

Regularized estimation of hemodynamic response function for fMRI data

CHUNMING ZHANG* AND ZHENGJUN ZHANG

One of the primary goals in analyzing fMRI data is to estimate the Hemodynamic Response Function (HRF), which is a large-dimensional parameter vector possessing some form of sparsity. This paper introduces a varying-dimensional model for the HRF, and develops novel regularization methods for estimating the HRF from fMRI time series via incorporating the sparsity feature. Particularly, we present three types of penalty choice methods: the Lasso, the adaptive Lasso and the SCAD. Simulation studies demonstrate the advantages of regularization methods, in terms of sparsity recovery, over conventional non-regularized approaches which restrict the HRF to be fixed low dimensional without capturing the sparsity structure. We illustrate the regularized methods for estimating the HRF using a real fMRI data set and compare with results offered by a popular imaging analysis tool AFNI.

KEYWORDS AND PHRASES: Covariance matrix, Linear model, Loss function, Penalty, Sparsity, Stimuli, Time resolution.

1. INTRODUCTION

Functional magnetic resonance imaging (fMRI) is a recent and exciting method that allows investigators to determine which areas of the brain are involved in a cognitive task. Following Ward (2001) and Worsley et al. (2002), a single-voxel fMRI time-series $\{s(t_i), y(t_i)\}_{i=1}^n$, for a given scan and a given subject, can be captured by the convolution model,

$$(1.1) \quad y(t) = d(t) + s * h(t) + \epsilon(t), \quad t = t_1, \dots, t_n,$$

where $*$ denotes the convolution operator, $y(t)$ is the measured noisy fMRI signal, $s(t)$ is the external input stimulus (which could be from a design either block- or event-related and where $s(t) = 1$ or 0 indicates the presence or absence of a stimulus), $h(t)$ is the hemodynamic response function (HRF) at time t after neural activity, $d(t)$ is a slowly drifting baseline, and the errors $\epsilon(t_i)$ are zero-mean and temporally autocorrelated. Similar models can be found in Chen et al. (2006). Refer to Zhang and Yu (2008) and references therein for a recent review of statistical issues and methods in fMRI data analysis.

*Corresponding author.

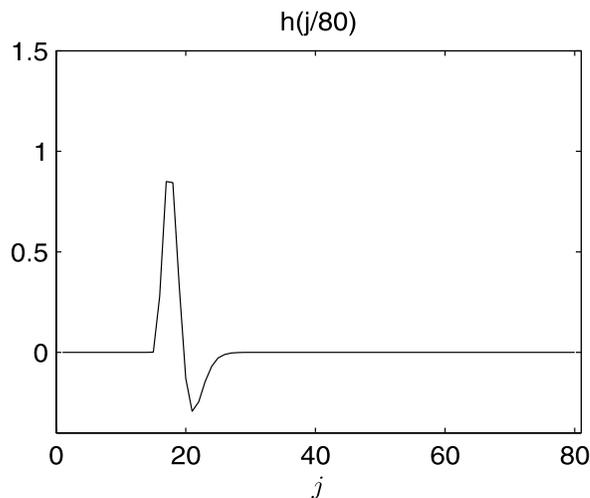


Figure 1. An illustrative plot of the HRF $h(t_j)$ with $t_j = j/n$ and $n = 80$.

One of the primary interests to neuroscientists in analyzing an event-related fMRI experiment is the estimation of the underlying HRF at time points t_j . Typically, the peak value of the HRF $h(\cdot)$ is reached after a short delay of the stimulus and drops quickly to zero. A typical example of $h(\cdot)$, given in Glover (1999), is plotted in Figure 1. If we look at the entire sequence of HRF coefficients $\{h(t_j) : j = 1, \dots, n\}$, the number of indices j for which $h(t_j)$ is non-zero is small. Thus, to obtain statistically more efficient estimates of the HRF associated with event-related fMRI experiments, the sparsity of the HRF needs to be taken into account. We thus suppose that exactly s_n values of $h(t_j)$ are non-zero, and $h(t_j) = 0$ for all $j > p_n$. By “sparsity of HRF coefficients”, we mean that s_n is small compared with the dimension p_n . Now the problem amounts to estimating the first p_n values of $h(t_j)$, where $s_n \leq p_n$ and p_n is less than n , the length of the fMRI time series. In neuroimaging studies, the temporal drift $d(\cdot)$ is a nuisance function and usually approximated by a k th (at most third order) polynomial; see for example, the popular imaging analysis tool AFNI at <http://afni.nimh.nih.gov/afni/> (Cox, 1996) and Worsley et al. (2002). As such, (1.1) is re-expressed as

$$(1.2) \quad \mathbf{y} = \tilde{\mathbf{T}}\tilde{\boldsymbol{\alpha}} + \mathbf{S}\mathbf{h}_n + \boldsymbol{\epsilon},$$

where $\mathbf{y} = (y(t_1), \dots, y(t_n))^T$,

$$\tilde{\mathbf{T}} = \begin{bmatrix} 1 & t_1 & \cdots & t_1^k \\ \vdots & \vdots & \ddots & \vdots \\ 1 & t_n & \cdots & t_n^k \end{bmatrix},$$

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

is the $n \times p_n$ Toeplitz matrix, $\boldsymbol{\epsilon} = (\epsilon(t_1), \dots, \epsilon(t_n))^T$, and $\tilde{\boldsymbol{\alpha}} = (\alpha_0, \alpha_1, \dots, \alpha_k)^T$ and $\mathbf{h}_n = (h(t_1), \dots, h(t_{p_n}))^T$ are both vectors of unknown parameters. (Note that the forms of models (1.1)–(1.2) also accommodate fMRI data with multiple types of stimuli and multiple runs. Details can be found in Zhang and Yu (2008), for example.)

In the framework of linear regression model (1.2), the p_n -dimensional vector \mathbf{h}_n can be estimated via weighted least-squares using the quadratic loss. Recall also that it is s_n that characterizes the features of the HRF that are “really there”. In practice, however, s_n , termed the “intrinsic dimensionality of fMRI data” (Cordes and Nandy, 2006), is unknown for real fMRI data. As far as we know, all published work assumes that

$$(1.3) \quad s_n = p_n = p$$

is a known fixed number, followed by traditional parametric inference based on asymptotic derivations of fixed-dimensional estimators. Indeed, neuroscientists are in general interested in the first 15–20, or 30 seconds at best, of the HRF, following the presentation of the stimulus (in an event-related experiment). There, the selection of the numbers 15, 20, 30 etc is made based more on experience or pilot studies. Some potential problems may arise from the use of (1.3): The larger the number p is specified, the more likely s_n will be over-estimated and overfitting will occur; conversely, the larger the modeling bias will be induced.

To reduce modeling biases due partly to a restrictive choice of p_n and to achieve effective feature selection, statistical modeling and estimation which allow flexible choices of the dimension p_n as well as data-driven methods for s_n are desired. This motivates us to consider a more flexible relation between s_n , p_n and n expressible as follows,

$$(1.4) \quad s_n \leq p_n < n \text{ or } p_n \text{ varies with } n.$$

We call the resulting model (1.2) a “varying-dimensional model” (VDM) relating fMRI signals to neural changes, at a given voxel. Clearly, the set-up (1.4) includes both the fixed-dimensional case (1.3) and various other cases where p_n depends on n .

This paper intends to explore the penalization methods for estimating the p_n -dimensional parameter \mathbf{h}_n . The proposed methods differ from existing methods in two aspects.

First, the choice of p_n is allowed to be either fixed or reasonably large, provided it embeds the “intrinsic dimension” s_n . Second, the use of penalization methods better integrates the sparsity of the HRF coefficients. In the statistical literature, many asymptotic results, such as consistency, sparsity property and limit law, have been established for the conventional penalized estimators of regression parameters in the traditional linear regression model with i.i.d. observations. To our knowledge, this is the first paper in the literature on examining the practical performance of the penalized estimates (in the temporal domain) of a varying-dimensional HRF for fMRI time series. Our simulation studies lend support to the regularization methods applied to the important application of fMRI time series with correlated data. Particularly, the results demonstrate the advantages of regularization methods, in terms of sparsity recovery, over conventional non-regularized approaches which restrict the HRF to be fixed low dimensional without capturing the sparsity structure.

More generally, penalization arises as a very natural approach to making inferences for sparse signals in various large-dimensional statistical problems. As stated in Johnstone and Silverman (2004, p. 1595), “*Estimating a sparse signal is like finding needles in a haystack; it will be necessary to find which are the very few signal values that are nonzero, as well as to estimate them. On the other hand, estimating a dense signal is more like finding straw in a haystack; no longer will we be surprised if a particular μ_i is nonzero.*” In their set-up, the signals $\{\mu_i\}$ correspond to the HRF coefficients $\{h(t_i)\}$ in the convolution model. Our simulation studies indicate that the penalization methods for estimating the HRF are adaptive to sparsity when it is advantageous to do so: For a sparse \mathbf{h}_n , the penalized methods produce estimates in which many are zero, while for the dense \mathbf{h}_n , the penalized methods perform as good as the non-penalized approach.

The rest of the paper is organized as follows. Section 2 introduces the regularization methods for estimating parameters in regression models with independent data. Section 3 develops regularized estimation of the HRF coefficients in varying-dimensional convolution models for fMRI time series data. Section 4 performs numerical assessment of the regularization methods. Section 5 analyzes a real fMRI data and compares estimates of the HRF using the non-regularized method of AFNI and regularized methods. Section 6 concludes this paper with a brief discussion.

2. REGULARIZED ESTIMATION WITH INDEPENDENT DATA

Regularization is a technique aiming at obtaining well behaved solutions to overparameterized estimation problems. This technique is particularly appealing for dealing with large-dimensional problems. A review article on regularization in statistics can be found in Bickel and Li (2006). This

section overviews some regularized estimation methods. Let (\mathbf{X}_n, Y_n) be the generic pair of a random realization from some underlying population, where $\mathbf{X}_n = (X_1, \dots, X_{p_n})^T$ is the input vector and Y_n is the output variable. The dimension p_n is related to n according to (1.4). Suppose that the mean regression function is modeled by

$$(2.1) \quad m(\mathbf{x}_n) = E(Y_n | \mathbf{X}_n = \mathbf{x}_n) = F^{-1}(\beta_{n,0;0} + \mathbf{x}_n^T \boldsymbol{\beta}_{n;0}),$$

where F is a known link function, $\beta_{n,0;0} \in \mathbb{R}^1$ and $\boldsymbol{\beta}_{n;0} = (\beta_{n,1;0}, \dots, \beta_{n,p_n;0})^T \in \mathbb{R}^{p_n}$ are the unknown true parameters. For sparse models, it is assumed that part of the true parameters in $\boldsymbol{\beta}_{n;0}$ are exactly zero. Without loss of generality, write $\boldsymbol{\beta}_{n;0} = (\boldsymbol{\beta}_{n;0}^{(I)T}, \boldsymbol{\beta}_{n;0}^{(II)T})^T$, in which the $\boldsymbol{\beta}_{n;0}^{(I)}$ part collects all nonzero coefficients, and $\boldsymbol{\beta}_{n;0}^{(II)} = \mathbf{0}$. Let s_n be the number of nonzero elements in $\boldsymbol{\beta}_{n;0}$, i.e., the length of the vector $\boldsymbol{\beta}_{n;0}^{(I)}$.

