Nonparametric Inference of the Hemodynamic Response using Multi-Subject fMRI Data

Tingting Zhang[a,1], Fan Li[b], Lane Beckes[a], Casey Brown[a], James A. Coan[a]

[a] *University of Virginia*
[b] *Duke University*

Abstract

Estimation and inferences for the hemodynamic response functions (HRF) using multi-subject fMRI data are considered. Within the context of the General Linear Model, two new nonparametric estimators for the HRF are proposed. The first is a kernel-smoothed estimator, which is used to construct hypothesis tests on the entire HRF curve, in contrast to only summaries of the curve as in most existing tests. To cope with the inherent large data variance, we introduce a second approach which imposes Tikhonov regularization on the kernel-smoothed estimator. An additional bias-correction step, which uses multi-subject averaged information, is introduced to further improve efficiency and reduce the bias in estimation for individual HRFs. By utilizing the common properties of brain activity shared across subjects, this is the main improvement over the standard methods where each subject's data is usually analyzed independently. A fast algorithm is also developed to select the optimal regularization and smoothing parameters. The proposed methods are compared with several existing regularization methods through simulations. The methods are illustrated by an application to the fMRI data collected under a psychology design employing the Monetary Incentive Delay (MID) task.

Key words: bias correction, general linear model, hemodynamic response function, kernel smoothing, regularization.

[1]Corresponding author at: Halsey Hall 111, University of Virginia, Charlottesville, VA, 22904, USA. Fax: (434) 924-3076. E-mail address: tz3b@virginia.edu (Tingting Zhang)

1. Introduction

There is a vast literature in functional magnetic resonance imaging (fMRI) data analysis on estimating the hemodynamic response function (HRF) within the framework of the General Linear Model (GLM) (Friston et al., 1995a; Friston et al., 1995b; Worsley and Friston, 1995). These methods differ in their assumptions about the shape of the HRFs. Standard parametric approaches assume a functional form for the HRF with a number of free parameters, such as the canonical form of mixtures of gamma functions (Friston et al., 1998; Glover, 1999; Worsley et al., 2002), poisson function (Friston et al., 1994), inverse logit function (Lindquist and Wager, 2007), and radial basis functions (Riera et al., 2004). Except for the model using the canonical form and its derivatives, estimation for parametric models with even a moderate number of parameters often relies on computationally-intensive iterative methods (such as the Gauss-Newton method), which can lead to unstable estimates when the algorithms do not converge (Liao, et al., 2002). This paper alternatively focuses on nonparametric approaches, which are flexible and usually fast to compute. Bai et al. (2009) and Wang et al. (2011) constructed nonparametric estimates of the HRF in the frequency domain. Nonparametric methods in the time domain mainly fall into two types: representing the HRF with a linear combination of functional bases (Aguirre et al., 1998; Woolrich et al., 2004; Zarahn, 2002; Vakorin et al., 2007), or treating the HRF at every unit time point as a free parameter (Dale, 1999; Lange et al., 1999). In this paper we adopt the latter approach in the time domain to develop nonparametric estimation and inferences for HRFs.

Since nonparametric methods for HRF estimation involve many free parameters and the HRF is generally believed to be smooth (Buxton et al., 2004), smoothing techniques are often employed. Kernel smoothing is a popular nonparametric statistical method for increasing temporal continuity of functional estimates (Eubank, 1988; Härdle, 1990). It has been used for temporal smoothing of fMRI time series (e.g., Friston et al., 1994; Worsley and Friston, 1995), but has rarely been used for HRF estimation. In this paper, we first introduce a kernel-smoothed HRF estimator, based on which we construct hypothe-

sis tests on the entire HRF curve, in contrast to the common practice of testing only some characteristics of the HRF.

Regularization is another increasingly popular technique used in nonparametric estimation that allows smoothness constraints to be imposed on the HRF estimates. One example is the smooth finite impulse response method (SFIR, Glover, 1999; Goutte et al., 2000; Ollinger et al., 2001), which exploits a regularization term to obtain smooth estimates that satisfy a boundary condition. Another example is given in Marrelec et al. (2001, 2003), where the HRF is represented by orthogonal functional bases and a smoothness constraint is imposed through regularizing the norm of its second order derivative. Similarly, representing the HRF by spline bases, Vakorin et al. (2007) and Zhang et al. (2007) used Tikhonov regularization (Tikhonov and Arsenin, 1977). The estimator proposed by Casanova et al. (2008, 2009) combines Tikhonov regularization and generalized cross validation (Wahba, 1990) (referred to Tik-GCV hereafter), greatly reducing the computational burden involved in parameter selection. Motivated by these developments, a second goal of this paper is to propose a new nonparametric estimator that combines kernel smoothing with Tikhonov regularization. Distinct from previous methods, this approach controls the degree of temporal smoothness and the norm of the estimates by two separate parameters. This separation makes the estimator more adaptive to different combinations of HRF temporal resolution and signal-to-noise ratio (SNR).

In analyzing multi-subject fMRI data, many existing methods, both parametric and nonparametric, estimate each subject's HRF independently to account for its variability across subjects (Aguirre et al., 1998; Handwerker et al., 2004). When data from each individual has a low SNR, utilizing the common characteristics of the HRFs shared across the population may improve the estimation efficiency. Moreover, for such data, though a strong scale of regularization is effective in stabilizing estimates, it also introduces additional biases. Thus, bias correction can be considered to improve over the regularized estimates (e.g., Zhang et al., 2007). Assuming that, under the same stimulus and in the same brain regions, the HRFs have similar functional shapes across subjects (Friston et al.,

1998; Handwerker et al., 2004), we propose to use sample-averaged HRF estimates to conduct bias correction for the regularization-based estimates. A fast algorithm is developed to select regularization and smoothing parameters and to evaluate the new estimators. Through simulations, the proposed bias-corrected estimator demonstrates significant improvement over the estimators without the bias-correction step.

The article is organized as follows. In Section 2, we briefly review the GLM framework and propose the nonparametric kernel-smoothed estimator for hypothesis testing on the whole curve of the HRF. We then refine the estimator by adding Tikhonov regularization and applying bias correction. Two fast algorithms for parameter selection are also developed. Section 3 presents results from applying the proposed methods to both simulated data and real fMRI data, and comparisons are drawn to several existing methods. Section 4 concludes with a discussion.

2. Materials and Methods

2.1. The GLM

We conduct massive univariate analysis of fMRI data in the context of the GLM. Since the same approach applies to each voxel, the subscript for voxel is omitted here. Let $y_i(t)$ for $t = 1, \cdots, T$ and $i = 1, \cdots, N$ be the fMRI time series for a pre-specified voxel of subject i, where T is the total observation time and N is the number of subjects. Suppose the design has K stimuli. Let $v_{i,k}(t)$ be the kth ($k = 1, \cdots, K$) stimulus function for subject i with $v_{i,k}(t) = 1$ if the stimulus is evoked at time t and 0 otherwise. The GLM represents the observed fMRI time series as a convolution of the HRF and the stimuli: $y_i(t) = \sum_{k=1}^{K} \int_0^m h_{i,k}(u) v_{i,k}(t-u) du + \varepsilon_i(t)$, where $h_{i,k}$ is the HRF of the pre-specified voxel in subject i under stimulus k, m is a known positive constant beyond which the HRF equals zero, and $\varepsilon_i(t)$ is an identically-distributed error term. The blood oxygen level dependent (BOLD) fMRI signal often contains a low-frequency drift due to physiological noise or subject motion (Smith et al., 1999; Brosch et al., 2002; Luo and Puthusserypady, 2008); this can be modeled by adding a polynomial term of time t (Mattay et al., 1996;

Worsley et al., 2002; Lindquist, 2008) to the above GLM as

$$y_i(t) = d_{0,i} + d_{1,i} \cdot t + d_{2,i} \cdot t^2 + \sum_{k=1}^{K} \int_0^m h_{i,k}(u) v_{i,k}(t-u) du + \varepsilon_i(t), \tag{1}$$

where the drift parameters $d_{0,i}, d_{1,i}$, and $d_{2,i}$ are allowed to vary across subjects.

2.2. Kernel-Smoothed Nonparametric Estimator

We treat each HRF at every unit time as a free parameter. Let Δ be the time unit representing the discretization of the HRF temporal resolution. Since it is possible to have the temporal resolution of the HRF shorter than that of the fMRI data (Ciuciu et al., 2003; Casanova et al., 2008), Δ can be smaller than the repetition time unit (TR) of the experimental design. For each subject i, let $\boldsymbol{Y}_i = (y_i(1), ..., y_i(T))'$ be the observed fMRI time series. Denote the discretized values of the HRF under stimulus k by $\boldsymbol{\beta}_{i,k} = (\beta_{i,k}(1), \cdots, \beta_{i,k}(m))'$, where $\beta_{i,k}(t) = \int_{(t-1)\cdot\Delta}^{t\cdot\Delta} h_{i,k}(u)du$ in a block design or $\beta_{i,k}(t) = h_{i,k}(t \cdot \Delta)$ in an event-related design (Josephs et al., 1997). Let $\boldsymbol{\beta}_i = (\boldsymbol{\beta}'_{i,1}, \cdots, \boldsymbol{\beta}'_{i,K})'$. Denoting all the coefficients $(d_{0,i}, d_{1,i}, d_{2,i}, \boldsymbol{\beta}'_i)'$ by $\boldsymbol{\eta}_i$, the GLM (1) can be written in a matrix form as

$$\boldsymbol{Y}_i = \mathbf{X}_i \boldsymbol{\eta}_i + \boldsymbol{\varepsilon}_i, \tag{2}$$

where $\mathbf{X}_i$ is the design matrix corresponding to the time covariates and the stimulus functions for subject i, and $\boldsymbol{\varepsilon}_i = (\varepsilon_1, \cdots, \varepsilon_T)' \sim N(0, \sigma_i^2 \Sigma_i)$ with unknown variance σ_i^2 and correlation matrix Σ_i. Since $h_{i,k}(t)$ is random across subjects, the coefficients $\boldsymbol{\beta}_i$ are also random. As a result, model (2) is a linear random-effect model. For each subject, we can remove the drift term through ordinary least square (OLS) regression and obtain an unbiased OLS estimate of $\boldsymbol{\beta}_i$, denoted by $\hat{\boldsymbol{\beta}}_i = (\hat{\beta}_{i,1}(1), \cdots, \hat{\beta}_{i,1}(m), \hat{\beta}_{i,2}(1), \cdots, \hat{\beta}_{i,K}(m))'$. As noted in Goutte et al. (2000), $\hat{\boldsymbol{\beta}}_i$ usually has an artificial high-frequency noise due to the large number of parameters under estimation and experimental designs with interleaved stimuli and inter-stimulus intervals. This can be clearly seen from the simulation example in Figure 2(a). Therefore, smoothing techniques are often employed to reduce

the unnatural ruggedness of the estimates.

