Supplementary Document of "Spatially Varying Coefficient Model for Neuroimaging Data with Jump Discontinuities"

Theoretical Results with Detailed Discussions

To investigate the asymptotic properties of $\hat{\beta}_j(\mathbf{d}_0; h_s)$, we need to characterize points close to and far from the boundary set $\partial\mathcal{D}^{(j)}$. For a given bandwidth h_s, we first define h_s-boundary sets:

$$\partial\mathcal{D}^{(j)}(h_s) = \{\mathbf{d} \in \mathcal{D} : B(\mathbf{d}; h_s) \cap \partial\mathcal{D}^{(j)} \neq \emptyset\} \text{ and } \partial\mathcal{D}_0^{(j)}(h_s) = \partial\mathcal{D}^{(j)}(h_s) \cap \mathcal{D}_0. \quad (1)$$

Thus, $\partial\mathcal{D}^{(j)}(h_s)$ can be regarded as a band with radius h_s covering the boundary set $\partial\mathcal{D}^{(j)}$, while $\partial\mathcal{D}_0^{(j)}(h_s)$ contains all grid points within such band. It is easy to show that for a sequence of bandwidths $h_0 = 0 < h_1 < \cdots < h_S$, we have

$$\partial\mathcal{D}^{(j)}(h_0) = \partial\mathcal{D}^{(j)} \subset \cdots \subset \partial\mathcal{D}^{(j)}(h_S) \quad \text{and} \quad \partial\mathcal{D}_0^{(j)}(h_0) \subset \cdots \subset \partial\mathcal{D}_0^{(j)}(h_S). \quad (2)$$

Therefore, for a fixed bandwidth h_s, any point $\mathbf{d} \in \mathcal{D}$ belongs to either $\mathcal{D} \setminus \partial\mathcal{D}^{(j)}(h_s)$ or $\partial\mathcal{D}^{(j)}(h_s)$. For each $\mathbf{d}_0 \in \mathcal{D} \setminus \partial\mathcal{D}^{(j)}(h_s)$, there exists one and only one $\mathcal{D}_{j,l}$ such that

$$B(\mathbf{d}_0; h_0) \subset \cdots \subset B(\mathbf{d}_0; h_s) \subset \mathcal{D}_{j,l}^o. \quad (3)$$

For any $\mathbf{d}_0 \in \partial\mathcal{D}^{(j)}(h_s)$, it follows from the local patch assumption that $B(\mathbf{d}_0, h_s) = P_j(\mathbf{d}_0, h_s) \cup P_j(\mathbf{d}_0, h_s)^c$ and $P_j(\mathbf{d}_0, h_s)^c$ contains all possible jump points. The performance of MASS strongly depends on $K_{st}(D_{\beta_j}(\mathbf{d}_0, \mathbf{d}_0'; h_{s-1})/C_n)$ and the degree of jumps as $\beta_{j*}(\mathbf{d}_0')$ varies in $P_j(\mathbf{d}_0, h_s)^c$ relative to $\beta_{j*}(\mathbf{d}_0)$. To have a better understanding of MASS, we examine the behavior of $K_{st}(D_{\beta_j}(\mathbf{d}_0, \mathbf{d}_0'; h_{s-1})/C_n)$ as $s = 1$. Let $\hat{\Delta}_j(\mathbf{d}_0) = \hat{\beta}_j(\mathbf{d}_0) - \beta_{j*}(\mathbf{d}_0)$ and $\Delta_{j*}(\mathbf{d}_0, \mathbf{d}_0') = \beta_{j*}(\mathbf{d}_0) - \beta_{j*}(\mathbf{d}_0')$. It follows from Theorem 1 that $D_{\beta_j}(\mathbf{d}_0, \mathbf{d}_0'; h_0)/C_n$ can be written as

$$\begin{aligned} D_{\beta_j}(\mathbf{d}_0, \mathbf{d}_0'; h_0)/C_n &= C_n^{-1} n\{\hat{\Delta}_j(\mathbf{d}_0) - \hat{\Delta}_j(\mathbf{d}_0') + \beta_{j*}(\mathbf{d}_0) - \beta_{j*}(\mathbf{d}_0')\}^2 / \Sigma_n(\sqrt{n}\hat{\beta}_j(\mathbf{d}_0)) \\ &= O_p(\{\sqrt{\log(1+N_D)/C_n} + \Delta_{j*}(\mathbf{d}_0, \mathbf{d}_0')\sqrt{n/C_n}\}^2). \end{aligned} \quad (4)$$

