}, \mathbf{V}_{(j)}^{\text{vol}}), \quad \sqrt{n}\{\text{vec}(\hat{\boldsymbol{\beta}}_{(j)}^{\text{ec}}) - \text{vec}(\boldsymbol{\beta}_{(j)})\} \xrightarrow{d} N(\mathbf{0}, \mathbf{V}_{(j)}^{\text{ec}}), \quad \text{and} \quad \mathbf{V}_{(j)}^{\text{ec}} \le \mathbf{V}_{(j)}^{\text{vol}}.$$

Under Assumption 1, response variables in different clusters are conditionally independent. Then we have

$$\sqrt{n}\{\text{vec}(\hat{\boldsymbol{\beta}}^{\text{vol}}) - \text{vec}(\boldsymbol{\beta})\} \xrightarrow{d} N(\mathbf{0}, \mathbf{V}^{\text{vol}}), \quad \sqrt{n}\{\text{vec}(\hat{\boldsymbol{\beta}}^{\text{ec}}) - \text{vec}(\boldsymbol{\beta})\} \xrightarrow{d} N(\mathbf{0}, \mathbf{V}^{\text{ec}})$$

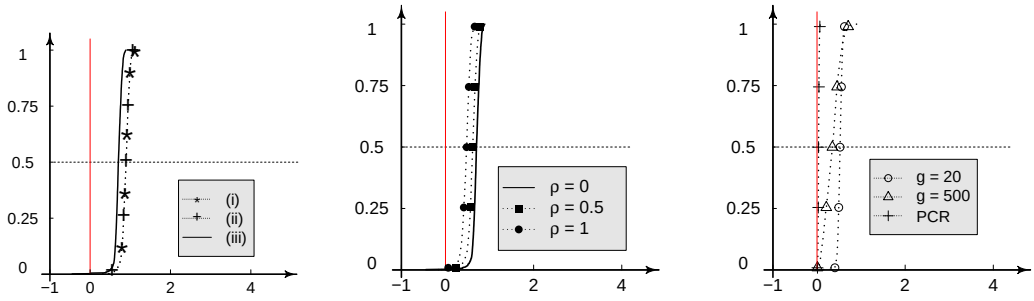
for $\mathbf{V}^{\text{vol}} = \text{bdiag}(\mathbf{V}_{(1)}^{\text{vol}}, \ldots, \mathbf{V}_{(g)}^{\text{vol}})$ and $\mathbf{V}^{\text{ec}} = \text{bdiag}(\mathbf{V}_{(1)}^{\text{ec}}, \ldots, \mathbf{V}_{(g)}^{\text{ec}})$. By noting that $\mathbf{V}_{(j)}^{\text{ec}} \le \mathbf{V}_{(j)}^{\text{vol}}$ for each j, we have $\mathbf{V}^{\text{ec}} \le \mathbf{V}^{\text{vol}}$.

| Proof of Proposition 3

Proof Suppose that there are $\tilde{g}$ clusters for the given clustering structure. Because BIC will select the correct envelope dimension with probability tending to 1 as $n \to \infty$ [26, 31, 32], then according to Theorem 5.1 in Cook et al. [20], for each cluster $j = 1, \ldots, \tilde{g}$, we have $\sqrt{n}\{\text{vec}(\hat{\boldsymbol{\beta}}_{(j)}^{\text{ec}}) - \text{vec}(\boldsymbol{\beta}_{(j)})\}$ converges to a normal distribution. This implies that $\hat{\boldsymbol{\beta}}_{(j)}^{\text{ec}}$ is a $\sqrt{n}$-consistent estimator for $\boldsymbol{\beta}_{(j)}$. By collecting the result for each j, we conclude that $\hat{\boldsymbol{\beta}}^{\text{ec}}$ is a $\sqrt{n}$-consistent estimator for $\boldsymbol{\beta}$.