Previous approaches have typically applied temporal smoothing directly to $y_i(t)$ to increase the statistical power for detecting responsive regions (e.g., Friston et al., 1994; Friston et al., 1995b; Worsley and Friston, 1995). The HRFs are generally believed to be smooth (e.g., Buxton et al., 2004); while smoothing the HRF estimated from the fMRI time series guarantees the smoothness of the resulting curve, directly smoothing the original fMRI times series does not, especially in complex designs with multiple stimuli. Since our interest lies in estimating the HRF and the degree of smoothness may vary across HRFs under different stimuli, we choose to conduct kernel smoothing on the OLS estimates $\hat{\boldsymbol{\beta}}_i$. Specifically, we propose to use the Nadaraya-Watson kernel estimator:

$$\tilde{\beta}_{i,k}(t) = \sum_{u=t-l}^{t+l} W_{t,u} \cdot \hat{\beta}_{i,k}(u), \text{ with } W_{t,u} = \frac{f(\frac{t-u}{h})/h}{\sum_{u=t-l}^{t+l} f(\frac{t-u}{h})/h}. \tag{3}$$

Here h is a pre-specified bandwidth controlling the degree of smoothing, $f(t)$ is a given symmetric density function (kernel), and l is a pre-specified constant giving an upper bound on the number of data points used for the estimation. In this article, we let $f(t)$ be a standard Gaussian density and $l = m$. Existing results suggest that the choice of these two values has only a small effect on the estimation (Eubank, 1988; Härdle, 1990). The choice of the key bandwidth parameter h is elaborated in Section 2.5. The underlying idea of kernel smoothing is to borrow information from the neighboring data: the estimate $\tilde{\beta}_{i,k}(t)$ is a weighted average of the neighboring OLS estimates and the weight $W_{t,u}$ is negatively correlated with the distance $|u - t|$. The boundary condition of $\beta_{i,k}(t) = 0$ for $t < 0$ and $t > m$ is imposed by setting $\hat{\beta}_{i,k}(u) = 0$ for $u < 1$ and $u > m$ in the estimator (3). Letting $\hat{\boldsymbol{\beta}}_{i,k} = (\hat{\beta}_{i,k}(1), \cdots, \hat{\beta}_{i,k}(m))'$ and $\tilde{\boldsymbol{\beta}}_{i,k} = (\tilde{\beta}_{i,k}(1), \cdots, \tilde{\beta}_{i,k}(m))'$, the first estimator we propose is a kernel-smoothed estimator that can be expressed as a matrix transformation of the OLS estimators as follows:

$$\tilde{\boldsymbol{\beta}}_{i,k} = \mathbf{B}_h \hat{\boldsymbol{\beta}}_{i,k} \text{ and } \tilde{\boldsymbol{\beta}}_i = \mathbf{A}_h \hat{\boldsymbol{\beta}}_i, \tag{4}$$

where $\mathbf{B}_h$ is an $m \times m$ matrix such that $\mathbf{B}_h(t, u_1) = W_{t,u_1}$ for $u_1, t = 1, \cdots, m$, $\mathbf{A}_h = \mathbf{I}_K \bigotimes \mathbf{B}_h$ with $\bigotimes$ denoting the Kronecker product, and $\mathbf{I}_K$ is a $K \times K$ identity matrix.

2.3. Tikhonov-Regularized Smoothed Estimator with Bias-Correction

The above kernel smoothing procedure increases the temporal continuity of HRF estimates. However, with the large number of free parameters to be evaluated, large variation in the magnitude of $\tilde{\beta}_{i,k}(t)$ across time may still persist. Tikhonov regularization (Tikhonov and Arsenin, 1977) is a common statistical technique used to address ill-posed inverse problems, and is effective in reducing the variation of regression estimates. A Tikhonov-regularized estimate of the regression coefficient $\boldsymbol{\eta}_i$ in model (2) is obtained through solving the following optimization problem

$$\min_{\boldsymbol{\eta}_i} \parallel \boldsymbol{Y}_i - \mathbf{X}_i \boldsymbol{\eta}_i \parallel^2 + \parallel \Gamma \boldsymbol{\eta}_i \parallel^2, \tag{5}$$

for some suitably-chosen matrix Γ, where $|| \cdot ||$ is the L^2 norm. The minimizer $\hat{\boldsymbol{\eta}}_i^T$ of (5) is given by $(\mathbf{X}_i' \mathbf{X}_i + \Gamma' \Gamma)^{-1} \mathbf{X}_i' \boldsymbol{Y}_i$.

Different choices of Γ defines different Tikhonov-regularized estimators. One choice of Γ in the fMRI literature is the discrete second derivative matrix, as adopted in Marrelec et al. (2003), and Casanova et al. (2008, 2009). Another choice of Γ is the scalar matrix $\alpha \mathbf{I}_{\text{dim}(\boldsymbol{\eta}_i)}$, where $\mathbf{I}_{\text{dim}(\boldsymbol{\eta}_i)}$ is an identity matrix with the dimension of $\boldsymbol{\eta}_i$. Solution from this Γ is equivalent to that from a ridge regression, a special case of Tikhonov regularization. Our choice of Γ is slightly modified from the above. Specifically, the dimension of $\boldsymbol{\eta}_i$ in our application is $3 + K \times m$, but regularization is imposed only on the estimates of the subvector $\boldsymbol{\beta}_i$ of $\boldsymbol{\eta}_i$, excluding the drift parameters ds. We let $\Gamma = \sqrt{\lambda}\text{D}$, where D is a $(3 + K \times m)$-by-$(3 + K \times m)$ diagonal matrix whose first 3 diagonal entries equal zero and the rest equal 1, and λ is a given positive constant controlling the degree of regularization. Letting $\mathbf{R}_\lambda^i$ be the lower $(K \times m)$-by-$(K \times m)$ square sub-matrix of $(\mathbf{X}_i' \mathbf{X}_i + \lambda \text{D})^{-1} (\mathbf{X}_i' \mathbf{X}_i)$, the Tikhonov-regularized estimator of $\boldsymbol{\beta}_i$ corresponding to this Γ is $\mathbf{R}_\lambda^i \hat{\boldsymbol{\beta}}_i$ (detailed derivation and explanation is given in Appendix). Since this regularization does not impose

any smoothness constraint, $\mathbf{R}_\lambda^i \hat{\boldsymbol{\beta}}_i$ can still be rough. Therefore, we once again use kernel smoothing as in (4) to increase its temporal continuity and define a new estimator: $\tilde{\boldsymbol{\beta}}_i^r = \mathbf{A}_h \mathbf{R}_\lambda^i \hat{\boldsymbol{\beta}}_i$, which we call the Tik-Kern estimator. Comparing to the existing regularization methods that directly apply Tikhonov regularization, such as SFIR and Tik-GCV, the Tik-Kern estimator $\tilde{\boldsymbol{\beta}}_i^r$ separately impose the constraints for regularization and for smoothing.

Tikhonov regularization and kernel smoothing together greatly reduce the variances in estimating HRFs, but they can also lead to large biases without further adjustment. This motivates us to propose an additional bias-correction step to $\tilde{\boldsymbol{\beta}}_i^r$. It is easy to show that the bias of $\tilde{\boldsymbol{\beta}}_i^r$ equals $[\mathbf{A}_h \mathbf{R}_\lambda^i - \mathbf{I}_{K \cdot m}]\, \boldsymbol{\beta}_i$, depending on the underlying true $\boldsymbol{\beta}_i$. A close approximation to $\boldsymbol{\beta}_i$ for each i is usually unavailable in practice. Intuitively, if the HRFs of sampled subjects for a given stimulus and a fixed brain region have similar functional shapes, each $\boldsymbol{\beta}_{i,k}$ should be reasonably close to $\sum_{i=1}^N \boldsymbol{\beta}_{i,k}/N$ (Liao et al., 2002; Henson et al., 2002). As such, one can use a sample-averaged estimate to approximate $\boldsymbol{\beta}_i$ in the presence of multiple subjects. Specifically, building on top of $\tilde{\boldsymbol{\beta}}^r$, we propose the following bias-corrected estimator $\tilde{\boldsymbol{\beta}}^{cor}$ (hereafter referred to as the BTik-Kern estimator):

$$\tilde{\boldsymbol{\beta}}_i^{cor} \;=\; \tilde{\boldsymbol{\beta}}_i^r - \left(\mathbf{A}_h \mathbf{R}_\lambda^i - \mathbf{I}_{K \cdot m}\right) \tilde{\boldsymbol{\beta}}^0, \tag{6}$$

where $\tilde{\boldsymbol{\beta}}^0 = \mathbf{A}_{h_0} \sum_{i=1}^N \hat{\boldsymbol{\beta}}_i / N$, and the initial smoothing bandwidth h_0 is $1/\sqrt{7/\mathrm{TR}}$ (Goutte et al., 2000). Since h_0 is relatively small, $\tilde{\boldsymbol{\beta}}^0$ usually has small variance and bias. In our analysis, estimation of the HRFs will all base on the BTik-Kern estimator instead of the Tik-Kern estimator. Upon obtaining the estimates $\tilde{\boldsymbol{\beta}}^{cor}$, we extract the summary statistics of the HRF, such as time to peak (TTP), width (W), and height (HR), using procedures as described in Lindquist and Wager (2007) for further analysis. The various estimators introduced above is summarized in Table 1.

Estimator	Description
$\hat{\boldsymbol{\beta}}$	The OLS estimates
$\tilde{\boldsymbol{\beta}}$	The kernel-smoothed OLS estimator, used for hypothesis testing
$\tilde{\boldsymbol{\beta}}^{r}$ (Tik-Kern)	The Tikhonov-regularized estimator based on $\tilde{\boldsymbol{\beta}}$
$\tilde{\boldsymbol{\beta}}^{cor}$ (BTik-Kern)	The bias-corrected Tik-Kern estimator, used for estimation

Table 1: Summaries of the estimators in Sections 2.2 and 2.3.

2.4. Kernel Estimate-Based Hypothesis Testing and Confidence Interval

We now use the proposed estimators to perform statistical inferences for population-wide brain activity. First, we identify brain voxels that are responsive to a specific stimulus. Next, we identify brain voxels that function differently in response to different stimuli. We formulate these two goals as two corresponding hypothesis tests for each voxel: (1) $H_0 : \mathbb{E}(\boldsymbol{\beta}_{i,k}) = \boldsymbol{\mu}_k$ for some given stimulus k, where the expectation applies to all subjects, and $\boldsymbol{\mu}_k$ is a pre-specified vector of constants which is usually a zero vector; and (2) $H_0 : \mathbb{E}(\boldsymbol{\beta}_{i,k}) = \mathbb{E}(\boldsymbol{\beta}_{i,k'})$ for $k \neq k'$.

Distinct from the existing tests on a single aspect of the HRF (e.g., latency or magnitude), hypotheses (1) and (2) aim to detect any deviation from the null hypothesized HRF on the whole time domain, as they involve the entire vector of the HRF. We choose to construct test statistics based on the kernel-smoothed estimator $\tilde{\boldsymbol{\beta}}$ instead of the BTik-Kern estimator $\tilde{\boldsymbol{\beta}}^{cor}$. This is because (1) $\mathbb{E}(\tilde{\boldsymbol{\beta}}_{i,k}) = 0$ as long as $\mathbb{E}(\boldsymbol{\beta}_{i,k}) = 0$, and (2) $\mathbb{E}(\tilde{\boldsymbol{\beta}}_{i,k}) = \mathbb{E}(\tilde{\boldsymbol{\beta}}_{i,k'})$ as long as $\mathbb{E}(\boldsymbol{\beta}_{i,k}) = \mathbb{E}(\boldsymbol{\beta}_{i,k'})$ for any fixed bandwidth h. However, this is not true for $\tilde{\boldsymbol{\beta}}^{cor}$, because regularization renders bias dependent of all the HRFs in the model under non-orthogonal designs (see Section 3). Thus, tests constructed using regularization-based estimators, including the SFIR and Tik-GCV estimators, will have an erroneous type I error as shown in the simulations later. Conceptually, the test statistics can be also constructed directly based on the OLS estimates $\hat{\boldsymbol{\beta}}$. However, since the kernel-smoothed estimator $\tilde{\boldsymbol{\beta}}$ has a much smaller variance, test statistics based on it will have a higher statistical power.