The goal is to estimate, via penalization, the true parameters $(\beta_{n,0;0}, \boldsymbol{\beta}_{n;0})$ in the presence of sparsity. Let $\mathcal{T}_n = \{(\mathbf{X}_{n1}, Y_{n1}), \dots, (\mathbf{X}_{nn}, Y_{nn})\}$ be the set of independent samples, where $\mathbf{X}_{ni} = (X_{i1}, \dots, X_{ip_n})^T$. To quantify the error measures for different types of response variables, Zhang, Jiang and Chai (2008) considered a broad class of loss functions $Q(\cdot, \cdot)$, called Bregman divergence (BD). The penalized-BD estimator $(\widehat{\beta}_{n,0}, \widehat{\boldsymbol{\beta}}_n)$ is defined as the minimizer of the following criterion function,

$$(2.2) \quad \ell_{n,\lambda_n}(\beta_{n,0}, \boldsymbol{\beta}_n) = \frac{1}{n} \sum_{i=1}^n Q(Y_{ni}, F^{-1}(\beta_{n,0} + \mathbf{X}_{ni}^T \boldsymbol{\beta}_n)) + \sum_{j=1}^{p_n} P_{\lambda_n}(|\beta_{n,j}|),$$

where $\boldsymbol{\beta}_n = (\beta_{n,1}, \dots, \beta_{n,p_n})^T$ is a vector of parameters for non-constant covariates, the loss function $Q(\cdot, \cdot)$ is a BD, and $P_{\lambda_n}(\cdot)$ represents a nonnegative penalty function, imposed on the regression coefficients, indexed by a tuning parameter $\lambda_n > 0$. Set $\widetilde{\boldsymbol{\beta}}_n = (\beta_{n,0}, \boldsymbol{\beta}_n^T)^T$, and correspondingly $\widetilde{\mathbf{X}}_{ni} = (1, \mathbf{X}_{ni}^T)^T$. Then the criterion function above can be written as follows,

$$(2.3) \quad \ell_{n,\lambda_n}(\widetilde{\boldsymbol{\beta}}_n) = \frac{1}{n} \sum_{i=1}^n Q(Y_{ni}, F^{-1}(\widetilde{\mathbf{X}}_{ni}^T \widetilde{\boldsymbol{\beta}}_n)) + \sum_{j=1}^{p_n} P_{\lambda_n}(|\beta_{n,j}|).$$

The penalized-BD estimator $\widehat{\boldsymbol{\beta}}_n = (\widehat{\beta}_{n,0}, \widehat{\beta}_{n,1}, \dots, \widehat{\beta}_{n,p_n})^T$ amounts to the solution,

$$\widehat{\boldsymbol{\beta}}_n = \arg \min_{\boldsymbol{\beta}_n} \ell_{n,\lambda_n}(\widetilde{\boldsymbol{\beta}}_n).$$

In the particular case of an identity link function using $F(\mu) = \mu$, model (2.1) reduces to the linear regression model,

$$(2.4) \quad Y_n = \beta_{n,0;0} + \mathbf{X}_n^T \boldsymbol{\beta}_{n;0} + \epsilon_n,$$

where ϵ_n is the random noise satisfying $E(\epsilon_n | \mathbf{X}_n) = 0$. Moreover, using a quadratic loss, $Q(Y, \mu) = (Y - \mu)^2$ in (2.2), the penalized least-squares estimator $(\widehat{\beta}_{n,0}, \widehat{\boldsymbol{\beta}}_n)$ minimizes the criterion function,

$$\ell_{n,\lambda_n}(\beta_{n,0}, \boldsymbol{\beta}_n) = \frac{1}{n} \sum_{i=1}^n (Y_{ni} - \beta_{n,0} - \mathbf{X}_{ni}^T \boldsymbol{\beta}_n)^2 + \sum_{j=1}^{p_n} P_{\lambda_n}(|\beta_{n,j}|).$$

Accordingly, the criterion function (2.3) can be written as

$$(2.5) \quad \ell_{n,\lambda_n}(\widetilde{\boldsymbol{\beta}}_n) = \frac{1}{n} \sum_{i=1}^n (Y_{ni} - \widetilde{\mathbf{X}}_{ni}^T \widetilde{\boldsymbol{\beta}}_n)^2 + \sum_{j=1}^{p_n} P_{\lambda_n}(|\beta_{n,j}|).$$

Recently, the regularization approach has drawn a good deal of attention as a tool for statistical estimation and modeling especially in dealing with large-dimensional data. We will describe three types of penalty choice methods that are of particular relevance to the application in fMRI data.

2.1 Conventional Lasso

For linear regression model estimation with a fixed number p of parameters, Tibshirani (1996) introduced the L_1 -penalty, $P_{\lambda_n}(|\beta_{n,j}|) = \lambda_n |\beta_{n,j}|$, for the proposed Lasso (least absolute shrinkage and selection operator) method, where the quadratic loss is in use. Theoretical properties related to the Lasso have been intensively studied. Particularly, Knight and Fu (2000) showed that, under proper conditions, the Lasso is consistent for estimating the regression parameters and that its limit distribution has positive probability mass at zero if the true value of the parameter is zero. Conditions for the Lasso to be variable selection consistent were studied in Meinshausen and Bühlmann (2006). Computationally, the Lasso solution for $\widehat{\boldsymbol{\beta}}_n$ can be obtained from the LARS algorithm (Efron et al., 2004).

2.2 SCAD

Using the smoothly clipped absolute deviation (SCAD) penalty,

$$(2.6) \quad P_{\lambda}(|x|) = \begin{cases} \lambda|x|, & \text{if } |x| \leq \lambda, \\ -\frac{(|x| - a\lambda)^2 - (a^2 - 1)\lambda^2}{2(a-1)}, & \text{if } \lambda < |x| \leq a\lambda, \\ \frac{(a+1)\lambda^2}{2}, & \text{if } |x| > a\lambda, \end{cases}$$

for some constant $a > 2$, Fan and Li (2001) showed that the penalized-likelihood estimator achieved the oracle property: the resulting estimator is asymptotically as efficient as the oracle estimator, which supplies perfect information about which coefficients were non-zero. In their treatment, the number p_n of model parameters is fixed at p , and the loss function equals the negative log-likelihood. Fan and Peng (2004) further extended the oracle property from the $p_n = p$ fixed setting to p_n diverging with n at a certain lower rate.

Unlike the convex L_1 penalty in Lasso, the SCAD penalty is non-convex.

Under the class of Bregman divergence as the general loss functions, Zhang, Jiang and Chai (2008) studied the consistency rate, sparsity property and asymptotic normality of the penalized-BD estimator in the relatively low-dimensional case,

Case I: $p_n \ll n$,

where the rate of p_n is similar to that in Fan and Peng (2004). Computationally, the SCAD penalty can be approximated by the local linear approximation (LLA) method (containing derivative calculation) and the resulting estimator $\widehat{\beta}_n$ can be obtained via the LARS algorithm.

2.3 Adaptive Lasso

The adaptive Lasso (aLasso) method, introduced by Zou (2006), uses adaptive weights for the L_1 penalties to penalized least-squares estimates in fixed-dimensional linear regression models. The penalty corresponds to the weighted L_1 -penalties, $P_{\lambda_n}(|\beta_{n,j}|) = \lambda_n w_{n,j} |\beta_{n,j}|$, with non-negative weights $\{w_{n,j}\}$. Zou mentioned that the Lasso is in general not variable selection consistent, but the adaptive Lasso via combining appropriately weighted L_1 -penalties is consistent.

Under the class of Bregman divergence as the general loss functions, Zhang, Jiang and Chai (2008) investigated the consistency rate, sparsity property and asymptotic normality of the penalized-BD estimator in the medium-dimensional case,

Case II: $p_n \approx n$.

Namely, the dimension p_n is allowed to be as large as nearly comparable to n . From a theoretical perspective, the sparsity property indicates that with probability converging to one as n increases, the estimators, $\widehat{\beta}_n^{(II)}$, of the zero parameters take exactly zero values. This issue was addressed assuming $s_n^4/n \rightarrow 0$ and $s_n(p_n - s_n) = o(n)$ as $n \rightarrow \infty$, the conditions of which are indeed applicable to (1.3) and (1.4). Analogous to the SCAD method, the adaptive Lasso solution for $\widehat{\beta}_n$ can be obtained from the LARS algorithm.

As to the selection of the weights $\{w_{n,j}\}_{j=1}^{p_n}$, Jiang and Zhang (2008) developed two selection methods, called the componentwise regression (CR) method and the penalized componentwise regression (PCR) method, respectively. They showed the oracle properties of the resulting adaptive Lasso estimators in Case I–Case II. In the CR method, the weights are selected according to

$$\widehat{w}_{n,j}^{\text{CR}} = |\widehat{\beta}_{n,j}^{\text{CR}}|^{-1}, \quad j = 1, \dots, p_n,$$

based on some initial estimator, $\widehat{\beta}_n^{\text{CR}} = (\widehat{\beta}_{n,1}^{\text{CR}}, \dots, \widehat{\beta}_{n,p_n}^{\text{CR}})^T$, minimizing the criterion function,

$$\ell_n^{\text{CR}}(\beta_n) = \sum_{j=1}^{p_n} \left\{ \frac{1}{n} \sum_{i=1}^n Q(Y_{ni}, F^{-1}(X_{ij}\beta_{n,j})) \right\},$$

for the componentwise regression. In the PCR method, they selected the weights by

$$\widehat{w}_{n,j}^{\text{PCR}} = |\widehat{\beta}_{n,j}^{\text{PCR}}|^{-1}, \quad j = 1, \dots, p_n,$$

where $\widehat{\beta}_n^{\text{PCR}} = (\widehat{\beta}_{n,1}^{\text{PCR}}, \dots, \widehat{\beta}_{n,p_n}^{\text{PCR}})^T$ is chosen to be the minimizer of the criterion function,

$$\ell_n^{\text{PCR}}(\beta_n) = \sum_{j=1}^{p_n} \left\{ \frac{1}{n} \sum_{i=1}^n Q(Y_{ni}, F^{-1}(X_{ij}\beta_{n,j})) + \kappa_n |\beta_{n,j}| \right\},$$

for the penalized componentwise regression, with some sequence $\kappa_n > 0$.

Interestingly, when the loss function is a quadratic function and the link function is an identity link function, both minimizers $\widehat{\beta}_n^{\text{CR}}$ and $\widehat{\beta}_n^{\text{PCR}}$ have explicit expressions, which greatly facilitate the computation. These results will be utilized in our numerical algorithm.

3. REGULARIZED ESTIMATION OF HRF FOR FMRI DATA

In fMRI data applications, the errors $\{\epsilon(t_i)\}$ in the convolution model (1.2) are serially dependent, thus some decorrelation procedure is needed before estimating the HRF coefficients. Denote by $\text{cov}(\epsilon, \epsilon) = \sigma^2 R_n$ the error covariance matrix, with R_n the true correlation matrix. We transform model (1.2) to

$$(3.1) \quad \mathbf{y}^* = \widetilde{\mathbf{T}}^* \widetilde{\boldsymbol{\alpha}} + \mathbf{S}^* \mathbf{h}_n + \boldsymbol{\epsilon}^*,$$

where $\mathbf{y}^* = R_n^{-1/2} \mathbf{y} = (Y_{n1}^*, \dots, Y_{nn}^*)^T$, $\widetilde{\mathbf{T}}^* = R_n^{-1/2} \widetilde{\mathbf{T}}$, $\mathbf{S}^* = R_n^{-1/2} \mathbf{S}$, and $\boldsymbol{\epsilon}^* = R_n^{-1/2} \boldsymbol{\epsilon} = (\epsilon^*(t_1), \dots, \epsilon^*(t_n))^T$. The transformed errors $\{\epsilon^*(t_i)\}$ are homoscedastic and uncorrelated. Model (3.1) can be re-written as

$$(3.2) \quad \mathbf{y}^* = \widetilde{\mathbf{X}}^* \widetilde{\boldsymbol{\beta}}_n + \boldsymbol{\epsilon}^*,$$

where

$$\widetilde{\mathbf{X}}^* = \begin{bmatrix} \widetilde{\mathbf{T}}^* & \mathbf{S}^* \end{bmatrix} \quad \text{and} \quad \widetilde{\boldsymbol{\beta}}_n = \begin{bmatrix} \widetilde{\boldsymbol{\alpha}} \\ \mathbf{h}_n \end{bmatrix}$$

are an $n \times (k+1+p_n)$ matrix and a $(k+1+p_n) \times 1$ vector respectively. Then the penalized least-squares estimator $\widehat{\widetilde{\boldsymbol{\beta}}}_n$ minimizes the criterion function,

$$(3.3) \quad \ell_{n,\lambda_n}(\widetilde{\boldsymbol{\beta}}_n) = \frac{1}{n} \sum_{i=1}^n (Y_{ni}^* - \widetilde{\mathbf{x}}_{ni}^{*T} \widetilde{\boldsymbol{\beta}}_n)^2 + \sum_{j=k+1}^{k+p_n} P_{\lambda_n}(|\beta_{n,j}|),$$

where $\widetilde{\mathbf{x}}_{ni}^{*T}$ is the i th row vector of $\widetilde{\mathbf{X}}^*$. The similarity between (2.5) and (3.3) enables us to apply the regularization

methods discussed in Section 2 to obtain the penalized estimators of \mathbf{h}_n .