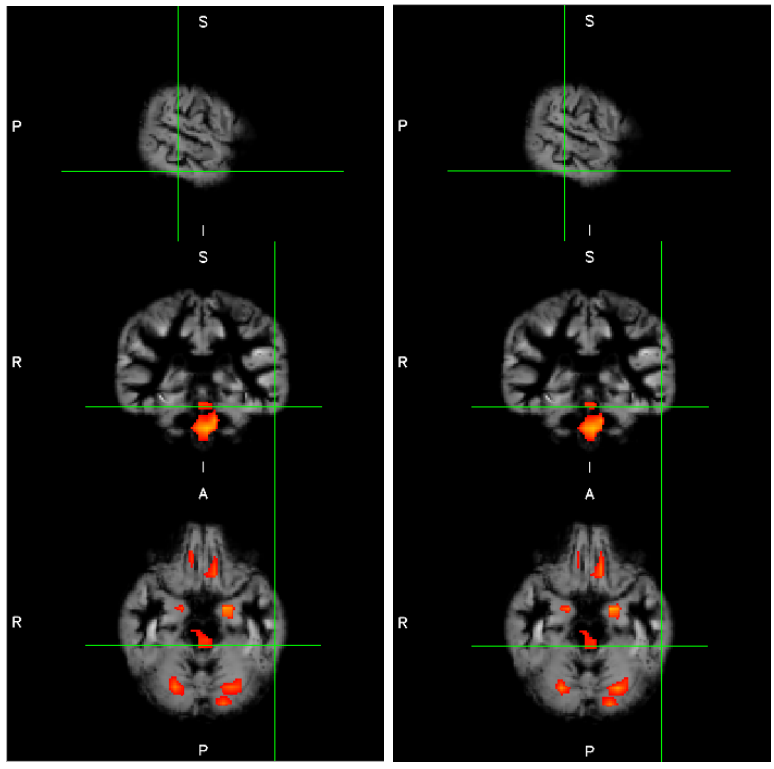
| Additional simulation and data analysis results

Figure 9 shows the prediction results of the voxel-wise, PCR and envelope chain estimators on test data when the envelope structure is correct, moderately and seriously violated. Figure 9 shows the significantly different brain regions among ADHD versus normal individuals that are identified by PCR.



(a) Blockwise structure holds for variance. **(b)** Blockwise structure is moderately violated. **(c)** Blockwise structure is seriously violated.

FIGURE 9 Empirical cumulative distribution function of the log ratio of squared prediction biases between the voxel-wise and the envelope chain estimators when (a) the envelope structure holds: (i) both clustering membership and u_j are known; (ii) only clustering membership is known; (iii) neither clustering membership or u_j is known; (b) the envelope structure is moderately violated: sensitivity parameter ρ is chosen as 0.5 and 1; and (c) the envelope structure is seriously violated: the specified number of clusters are 20 and 500. PCR results are also included in (c).



(a) PCR estimate with 90% variance explained **(b)** PCR estimate with 70% variance explained

FIGURE 10 Brain areas that are significantly different in ADHD versus normal individuals using the principal component regression (PCR) estimators. The high lighted regions are those whose Z values are larger than 3 in magnitude and the regions with larger values (up to 5) have a yellower tone.

REFERENCES

[1] Ogawa S, Menon RS, Tank DW, Kim SG, Merkle H, Ellermann JM, et al. Functional brain mapping by blood oxygenation level-dependent contrast magnetic resonance imaging. A comparison of signal characteristics with a biophysical model. Biophysical Journal 1993;64:803–812.

[2] Alexander GE, Chen K, Merkley TL, Reiman EM, Caselli RJ, Aschenbrenner M, et al. Regional network of magnetic resonance imaging gray matter volume in healthy aging. Neuroreport 2006;17:951–956.

[3] Biswal B, Zerrin Yetkin F, Haughton VM, Hyde JS. Functional connectivity in the motor cortex of resting human brain using echo-planar MRI. Magnetic Resonance in Medicine 1995;34:537–541.