Due to the inhomogeneous variances across subjects, for the first hypothesis, we use

$\mathbf{z}_{i,k} = (\tilde{\boldsymbol{\beta}}_{i,k} - \mathbf{B}_h\boldsymbol{\mu}_k)/\hat{\sigma}_i$, where $\hat{\sigma}_i$ is the OLS estimate of σ_i, to construct the Hotelling's T-square test statistic:

$$U_k = N\bar{\mathbf{z}}_k'\Lambda_k^{-1}\bar{\mathbf{z}}_k, \tag{7}$$

where $\bar{\mathbf{z}}_k$ and Λ_k are the sample average and the sample variance-covariance matrix of $\mathbf{z}_{i,k}$, respectively. Assuming the fMRI data for a fixed voxel across different subjects are independent and identically distributed (i.i.d.), under the null hypothesis, $(N-m)U_k/(m(N-1))$ follows an F distribution with degrees of freedom $(m, N-m)$ (Hotelling, 1931). The i.i.d. assumption may not exactly hold in fMRI data; nevertheless, with large enough sample size N, the distribution of $(N-m)U_k/(m(N-1))$ under the null is still expected to closely approximate the F distribution. Then, given the significance level α, the null is rejected if $(N-m)U_k/(m(N-1))$ is larger than $100(1-\alpha)\%$ percentile of $F(m, N-m)$. Similarly for the second hypothesis, let $\mathbf{Z}_{k,k'}^i = (\tilde{\boldsymbol{\beta}}_{i,k} - \tilde{\boldsymbol{\beta}}_{i,k'})/\hat{\sigma}_i$, for $i = 1, \cdots, N$, and we use the following test statistic

$$Q_{k,k'} = N\bar{\mathbf{Z}}_{k,k'}'\Omega_{k,k'}^{-1}\bar{\mathbf{Z}}_{k,k'}, \tag{8}$$

where $\bar{\mathbf{Z}}_{k,k'}$ and $\Omega_{k,k'}$ are the sample average and variance-covariance matrix of $\mathbf{Z}_{k,k'}^i$. The null hypothesis is rejected at α level if $(N-m)Q_{k,k'}/(m(N-1))$ is larger than $100(1-\alpha)\%$ percentile of $F(m, N-m)$.

Confidence intervals can also be constructed for the kernel smoothed HRF estimates. From equation (4), it is easy to show that $\mathbb{V}(\tilde{\boldsymbol{\beta}}_i) = \sigma_i^2\mathbf{A}_h\Psi_i\mathbf{A}_h'$, where Ψ_i is the lower $(K\times m)$-by-$(K\times m)$ square sub-matrix of $(\mathbf{X}_i'\mathbf{X}_i)^{-1}\mathbf{X}_i'\Sigma_i\mathbf{X}_i(\mathbf{X}_i'\mathbf{X}_i)^{-1}$. Here, Σ_i is estimated based on the model assumption such as AR(1) or AR(2) for $\boldsymbol{\varepsilon}_i$ (Worsley et al., 2002), and σ_i^2 is estimated by its OLS estimate $\hat{\sigma}_i^2$.

Let $\boldsymbol{\tau}_i = (\tau_{1,1}^i, \ldots \tau_{1,m}^i, \ldots, \tau_{K,1}^i, \ldots, \tau_{K,m}^i)'$ be $\text{diag}(\mathbf{A}_h\Psi_i\mathbf{A}_h')$, where $\text{diag}(\cdot)$ represents the diagonal vector of a square matrix. A 95% confidence band for the individual i's and

the population-averaged HRF estimates are

$$\tilde{\boldsymbol{\beta}}_i \pm 2\,\hat{\sigma}_i\sqrt{\boldsymbol{\tau}}_i, \quad \text{and} \quad \frac{1}{N}\sum_{i=1}^{N}\tilde{\boldsymbol{\beta}}_i \pm 2\,\sqrt{\frac{1}{N}\sum_{i=1}^{N}\hat{\sigma}_i^2\boldsymbol{\tau}_i},$$

respectively. Based on the variance matrix for the entire curve of the HRFs for all stimuli, confidence intervals for the difference between the HRFs corresponding to different stimuli can also be constructed. Specifically, let $\boldsymbol{\varsigma}^i_{k,k'} = \text{diag}(\mathbf{S}_{k,k'}\mathbf{A}_h\Psi_i\mathbf{A}_h'\mathbf{S}_{k,k'}')$, where $\mathbf{S}_{k,k'}$ is an m-by-$K\times m$ matrix, whose $(l,(k-1)m+l)$th and $(l,(k'-1)m+l)$th entries equal 1 and -1 respectively for $l=1,\ldots,m$, and the rest entries equal zero. Then a 95% confidence band for the difference between the kth and k'th sample-averaged HRFs is given by

$$\bar{\mathbf{Z}}_{k,k'} \pm 2\,\sqrt{\frac{1}{N}\sum_{i=1}^{N}\hat{\sigma}_i^2\boldsymbol{\varsigma}^i_{k,k'}}.$$

2.5. Algorithms for Parameter Selection

We propose two algorithms for selecting the optimal parameters respectively for the kernel-smoothed estimator $\tilde{\boldsymbol{\beta}}_{i,k}$ and the regularized estimator $\tilde{\boldsymbol{\beta}}^{cor}_{i,k}$. Mean squared error (MSE) is typically used as the criterion for parameter selection. The σ_i^2's are highly heteroscedastic across subjects in real applications; therefore, it is appropriate to weight each subject's fMRI data inversely proportional to σ_i^2 in population-wide inferences: the weighted MSE (WMSE) is used as the criterion instead. For the kernel-smoothed estimator $\tilde{\boldsymbol{\beta}}_{i,k}$, only bandwidth h needs to be selected. Define the WMSE of $\tilde{\boldsymbol{\beta}}_{i,k}$ as

$$\text{WMSE}_k(h) = \frac{1}{N}\sum_{i=1}^{N}\mathbb{E}\,||\tilde{\boldsymbol{\beta}}_{i,k} - \boldsymbol{\beta}_{i,k}||^2/\sigma_i^2.$$

Since the true $\boldsymbol{\beta}_{i,k}$'s are unknown, we approximate $\text{WMSE}_k(h)$ for each candidate h and select the one that minimizes the WMSE estimate. We describe the technical derivations of $\text{WMSE}_k(h)$ in the Appendix and present the selection algorithm below.

Algorithm 1 for selecting bandwidth h for the kernel-smoothed estimator $\tilde{\boldsymbol{\beta}}$.

1. Starting from an initial bandwidth $h_0 = 1/\sqrt{7/\text{TR}}$, for each stimulus k calculate $\tilde{\boldsymbol{\beta}}_{i,k}$ for $i = 1, \ldots, N$, and their average, which is denoted by $\tilde{\boldsymbol{\beta}}^0_{\cdot k}$.

2. For every subject i, calculate the OLS estimate $\hat{\sigma}_i^2$ of the variance σ_i^2 of the regression error in model (2). For each candidate h, calculate matrix $\mathbf{B}_h$ and $\mathbf{A}_h$. Assuming $\Sigma_i = \mathbf{I}_T$, let

$$(\tau^i_{1,1}, \ldots \tau^i_{1,m}, \ldots, \tau^i_{K,1}, \ldots, \tau^i_{K,m})' = \text{diag}(\mathbf{A}_h \Psi_i \mathbf{A}_h'),$$

   where Ψ_i is the lower $(K \times m)$-by-$(K \times m)$ square sub-matrix of $(\mathbf{X}_i'\mathbf{X}_i)^{-1}$.

3. For each candidate h, get an estimate of $\text{WMSE}_k(h)$ as

$$\hat{\text{WMSE}}_k(h) = \sum_{i=1}^{N} \frac{1}{N} \sum_{t=1}^{m} \tau^i_{k,t} + \sum_{i=1}^{N} \frac{1}{N\hat{\sigma}_i^2} \tilde{\boldsymbol{\beta}}'_{\cdot k(h_0)} (\mathbf{B}_h - \mathbf{I}_m)'(\mathbf{B}_h - \mathbf{I}_m) \tilde{\boldsymbol{\beta}}^0_{\cdot k}.$$

4. Choose the h that leads to the smallest $\hat{\text{WMSE}}_k(h)$ for each k, or choose a universal h that minimizes $\sum_{k=1}^{K} \hat{\text{WMSE}}_k(h)$.

Choosing the proper values of h and λ that control the extent of smoothing and regularization is crucial in balancing the variance and bias of the BTik-Kern estimator $\tilde{\boldsymbol{\beta}}^{cor}_{i,k}$. We note that even though $\tilde{\boldsymbol{\beta}}^{cor}_{i,k}$ is the estimator we use in analysis, parameter selection is easier to be conducted on the intermediate Tik-Kern estimator $\tilde{\boldsymbol{\beta}}^{r}_{i,k}$ (here $\tilde{\boldsymbol{\beta}}^{r}_{i,k}$ denotes the sub-vector of $\tilde{\boldsymbol{\beta}}^{r}_{i}$ corresponding to $\boldsymbol{\beta}_{i,k}$). Generalized cross-validation (GCV) (Wahba, 1990) is a standard method for choosing the regularization parameter, and was employed by Casanova et al. (2008, 2009). GCV improves upon the time-consuming leave-one-out ordinary cross-validation (OCV). However, GCV cannot be applied to select parameters for the Tik-Kern estimator, whose prediction errors are different from those of the purely Tikhonov-regularized estimators. Moreover, GCV cannot be used to choose different optimal parameters for different HRFs. To avoid conducting the computationally intensive OCV, we developed a fast algorithm for choosing h and λ that minimize the WMSE of $\tilde{\boldsymbol{\beta}}^{r}_{i,k}$, and its computational time is of the same order as $T \times N$ for each voxel given the length of HRF and the number of stimuli.

We define the WMSE of $\tilde{\boldsymbol{\beta}}^r_{i,k}$ as $\text{WMSE}_k(h,\lambda) = \sum_{i=1}^N \mathbb{E}\,||\tilde{\boldsymbol{\beta}}^r_{i,k} - \boldsymbol{\beta}_{i,k}||^2/(N\sigma_i^2)$. The algorithm is presented below.

Algorithm 2 for selecting h and λ for the Tik-Kern estimator $\tilde{\boldsymbol{\beta}}^r$.

1. Starting from an initial bandwidth $h_0 = 1/\sqrt{7/\text{TR}}$, compute $\tilde{\boldsymbol{\beta}}^0_{\cdot k}$ as in Algorithm 1.

2. Estimate σ_i^2 in model (2) by its OLS estimate $\hat{\sigma}_i^2$ for each subject i.

3. For each candidate h, calculate matrix $\mathbf{B}_h$ and $\mathbf{A}_h$. Let

$$(b_1^i(1),\ldots,b_1^i(m),b_2^i(1),\ldots,b_K^i(m))' = \left(\mathbf{A}_h\mathbf{R}_\lambda^i - \mathbf{I}_{K\cdot m}\right)\tilde{\boldsymbol{\beta}}^0_{\cdot k},$$

   and $(\tau^i_{1,1},\ldots\tau^i_{1,m},\ldots,\tau^i_{K,1},\ldots,\tau^i_{K,m})' = \text{diag}(\mathbf{A}_h\mathbf{R}_\lambda^i\Psi_i(\mathbf{R}_\lambda^i)'\mathbf{A}_h')$.