Remark 1. Long et al. (2004) adopted the wavelet domain L_1 spatial constraint on the activation map. The current paper develops the temporal regularization of HRF coefficients via Lasso, adaptive Lasso and SCAD penalties.

3.1 Unknown noise covariance matrix

In practice, the true error correlation matrix R_n is unknown and needs to be estimated before applying the transform in (3.1). Since the dimension of R_n is large (increasing with n), not all commonly assumed structures on the fMRI error process are computationally feasible in estimating the large error covariance matrix. Refer to Bickel and Levina (2008) for further discussions on large covariance matrix estimation. Since the temporal correlation decreases significantly as the time interval between measurements increases, we assume that $\{\epsilon(t_i)\}$ is a zero-mean stationary g -dependent sequence (Sen 1968, p. 1724) with variance equal to σ^2 , where g is a non-negative integer. Namely, the random vectors $(\epsilon(t_1), \dots, \epsilon(t_i))$ and $(\epsilon(t_j), \epsilon(t_{j+1}), \dots)$ are stochastically independent if $j - i > g$. Denote by $\rho(j)$ and $\gamma(j)$ the auto-correlation and auto-covariance of $\{\epsilon(t_i)\}$ respectively, with time lag $j \geq 0$. It is readily seen that

$$(3.4) \quad \rho(0) = 1, \text{ and } \rho(|j|) = 0 \text{ for } |j| > g,$$

$$(3.5) \quad \gamma(j) = \sigma^2 \rho(j),$$

and

$$R_n = \begin{bmatrix} 1 & \rho(1) & \cdots & \rho(n-1) \\ \rho(1) & 1 & \cdots & \rho(n-2) \\ \vdots & \vdots & \ddots & \vdots \\ \rho(n-1) & \rho(n-2) & \cdots & 1 \end{bmatrix}.$$

To balance the trade off between the bias and variance of higher-order difference-based estimators, as exemplified in Fan and Zhang (2003), and to expedite computation, we adopt the first-order difference-based method (Zhang et al., 2008) for estimating R_n . Define by $\mathcal{D}\{y\}(t_i) = y(t_i) - y(t_{i-1})$ the operator for the first-order difference of $\{y(t_i)\}$. Denote by

$$(3.6) \quad e(t_i) \equiv \mathcal{D}\left(y - \sum_{j=1}^r s_j * h_j\right)(t_i)$$

the transformed data after applying the first-order difference, where r is the number of types of stimuli. Let $\gamma_e(j)$ be the auto-covariance of $\{e(t_i)\}$, with time lag $j \geq 0$. Since $e(t_i) = \mathcal{D}d(t_i) + \epsilon(t_i) - \epsilon(t_{i-1})$, direct calculations give that

$$\gamma_e(j) = -\gamma(j-1) + 2\gamma(j) - \gamma(j+1), \quad j = 0, 1, \dots$$

This, along with (3.4)–(3.5), leads to the identities,

$$(3.7) \quad \begin{bmatrix} \gamma_e(0) \\ \gamma_e(1) \\ \gamma_e(2) \\ \vdots \\ \gamma_e(g-1) \\ \gamma_e(g) \end{bmatrix} = \begin{bmatrix} 2 & -2 & 0 & 0 & 0 & \cdots & 0 & 0 & 0 & 0 \\ -1 & 2 & -1 & 0 & 0 & \cdots & 0 & 0 & 0 & 0 \\ 0 & -1 & 2 & -1 & 0 & \cdots & 0 & 0 & 0 & 0 \\ \vdots & \vdots \\ 0 & 0 & 0 & 0 & 0 & \cdots & 0 & -1 & 2 & -1 \\ 0 & 0 & 0 & 0 & 0 & \cdots & 0 & 0 & -1 & 2 \end{bmatrix} \times \begin{bmatrix} \gamma(0) \\ \gamma(1) \\ \gamma(2) \\ \vdots \\ \gamma(g-1) \\ \gamma(g) \end{bmatrix}.$$

Substituting $\{\gamma_e(j)\}$ with their empirical moment estimates and solving equation (3.7), we can obtain the estimates $\{\hat{\gamma}(j)\}$ of $\{\gamma(j)\}$. Utilizing (3.4)–(3.5), the noise variance and auto-correlation can be estimated by $\hat{\sigma}^2 = \hat{\gamma}(0)$ and $\hat{\rho}(j) = \hat{\gamma}(j)/\hat{\gamma}(0)$ respectively. It is then immediate to acquire an estimate \hat{R}_n of R_n . In our numerical implementation, $g = 2$ is adopted for computational stability and efficiency. For different problems, other choices of g can also be used.

In practice, we need a preliminary estimate of the HRF in (3.6). Again, this initial estimate $\hat{\mathbf{h}}_{\text{DBE}}$ can be obtained via a difference-based method: *apply the first-order difference to \mathbf{y} and \mathbf{S} simultaneously, ignore the drift term in (1.2), followed by an ordinary least-squares estimation of the HRF.* To illustrate the idea, let us take the difference of (1.2). Under the smoothness assumption on the drift function $d(\cdot)$, it follows that $d(t_i) - d(t_{i-1}) = d'(t_{i-1})n^{-1} + O(n^{-2}) = O(n^{-1})$. Hence the difference of (1.2) yields

$$y(t_i) - y(t_{i-1}) = (\mathbf{e}_i - \mathbf{e}_{i-1})^T \mathbf{S} \mathbf{h}_n + \epsilon(t_i) - \epsilon(t_{i-1}) + O(n^{-1}), \quad i = 2, \dots, n,$$

where \mathbf{e}_j denotes the j th column of an identity matrix. As the sequence length n grows, the magnitude of the difference (and higher-order difference) of $d(\cdot)$ will be negligible. This example lends support to the root- n consistency of $\hat{\mathbf{h}}_{\text{DBE}}$ under mild conditions.

Remark 2. Carew et al. (2003) characterized the error autocorrelation by assuming that $\epsilon \sim N(\mathbf{0}, \sigma^2 K K^T)$ for an unknown matrix K . Instead of estimating the error covariance matrix, they examined the smoothed fMRI by multiplying both sides of (1.2) by a particular choice of smoothing matrix, A , of the spline smoothing; they used AA^T to approximate the covariance matrix, $AKK^T A^T$, of the smoothed

errors $A\epsilon$, and to ease the inference procedure, but the smoothed errors continued to be correlated.

3.2 Computation

We would like to make some remarks on the computational aspects of the penalization procedure. For an entire brain data, the penalized estimation of the HRF (via minimizing (3.3)) will be applied voxel-wise, involving large-scale computational power. On the other hand, for each fixed voxel, minimizing (3.3) via modifying the LARS algorithm is indeed computationally efficient. As discussed in Efron et al. (2004), the computational cost of the LARS algorithm is the same as that of the conventional least-squares estimation. This makes the penalization method computationally feasible and efficient for real brain data.

4. SIMULATION STUDY

To illustrate the application of penalized least-squares estimators of the HRF in fMRI data, we conduct the simulation study. For illustrative purpose, 6 numerical procedures for obtaining these estimators are compared: (I) the ordinary least-squares (OLS) and weighted least-squares (WLS) estimators; (II) the SCAD penalty combined with the local linear approximation (LLA) and the LARS algorithm; (III) the L_1 penalty combined with the LARS algorithm; (IV) the adaptive Lasso combined with the CR weight selection method and the LARS algorithm; (V) the adaptive Lasso combined with the PCR weight selection method and the LARS algorithm; (VI) the Oracle estimator using the set of significant variables (whose coefficients are non-zero). Throughout the numerical work in the paper, the SCAD penalty in (2.6) uses an accompanying parameter $a = 3.7$. Refer to Fan and Li (2001) for a detailed discussion on this choice of a .

4.1 Example 1: Regression model for i.i.d. Gaussian responses

We generate i.i.d. Gaussian responses from the linear regression model (2.4), with

$$\mathbf{x}_n \sim N(\mathbf{0}, (\sqrt{2})^2 \mathbf{I}_{p_n}), \text{ and } \epsilon_n \sim N(0, 1),$$

where \mathbf{I}_{p_n} denotes a $p_n \times p_n$ identity matrix, and \mathbf{x}_n and ϵ_n are independent. The regression coefficients are $\beta_{n,0;0} = 2.5$, and $\beta_{n;0} = [0, 0, 0, 0, .0014, .2788, .8526, .8698, .4031, -.0310, -.2017, -.1842, -.1121, -.0541, -.0221, -.0080, -.0026, -.0008, 0, \dots, 0]^T$. Note that the number s_n of non-zero entries of $\beta_{n;0}$ equals 14 and the number of zero entries of $\beta_{n;0}$ is p_n minus s_n . The penalization problem (2.5) is considered. For illustrative purpose, we focus on $n = 200$, with the dimensions p_n varying from $n/8 = 25$, $n/2 = 100$ to $n - 1 = 199$ representing small-, medium- and large-dimensions, respectively. By varying p_n , sequences of $\beta_{n;0}$

possessing different amount of sparsity can also be generated. The top left panels of Figures 2, 3 and 4 (corresponding to p_n equal to $n/8 = 25$, $n/2 = 100$ and $n - 1 = 199$ respectively) plot the magnitude of the components of $\beta_{n;0}$, which mimics the shape of an HRF in Figure 1.

First, to examine the effect of penalized regression estimates on model fitting, for each set of simulated data $\{(\mathbf{x}_{ni}, Y_{ni})\}_{i=1}^n$, the model error (ME) is calculated by the mean squared error of the regression function estimate, $\sum_{i=1}^n \{\widehat{m}(\mathbf{x}_{ni}) - m(\mathbf{x}_{ni})\}^2/n$, and the relative model error (RME) is the ratio of ME using penalized estimators and ME using the non-penalized OLS estimator. The tuning parameters λ_n in each simulation for methods (II)–(V) are selected separately by 3-fold cross validation; the tuning parameters λ_n and κ_n for method (V) are searched on a surface of grid points and selected by minimizing the 3-fold cross-validated residual sum of squares. The MRME is the median of RMEs obtained from those 100 sets of simulation. Table 1 summarizes the penalized estimates of parameters, in small-, medium- and large-dimensional cases. The gain achieved by the penalized estimators in reducing the function estimation error over the OLS estimator is clearly seen. In the low-dimensional case, i.e. where $\beta_{n;0}$ is dense, the reduction of function estimation errors by the penalized estimators is close to that of the Oracle estimator.