That is, $D_{\beta_j}(\mathbf{d}_0, \mathbf{d}_0'; h_0)/C_n$ is determined by the size of $\hat{\Delta}_j(\mathbf{d}_0) - \hat{\Delta}_j(\mathbf{d}_0')$ relative to $\Delta_{j*}(\mathbf{d}_0, \mathbf{d}_0')$. If $\log(1 + N_D) = o(C_n)$, $C_n = o(n)$, and $\lim_{u\to\infty} K_{st}(u) = 0$, then $K_{st}(D_{\beta_j}(\mathbf{d}_0, \mathbf{d}_0'; h_0)/C_n)$ converges to 0, when $\Delta_{j*}(\mathbf{d}_0, \mathbf{d}_0')\sqrt{n/C_n}$ diverges. Therefore,

if the jump $\Delta_{j*}(\mathbf{d}_0, \mathbf{d}_0')$ is an order larger than $\sqrt{C_n/n}$, the voxel $\mathbf{d}_0' \in P_j(\mathbf{d}_0, h_s)^c$ has a small impact on $\hat{\beta}_j(\mathbf{d}_0; h_1)$. We will show below that the above discussions are also valid even for $s > 1$.

Due to the discontinuity of $\beta_{j*}(\mathbf{d}_0)$ in $\partial\mathcal{D}^{(j)}(h_s)$, we need a better refinement (or decomposition) of $\mathcal{D}$ according to the value of $\beta_{j*}(\mathbf{d}_0)$. Specifically, for each $\mathbf{d} \in \mathcal{D}$ and $\delta_2 > \delta_1 \geq 0$, we define a (δ_1, δ_2)-neighborhood set of $\beta_{j*}(\mathbf{d})$ as follows:

$$I_j(\mathbf{d}, \delta_1, \delta_2) = \{\mathbf{d}' : \mathbf{d}' \in \mathcal{D}, \delta_1 \leq |\beta_{j*}(\mathbf{d}) - \beta_{j*}(\mathbf{d}')| < \delta_2\}. \tag{5}$$

If $\delta_1 = 0$ and $\delta_2 = \infty$, then $I_j(\mathbf{d}_0, \delta_1, \delta_2) = \mathcal{D}$. For $\mathbf{d}_0 \in \mathcal{D}_{j,l}^o$, since $\beta_{j*}(\mathbf{d}_0)$ is a smooth function in $\mathcal{D}_{j,l}^o$, there always exists a sufficiently small bandwidth $h > 0$ such that $B(\mathbf{d}_0, h) \subset I_j(\mathbf{d}_0, 0, \delta)$ for a given $\delta > 0$. Particularly, if $\beta_{j*}(\mathbf{d}_0)$ is constant in $\mathcal{D}_{j,l}^o$, then $\mathcal{D}_{j,l}^o \subset I_j(\mathbf{d}_0, 0, \delta)$ for any $\delta > 0$.

To further delineate the structure of $P_j(\mathbf{d}_0, h_s)^c$, we introduce a lower threshold and an upper threshold, which are denoted by δ_L and δ_U, respectively. For any $0 \leq \delta_L < \delta_U$, $P_j(\mathbf{d}_0, h_s)^c$ is a union of three sets including $P_j(\mathbf{d}_0, h_s)^c \cap I_j(\mathbf{d}_0, 0, \delta_L)$, $P_j(\mathbf{d}_0, h_s)^c \cap I_j(\mathbf{d}_0, \delta_L, \delta_U)$, and $P_j(\mathbf{d}_0, h_s)^c \cap I_j(\mathbf{d}_0, \delta_U, \infty)$. We consider $\delta_L = O(\sqrt{\log(1 + N_D)/n}) = o(1)$ and $\delta_U = \sqrt{C_n/n} M_n = o(1)$, in which $\lim_{n\to\infty} M_n = \infty$. For any $\mathbf{d}_0' \in I_j(\mathbf{d}_0, 0, \delta_L)$ and $\mathbf{d}_0'' \in I_j(\mathbf{d}_0, \delta_U, \infty)$, it follows from (4) that

$$\begin{aligned} K_{st}(D_{\beta_j}(\mathbf{d}_0, \mathbf{d}_0'; h_s)/C_n) &= K_{st}(O_p(\log(1 + N_D)/C_n)) \approx K_{st}(0) > 0, \\ K_{st}(D_{\beta_j}(\mathbf{d}_0, \mathbf{d}_0''; h_s)/C_n) &= K_{st}(O_p(M_n^2)) \approx K_{st}(\infty) = 0. \end{aligned}$$