4. For each combination of candidate (h,λ), get an estimate of $\text{WMSE}_k(h,\lambda)$ as

$$\hat{\text{WMSE}}_k(h,\lambda) = \sum_{i=1}^N \frac{1}{N}\sum_{t=1}^m \tau_{k,t}^i + \sum_{i=1}^N \frac{1}{N\hat{\sigma}_i^2}\sum_{t=1}^m (b_k^i(t))^2.$$

5. Choose the h and λ that minimize $\hat{\text{WMSE}}_k(h,\lambda)$ for each stimulus k or select a universal pair (h,λ) that leads to the smallest $\sum_{k=1}^K \hat{\text{WMSE}}_k(h,\lambda)$.

In the above two algorithms, the OLS estimates $\hat{\sigma}_i^2$ are calculated assuming that the error terms $\boldsymbol{\varepsilon}_i$ are i.i.d. The simulations in Section 3.1 suggest that the proposed methods are robust to this assumption. In practice, we do not select the optimal parameters for each voxel, instead we select one set of optimal parameters for each region of interest (ROI) based on the data from one representative voxel or averaged data of the ROI.

3. Results

In this section, we present the results obtained from simulated data and an event-related fMRI dataset, where the proposed estimation and testing methods are compared with three existing methods with similar order of computation: the Tik-GCV (Casanova

et al., 2008), the SFIR (Goutte et al., 2000), and the basis set method (Friston et al., 1998) that represents the HRF by a linear combination of the canonical HRF and its temporal derivative (referred to as the canonical method hereafter).

3.1. Simulated data

Signal Generation. The simulated data follow the experimental design of the real data — the MID experiment. In summary, fMRI data for $N = 19$ subjects (the sample size of the real data) were simulated. For each subject, we generated 223 frames of fMRI data separated by 2s (TR) from an event-related design with six different stimuli, excluding the first four frames ($T = 219$). More details of the design are described in Section 3.2. Two simulations are conducted: the first focuses on examining the performance of the proposed method for various possible HRFs and the second focuses on a scenario closely matching the real MID data.

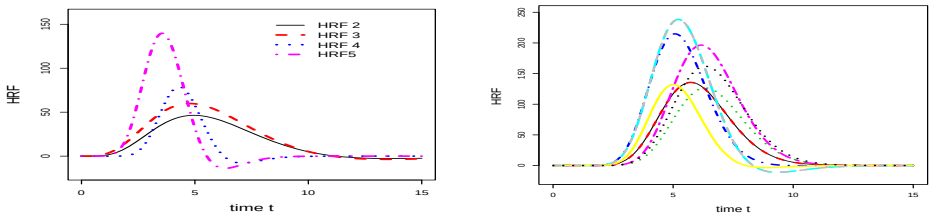
Simulation 1. Six HRFs were considered: the first is a zero function, corresponding to non-responsive voxels; and the remaining five ($k = 2, ..., 6$) follow the form $h_{i,k}(t) = A_{i,k} \cdot f_{i,k}(t+\delta_{i,k})$ for $i = 1, \ldots, N$, where $A_{i,k}$ and $\delta_{i,k}$ are the subject-specific magnitude and latency, respectively. The function $f_{i,k}(t)$ controls the shape of the HRFs and is assumed to be the difference of two gamma density functions (Worsley et al., 2002) as follows (subscript k is dropped here):

$$f_i(t) = b_{1,i}^{a_{1,i}} \frac{t^{a_{1,i}-1} \exp(-b_{1,i}t)}{\Gamma(a_{1,i})} - c \cdot b_{2,i}^{a_{2,i}} \frac{t^{a_{2,i}-1} \exp(-b_{2,i}t)}{\Gamma(a_{2,i})}. \tag{9}$$

The parameters for each simulated HRF are given in Table 2 and examples of the simulated HRFs are displayed in Figure 1. Specifically, the second and third HRFs both follow the canonical form in SPM, differing in subject-specific magnitude and latency. The fourth and fifth HRFs across subjects also differ in magnitude and latency, but both have different functional forms from the canonical one. In addition, to investigate the limitation of the proposed method, we purposely set the range of the latency variation of the fifth HRF comparable to its W (2s). The functional form of the sixth HRF is distinct from the

k	$A_{i,k}$	$\delta_{i,k}$	$a_{1,i}$	$a_{2,i}$	$b_{1,i}$	$b_{2,i}$	c
2	N(300, 50^2)	0	6	16	1	1	1/6
3	$A_{i,2}$+U(30, 50)	U(-0.2,0.2)	6	16	1	1	1/6
4	U(200, 700)	0	20	22	4	4	2/3
5	$A_{i,4}$+U(100, 200)	U(-1,1)	20	22	4	4	2/3
6	U(300, 800)	0	U(18,22)	U(20,24)	U(3,4)	U(3,4)	1/6

Table 2: The parameters of the simulated HRFs $h_{i,k}$.



(a) HRF 2 to HRF 5 (b) HRF 6

Figure 1: Examples of simulated HRFs: (a) one simulated HRF for each of stimuli 2 to 5. (b) several simulated HRF 6

canonical form and differs in shape across subjects.

Following Casanova et al. (2008), we simulated the error term ε_i from an autoregressive model of order 4 (AR(4)), representing a structure with a strong autocorrelation (the lag-1 and lag-2 correlations of the errors are 0.73 and 0.53, respectively; the more detailed autocorrelation function is provided in the supplementary file):

$$\varepsilon_i(t) = 0.37\,\varepsilon_i(t-1) + 0.14\,\varepsilon_i(t-2) + 0.05\,\varepsilon_i(t-3) + 0.02\,\varepsilon_i(t-4) + e_i(t),$$

where $e_i(t) \overset{i.i.d}{\sim} \mathrm{N}(0, \sigma_i^2)$. To reflect the heteroscedastic variances across subjects in practice, we let $\sigma_i \sim \mathrm{Gamma}(1, 1/10) + 10$. For individual subject's fMRI, the signal-to-noise ratio (SNR) defined as $10\log_{10}\left[\frac{\mathrm{var(signal)}}{\mathrm{var(noise)}}\right]$ ranges between -3 to 16 with 99% of probability. Independent errors were also simulated and the results were very similar (details are presented in the online supplementary document). This suggests that the proposed methods are robust to the noise autocorrelation structure under this experimental design.

We simulated 100 i.i.d. fMRI data sets. Within each simulation, we first generated $N =$

19 sets of random functions $h_{i,k}(t)$ for $k = 1, \ldots, 6$, and then computed the observed fMRI time series $\boldsymbol{Y}_i$ from the GLM (2) using the design matrices $\mathbf{X}_i$, the simulated error term $\varepsilon_i(t)$, and a simulated quadratic drift term $d_{0,i} + d_{1,i} \cdot t + d_{2,i} \cdot t^2$ with $d_{0,i} \sim \text{U}(-1, 1)$, $d_{1,i} \sim \text{U}(-0.1, 0.1)$, and $d_{2,i} \sim \text{U}(-0.05, 0.05)$, where U(a,b) denotes uniform distribution with minimum a and maximum b.

Simulation 2. The model for Simulation 2 is chosen to better resemble the real MID data with the following key properties: (1) brain responses to most of the stimuli are inactive, (2) neuro-activities across subjects have a large variation, (3) the magnitudes are positively correlated with the error variance, and (4) the drift terms are in a much larger scale than HRFs. Specifically, the same experimental design as in Simulation 1 is used, and the first four HRFs are set to zero. We let $h_{i,5}(t) = h_{i,6}(t) = A_{i,6} f_i(t)$ using the same $f_i(t)$ as that of the sixth stimulus in Simulation 1, and let $A_{i,6}$ be simulated from a mixture of uniforms: four out of N from Unif$(100, 200)$ and the rest from Unif$(6000, 8000)$. The error terms $\varepsilon_i(t)$ are generated under the same AR(4) model as above except that the σ_i's are from $\Gamma(2, 1/10) + 20$, and are ordered to have the same rank as that of $A_{i,6}$, representing a strong positive correlation with the magnitudes. The fMRI data were simulated in the same manner as Simulation 1 with $d_{0,i} \sim \text{U}(8000, 15000)$, $d_{1,i} \sim \text{U}(-2, 3)$, and $d_{2,i} \sim \text{U}(-0.01, 0.01)$.

Statistical Analysis and Discussion

To compare the different estimation methods, we used the criterion of average relative error (ARE):

$$e(S_k) = \frac{1}{N} \sum_{i=1}^{N} \frac{|S_{i,k} - S_{i,k}^{est}|}{S_{i,k}}, \quad e(\text{RMSE}_k) = \frac{1}{N} \sum_{i=1}^{N} \frac{\| \boldsymbol{\beta}_{i,k} - \tilde{\boldsymbol{\beta}}_{i,k}^{est} \|}{\| \boldsymbol{\beta}_{i,k} \|},$$

where S stands for a summary statistic of the HRF, including HR, TTP, and W, and RMSE stands for root mean squared error.

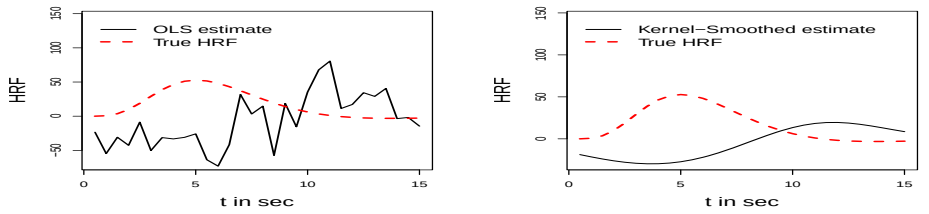
Simulation 1. The median AREs of the estimates for the HRFs by different methods in Simulation 1 are summarized in Table 3. In the estimation of the first five HRFs, the BTik-

	HRF k	BTik-Kern	Tik-GCV	SFIR ($g = 1$)	SFIR ($g = 10$)	Canonical
HR	2	0.34	0.54	0.66	0.75	0.83
	3	0.25	0.60	0.27	0.36	1.70
	4	0.47	0.67	0.75	0.92	0.83
	5	0.36	0.57	0.65	0.83	0.58
	6	0.36	0.58	0.57	0.73	0.82
TTP	2	0.21	0.62	0.67	0.75	0.50
	3	0.19	0.43	0.43	0.48	0.28
	4	0.19	0.76	0.52	0.76	1.20
	5	0.14	0.50	0.26	0.47	0.45
	6	0.11	0.17	0.07	0.06	0.68
W	2	0.29	0.52	0.66	0.60	0.27
	3	0.19	0.45	0.30	0.33	0.07
	4	0.24	1.34	0.28	0.31	1.28
	5	0.50	1.50	0.14	0.13	1.50
	6	0.20	0.58	0.13	0.13	0.70
RMSE	2	0.78	1.28	1.09	1.11	1.65
	3	0.60	1.19	0.71	0.75	1.68
	4	0.89	1.68	1.23	1.18	2.36
	5	0.79	1.24	0.84	0.92	1.58
	6	0.61	0.86	0.69	0.78	1.34

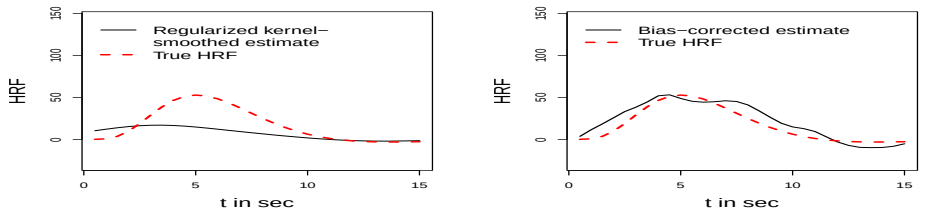
Table 3: Median AREs for estimating HR, TTP, W and RMSE of the simulated HRFs from different methods in Simulation 1.