Second, to study the utility of penalized estimators in revealing the effects in variable selection, Table 1 lists a column labeled “#CZ” as the average number of coefficients which are correctly estimated to be zero when the true coefficients are zero, and a column labeled “#IZ” as the average number of coefficients which are incorrectly estimated to be zero when the true coefficients are nonzero. The standard deviations of the corresponding estimations across 100 simulations are given in brackets. For instance, the Oracle method invariably produces estimates with “#CZ” equal to p_n minus s_n and “#IZ” equal to 0. For a better visual assessment, Figures 2–4 each display a path of the coefficient estimates for $\beta_{n;0}$ from a typical sample for each method. The typical sample is selected in such a way that the number of coefficients which are estimated to be zero is equal to the 50th percentile of the numbers in the 100 simulations. In the presence of sparsity, the penalized estimators help yield a sparse solution and build a parsimonious model. It is clearly seen that as the dimension p_n grows, the performance of the OLS estimator dramatically deteriorates, without capturing the sparsity structure.

4.2 Example 2: Convolution model for fMRI with known R_n

Typically, the analysis of the entire brain fMRI data is conducted by a two-step procedure: voxelwise statistical analysis in the first step, followed by multiple comparison in the second step. Since this paper aims at developing statistical estimation of the HRF in the first step, illustration of single-voxel analysis is focused on.

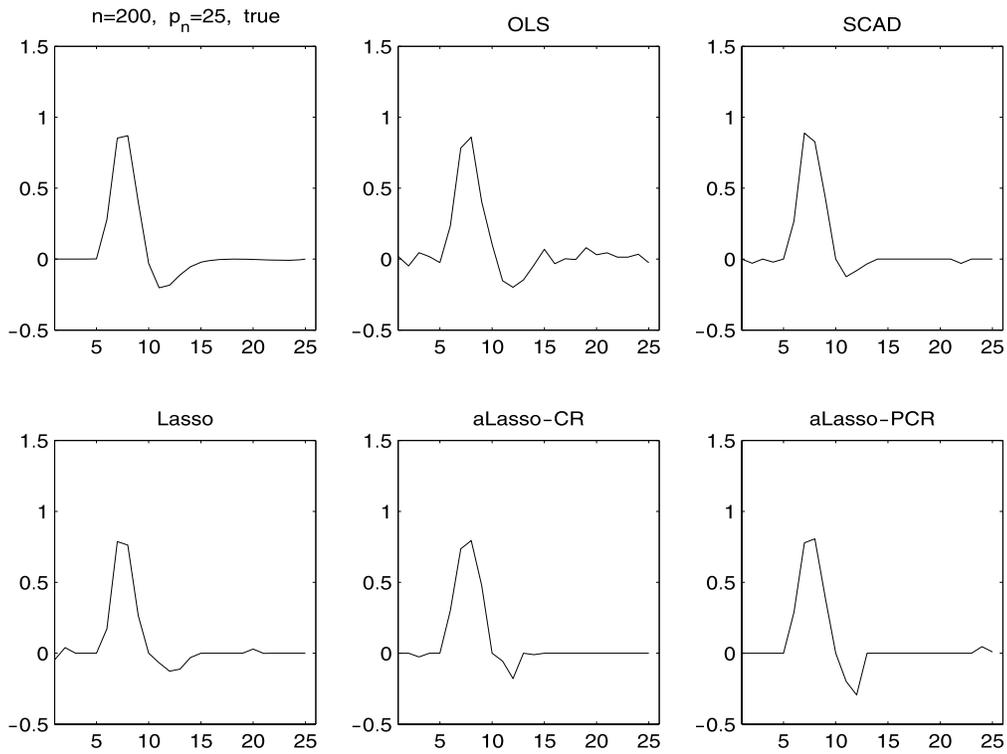


Figure 2. Penalized estimates of $\beta_{n,0}$ for i.i.d. Gaussian responses, where $p_n = n/8$.

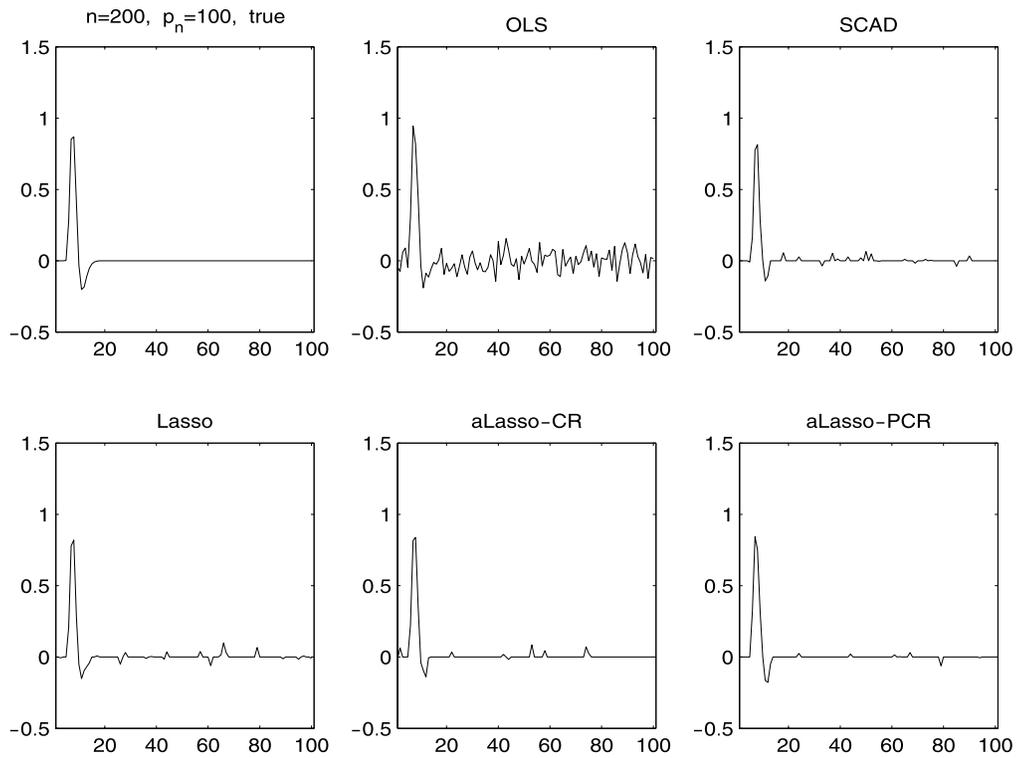


Figure 3. Penalized estimates of $\beta_{n,0}$ for i.i.d. Gaussian responses, where $p_n = n/2$.

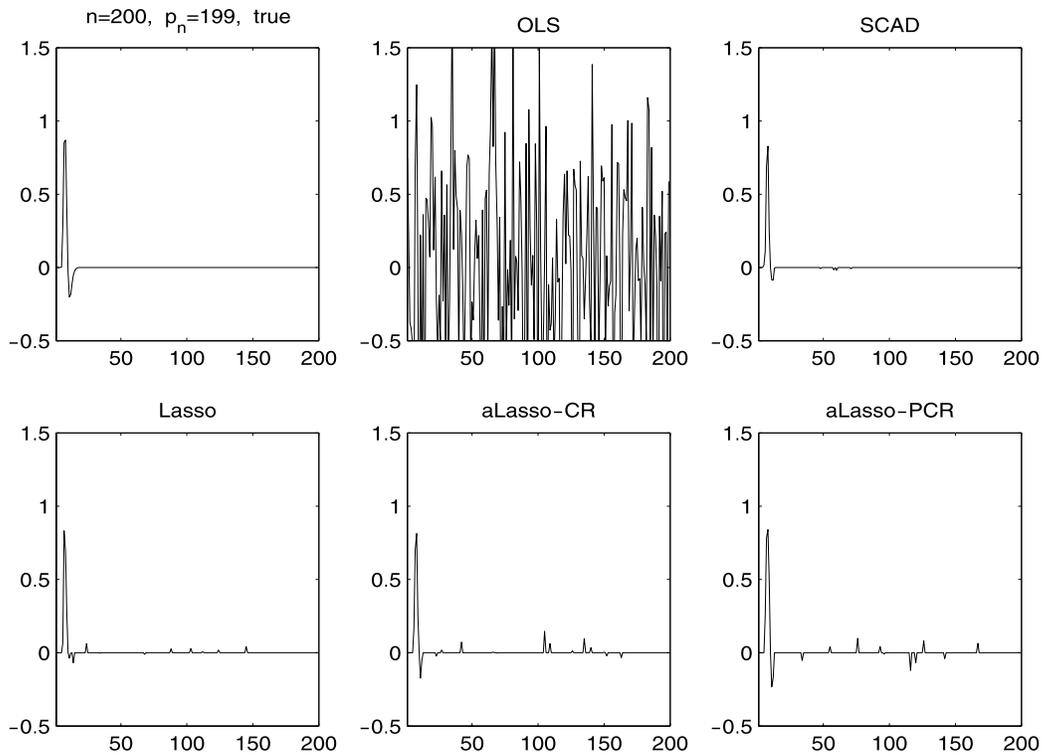


Figure 4. Penalized estimates of $\beta_{n;0}$ for i.i.d. Gaussian responses, where $p_n = n - 1$.

Table 1. Penalized estimates of $\beta_{n;0}$ for i.i.d. Gaussian responses

n	p_n	Method	Regression	Variable Selection	
			MRME	#CZ (std)	#IZ (std)
200	$n/8 = 25$	OLS	1.00	0.00 (0.00)	0.00 (0.00)
		SCAD; LLA-LARS	0.62	7.89 (2.26)	4.55 (1.66)
		L_1 ; LARS	0.74	6.54 (2.64)	3.81 (1.81)
		aLasso-CR	0.68	8.60 (2.10)	5.52 (1.93)
		aLasso-PCR	0.73	8.74 (2.09)	5.94 (1.95)
		Oracle	0.58	11.00 (0.00)	0.00 (0.00)
		$n/2 = 100$		OLS	1.00
SCAD; LLA-LARS	0.25			74.14 (8.66)	6.09 (1.61)
L_1 ; LARS	0.27			69.41 (8.43)	5.67 (1.53)
aLasso-CR	0.25			76.98 (5.64)	6.81 (1.29)
aLasso-PCR	0.27			78.00 (5.55)	7.27 (1.35)
Oracle	0.13			86.00 (0.00)	0.00 (0.00)
$n - 1 = 199$		OLS	1.00	0.00 (0.00)	0.00 (0.00)
		SCAD; LLA-LARS	0.16	173.03 (13.09)	7.11 (1.15)
		L_1 ; LARS	0.20	167.36 (15.11)	6.72 (1.42)
		aLasso-CR	0.16	171.66 (10.18)	7.27 (1.32)
		aLasso-PCR	0.17	172.60 (8.87)	7.62 (1.43)
		Oracle	0.06	185.00 (0.00)	0.00 (0.00)