For any $\mathbf{d}_0''' \in I_j(\mathbf{d}_0, \delta_L, \delta_U)$, we have $K_{st}(D_{\beta_j}(\mathbf{d}_0, \mathbf{d}_0'''; h_s)/C_n) \in [0, K_{st}(0)]$. Generally, MASS discards almost all information contained in voxels in $P_j(\mathbf{d}_0, h_s)^c \cap I_j(\mathbf{d}_0, \delta_U, \infty)$, whereas it incorporates almost all information contained in voxels in $P_j(\mathbf{d}_0, h_s)^c \cap I_j(\mathbf{d}_0, 0, \delta_L)$ and partial information contained in voxels in $P_j(\mathbf{d}_0, h_s)^c \cap I_j(\mathbf{d}_0, \delta_L, \delta_U)$. For voxels in $P_j(\mathbf{d}_0, h_s)^c \cap I_j(\mathbf{d}_0, 0, \delta_U)$, MASS has difficulty in preventing biases in estimating $\beta_{j*}(\mathbf{d}_0)$. In practice, δ_U can be regarded as the *sensitivity* (or *capability*) of MASS to respond to jumps in $\partial\mathcal{D}^{(j)}$.

We first investigate the asymptotic behavior of $\hat{\beta}_j(\mathbf{d}_0; h_s)$ when $\beta_{j*}(\mathbf{d}_0)$ is piecewise constant. Let $\Delta_{j*}(\mathbf{d}_0, \mathbf{d}_0') = \beta_{j*}(\mathbf{d}_0) - \beta_{j*}(\mathbf{d}_0')$ and $\tilde{\beta}_{j*}(\mathbf{d}_0; h_s) = \sum_{\mathbf{d}_m \in B(\mathbf{d}_0, h_s)} \tilde{\omega}_j(\mathbf{d}_0, \mathbf{d}_m; h_s)\beta_{j*}(\mathbf{d}_m)$ be the pseudo-true value of $\beta_j(\mathbf{d}_0)$ at scale h_s in voxel $\mathbf{d}_0$. For all $\mathbf{d}_0 \in \mathcal{D} \setminus \partial\mathcal{D}^{(j)}(h_S)$, we have $\tilde{\beta}_{j*}(\mathbf{d}_0; h_s) = \cdots = \tilde{\beta}_{j*}(\mathbf{d}_0; h_0) = \beta_{j*}(\mathbf{d}_0)$ due to (3). In contrast, for $\mathbf{d}_0 \in \partial\mathcal{D}^{(j)}(h_S)$, let $\boldsymbol{u}^{(j)}(h_s) = \min_{(\mathbf{d}_0, \mathbf{d}_0'): \Delta_{j*}(\mathbf{d}_0, \mathbf{d}_0') \neq 0, \mathbf{d}_0' \in B(\mathbf{d}_0, h_s)} |\Delta_{j*}(\mathbf{d}_0, \mathbf{d}_0')|$ be the smallest absolute value of all possible jumps at scale h_s. In this case, $\tilde{\beta}_{j*}(\mathbf{d}_0; h_s)$ may vary from h_0 to h_S. However, we can show below that $\tilde{\Delta}_{j*}(\mathbf{d}_0; h_s) = \tilde{\beta}_{j*}(\mathbf{d}_0; h_s) - \beta_{j*}(\mathbf{d}_0) = o_p(\sqrt{\log(1 + N_D)/n})$ under some mild conditions on h_s and $\boldsymbol{u}^{(j)}(h_s)$, which will be detailed below. A remarkable property of MASS is that h_S is not required to converge to zero when $\boldsymbol{u}^{(j)}(h_s)$ is relatively large and $K_{st}(t)$ satisfies certain tail property.