Kern estimator clearly outperforms the other methods except for SFIR in approximating W of the fifth HRF. This is because when the horizontal shifts of HRFs are comparable to their W, the sample-averaged (averaged for each time point) HRF with an enlarged W has a very different shape from the individual ones, thus BTik-Kern leads to a biased estimate of W. Similarly, in the sixth simulation scenario, when very large variations in HRF curves exist across subjects, an over-shrinkage by BTik-Kern is possible for curves far away from the average curve, which explains why BTik-Kern is slightly inferior to SFIR in estimating TTP and W of the sixth HRF. Nevertheless, the advantage of BTik-Kern in estimating the whole HRF curve (measured by the RMSE) is still evident even under the sixth scenario. Interestingly, the canonical method leads to larger errors than BTik-Kern even when the underlying HRFs follow the canonical form ($k = 2, 3$). This is possibly due to a poor overall model fitting when the other HRFs are distinct from the canonical form, since

the HRFs corresponding to different stimuli are estimated simultaneously. The separate effects of kernel-smoothing, regularization, and bias correction on the HRF estimation are clearly illustrated in Figure 2, where the estimation of one $h_{i,2}(t)$ is used as an example: kernel-smoothing increases the temporal continuity of the OLS estimate; regularization further reduces the variability of the estimate in magnitude, but also introduces a large bias; the bias-correction step adjusts this bias and nearly recovers the underlying truth.



(a) Least Square Estimate $\hat{\boldsymbol{\beta}}_{i,2}$ (b) Kernel Smoothed Estimate $\tilde{\boldsymbol{\beta}}_{i,2}$



(c) Regularized Estimate $\tilde{\boldsymbol{\beta}}^{r}_{i,2}$ (d) Bias Corrected Estimate $\tilde{\boldsymbol{\beta}}^{cor}_{i,2}$

Figure 2: Four estimates of a random $h_{i,2}$: (a) the OLS estimate of the HRF with randomly generated fMRI data; (b) the kernel-smoothed estimate using the selected h; (c) the Tikhonov-regularized and smoothed estimate (Tik-Kern) using the selected h and λ; (d) the bias-corrected (BTik-Kern) estimate using the selected h and λ.

For hypothesis testing, the proposed kernel-smoothed tests, based on statistics (7) and (8), are compared with the Hotelling's t-tests of HRF estimates obtained from Tik-GCV and SFIR. The parametric canonical method is excluded from the comparison, because the resulting estimates with similar shapes are not suitable for Hotelling's tests. Readers are referred to Calhoun et al. (2004) and Lindquist et al. (2009) for details of hypothesis testing based on the canonical method. Here we first evaluated the type I error of the tests by testing the significance of the first HRF — since $h_1 = 0$, any significant result detected by a test is a type I error. The proposed test (7) has a much smaller type I error

than the tests by the other methods as shown in the first column of Figure 3. In fact, tests based on the Tik-GCV and SFIR with inflated type I error resulted from the bias induced by the regularization incorrectly rejected the null (no signal) at the 5% level in almost all of the simulations, while the kernel-smoothed tests did so in 20% of the simulations due to the almost singular design matrix. We found that if the design matrix is non-singular, the kernel-smoothed tests will achieve the nominal 5% significance level, while Tik-GCV and SFIR still have inflated type-I error (the simulation result is illustrated in the supplementary file).

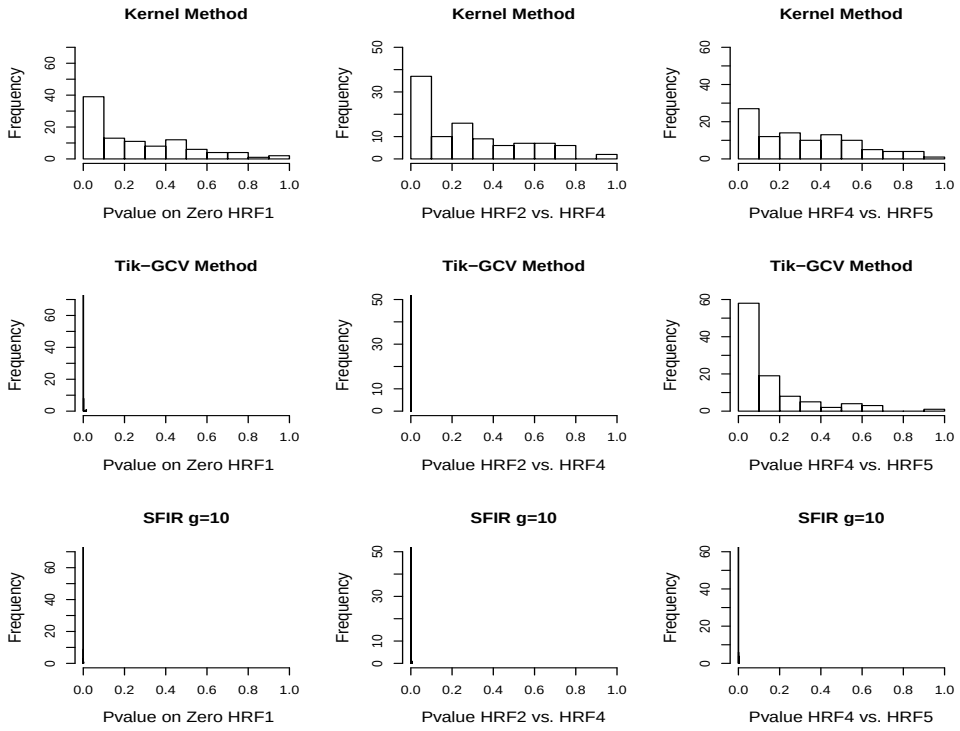


Figure 3: **The P-values of testing HRFs using different methods in Simulation 1.**

To evaluate the statistical power of the tests, we conduct tests on the difference between pairs of the HRFs—since the underlying truth is $h_{i,k} \neq h_{i,k'}$, any failure in rejecting the null is a type II error. The proposed method is based on the test statistic $Q_{k,k'}$ in (8). The second and third columns of Figure 3 display the histograms of the P-values of testing $h_{i,2}$ vs. $h_{i,4}$ and $h_{i,4}$ vs. $h_{i,5}$, respectively, using the different methods. SFIR appears

to have the largest power, as it correctly rejected the null in all simulations, while BTik-Kern has a smaller power. However, in another simulation where $h_{i,5}$ was set to equal $h_{i,4}$ with the rest of HRFs unchanged, both SFIR and Tik-GCV incorrectly rejected the null hypothesis most of time. We suppose the power gain of these two methods is at the cost of a large type two error for this example. Comparisons of the other pairs of HRFs show a similar pattern.

Simulation 2. Due to the large variation of HRFs in shape and magnitude across subjects, SFIR performed best in estimating TR and the entire function. However, BTik-Kern beat the other two methods in estimating TTP and W (the numerical results are omitted here). In addition, kernel method achieved a much better power in hypothesis testing than SFIR and Tik-GCV with the type I error kept at a nominal significance level. Figure 4 shows the histograms of P-values of testing (1) $\mathbb{E}\, h_{i,1} = 0$, (2) $\mathbb{E}\, h_{i,5} = 0$, (3) $\mathbb{E}\, h_{i,6} = \mathbb{E}\, h_{i,5}$ and (4) $\mathbb{E}\, h_{i,5} = \mathbb{E}\, h_{i,4}$ in 100 i.i.d. simulations through the three methods. In the first and third tests with true null hypotheses, the smoothed method has a reasonably flat histogram, indicating its P-value close to the nominal level, and it also demonstrates a larger power than SFIR and Tik-GCV in the other two tests. Because most HRFs are not significant, the inflated type I errors of SFIR and Tik-GCV no longer exist in this example (more discussions on this issue are given in Conclusion). However, possibly due to the shrinkage imposed by Tikhonov regularization in these two methods, the resulting two tests are not sensitive to the deviation from the null hypothesis, compared to the kernel-smoothed method.

3.2. Illustrative example

Subjects. The data were collected from the Monetary Incentive Delay (MID) Experiment, which measures subjects' brain activity related to reward and penalty processing (Knutson et al., 2000). In total, 19 subjects (10 male, 9 female) participated in exchange for financial payment ($40.00 minimum, plus whatever money they managed to win during the study task). Subjects were recruited from a larger representative longitudinal community sample (Allen et al., 2007). All participants were between 22 and 25 years of age

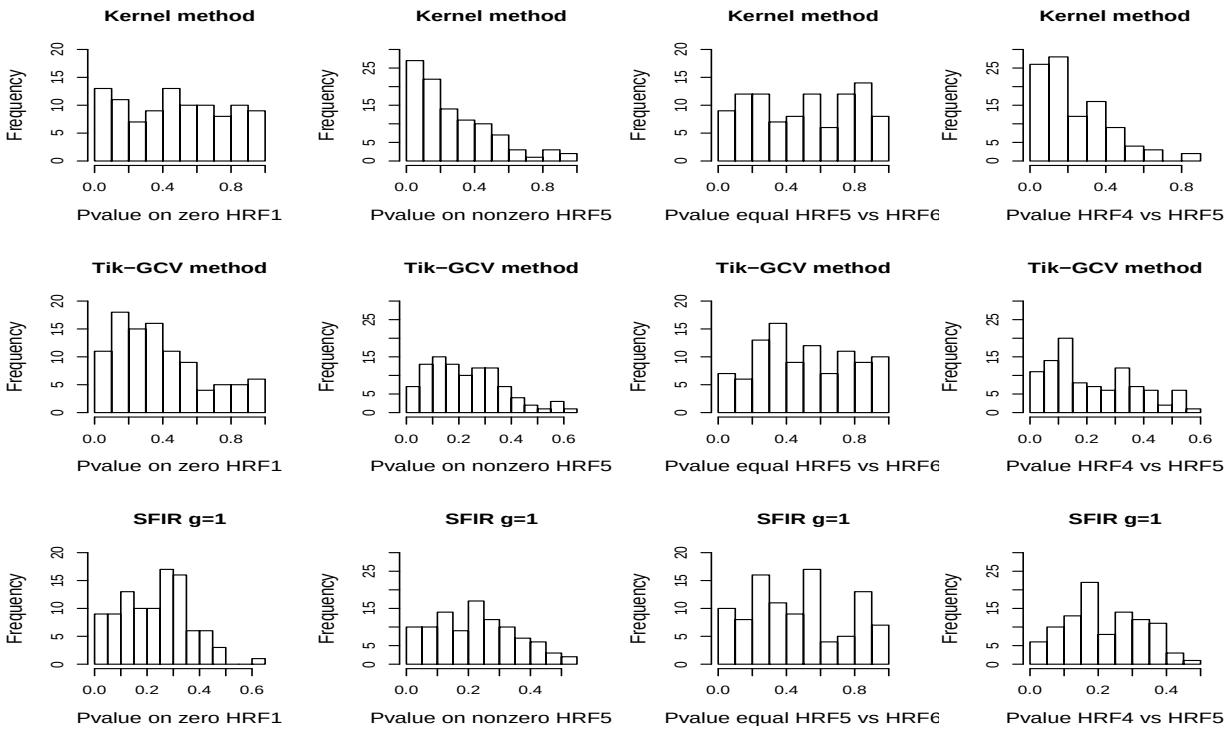


Figure 4: **The P-values of testing HRFs using different methods in Simulation 2.**

at the time of participation, with 37% identified as black and 63% identified as white.