Table 2. Penalized estimates of \mathbf{h}_n for the simulated fMRI with true R_n

n	p_n	Method	Variable Selection	
			#CZ (std)	#IZ (std)
200	$n/8 = 25$	WLS	0.00 (0.00)	0.00 (0.00)
		SCAD; LLA-LARS	10.65 (2.61)	3.98 (1.19)
		L_1 ; LARS	7.21 (3.35)	2.37 (1.50)
		aLasso-CR	8.38 (3.01)	3.50 (1.70)
		aLasso-PCR	10.60 (2.48)	5.41 (1.73)
		Oracle	14.00 (0.00)	0.00 (0.00)
$n/2 = 100$		WLS	0.00 (0.00)	0.00 (0.00)
		SCAD; LLA-LARS	76.06 (4.26)	4.60 (0.99)
		L_1 ; LARS	74.10 (6.78)	4.12 (1.08)
		aLasso-CR	77.61 (9.12)	6.80 (1.32)
		aLasso-PCR	79.94 (7.23)	7.43 (1.24)
		Oracle	89.00 (0.00)	0.00 (0.00)
$n - 5 - k = 193$		WLS	0.00 (0.00)	0.00 (0.00)
		SCAD; LLA-LARS	163.45 (7.05)	4.75 (1.08)
		L_1 ; LARS	160.53 (9.49)	4.33 (1.36)
		aLasso-CR	168.44 (10.44)	7.07 (1.23)
		aLasso-PCR	169.47 (8.80)	7.59 (1.12)
		Oracle	182.00 (0.00)	0.00 (0.00)

We simulate an fMRI experiment with a single run and a single type of stimulus. In the simulation, $n = 200$, $t_i = i/n$, $i = 1, \dots, n$, and 100 realizations are conducted. **(I)** The time-varying stimuli are generated from independent Bernoulli trials such that $P\{s(t_i) = 1\} = .5$. **(II)** Following Glover (1999), the HRF is $h(t_j) = g_1(1.5(j-1) - 5.5)/a_1 - 0.4 \times g_2(1.5(j-1) - 5.5)/a_2$ for $j = 1, \dots, [9(n^{1/5.5} - 1)] + 1$, and $h(t_j) = 0$ for $j = [9(n^{1/5.5} - 1)] + 2, \dots, p_n$, where $g_1(t) = t^5 \exp(-t/.9)I(t \geq 0)$ and $g_2(t) = t^{12} \exp(-t/.7)I(t \geq 0)$, $a_1 = \max\{g_1(t)\}$ and $a_2 = \max\{g_2(t)\}$. **(III)** The drift function is $d(t_i) = \alpha_0 + \alpha_1 t_i + \alpha_2 t_i^2$, $i = 1, \dots, n$, where $(\alpha_0, \alpha_1, \alpha_2) = (-7.8737, 47.5836, -32.6734)/5$. **(IV)** The noise process ϵ is the sum of independent noise processes ϵ_1 and ϵ_2 (see Purdon et al., 2001); $\{\epsilon_1(t_i)\}$ are i.i.d. normal with mean zero and variance $.5216^2$, $.3689^2$, $.2608^2$ and $.1844^2$ respectively; ϵ_2 is AR(1), i.e., $\epsilon_2(t_i) = \rho \epsilon_2(t_{i-1}) + z(t_i)$ with $\rho = .638$ and $z(t_i)$ follows the normal distribution with mean zero and variance $.5216^2$, $.3689^2$, $.2608^2$ and $.1844^2$ respectively. These choices give the noise lag-one auto-correlation equal to .4 and the signal-to-noise-ratio (SNR) about 1, 2, 4 and 8, where $\text{SNR} = \text{variance}(\mathbf{S}\mathbf{h}_n)/\text{variance}(\epsilon)$.

The estimation problem (3.1) is studied. For lack of space, we only present the results with $\text{SNR} = 4$; see Table 2 and Figures 5–7 for the estimates of the HRF coefficients, where the number s_n of non-zero HRF coefficients equals 11, and the order $k = 2$ for the drift is used in the estimation throughout the computation. Three things can be observed. First, similar to the results of Example 1 with i.i.d. Gaussian responses, the WLS estimator fails to produce zero estimates when the true HRF coefficients are equal or close to zero. In terms of the sparsity property, the SCAD, aLasso-CR and aLasso-PCR continue to outperform the Lasso. Second, for both the WLS and penalized methods, the larger the dimen-

sion p_n is, the less efficient the estimates are. Third, for Figures 6–7 with medium- and large-dimensions p_n , the HRF estimates obtained by the penalized methods miss some dip point of the HRF, partly because the penalization methods tend to shrink small estimates to zero. Nonetheless, for the counterparts in Figures 3–4 with i.i.d. data, all penalized estimation methods maintain the dip point. In this respect, it will be useful to devise penalization methods that are better able to account for the time series dependence.

4.3 Example 3: Convolution model for fMRI with unknown R_n

In this study, the simulation set-up is identical to that of the previous Example 2, except that the true correlation matrix R_n in (3.1) is unknown and replaced by the estimate \widehat{R}_n . We would like to remark that since entries of the matrix \mathbf{S} are either 0 or 1, \mathbf{S} itself forms a sparse matrix. Henceforth, we set p_n in the large-dimensional case to be $n - 5 - k$ to ensure that the difference-based estimator of R_n is computationally feasible. For the sake of comparison, $n - 5 - k$ is also used in Example 2.

The results analogous to Table 2 and Figures 5–7 are given in Table 3 and Figures 8–10. They lend support to the validity of the covariance matrix estimation method proposed in Section 3.1.

5. APPLICATION TO A REAL FMRI DATA

5.1 Experiment design, data description and analysis aim

In an emotional control study, subjects saw a series of negative or positive emotional images, and were asked to either suppress or enhance their emotional responses to the

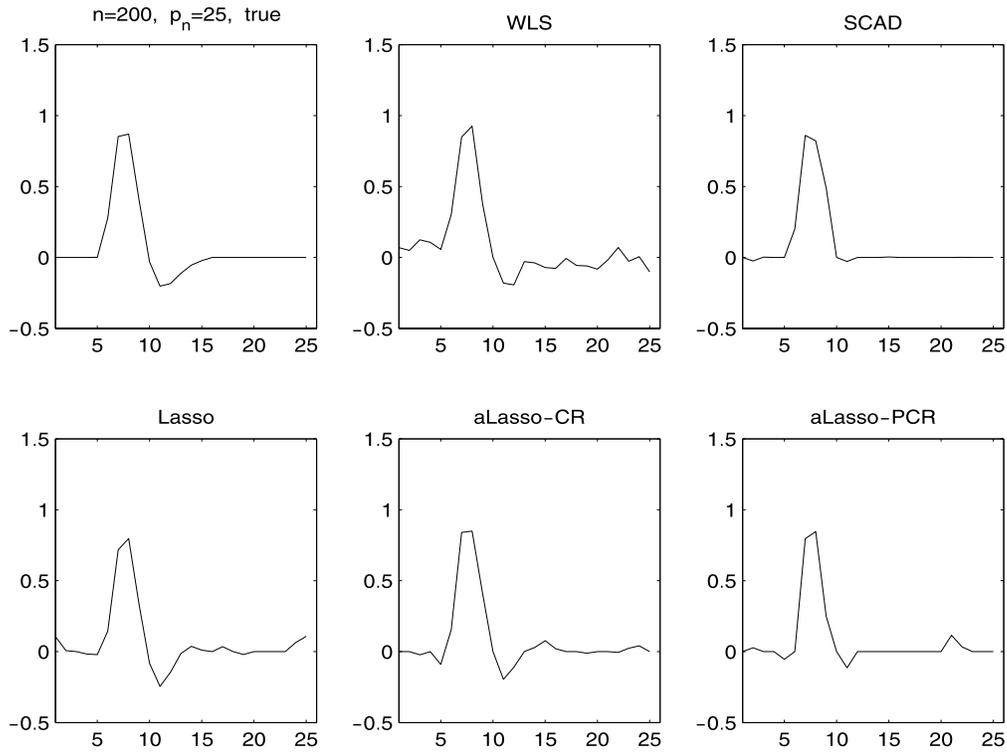


Figure 5. Penalized estimates of \mathbf{h}_n for the simulated fMRI with the true R_n , where $p_n = n/8$.

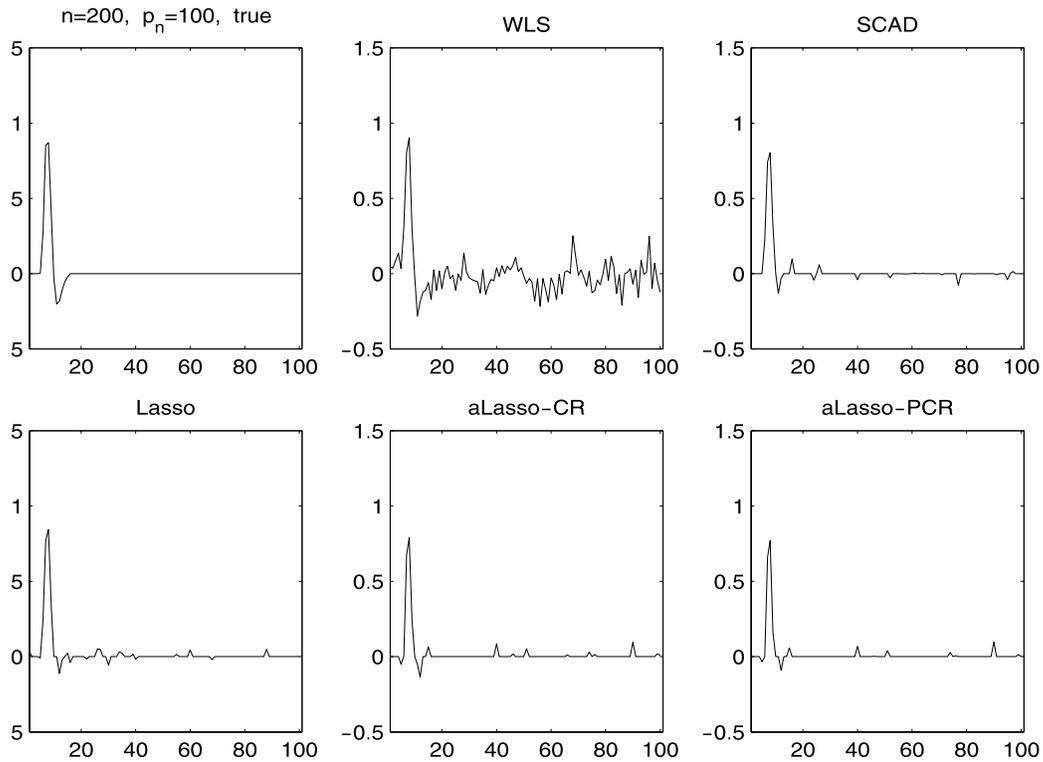


Figure 6. Penalized estimates of \mathbf{h}_n for the simulated fMRI with the true R_n , where $p_n = n/2$.

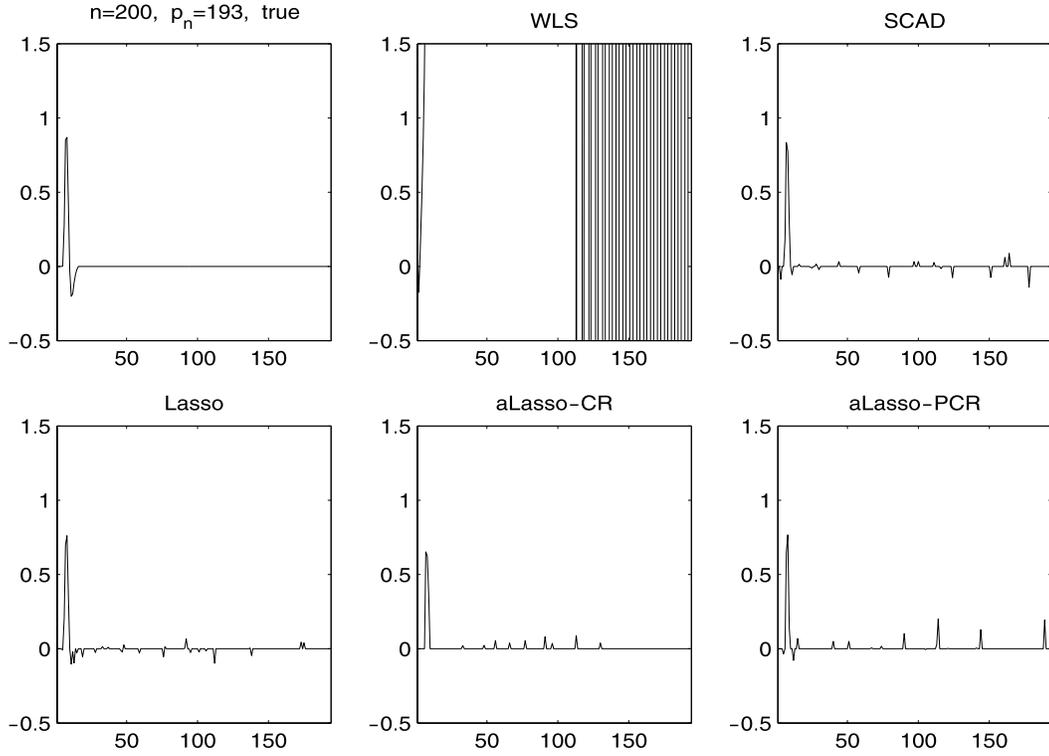


Figure 7. Penalized estimates of \mathbf{h}_n for the simulated fMRI with the true R_n , where $p_n = n - 5 - k$.