For a fixed $S > 0$ and piecewise constant $\beta_{j*}(\mathbf{d}_0)$, we can establish several important theoretical results to characterize the asymptotic behavior of $\hat{\boldsymbol{\beta}}(\mathbf{d}_0; h_s)$. We need to introduce some additional notation as follows:

$$
\begin{aligned}
&\omega_j^{(0)}(\mathbf{d}_0, \mathbf{d}_0'; h_s) = K_{loc}(||\mathbf{d}_0 - \mathbf{d}_0'||_2/h_s)K_{st}(0)\mathbf{1}(\Delta_{j*}(\mathbf{d}_0, \mathbf{d}_0') = 0), \qquad (6)\\
&\omega_j^{(1)}(\mathbf{d}_0, \mathbf{d}_0'; h_s) = K_{loc}(||\mathbf{d}_0 - \mathbf{d}_0'||_2/h_s)K_{st}(0)\mathbf{1}(\mathbf{d}_0' \in P_j(\mathbf{d}_0, h_s) \cup I_j(\mathbf{d}_0, 0, \delta_L)),\\
&\tilde{\omega}_j^{(k)}(\mathbf{d}_0, \mathbf{d}_0'; h_s) = \omega_j^{(k)}(\mathbf{d}_0, \mathbf{d}_0'; h_s) / \sum_{\mathbf{d}_m \in B(\mathbf{d}_0, h_s) \cap \mathcal{D}_0} \omega_j^{(k)}(\mathbf{d}_0, \mathbf{d}_m; h_s),\\
&\hat{\Sigma}^{(k)}(\sqrt{n}\hat{\beta}_j(\mathbf{d}_0; h_s)) = \mathbf{e}_{j,p}^T \Omega_{X,n}^{-1} \mathbf{e}_{j,p} \sum_{\mathbf{d}_m, \mathbf{d}_m' \in B(\mathbf{d}_0, h_s) \cap \mathcal{D}_0} \tilde{\omega}_j^{(k)}(\mathbf{d}_0, \mathbf{d}_m; h_s)\tilde{\omega}_j^{(k)}(\mathbf{d}_0, \mathbf{d}_m'; h_s)\hat{\Sigma}_y(\mathbf{d}_m, \mathbf{d}_m'),\\
&\Sigma_j^{(k)}(\mathbf{d}_0; h_s) = \mathbf{e}_{j,p}^T \Omega_X^{-1} \mathbf{e}_{j,p} \sum_{\mathbf{d}_m, \mathbf{d}_m' \in B(\mathbf{d}_0, h_s) \cap \mathcal{D}_0} \omega_j^{(k)}(\mathbf{d}_0, \mathbf{d}_m; h_s)\omega_j^{(k)}(\mathbf{d}_0, \mathbf{d}_m'; h_s)\Sigma_y(\mathbf{d}_m, \mathbf{d}_m').
\end{aligned}
$$

Theorem 3. *Under assumptions (C1)-(C10) in Section 6 for piecewise constant* $\{\beta_{j*}(\mathbf{d}) : d \in \mathcal{D}\}$*, we have the following results for all* $0 \leq s \leq S$*:*

(i) $\sup_{d \in \mathcal{D}_0} |\tilde{\Delta}_{j*}(\mathbf{d}_0; h_s)| = o_p(\sqrt{\log(1 + N_D)/n})$*, where* $\tilde{\Delta}_{j*}(\mathbf{d}_0; h_s) = \tilde{\beta}_{j*}(\mathbf{d}_0; h_s) - \beta_{j*}(\mathbf{d}_0)$*;*

(ii) $\hat{\beta}_j(\mathbf{d}_0; h_s) - \beta_{j*}(\mathbf{d}_0) = \sum_{\mathbf{d}_m \in B(\mathbf{d}_0, h_s) \cap \mathcal{D}_0} \tilde{\omega}_j^{(0)}(\mathbf{d}_0, \mathbf{d}_m; h_s)\hat{\Delta}_j(\mathbf{d}_m)[1 + o_p(1)]$*;*

(iii) $\sup_{\mathbf{d}_0 \in \mathcal{D}_0} |\hat{\Sigma}(\sqrt{n}\tilde{\beta}_{j*}(\mathbf{d}_0; h_s)) - \Sigma_j^{(0)}(\mathbf{d}_0; h_s)| = o_p(1)$*;*