[4] Van Dijk KR, Hedden T, Venkataraman A, Evans KC, Lazar SW, Buckner RL. Intrinsic functional connectivity as a tool for human connectomics: theory, properties, and optimization. Journal of Neurophysiology 2010;103:297–321.

[5] Cordes D, Haughton VM, Arfanakis K, Wendt GJ, Turski PA, Moritz CH, et al. Mapping functionally related regions of brain with functional connectivity MR imaging. American Journal of Neuroradiology 2000;21:1636–1644.

[6] Buckner RL, Andrews-Hanna JR, Schacter DL. The brain's default network: anatomy, function, and relevance to disease. Annals of the New York Academy of Sciences 2008;1124:1–38.

[7] Hampson M, Peterson BS, Skudlarski P, Gatenby JC, Gore JC. Detection of functional connectivity using temporal correlations in MR images. Human Brain Mapping 2002;15:247–262.

[8] Fox MD, Snyder AZ, Vincent JL, Corbetta M, Van Essen DC, Raichle ME. The human brain is intrinsically organized into dynamic, anticorrelated functional networks. Proceedings of the National Academy of Sciences 2005;102:9673–9678.

[9] Vincent JL, Kahn I, Snyder AZ, Raichle ME, Buckner RL. Evidence for a frontoparietal control system revealed by intrinsic functional connectivity. Journal of Neurophysiology 2008;100:3328–3342.

[10] Newton AT, Morgan VL, Gore JC. Task demand modulation of steady-state functional connectivity to primary motor cortex. Human Brain Mapping 2007;28:663–672.

[11] Worsley KJ, Friston KJ. Analysis of fMRI time-series revisited–again. Neuroimage 1995;2:173–181.

[12] Friston KJ, Penny W, Phillips C, Kiebel S, Hinton G, Ashburner J. Classical and Bayesian inference in neuroimaging: theory. NeuroImage 2002;16:465–483.

[13] Frackowiak RS. Human Brain Function. Elsevier; 2004.

[14] Calhoun VD, Liu J, Adalı T. A review of group ICA for fMRI data and ICA for joint inference of imaging, genetic, and ERP data. NeuroImage 2009;45(1):S163 – S172. http://www.sciencedirect.com/science/article/pii/S1053811908012032.

[15] Mwangi B, Tian TS, Soares JC. A review of feature reduction techniques in neuroimaging. Neuroinformatics 2014;12(2):229–244.

[16] Mahmoudi A, Takerkart S, Regragui F, Boussaoud D, Brovelli A. Multivoxel pattern analysis for fMRI data: a review. Computational and Mathematical Methods in Medicine 2012;2012:1–14.

[17] Krishnan A, Williams LJ, McIntosh AR, Abdi H. Partial least squares (PLS) methods for neuroimaging: a tutorial and review. Neuroimage 2011;56(2):455–475.

[18] Gottfries J, Blennow K, Wallin A, Gottfries C. Diagnosis of dementias using partial least squares discriminant analysis. Dementia and Geriatric Cognitive Disorders 1995;6(2):83–88.

[19] Lehmann C, Koenig T, Jelic V, Prichep L, John RE, Wahlund LO, et al. Application and comparison of classification algorithms for recognition of Alzheimer's disease in electrical brain activity (EEG). Journal of Neuroscience Methods 2007;161(2):342–350.

[20] Cook RD, Li B, Chiaromonte F. Envelope models for parsimonious and efficient multivariate linear regression. Statistica Sinica 2010;20:927–960.

[21] Su Z, Cook RD. Inner envelopes: efficient estimation in multivariate linear regression. Biometrika 2012;99:687–702.

[22] Cook RD, Forzani L, Zhang X. Envelopes and reduced-rank regression. Biometrika 2015;102:439–456.