Experimental Design. In the MID task, each participant completed a protocol comprised of 2 blocks of 72 6-s trials involving either no monetary outcome (control task), a potential reward (reward task), or a potential penalty (penalty task). During each trial, participants are shown a shape for 500 ms (anticipation condition), a variable interval delay of between 4000 and 4500 ms, and a white target square lasting between 160 and 260 ms (response condition). Participants are then instructed to respond with a button press. The cue shape (circle, square or triangle) shown at the start of each trial signals the type of the trial to be implemented, i.e., reward, penalty or no incentive respectively. Additionally, each reward and penalty shape included lines across the shape, which indicated the amount of money the participant could win or lose during the trial (i.e., 3 lines = \$5.00, 2 lines = \$1.00, and 1 line = \$0.20). Participants were also told that their reaction times to the white target would be recorded, and that receiving the monetary reward or preventing

punishment depended on whether they responded within a given window of time. The order of trials in the protocol for each participant was randomized, with 25% of them control trials, 37.5% reward trials, and 37.5% punishment trials. In addition to the fMRI data, measures of each subject's state anxiety were collected using the state-trait anxiety inventory (Spielberger and Vagg, 1984).

Data Acquisition and Preprocessing. Functional images were acquired using a Siemens 3.0 Tesla MAGNETOM Trio high-speed magnetic imaging device at UVA's Fontaine Research Park, with a CP transmit/receive head coil with integrated mirror. Two hundred twenty-four functional T2*-weighted Echo Planar images (EPIs) sensitive to BOLD contrast were collected per block, in volumes of 28 3.5-mm transversal echo-planar slices (1-mm slice gap) covering the whole brain (1-mm slice gap, TR=2000ms, TE=40ms, flip angle=90°, FOV= 192 mm, matrix= 64×64, voxel size= 3×3×3.5mm). Prior to the collection of functional images, 176 high-resolution T1-magnetization-prepared rapid-acquisition gradient echo images were acquired to determine the localization of function (1-mm slices, TR=1900 ms, TE=2.53ms, flip angle= 9°, FOV=250mm, voxel size= 1×1×1mm). Data were preprocessed and analyzed using FMRIB's Software Library (FSL) software (Version 5.98; www.fmrib.ox.ac.uk/fsl, Smith et al., 2004; Woolrich et al., 2009). Motion correction involved FMRIB's Linear Image Registration Tool, an intra-modal correction algorithm tool (MCFLIRT; Jenkinson et al., 2002), with slice scan-time correction and a high-pass filtering cutoff point of 100 seconds, removing irrelevant signals. We used BET (Smith, 2002) brain extraction, eliminating non-brain material voxels in the fMRI data, and a 5-mm full width at half maximum Gaussian kernel for smoothing. Images were registered to the Montreal Neurological Institute (MNI) space by FLIRT (Jenkinson et al., 2002). Regions of interest (ROIs) were determined structurally using the Harvard subcortical brain atlas, and were chosen for their likely involvement in affective processing based on previous studies of affective neural processes (e.g., Knutson et al., 2000). The ROIs chosen for analysis were the right putamen, right caudate, right nucleus accum-

bens, right pallidum, and right amygdala.

Statistical Analysis and Discussion We included six stimuli in the GLM for the MID data: the three signal stimuli for the three types of monetary outcomes and the corresponding three reaction stimuli to which the participants are required to respond. The six stimuli are henceforth referred to as neutral anticipation, reward anticipation, penalty anticipation, neutral response, reward response, and penalty response. Our analysis focused on three goals: first, identifying the brain voxels responsive to each stimulus, especially those involving monetary outcomes; second, identifying the voxels that react differentially to monetary reward and punishment stimuli; and third, modelling the relationship between subjects' brain functions related to reward and punishment processing measured by the fMRI data and self-reported state anxiety.

To identify the brain voxels responsive to each stimulus, we conducted the proposed kernel-smoothed hypothesis tests in Section 2.4 using the test statistic (7). Comparisons were drawn to the Tik-GCV and the SFIR. Analysis results show a similar pattern across the five ROIs; thus, results for only the right caudate and putamen are presented below. Rows 1-3 in Table 4 summarize the percentages of the voxels in each ROI responsive, respectively, to the neutral response, the reward anticipation, and the penalty anticipation stimuli, identified at a significance level of .05 by tests based on the three methods. Among the six stimuli, the kernel-smoothed tests identified the most responsive voxels corresponding to these three stimuli. By contrast, SFIR and Tik-GCV detected much fewer responsive voxels than the kernel-smoothed test (7) to all the stimuli. In Figure 5, panels (a), (b) and (c), respectively, show the significant voxels identified by the kernel-smoothed tests in the right caudate responsive to neutral response, reward anticipation, and penalty anticipation cues, and panels (d), (e), (f) respectively show the corresponding voxels in the right putamen. These findings are consistent with previous MID findings that the right caudate and putamen are sensitive to the motivational value (reward response and penalty anticipation) of stimuli (Knutson et al., 2000; Knutson et al., 2001; Bjork et al., 2004).

subject of test	ROI	BTik-Kern	Tik-GCV	SFIR
penalty	right caudate	32.3	5.4	3.3
anticipation	right putamen	43.9	2.9	4.2
reward	right caudate	29.4	14.0	6.1
anticipation	right putamen	40.0	14.7	4.2
neutral	right caudate	27.6	5.1	5.5
response	right putamen	45.3	5.4	8.2
neutral vs. penalty	right caudate	19.4	3.4	4.7
anticipation[1]	right putamen	13.2	2.1	4.7
neutral vs. penalty	right caudate	17.1	5.0	7.8
response[2]	right putamen	20.4	3.8	15.4
regression of	right caudate	6.6	0.6	3.0
state anxiety[3]	right putamen	14.3	1.0	2.5

Table 4: The percentages of significant voxels in the right caudate and the right putamen identified by different tests. [1] Voxels react differentially to neutral anticipation stimulus versus penalty anticipation stimulus; [2] Voxels react differentially to neutral response stimulus versus penalty response stimulus; [3] Voxels with difference in HRF magnitudes between penalty anticipation and neutral anticipation stimuli that are significantly predictive of the individual state anxiety score.

To identify the brain voxels that react differentially to the monetary and the neutral stimuli, we conducted tests on six pairs of HRFs: neutral vs. reward, neutral vs. penalty, and reward vs. penalty for both anticipatory and response stimuli. The percentages of significant voxels in the ROIs from the three methods are displayed in rows 4-5 in Table 4. The kernel-smoothed tests, based on the statistic (8), identified a sizable number of voxels at a significance level of .05 that react differentially in two among the six pairs of comparisons: neutral anticipation vs. penalty anticipation, and neutral response vs. penalty response. The corresponding significant voxels in the right caudate and putamen are shown in Figures 6. By contrast, the Tik-GCV and SFIR identified very few voxels that function differentially for these two pairs of comparisons.

To explore the connection between reward-processing brain activity and subjective experience, as measured by the state anxiety score, we conducted a multiple linear regression with the state anxiety as the dependent variable, and the four magnitude differences between the HRFs of monetary and neutral stimuli as predictors. The BTik-Kern

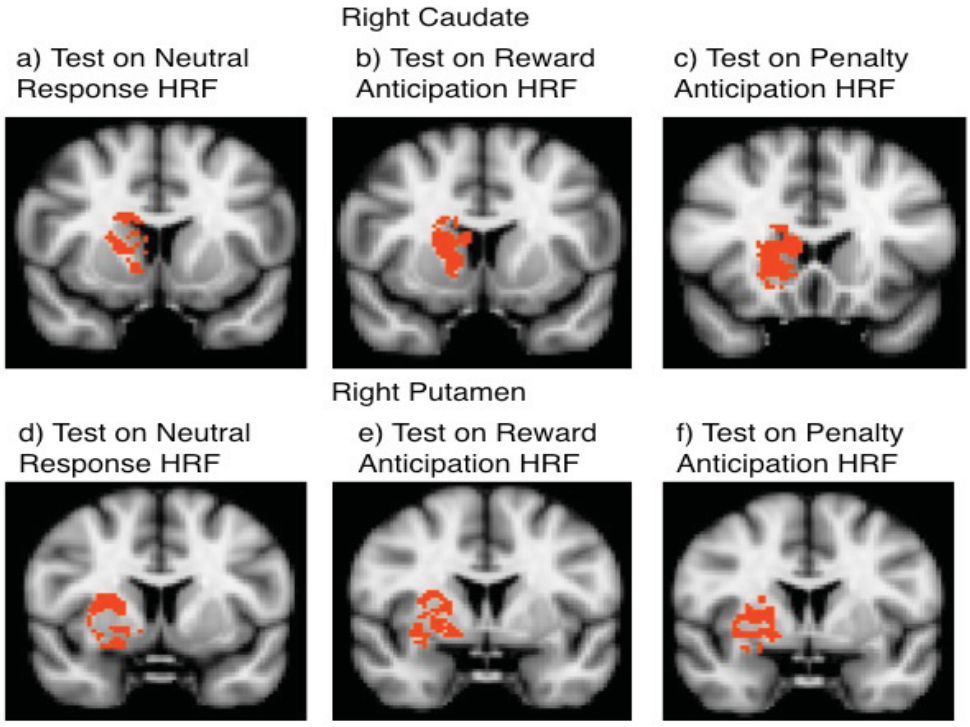


Figure 5: **Tests on neutral reaction, reward anticipation and penalty anticipation stimuli** 5(a),5(b), 5(c): Tests based on kernel-smoothed HRF estimates for voxels in the right caudate. 5(d),5(e), 5(f): Tests based on kernel-smoothed HRF estimates for voxels in the right putamen.

estimator $\tilde{\beta}_{i,k}^{cor}$ was used to estimate the HRFs, the HR of which was extracted for constructing the predictors. For both the right caudate and putamen, across the four predictors, the BTik-Kern identified a sizeable number of voxels that are significantly (P-value .05) associated with the magnitude difference between the penalty and neutral anticipation stimuli, while none of the other methods showed significant results for any of the four predictors (row 6 in Table 4). Figure 7 shows the significant voxels in the right caudate and putamen detected by the BTik-Kern. In addition, state anxiety was found to be positively correlated with penalty stimuli and negatively correlated with reward stimuli. Given that penalty anticipation is expected to co-vary with state anxiety, the finding that penalty anticipation in the right putamen is significantly positively correlated with state anxiety provides compelling evidence that the HRFs estimated by the BTik-Kern are

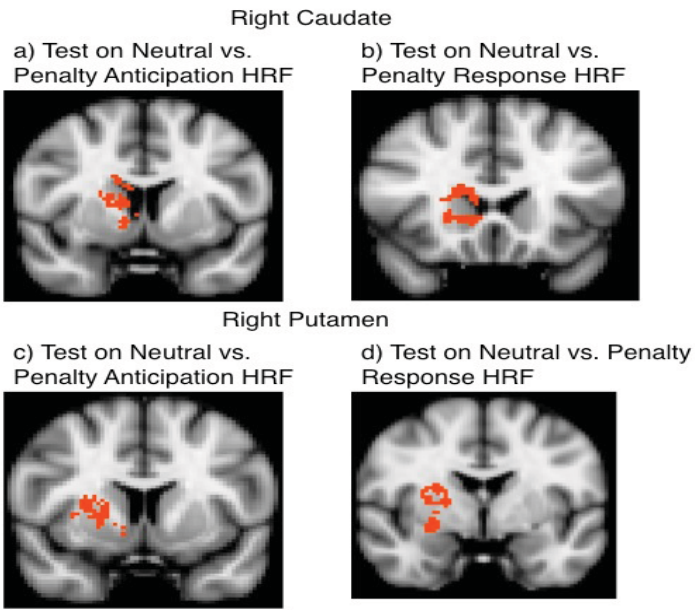


Figure 6: Figure 6(a) and 6(c) compare the HRFs of neutral anticipation and penalty anticipation through the kernel-smoothed method for voxels in the right caudate and right putamen, respectively. 6(b) and 6(d) compare the HRFs of neutral response and penalty response through the kernel-smoothed method for voxels in the right caudate and right putamen, respectively.

picking up on meaningful variance in the data.