Table 3. Penalized estimates of \mathbf{h}_n for the simulated fMRI with the estimated R_n

n	p_n	Method	Variable Selection	
			#CZ (std)	#IZ (std)
200	$n/8 = 25$	WLS	0.00 (0.00)	0.00 (0.00)
		SCAD; LLA-LARS	10.48 (2.99)	3.50 (1.45)
		L_1 ; LARS	7.51 (3.43)	2.43 (1.54)
		aLasso-CR	7.48 (3.52)	3.92 (2.14)
		aLasso-PCR	9.73 (3.13)	5.79 (2.04)
		Oracle	14.00 (0.00)	0.00 (0.00)
		$n/2 = 100$		WLS
SCAD; LLA-LARS	77.10 (4.87)			4.48 (1.29)
L_1 ; LARS	74.64 (8.52)			4.28 (1.40)
aLasso-CR	75.55 (10.11)			7.43 (1.05)
aLasso-PCR	78.16 (9.56)			7.84 (0.97)
Oracle	89.00 (0.00)			0.00 (0.00)
$n - 5 - k = 193$		WLS	0.00 (0.00)	0.00 (0.00)
		SCAD; LLA-LARS	160.96 (8.40)	4.31 (1.54)
		L_1 ; LARS	153.97 (18.40)	4.12 (1.50)
		aLasso-CR	160.54 (15.85)	6.97 (1.47)
		aLasso-PCR	162.96 (15.26)	7.38 (1.34)
		Oracle	182.00 (0.00)	0.00 (0.00)

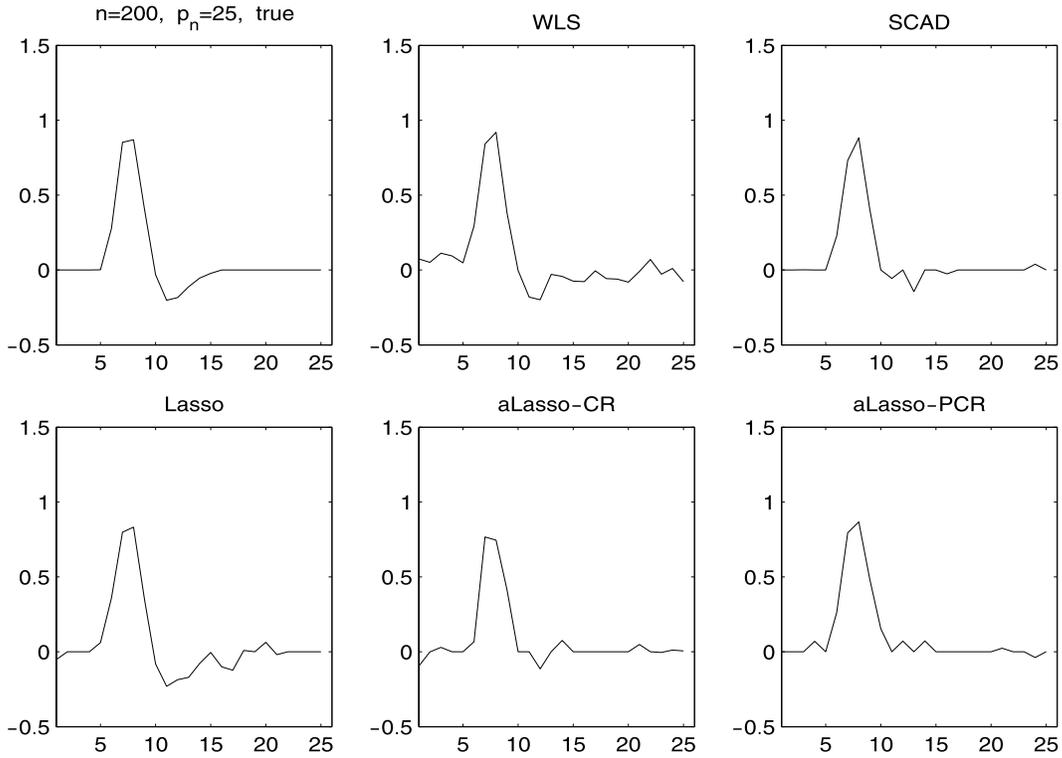


Figure 8. Penalized estimates of \mathbf{h}_n for the simulated fMRI with the estimated R_n , where $p_n = n/8$.

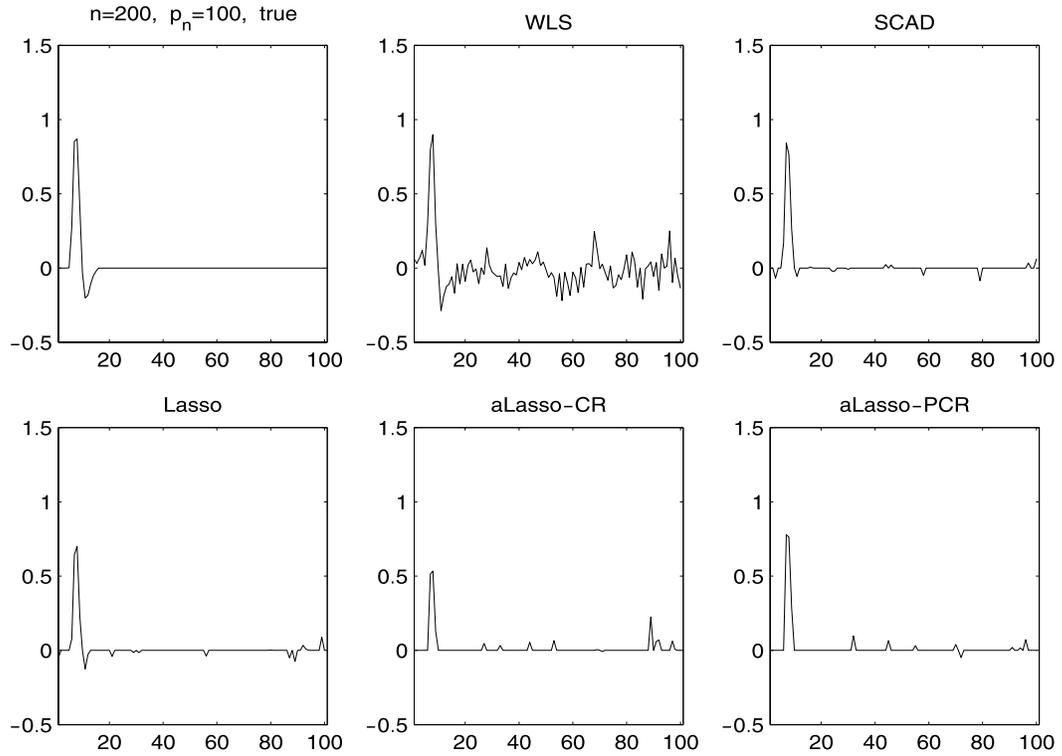


Figure 9. Penalized estimates of \mathbf{h}_n for the simulated fMRI with the estimated R_n , where $p_n = n/2$.

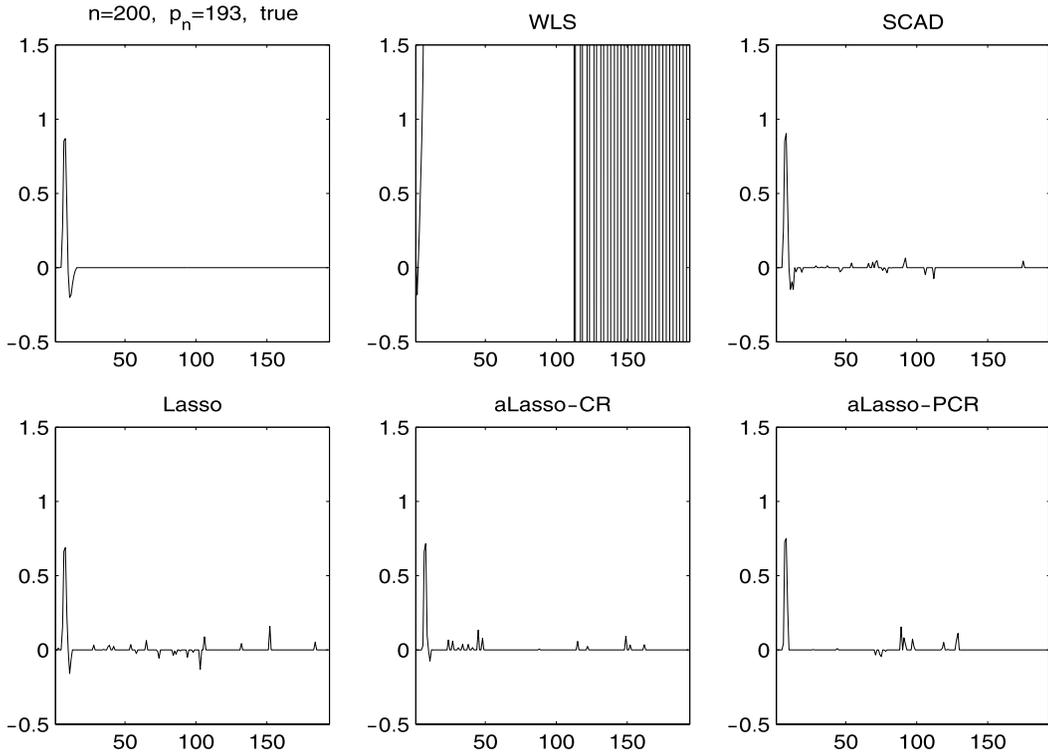


Figure 10. Penalized estimates of \mathbf{h}_n for the simulated fMRI with the estimated R_n , where $p_n = n - 5 - k$.

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

The size of the whole brain dataset is $64 \times 64 \times 30$. At each voxel, the time series has 6 runs, each containing 185 observations with a time resolution of 2 secs, thus $TR = 2$ secs and the total length is 1,110. In contrast, the length of stimuli is 2,220; the timing of the stimuli has a time resolution of 1 second, and thus each HRF output will also be sampled at 1 second. Hence, the odd rows of the design matrix \mathbf{S} in (1.2) suffice for analysis.

The study aims to estimate the BOLD (Blood Oxygenation Level-Dependent) response to each of the trial types following the image onset. We analyze the fMRI dataset containing one subject. The length of the estimated HRF is set equal to $\lceil n/10 \rceil = 18$. This choice mimics the low-dimensional cases in the simulation study of Section 4.