(iv) $\sqrt{n}\{\hat{\beta}_j(\mathbf{d}_0; h_s) - \beta_{j*}(\mathbf{d}_0)\}$ *converges in distribution to a normal distribution with mean zero and variance* $\Sigma_j^{(0)}(\mathbf{d}_0; h_s)$ *as* $n \to \infty$*.*

We now consider a much complex scenario when $\beta_{j*}(\mathbf{d}_0)$ is piecewise smooth. In

this case, $\tilde{\beta}_{j*}(\mathbf{d}_0; h_s)$ may vary from h_0 to h_S for all voxels $\mathbf{d}_0 \in \mathcal{D}$ regardless whether $\mathbf{d}_0$ belongs to $\partial\mathcal{D}^{(j)}(h_s)$ or not. In this case, if $\beta_{j*}(\mathbf{d}_0)$ is Lipschitz continuous for each piece, it will be shown below that the bias of $\tilde{\beta}_{j*}(\mathbf{d}_0; h_s)$ is always at the order of h_s for $\mathbf{d}_0 \in \mathcal{D} \setminus \partial\mathcal{D}^{(j)}(h_s)$. However, for $\mathbf{d}_0 \in \partial\mathcal{D}^{(j)}(h_s)$, only when $P_j(\mathbf{d}_0, h_s)^c \cap I_j(\mathbf{d}_0, \delta_L, \delta_U)$ is an empty set, we can control the bias of $\tilde{\beta}_{j*}(\mathbf{d}_0; h_s)$ to be at the order of h_s. Therefore, to control the bias of $\tilde{\beta}_{j*}(\mathbf{d}_0; h_s)$ across all voxels, h_s must converge to zero. Moreover, as shown below, we can only establish the asymptotic normality of $\hat{\beta}_j(\mathbf{d}_0; h_s)$ relative to $\tilde{\beta}_{j*}(\mathbf{d}_0; h_s)$, not $\beta_{j*}(\mathbf{d}_0)$. These results differ significantly from those for piecewise smooth $\beta_{j*}(\mathbf{d}_0)$. Generally, for a fixed $S > 0$, we can establish important theoretical results to characterize the asymptotic behavior of $\hat{\boldsymbol{\beta}}(\mathbf{d}_0; h_s)$ as follows.

Theorem 4. *Suppose assumptions (C1)-(C9) and (C11) in Section 6 hold for piecewise continuous* $\{\beta_{j*}(\mathbf{d}) : d \in \mathcal{D}\}$. *For all* $0 \le s \le S$, *we have the following results:*

(i) $\sup_{\mathbf{d}_0 \in \mathcal{D}_0} |\tilde{\Delta}_{j*}(\mathbf{d}_0; h_s)| = O_p(h_s)$;

(ii) $\hat{\beta}_j(\mathbf{d}_0; h_s) - \tilde{\beta}_{j*}(\mathbf{d}_0; h_s) = \sum_{\mathbf{d}_m \in B(\mathbf{d}_0, h_s) \cap \mathcal{D}_0} \tilde{\omega}_j^{(1)}(\mathbf{d}_0, \mathbf{d}_m; h_s) \hat{\Delta}_j(\mathbf{d}_m)[1 + o_p(1)]$;

(iii) $\sup_{\mathbf{d}_0 \in \mathcal{D}_0} |\hat{\Sigma}(\sqrt{n}\tilde{\beta}_{j*}(\mathbf{d}_0; h_s)) - \Sigma_j^{(1)}(\mathbf{d}_0; h_s)| = o_p(1)$.

(iv) $\sqrt{n}\{\hat{\beta}_j(\mathbf{d}_0; h_s) - \tilde{\beta}_{j*}(\mathbf{d}_0; h_s)\}$ *converges in distribution to a normal distribution with mean zero and variance* $\Sigma_j^{(1)}(\mathbf{d}_0; h_s)$ *as* $n \to \infty$.