[23] Su Z, Zhu G, Chen X, Yang Y. Sparse envelope model: efficient estimation and response variable selection in multivariate linear regression. Biometrika 2016;103:579–593.

[24] Li L, Zhang X. Parsimonious tensor response regression. Journal of the American Statistical Association 2017;112:1131–1146.

[25] Cook RD, Zhang X. Foundations for envelope models and methods. Journal of the American Statistical Association 2015;110:599–611.

[26] Cook RD. An Introduction to Envelopes: Dimension Reduction for Efficient Estimation in Multivariate Statistics. John Wiley & Sons; 2018.

[27] Wold S, Kettaneh N, Tjessem K. Hierarchical multiblock PLS and PC models for easier model interpretation and as an alternative to variable selection. Journal of Chemometrics 1996;10:463–482.

[28] Rokach L, Maimon O. Clustering methods. In: Data mining and knowledge discovery handbook Springer; 2005.p. 321–352.

[29] Cook RD, Su Z, Yang Y. envlp: A MATLAB toolbox for computing envelope estimators in multivariate analysis. Journal of Statistical Software 2015;62:1–20.

[30] Lee M, Su Z. Renvlp: computing envelope estimators; 2018, `https://CRAN.R-project.org/package=Renvlp`, r package version 2.5.

[31] Eck DJ, Cook RD. Weighted envelope estimation to handle variability in model selection. Biometrika 2017;104:743–749.

[32] Yang Y. Can the strengths of AIC and BIC be shared? A conflict between model indentification and regression estimation. Biometrika 2005;92(4):937–950.

[33] Suzuki R, Shimodaira H. An application of multiscale bootstrap resampling to hierarchical clustering of microarray data: How accurate are these clusters. In: The Fifteenth International Conference on Genome Informatics, vol. 34 Pacifico Convention Plaza Yokohama Japan; 2004. .

[34] Suzuki R, Shimodaira H. Pvclust: an R package for assessing the uncertainty in hierarchical clustering. Bioinformatics 2006;22(12):1540–1542.

[35] Bellec P, Chu C, Chouinard-Decorte F, Benhajali Y, Margulies DS, Craddock RC. The Neuro Bureau ADHD-200 Preprocessed repository. NeuroImage 2017;144:275 – 286.

[36] Zhou H, Li L, Zhu H. Tensor regression with applications in neuroimaging data analysis. Journal of the American Statistical Association 2013;108:540–552.

[37] Bennet C, Baird A, Miller M, Wolford G. Neural correlates of interspecies perspective taking in the post-mortem Atlantic salmon: An argument for proper multiple comparisons correction. Journal of Serendipitous and Unexpected Results 2010;1:1–5.

[38] Castellanos FX, Giedd JN, Eckburg P, Marsh WL, Vaituzis AC, Kaysen D, et al. Quantitative morphology of the caudate nucleus in attention deficit hyperactivity disorder. The American Journal of Psychiatry 1994;151:1791–1796.

[39] Berquin PC, Giedd JN, Jacobsen LK, Hamburger SD, Krain AL, Rapoport JL, et al. Cerebellum in attention-deficit hyperactivity disorder: a morphometric MRI study. Neurology 1998;50:1087–1093.

[40] Al-Amin M, Zinchenko A, Geyer T. Hippocampal subfield volume changes in subtypes of attention deficit hyperactivity disorder. Brain Research 2018;1685:1–8.

[41] DeWeerd P, Peralta III MR, Desimone R, Ungerleider LG. Loss of attentional stimulus selection after extrastriate cortical lesions in macaques. Nature Neuroscience 1999;2:753–758.

[42] Max JE, Fox PT, Lancaster JL, Kochunov P, Mathews K, Manes FF, et al. Putamen lesions and the development of attention-deficit/hyperactivity symptomatology. Journal of the American Academy of Child & Adolescent Psychiatry 2002;41:563–571.