Recall from Section 3.2, the penalized estimates of the HRF coefficients can be obtained in each brain voxel, so that the voxel-wise comparison of the penalized estimates and the AFNI estimates can be made. For lack of space, we

present results conducted at two randomly selected voxels which are marked by AFNI as activated. Specifically, no procedure that corrects for multiple comparison is applied in AFNI analysis here. Similar to the specification in AFNI, the polynomial degree $k = 2$ for the drift part is used in the penalized estimation throughout the computation.

Remark 3. Although the paper compares results using AFNI, other popular tools in neuroimage study, such as SPM (Friston et al., 1997) and FSL (at <http://www.fmrib.ox.ac.uk/fsl/>) (Smith et al., 2004 and Woolrich et al., 2001), have been implicitly considered as well, since the AFNI method is being offered as an option under the name FIR.

5.2 Single-voxel fMRI analysis: Voxel 1

Figure 11 plots the fMRI time series of six runs from a brain voxel (24, 32, 7) that shows activation as given by AFNI analysis. The estimates of BOLD response to each of the 6 stimulus types, i.e. six sets (using different colors) of HRF estimates, as calculated by the AFNI program 3dDeconvolve and the regularized estimation methods, are compared in Figure 12. Note that since the drift and HRF are common to all runs, the coefficients of the drift and HRF are estimated based on data from the combined runs.

In this voxel, the detected activation effect by the regularized methods appears to be quite different from that of AFNI. The regularized estimates of the HRF, associated

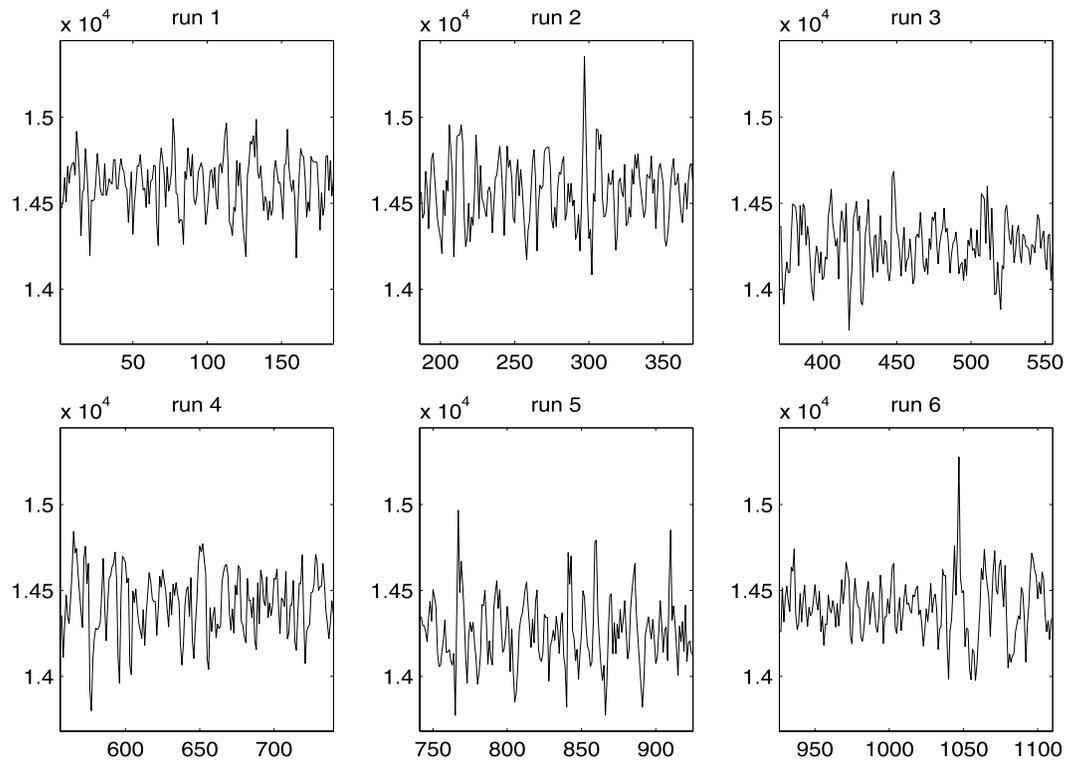


Figure 11. Raw fMRI time series from an activated brain voxel (24, 32, 7).

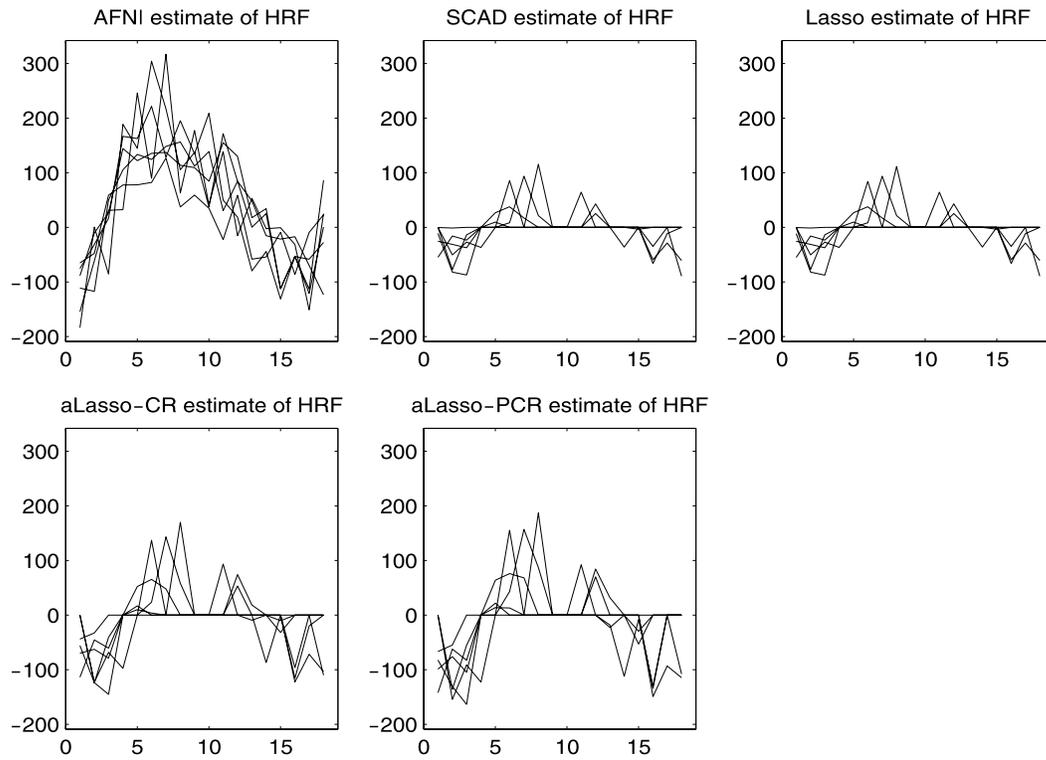


Figure 12. Six sets of HRF estimates (for six types of stimuli) at an activated brain voxel (24, 32, 7).

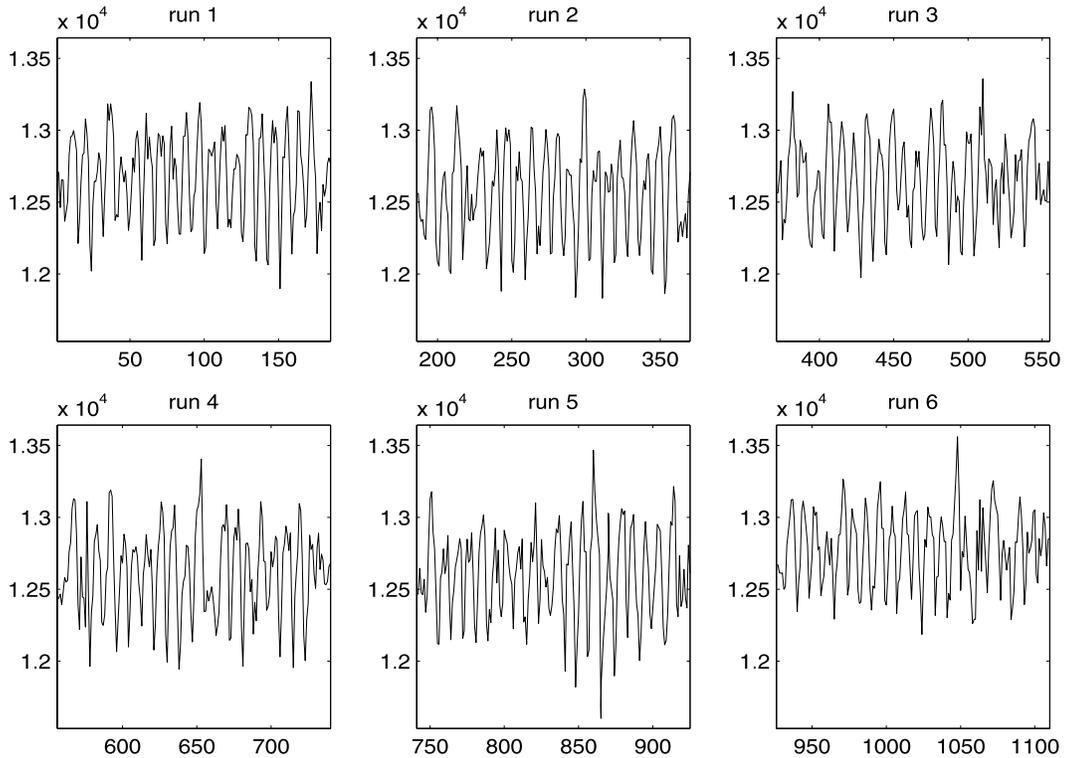


Figure 13. Raw fMRI time series from an activated brain voxel (49, 41, 10).

Table 4. The estimated number s_n of non-zero HRF coefficients for the real fMRI data

Voxel	Method	Stimulus Type					
		1	2	3	4	5	6
(24, 32, 7)	AFNI	18	18	18	18	18	18
	SCAD; LLA-LARS	5	8	9	6	2	7
	L_1 ; LARS	5	8	9	6	2	7
	aLasso-CR	6	8	10	9	3	7
	aLasso-PCR	6	8	11	8	3	7
(49, 41, 10)	AFNI	18	18	18	18	18	18
	SCAD; LLA-LARS	17	16	16	17	17	16
	L_1 ; LARS	17	18	17	16	17	17
	aLasso-CR	18	15	17	17	16	17
	aLasso-PCR	18	15	17	17	16	17

with certain stimulus types, exhibit some sparsity pattern at certain time points. See Table 4 for the estimated number s_n of non-zero HRF coefficients. Recall that the AFNI results are based on the OLS method with the errors specified to be i.i.d. Recall also that the OLS estimator invariably produces non-zero estimates at all time points, even if the sparsity of the HRF does exist at certain time points. In that case, some of the non-zero estimates obtained by OLS are spurious. Based on our previous simulation studies, the seemingly strong magnitude of the estimated HRF by AFNI calls for re-examination.

5.3 Single-voxel fMRI analysis: Voxel 2

Recall in model (1.2), the matrix \mathbf{S} for stimuli is common to all voxels, but the HRF \mathbf{h}_n varies from one voxel to another voxel. To examine the relative performance of the penalized estimates of the HRF at other voxels, we consider a brain voxel (49, 41, 10) that shows activation as given by AFNI analysis. Figure 13 plots the fMRI times series from 6 runs. The estimated HRFs using AFNI and penalized methods are graphed in Figure 14 and are tabulated in Table 4. Unlike the previous voxel, the profiles of the HRF estimated by the penalized methods compare favorably well with those calculated by AFNI. This is possibly due to the absence (or less presence) of sparsity of the HRF at this particular voxel. This indicates that when a model well reflects the realistic data and the parameter \mathbf{h}_n is not sparse, the penalized methods do not falsely produce zero estimates, thus are more adaptive to either sparse or non-sparse models.