4. Conclusions

Within the framework of the GLM, we propose two nonparametric HRF estimators that provide flexible modeling of brain activities across different brain regions, stimuli and subjects. The first kernel-smoothed estimator is developed to construct population-wide hypothesis tests on brain responses to stimuli. This test is on the whole HRF curve rather than only summaries of the HRF, as in standard methods; simulations suggest that it has a much smaller type I error than the standard t-tests based on several existing regularization-based estimators, such as SFIR and Tik-GCV.

It is important to point out that the larger power of the tests through SFIR and Tik-GCV methods is largely countered by an enormous type I error, while the kernel-smoothed test (7) has a much smaller type I error under the (close to) singular MID design. Ad-

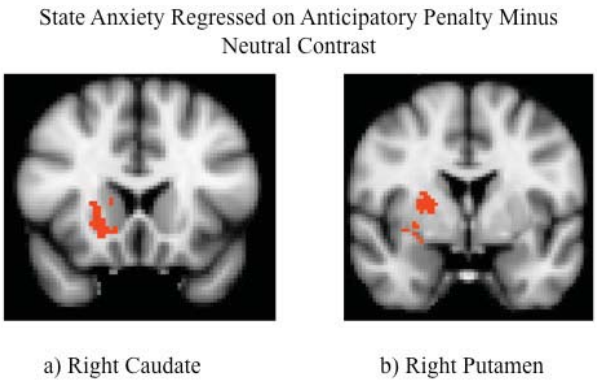


Figure 7: Voxels in the right caudate and right putamen, respectively, whose HRF-magnitude difference under penalty anticipation and neutral anticipation stimuli are significantly correlated with subjects' state anxiety measures.

ditional simulations (presented in the online supplementary document) show that the kernel-smoothed tests have a type I error very close to the nominal 5% level when the design matrix is non-singular. The inflated type I errors of SFIR and Tik-GCV are possibly due to the bias incurred by Tikhonov regularization. In the non-orthogonal design we are investigating, this bias depends on the values of all the HRFs in the model. We expect two scenarios under which the tests on HRFs of a single stimulus through SFIR and Tik-GCV have the nominal significance level: (1) most or all the HRFs are zero functions, or (2) the design is orthogonal such that HRF estimates for different stimuli are independent. However, under the second circumstance, if the number and the order of stimuli evoked in the experiment are not balanced, the biases due to Tikhonov regularization are possibly unequal for different HRFs. Consequently, inflated type I error may still persist in the SFIR or Tik-GCV based tests for pairwise comparisons between the HRFs if the HRFs of most stimuli are far from zero.

The second estimator (BTik-Kern) is constructed by applying both Tikhonov regularization and kernel-smoothing to the OLS estimate, where the kernel bandwidth and regularization parameter separately control the extent of smoothness and regularization constraints on the estimates. Generally, to achieve efficient estimation, the larger the HRF temporal resolution is, the smaller the smoothing bandwidth should be; the smaller the

SNR is, the stronger the regularization should be. Under the studied experimental design, we found that BTik-Kern generally outperforms Tik-GCV and SFIR in terms of estimation error. The proposed kernel smoothing method does not require any differentiability constraint and works well when the underlying function is non-differentiable and the observations are reasonably dense. However, if the temporal resolution of the HRF is small, for differentiable HRFs, kernel smoothing may be inferior to the regularization methods that directly utilize the differentiability assumption, such as Tik-GCV and SFIR.

A main thrust of this paper is to employ sample-averaged HRFs in multi-subject data for bias correction. As shown in the simulations, this is very effective in reducing estimation errors. The reasons for this improvement are two-fold: first, in principle, the common characteristics shared by HRFs across subjects can be evaluated most efficiently by using all the data; second, the sample-averaged HRF estimate approximates the shape of the HRFs across subjects well when they are not very different from each other. However, BTik-Kern may perform poorly when the underlying HRFs vary tremendously across subjects, but this is less of a concern for HRFs in the same brain region and under the same stimulus. Along the lines of borrowing information from the sample average, it is also possible to conduct bias correction towards SFIR and Tik-GCV; this is subject to further investigation.

Although our primary goal was not to investigate the choice of regularization parameters for different HRFs, we indeed observed that the optimal parameters can vary across stimuli. When some specific HRFs display distinct properties from the rest, it is more desirable to select the regularization parameter suitable for estimating only those HRFs. To this end, BTik-Kern is more adaptive than Tik-GCV, as it allows using different parameters for different HRFs.

As the HRF temporal resolution increases, the number of free parameters also increases. Consequently, the variability of nonparametric HRF estimates generally increases as well. In addition, the design matrix would become close to singular given a fixed observational time. Under this situation, a mild Tikhonov regularization with a small

parameter λ is recommended for calculating the subject-averaged HRF estimate.

The false discovery rate of the multiple hypothesis testing can be evaluated and controlled by the empirical Bayes estimate by Efron (2008) or the approaches by Benjamini and Hochberg (1995) and Genovese (2000). However, correction of multiple comparison by evaluating family wise error of the proposed test is difficult, because many factors, such as how the errors are correlated temporally and spatially, and how HRFs vary across subjects, stimulus types and voxels, may affect the error rate. This issue is beyond the scope of current paper, and is subject to more investigation in future research.

The online supplementary document is available at:
`http://www.stat.duke.edu/~ fl35/HRF/SupplHRF.pdf`.

Appendix

A. Evaluation of WMSE$_k(h)$. We first show the derivation of the WMSE of the nonparametric estimator $\tilde{\boldsymbol{\beta}}_i$, $i = 1, \ldots, N$. It is easy to see that

$$\mathbb{E}[\tilde{\beta}_{i,k}(t) - \beta_{i,k}(t)]^2/\sigma_i^2 = \mathbb{E}[\mathbb{V}(\tilde{\beta}_{i,k}(t)|i)]/\sigma_i^2 + \mathbb{E}[\mathbb{E}(\tilde{\beta}_{i,k}(t)|i) - \beta_{i,k}(t)]^2/\sigma_i^2.$$

As the OLS estimate $\hat{\boldsymbol{\eta}}_i = \boldsymbol{\eta}_i + (\mathbf{X}_i'\mathbf{X}_i)^{-1}\mathbf{X}_i'\boldsymbol{\epsilon}_i$, we have

$$\mathbb{E}(\hat{\boldsymbol{\beta}}_i|i) = \boldsymbol{\beta}_i \quad \text{and} \quad \mathbb{V}(\hat{\boldsymbol{\eta}}_i|i) = \sigma_i^2(\mathbf{X}_i'\mathbf{X}_i)^{-1}\mathbf{X}_i'\Sigma_i\mathbf{X}_i(\mathbf{X}_i'\mathbf{X}_i)^{-1}.$$

With $\Sigma_i = \mathbf{I}_T$, $\mathbb{V}(\hat{\boldsymbol{\eta}}_i|i) = \sigma_i^2(\mathbf{X}_i'\mathbf{X}_i)^{-1}$. Denote $\mathbb{V}(\hat{\boldsymbol{\beta}}_i|i)$ by $\sigma_i^2\Psi_i$, which is the sub-matrix of $\mathbb{V}(\hat{\boldsymbol{\eta}}_i|i)$ corresponding to $\hat{\boldsymbol{\beta}}_i$. Then we have

$$\tilde{\boldsymbol{\beta}}_i = \mathbf{A}_h\hat{\boldsymbol{\beta}}_i, \;\; \mathbb{E}(\tilde{\boldsymbol{\beta}}_i|i) = \mathbf{A}_h\boldsymbol{\beta}_i \;\text{ and }\; \mathbb{V}(\tilde{\boldsymbol{\beta}}_i|i) = \sigma_i^2\mathbf{A}_h\Psi_i\mathbf{A}_h'.$$

For estimating the bias $\mathbb{E}[\mathbb{E}(\tilde{\beta}_{i,k}(t)|i) - \beta_{i,k}(t)]^2$, since $\mathbb{E}(\tilde{\boldsymbol{\beta}}_{i,k}|i) - \boldsymbol{\beta}_{i,k} = (\mathbf{B}_h - \mathbf{I}_m)\boldsymbol{\beta}_{i,k}$, then

$$||\,\mathbb{E}(\tilde{\boldsymbol{\beta}}_{i,k}|i) - \boldsymbol{\beta}_{i,k}||^2 = \boldsymbol{\beta}_{i,k}'(\mathbf{B}_h - \mathbf{I}_m)'(\mathbf{B}_h - \mathbf{I}_m)\boldsymbol{\beta}_{i,k}.$$

Combining the above equation and the formula for $\mathbb{V}(\hat{\boldsymbol{\beta}}_i|i)$, we have

$$\text{WMSE}_k(h) = \sum_{i=1}^{N}\sum_{t=1}^{m}\tau_{k,t}^{i}/N + \sum_{i=1}^{N}\boldsymbol{\beta}_{i,k}'(\mathbf{B}_h - \mathbf{I}_m)'(\mathbf{B}_h - \mathbf{I}_m)\boldsymbol{\beta}_{i,k}/(N\sigma_i^2).$$

In practical approximation, σ_i^2 above is estimated by its OLS $\hat{\sigma}_i^2$, and the $\boldsymbol{\beta}_{i,k}$'s are all replaced with their sample average $\tilde{\boldsymbol{\beta}}_{\cdot k}^{0}$.

B. Evaluation of WMSE$_k(h,\lambda)$. The Tikhonov-regularized estimator of $\boldsymbol{\eta}_i$ is given by

$$\hat{\boldsymbol{\eta}}_i^T = (\mathbf{X}_i'\mathbf{X}_i + \lambda\mathrm{D})^{-1}\mathbf{X}_i'\boldsymbol{Y}_i = (\mathbf{X}_i'\mathbf{X}_i + \lambda\mathrm{D})^{-1}(\mathbf{X}_i'\mathbf{X}_i)\hat{\boldsymbol{\eta}}_i.$$

Because the lower left $(K\times m)$-by-3 sub-matrix of $(\mathbf{X}_i'\mathbf{X}_i + \lambda\mathrm{D})^{-1}(\mathbf{X}_i'\mathbf{X}_i)$ equals zero, then the sub-vector $\hat{\boldsymbol{\beta}}_i^T$ of $\hat{\boldsymbol{\eta}}_i^T$, which is corresponding to $\boldsymbol{\beta}_i$, equals $\mathbf{R}_\lambda^i\hat{\boldsymbol{\beta}}_i$. Consequently, the Tik-Kern estimator is essentially the kernel-smoothed $\hat{\boldsymbol{\beta}}_i^T$: $\tilde{\boldsymbol{\beta}}_i^r = \mathbf{A}_h\mathbf{R}_\lambda^i\hat{\boldsymbol{\beta}}_i$. We can easily get its bias and variance as

$$\mathbb{E}(\tilde{\boldsymbol{\beta}}_i^r|i) = \mathbf{A}_h\mathbf{R}_\lambda^i\boldsymbol{\beta}_i \quad \text{and} \quad \mathbb{V}(\tilde{\boldsymbol{\beta}}_i^r|i) = \sigma_i^2\mathbf{A}_h\mathbf{R}_\lambda^i\Psi_i(\mathbf{R}_\lambda^i)'\mathbf{A}_h'.$$

The WMSE of $\tilde{\boldsymbol{\beta}}_{i,k}^r$ can be derived in a similar way as that of $\tilde{\boldsymbol{\beta}}_{i,k}$.