Theorem 4 characterizes several key features of MASS for a piecewise continuous function $\beta_{j*}(\mathbf{d}_0)$. Theorem 4 (i) quantifies the bias of the pseudo true value $\tilde{\beta}_{j*}(\mathbf{d}_0; h_s)$ relative to the true value $\beta_{j*}(\mathbf{d}_0)$ across all $\mathbf{d}_0 \in \mathcal{D}_0$ for a fixed s. Even for voxels inside the smooth areas of $\beta_{j*}(\mathbf{d}_0)$, the bias $O_p(h_s)$ is still much higher than the standard bias at the rate of h_s^2 due to the presence of $K_{st}(D_{\beta_j}(\mathbf{d}_0, \mathbf{d}_0'; h_{s-1})/C_n)$. If we set $K_{st}(u) = \mathbf{1}(u \in [0, 1])$ and $\beta_{j*}(\mathbf{d}_0)$ is twice differentiable, then the bias of $\tilde{\beta}_{j*}(\mathbf{d}_0; h_s)$ relative to $\beta_{j*}(\mathbf{d}_0)$ may be reduced to be close to $O_p(h_s^2)$. Theorem 4 (ii) establishes the asymptotic equivalence between $\hat{\beta}_j(\mathbf{d}_0; h_s) - \tilde{\beta}_{j*}(\mathbf{d}_0; h_s)$ and $\sum_{\mathbf{d}_m \in B(\mathbf{d}_0, h_s) \cap \mathcal{D}_0} \tilde{\omega}_j^{(1)}(\mathbf{d}, \mathbf{d}_m; h_s) \hat{\Delta}_j(\mathbf{d}_m)$. Theorem 4 (iii) ensures that $\hat{\Sigma}(\sqrt{n}\tilde{\beta}_{j*}(\mathbf{d}_0; h_s))$ is a uniform consistent estimator of $\Sigma_j^{(1)}(\mathbf{d}_0; h_s)$ across $\mathbf{d}_0 \in \mathcal{D}_0$. Theorem 4 (iv) ensures that $\sqrt{n}\{\hat{\beta}_j(\mathbf{d}_0; h_s) - \tilde{\beta}_{j*}(\mathbf{d}_0; h_s)\}$ is asymptotically normally distributed.

Finally, we delineate the asymptotic variance of $\hat{\beta}_j(\mathbf{d}_0; h_s)$. For simplicity, we focus on $\mathbf{d}_0 \in \mathcal{D} \setminus \partial\mathcal{D}^{(j)}(h_s)$ and do not distinguish the piecewise constant case and the piecewise

continuous one. Let $\tilde{K}_{loc}(||\mathbf{d}_0-\mathbf{d}_m||_2/h) = K_{loc}(||\mathbf{d}_0-\mathbf{d}_m||_2/h)/[\sum_{\mathbf{d}'_m\in B(\mathbf{d}_0;h)\cap\mathcal{D}_0} K_{loc}(||\mathbf{d}_0-\mathbf{d}'_m||_2/h)]$. It follows from (30) that for $k=0$ and 1, $\Sigma_j^{(k)}(\mathbf{d}_0;h_s)/\mathbf{e}_{j,p}^T\Omega_X^{-1}\mathbf{e}_{j,p}$ equals the sum of terms (T1) and (T2), which are, respectively, given by

$$\begin{aligned}
\text{(T1)} &= \sum_{\mathbf{d}_m,\mathbf{d}'_m\in B(\mathbf{d}_0;h_s)\cap\mathcal{D}_0} \tilde{K}_{loc}(||\mathbf{d}_0-\mathbf{d}_m||^2/h_s)\tilde{K}_{loc}(||\mathbf{d}_0-\mathbf{d}'_m||^2/h_s)\Sigma_\eta(\mathbf{d}_m,\mathbf{d}'_m) \\
&= \sum_{l=1}^{\infty}\lambda_l[\sum_{\mathbf{d}_m\in B(\mathbf{d}_0;h_s)\cap\mathcal{D}_0} \tilde{K}_{loc}(||\mathbf{d}_0-\mathbf{d}_m||^2/h_s)\psi_l(\mathbf{d}_m)]^2, \qquad (7) \\
\text{(T2)} &= \sum_{\mathbf{d}_m\in B(\mathbf{d}_0;h_s)\cap\mathcal{D}_0} \tilde{K}_{loc}(||\mathbf{d}_0-\mathbf{d}_m||^2/h_s)^2\Sigma_\epsilon(\mathbf{d}_m,\mathbf{d}_m).
\end{aligned}$$

If $h_s\to 0$ and $N_D h_s^{3/2}\to\infty$, it can be shown that (T1) and (T2), respectively, converge to $\Sigma_\eta(\mathbf{d}_0,\mathbf{d}_0)$ and 0. Thus, both $\Sigma_j^{(0)}(\mathbf{d}_0;h_s)$ and $\Sigma_j^{(1)}(\mathbf{d}_0;h_s)$ converge to $\mathbf{e}_{j,p}^T\Omega_X^{-1}\mathbf{e}_{j,p}\Sigma_\eta(\mathbf{d}_0,\mathbf{d}_0)$, which is smaller than the asymptotic variance of the raw estimate $\hat{\beta}_j(\mathbf{d}_0)$. In general, for relatively small h_s, MASS leads to smaller standard deviations for estimating $\beta_j(\mathbf{d}_0)$.