6. DISCUSSION

In considering the choice between the regularized and non-regularized estimation methods for the HRF coefficients, both approaches have merit, depending on the context or goal of the experiment. For example, in cases where the prior information on the true dimension p_n is available and there is dense structure among the HRF coefficients, the least-squares method enjoys computational simplicity and

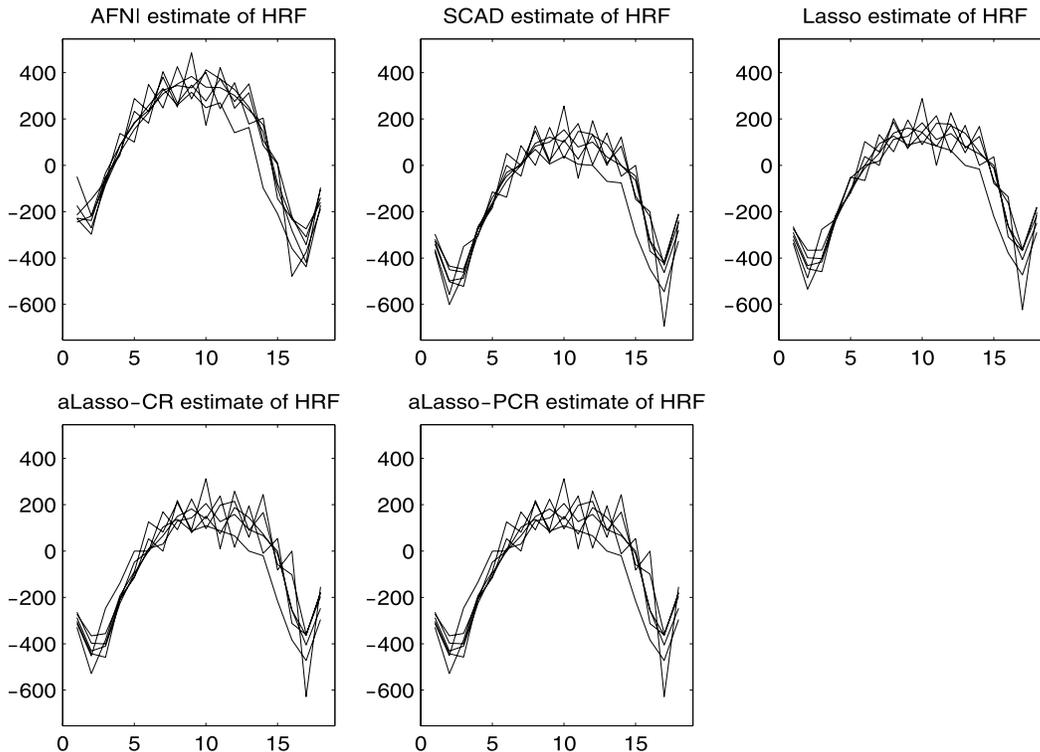


Figure 14. Six sets of HRF estimates (for six types of stimuli) at an activated brain voxel (49, 41, 10).

good practical performance. In this case, the oracle property of the penalized estimation methods guarantees that they perform as good as the least-squares method, though some difference may arise in finite-sample situations.

This paper focuses on the problem of estimating an unknown parameter HRF having a sparse representation. We introduce a varying-dimensional model for the HRF and present penalization schemes into the estimation of the HRF. The simulation studies demonstrate the improvement of the adaptive Lasso and SCAD compared with the Lasso, which performs better than the least-squares estimation method in terms of model fitting and sparsity recovery. The numerical results also indicate that, for large-dimensional HRF, the adaptive Lasso methods utilizing both CR and PCR for selecting weights are computationally more stable than the SCAD. For realistic applications, we recommend the adaptive Lasso and SCAD in the estimation of small- and medium-dimensional HRF; for large-dimensional HRF, the adaptive Lasso approach is preferable.

The penalized estimation methods could be improved in many aspects. For example, as the design matrix for fMRI data is sparse, more refined weight selection procedures for aLasso methods are desirable to enhance their performances in revealing the dip points. Regarding sources for sparsity, temporal sparsity in this paper seems to be more natural for the voxelwise estimation approach. Exploiting other physiological factors to be included in model (1.2) and taking into account the sparsity structure from both the temporal

and spatial domains will be an interesting future work. Areas for future research include (I) more efficient methods for estimating the large error covariance matrix, (II) more rigorous investigation of the sampling properties of penalized estimators under the convolution model, (III) confidence intervals for \mathbf{h}_n in spirit similar to that of Sara et al. (2004) and hypothesis testing of $\mathbf{h}_n = \mathbf{0}$ for detecting activation via the penalized estimators, and (IV) related multiple comparison procedures and activation maps. It's worthy mentioning that Zhang, Jiang and Chai (2008) developed statistical inference tools for testing the significance of large-dimensional parameters via penalized estimators, when the data are independent and identically distributed. More refined work is needed for generalizing those inference methods to the fMRI time series, which are serially correlated. Again, for issue (III) relevant to a more in-depth study of the brain activity, issues (I), (II) and (IV) are indispensable and very likely need to be well developed ahead of (III).

ACKNOWLEDGMENT

The research is supported in part by National Science Foundation grants DMS-03-53941, DMS-07-05209, DMS-08-04575, and the Wisconsin Alumni Research Foundation. The authors are grateful to the Editor, Associate Editor and two anonymous referees for insightful comments and suggestions.

Received 29 July 2009

REFERENCES

- BICKEL, P.J. AND LEVINA, E. (2008). Regularized estimation of large covariance matrices. *Ann. Statist.*, **36**, 199–227.
- BICKEL, P. AND LI, B. (2006). Regularization in statistics (with discussion). *Test*, **15**, 271–344.
- CAREW, J.D., WAHBA, G., XIE, X., NORDHEIM, E.V. AND MEYERAND, M.E. (2003). Optimal spline smoothing of fMRI time series by generalized cross-validation. *NeuroImage*, **18**, 950–961.
- CHEN, Y.H., BRESSLER, S.L., KNUTH, K.H., TRUCCOLO, W.A., AND DING, M.Z. (2006). Stochastic modeling of neurobiological time series: Power, coherence, Granger causality, and separation of evoked responses from ongoing activity. *Chaos*, **16**, article no. 026113.
- CORDES, D. AND NANDY, R.R. (2006). Estimation of the intrinsic dimensionality of fMRI data. *NeuroImage*, **29**, 145–154.
- COX, R.W. (1996). AFNI: software for analysis and visualization of functional magnetic resonance neuroimages. *Comput. Biomed. Res.*, **29**, 162–173.
- EFRON, B., HASTIE, T., JOHNSTONE, I., AND TIBSHIRANI, R. (2004). Least angle regression. *Ann. Statist.*, **32**, 407–499.
- FAN, J. AND LI, R. (2001). Variable selection via nonconcave penalized likelihood and its oracle properties. *J. Amer. Statist. Assoc.*, **96**, 1348–1360.
- FAN, J. AND PENG, H. (2004). Nonconcave penalized likelihood with a diverging number of parameters. *Ann. Statist.*, **32**, 928–961.
- FAN, J. AND ZHANG, C.M. (2003). A reexamination of diffusion estimators with applications to financial model validation. *J. Amer. Statist. Assoc.*, **98**, 118–134.
- FRISTON et al. (1997). SPM course notes. <http://www.fil.ion.ucl.ac.uk/spm/>
- GLOVER, G.H. (1999). Deconvolution of impulse response in event-related BOLD fMRI. *NeuroImage*, **9**, 416–429.
- JIANG, Y. AND ZHANG, C.M. (2008). The adaptive lasso for high-dimensional models under general loss. *Technical report #1149*, Department of Statistics, University of Wisconsin-Madison.
- JOHNSTONE, I.M. AND SILVERMAN, B.W. (2004). Needles and straw in haystacks: empirical Bayes estimates of possibly sparse sequences. *Ann. Statist.*, **32**, 1594–1649.
- KNIGHT, K. AND FU, W. J. (2000). Asymptotics for Lasso-type estimators. *Ann. Statist.*, **28**, 1356–1378.
- LONG, C., BROWN, E. N., MANOACH, D., AND SOLO, V. (2004). Spatiotemporal wavelet analysis for functional MRI. *NeuroImage*, **23**, 500–516.
- MEINSHAUSEN, N. AND BUHLMANN, P. (2006). High dimensional graphs and variable selection with the lasso. *Ann. Statist.*, **34**, 1436–1462.
- PURDON, P.L., SOLO, V., WEISSKO, R.M. AND BROWN, E. (2001). Locally regularized spatiotemporal modeling and model comparison for functional MRI. *NeuroImage*, **14**, 912–923.
- SAHA, S., LONG, C.J., BROWN, E.N., AMINOFF, E., BAR, M., AND SOLO, V. (2004). Hemodynamic transfer function estimation with Laguerre polynomials and confidence interval construction, from fMRI data. *Proc IEEE ICASSP*, vol III:109–112. Montreal, Canada.
- SEN, P.K. (1968). Asymptotic normality of sample quantiles for m -dependent processes. *The Annals of Mathematical Statistics*, **39**, 1724–1730.
- SMITH, S., JENKINSON, M., WOOLRICH, M., BECKMANN, C.F., BEHRENS, T.E.J., JOHANSEN-BERG, H., BANNISTER, P.R., DE LUCA, M., DROBNJAK, I. FLITNEY, D.E., NIAZY, R.K., SAUNDERS, J., VICKERS, J., ZHANG, Y., DE STEFANO, N., BRADY, J.M. AND MATTHEWS, P.M. (2004). Advances in functional and structural MR image analysis and implementation as FSL. *NeuroImage*, **23**, 208–219.
- TIBSHIRANI, R. (1996). Regression shrinkage and selection via the lasso. *J. Roy. Statist. Soc. Ser. B*, **58**, 267–288.
- WARD, B.D. (2001). Deconvolution analysis of fMRI time series data. *Technical Report*, Biophysics Research Institute, Medical College of Wisconsin.
- WOOLRICH, M.W., RIPLEY, B.D., BRADY, M., AND SMITH, S.M. (2001). Temporal autocorrelation in univariate linear modelling of fMRI data. *NeuroImage*, **14**, 1370–1386.
- WORSLEY, K.J., LIAO, C.H., ASTON, J., PETRE, V., DUNCAN, G.H., MORALES, F., AND EVANS, A.C. (2002). A general statistical analysis for fMRI data. *NeuroImage*, **15**, 1–15.
- ZHANG, C.M., LU, Y., JOHNSTONE, T., OAKES, T., AND DAVIDSON, R.J. (2008). Efficient modeling and inference for event-related fMRI data. *Computational Statistics and Data Analysis*, **52**, 4859–4871.
- ZHANG, C.M. AND YU, T. (2008). Semiparametric detection of significant activation for brain fMRI. *Ann. Statist.*, **36**, 1693–1725.
- ZHANG, C.M., JIANG, Y. AND CHAI, Y. (2008). Penalized Bregman divergence for large dimensional regression and classification. *Biometrika*, tentatively accepted. (*Technical report #1153*, Department of Statistics, University of Wisconsin-Madison.)
- ZOU, H. (2006). The adaptive lasso and its oracle properties. *J. Amer. Statist. Assoc.*, **101**, 1418–1429.

Chunming Zhang
Department of Statistics
University of Wisconsin-Madison
E-mail address: cmzhang@stat.wisc.edu

Zhengjun Zhang
Department of Statistics
University of Wisconsin-Madison
E-mail address: zjz@stat.wisc.edu