Acknowledgments

The authors are grateful to the editor and three reviewers for their constructive comments that greatly improve the paper. ZT's research is partially funded by the NSF-DMS grant 12-09118. FL's research is partially funded by the NSF-DMS grant 12-08983. Part of the project was conducted when ZT and FL were research fellows of the Object Data Analysis program of the U.S. Statistical and Applied Mathematical Sciences Institute (SAMSI). JAC's research was partially funded by a grant issued by the National Institute of Mental Health (NIMH). The project described was supported by Award Number R01MH080725 to JAC. The content is solely the responsibility of the authors and does not necessarily represent the official views of NIMH, the National Institutes of Health or SAMSI. Sup-

port from Karen Hasselmo, Alexander Tatum, and Zoe Englander is also acknowledged.

References

[1] Aguirre, G.K., Zarahn, E. and D'Esposito, M. (1998). The variability of human, BOLD hemodynamic responses. Neuroimage 8, 360-369.

[2] Allen, J. P., Porter, M., McFarland, F. C., McElhaney, K. B., Marsh, P. (2007). The relation of attachment security to adolescents' paternal and peer relationships, depression, and externalizing behavior. Child Development 78, 1222-1239.

[3] Bai, P., Truong, Y. and Huang, X. (2009). Nonparametric Estimation of Hemodynamic Response Function: A Frequency Domain Approach. IMS Lecture Notes-Monograph Series. Optimality: The Third Erich L. Lehmann Symposium. 57, 190-215.

[4] Benjamini, Y. and Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. Journal of the Royal Statistical Society. Series B (Methodological), 289-300.

[5] Bjork, J. M., Knutson, B., Fong, G. W., Caggiano, D. M., Bennett, S. M., and Hommer, D. (2004). Incentive-elicited brain activation in adolescents: Similarities and differences from young adults. Journal of Neuroscience 24, 1793-1802.

[6] Brosch, J., Talavage, T., Ulmer, J., and Nyenhuis, J. (2002). Simulation of human respiration in fMRI with a mechanical model. IEEE Transcations on Biomedical Engeering 49, 700-707.

[7] Buxton, R. B., Uludag, K., Dubowitz, D. J., and Liu, T. T. (2004). Modeling the hemodynamic response to brain activation. Neuroimage 23, 220-233.

[8] Calhoun, V. D., Stevens, M. C., Pearlson, G. D., and Kiehl, K.A. (2004). fMRI analysis with the general linear model: removal of latency-induced amplitude bias by incorporation of hemodynamic derivative terms. Neuroimage 22: 252-257.

[9] Casanova, R., Ryali, S., Serences, J., Yang, L., Kraft, R., Laurienti, P. J., Maldjian, J. A. (2008). The impact of temporal regularization on estimates of the BOLD hemodynamic response function: a comparative analysis. Neuroimage 40(4), 1606-1618.

[10] Casanova, R., Yang, L., Hairston, W. D., Laurienti, P. J., Maldjian, J. A. (2009). Evaluating the impact of spatio-temporal smoothness constraints on the BOLD hemodynamic response function estimation: an analysis based on Tikhonov regularization. Physiological Measurement 30(5), 37-51.

[11] Ciuciu, P., Poline, J. B., Marrelec, G., Idier, J., Pallier, C., Benali, H. (2003). Unsupervised robust nonparametric estimation of the hemodynamic response function for any fMRI experiment.IEEE Transactions on Medical Imaging 22, 1235-51.

[12] Dale, A. (1999) Optimal Experimental Design for Event-Related fMRI. Human Brain Mapping 8, 109-114.

[13] Efron, B. (2008) False discovery rates and the James-Stein estimator. Statistica Sinica 18, 805-816.

[14] Eubank, R. L.(1988). Nonparametric Regression and Spline Smoothing. Marcel Dekker, New York.

[15] Friston, K. J., Fletcher, P., Josephs, O., Holmes, A., Rugg, M. D., and Turner, R. (1998). Event-related fMRI: characterizing Differential Responses. Neuroimage 7, 30-40.

[16] Friston, K. J., Jezzard, P. J., and Turner R. (1994). Analysis of functional MRI time-series. Human Brain Mapping 1, 153-171.

[17] Friston, K. J., Holmes, A. P., Poline, J. B., Grasby, P. J., Williams, S. C., Frackowiak, R.S. and Turner, R. (1995a). Analysis of fMRI time-series revisited. Neuroimage 2, 45-53.

[18] Friston, K. J., Holmes, A. P., Worsley, K., Poline, P. J., Frith, C. and Frackowiak, R. (1995b). Statistical parametric maps in functional imaging: A general linear approach. Human Brain Mapping 2, 189-210.

[19] Genovese, C. R. (2000). A Bayesian time-course model for functional magnetic resonance imaging data (with comments). Journal of the American Statistical Association 95, 691-719.

[20] Glover, G. H. (1999). Deconvolution of impulse response in event-related BOLD fMRI. Neuroimage 9, 416-429.

[21] Goutte, C., Nielsen, F. A., and Hansen, L. K. (2000). Modeling the haemodynamic response in fMRI using smooth FIR filters. IEEE Transactions on Medical Imaging 19, 1188-1201.

[22] Handwerker DA, Ollinger JM, D'Esposito M. (2004). Variation of BOLD hemodynamic responses across subjects and brain regions and their effects on statistical analyses. Neuroimage 21, 1639-51.

[23] Härdle, W. (1990). Applied Nonparametric Regression. Cambridge University Press.

[24] Henson, R., Price, C. J., Rugg, M. D., Turner, R., and Friston, K. J. (2002). Detecting latency differences in event-related BOLD responses: application to words versus nonwords and initial versus repeated face presentations. Neuroimage 15, 83-97.

[25] Hotelling, H. (1931). The generalization of Student's ratio. Annals of Mathematical Statistics 2, 360-378.

[26] Jenkinson, M., Bannister, P., Brady, M., and Smith, S. (2002). Improved optimization for the robust and accurate linear registration and motion correction of brain images. Neuroimage 17, 825-41.

[27] Josephs, O., Turner, R., and Friston, K. J. (1997). Event-related fMRI. Human Brain Mapping 5, 243-48.

[28] Knutson, B., Fong, G. W., Adams, C. M., Varner, J. L., and Hommer, D. (2001). Dissociation of reward anticipation and outcome with event-related fMRI. Neuroreport 12(17), 3683-3687.

[29] Knutson, B., Westdorp, A., Kaiser, E., and Hommer, D. (2000). FMRI visualization of brain activity during a monetary incentive delay task. Neuroimage 12, 20-27.

[30] Lange, N., Strother, S. C., Anderson, J. R., Nielsen, F. A., Holmes, A. P., Kolenda, T., Savoy, R., and Hansen, L. K. (1999). Plurality and resemblance in fMRI data analysis. Neuroimage 10, 282-303.

[31] Liao, C. H., Worsley, K. J., Poline, J-B., Aston, J. A. D., Duncan, G. H., and Evans, A. C. (2002). Estimating the Delay of the fMRI response. Neuroimage 16, 593-606.

[32] Lindquist, M. A. (2008). The statistical analysis of fMRI data. Statistical Science 23, 439-464.

[33] Lindquist, M. A. and Wager, T. D. (2007). Validity and power in hemodynamic response modelling: a comparison study and a new approach. Human Brain Mapping 28, 764-784.

[34] Lindquist, M.A., Loh, J.M., Atlas, L. and Wager, T.D. (2009). Modeling the Hemodynamic Response Function in fMRI: Efficiency, Bias and Mis-modeling. Neuroimage 45, S187-S198.

[35] Luo, H. and Puthusserypady, S. (2008). Analysis of fMRI Data With Drift: Modified General Linear Model and Bayesian Estimator. IEEE Transcations on Biomedical Engeering 55, 1504-1511.

[36] Marrelec, G., Benali, H., Ciuciu, P., Pelegrini-Issac, M., Poline, J. B. (2003). Robust Estimation of the Hemodynamic Response Function in Event-Related BOLD fMRI Using Basic Physiological Information. Human Brain Mapping 19, 1-17.

[37] Marrelec, G., Benali, H., Ciuciu, P., and Poline, J. B. (2001). Bayesian estimation of the hemodynamic of the hemodynamic response function in functional MRI. AIP Conference Proceedings 617, 229-247.

[38] Mattay, V. S., Frank, J. A., Santha, A. K. S., Pekar, J. J., Duyn, J. H., McLaughlin, A. C., and Weinberger, D. R. (1996). Whole brain functional mapping with isotropic MR imaging. Radiology 201, 399-404.

[39] Ollinger, J. M., Corbetta, M., and Shulman, G. L. (2001). Separating processes within a trial in event-related functional MRI. Neuroimage 13 218-229.

[40] Riera, J. J., Watanabe, J., Kazuki, I., Naoki, M., Aubert, E., Ozaki, T., and Kawashima, R. (2004). A state-space model of the hemodynamic approach: Non-linear filtering of bold signals. Neuroimage 21, 547-567.

[41] Smith, A., Lewis, B., Ruttinmann, U., et al. (1999). Investigation of low frequency drift in fMRI signal. Neuroimage 9, 526-533.

[42] Smith. S.M. (2002). Fast robust automated brain extraction. Human Brain Mapping 17, 143-155.

[43] Smith, S.M., Jenkinson, M., Woolrich, M.W., Beckmann, C.F., Behrens, T.E.J., Johansen-Berg, H., Bannister, P.R., De Luca, M., Drobnjak, I., Flitney, D.E., Niazy, R., 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(S1), 208-219.

[44] Spielberger, C. D., and Vagg, P.R. (1984). Psychometric properties of the STAI: A reply to Ramanaiah, Franzen, and Schill. Journal of Personality Assessment 48, 95-97.

[45] Tikhonov, A. N., and Arsenin, V. Y. (1977). Solution of Ill-posed Problems. Washington: Winston & Sons.

[46] Vakorin, V. A., Borowsky, R., and Sarty, G. E. (2007). Characterizing the functional MRI response using Tikhonov regularization. Statistics in Medicine 26(21), 3830-3844.

[47] Wahba, G. (1990). Spline Models for Observational Data. Philadelphia: SIAM.

[48] Wang, J., Zhu, H., Fan, J.Q., Giovanello, K. and Lin, W. L. (2011). Multiscale Adaptive Smoothing Model for the Hemodynamic Response Function in fMRI. MICCAI, LNCS 6892, 269-276.

[49] Woolrich, M.W., Behrens, T.E., and Smith, S.M. (2004). Constrained linear basis sets for HRF modelling using Variational Bayes. Neuroimage 21, 1748-1761.

[50] Woolrich, M.W., Jbabdi, S., Patenaude, B., Chappell, M., Makni, S., Behrens, T., Beckmann, C., Jenkinson, M., and Smith, S.M. (2009). Bayesian analysis of neuroimaging data in FSL. Neuroimage 45, S173-186.

[51] Worsley, K. J. and Friston, K.J. (1995). Analysis of fMRI time-series revisited again. Neuroimage 2, 173-181.

[52] Worsley, K. J., Liao, C. H., Aston, J., Petre, V., Duncan, G. H., Morales, F., and Evans, A. (2002). A General Statistical Analysis for fMRI Data. Neuroimage 15, 1-15.

[53] Zarahn, E. (2002). Using larger dimensional signal subspaces to increase sensitivity in fMRI time series analyses. Human Brain Mapping 17, 13-16.

[54] Zhang, C. M., Jiang, Y. and Yu, T. (2007). A comparative study of one-level and two-level semiparametric estimation of hemodynamic response function for fMRI data. Statistics in Medicine 26, 3845-3861.