Proofs

Proof of Theorem 1. The proof of Theorem 1 (i) can be easily proved by using the standard asymptotic arguments (van der Vaar and Wellner, 1996), so we omit repeating them here. To prove Theorem 1(ii), we will show

$$\sup_{\mathbf{d}_0\in\mathcal{D}_0}||\hat{\boldsymbol{\beta}}(\mathbf{d}_0)-\boldsymbol{\beta}_*(\mathbf{d}_0)||_2 = O_p(n^{-1/2}\sqrt{\log(1+N_D)}). \qquad (8)$$

It is easy to show that

$$\begin{aligned}
\hat{\boldsymbol{\beta}}(\mathbf{d}_0) &= (\sum_{i=1}^n \mathbf{x}_i^{\otimes 2})^{-1}\sum_{i=1}^n \mathbf{x}_i y_i(\mathbf{d}_0) = \boldsymbol{\beta}_*(\mathbf{d}_0)+A_{n,\eta}(\mathbf{d}_0)+A_{n,\epsilon}(\mathbf{d}_0) \\
&= \boldsymbol{\beta}_*(\mathbf{d}_0)+(\sum_{i=1}^n \mathbf{x}_i^{\otimes 2})^{-1}\sum_{i=1}^n \mathbf{x}_i\eta_i(\mathbf{d}_0)+(\sum_{i=1}^n \mathbf{x}_i^{\otimes 2})^{-1}\sum_{i=1}^n \mathbf{x}_i\epsilon_i(\mathbf{d}_0). \qquad (9)
\end{aligned}$$

It follows from the law of the large number and assumption (C3) that $n^{-1}\sum_{i=1}^n \mathbf{x}_i^{\otimes 2}$ converges to Ω_X almost surely. It follows from assumption (C4) that $\{\mathbf{x}\eta(\mathbf{d}):\mathbf{d}\in\mathcal{D}\}$ is

a Donsker class and $\sum_{i=1}^{n} \mathbf{x}_i \eta_i(\mathbf{d})/\sqrt{n}$ converges to a Gaussian process with zero mean and covariance function $\Omega_X \Sigma_\eta(\mathbf{d}, \mathbf{d}')$ as $n \to \infty$ (van der Vaar and Wellner, 1996). Thus, we have

$$\sup_{\mathbf{d} \in \mathcal{D}} |\sum_{i=1}^{n} \mathbf{x}_i \eta_i(\mathbf{d})| = O_p(\sqrt{n}). \tag{10}$$

It follows from assumptions (C2) and (C3) that

$$P(|\sum_{i=1}^{n} x_{ij} \epsilon_i(\mathbf{d}_0)| > t) \leq C \exp(-\frac{Ct^2}{n||\mathbf{x}||_\infty^2 \max_{\mathbf{d}_0 \in \mathcal{D}_0} ||\epsilon_i(\mathbf{d}_0)||_{\psi_2}^2}),$$

where C is a generic constant and $||\cdot||_{\psi_l}$ denotes the Orlicz norm for $\psi_l(x) = \exp(x^l) - 1$. Then, we can apply Lemma 2.2.10 in van der Vaar and Wellner (1996) to get

$$|| \max_{\mathbf{d}_0 \in \mathcal{D}_0} |\sum_{i=1}^{n} x_{ij} \epsilon_i(\mathbf{d}_0)| ||_{\psi_1} \leq C\{\sqrt{n}(||\mathbf{x}||_\infty C)\sqrt{\log(1 + N_D)}\}.$$

Finally, we get

$$\max_{\mathbf{d}_0 \in \mathcal{D}_0} |\sum_{i=1}^{n} x_{ij} \epsilon_i(\mathbf{d}_0)| = O_p(\sqrt{n \log(1 + N_D)}). \tag{11}$$

By combining (9)-(11), we can finish the proof of (8).

We define some notation as follows:

$$\begin{aligned}
&\Delta_{j*}(\mathbf{d}, \mathbf{d}') = \beta_{j*}(\mathbf{d}) - \beta_{j*}(\mathbf{d}'), \\
&I_j(\mathbf{d}, \delta_1, \delta_2) = \{\mathbf{d}' : \mathbf{d}' \in \mathcal{D}, \delta_1 \leq |\Delta_{j*}(\mathbf{d}, \mathbf{d}')| < \delta_2\} \quad \text{for} \quad j = 1, \ldots, p, \\
&\tilde{K}_h^0(\mathbf{d}_m, \mathbf{d}) = (1, 0, 0, 0)\{\sum_{\mathbf{d}_m} K_h(\mathbf{d}_m - \mathbf{d}) \mathbf{z}_h(\mathbf{d}_m - \mathbf{d})^{\otimes 2}\}^{-1} K_h(\mathbf{d}_m - \mathbf{d}) \mathbf{z}_h(\mathbf{d}_m - \mathbf{d}), \\
&\hat{\eta}_i(\mathbf{d}) = (1, 0, 0, 0)\hat{C}_i(\mathbf{d}) = \sum_{\mathbf{d}_m} \tilde{K}_h^0(\mathbf{d}_m; \mathbf{d})\{y_i(\mathbf{d}_m) - \mathbf{x}_i^T \hat{\boldsymbol{\beta}}(\mathbf{d}_m)\}, \\
&\overline{\epsilon}_i(\mathbf{d}) = \sum_{\mathbf{d}_m} \tilde{K}_h^0(\mathbf{d}_m, \mathbf{d}) \epsilon_i(\mathbf{d}_m), \qquad \Delta\eta_i(\mathbf{d}) = \sum_{\mathbf{d}_m} \tilde{K}_h^0(\mathbf{d}_m, \mathbf{d})[\eta_i(\mathbf{d}_m) - \eta_i(\mathbf{d})], \\
&\Delta_i(\mathbf{d}) = \overline{\epsilon}_i(\mathbf{d}) + \Delta\eta_i(\mathbf{d}) + \mathbf{x}_i^T \Delta\boldsymbol{\beta}(\mathbf{d}), \quad \Delta\boldsymbol{\beta}(\mathbf{d}) = \sum_{\mathbf{d}_m} \tilde{K}_h^0(\mathbf{d}_m, \mathbf{d})[\boldsymbol{\beta}_*(\mathbf{d}_m) - \hat{\boldsymbol{\beta}}(\mathbf{d}_m)],
\end{aligned}$$

where $\delta_2 > 0$ and $\delta_1 \geq 0$ are non-negative scalars. Moreover, $\Pi_{N_D}(\cdot)$ is the sampling distribution function based on $\mathcal{D}_0$, and $\Pi(\cdot)$ is the distribution function of $\mathbf{d}$. We need the following lemmas to prove Theorem 2.

Lemma 1. Under Assumptions (C1)-(C7), we have the following results:

$$\sup_{\mathbf{d}\in\mathcal{D}}\left|\int K_h(\mathbf{u}-\mathbf{d})\frac{\prod_{k=1}^3(u_k-d_k)^r}{h^{3r}}d[\Pi_{N_D}(\mathbf{u})-\Pi(\mathbf{u})]\right|$$
$$= O_p((N_Dh^3)^{-1/2}\max(3|\log h|,\log\log N_D)^{1/2}), \tag{12}$$
$$\sup_{\mathbf{d}\in\mathcal{D}} n^{-1}|\sum_{i=1}^n \overline{\epsilon}_i(\mathbf{d})\mathbf{x}_i| = o_p(n^{-1/2}), \tag{13}$$
$$\sup_{(\mathbf{d},\mathbf{d}')\in\mathcal{D}^2} n^{-1}|\sum_{i=1}^n \overline{\epsilon}_i(\mathbf{d})\Delta\eta_i(\mathbf{d}')| = O_p(n^{-1/2}(\log n)^{1/2}), \tag{14}$$
$$\sup_{(\mathbf{d},\mathbf{d}')\in\mathcal{D}^2} n^{-1}|\sum_{i=1}^n \overline{\epsilon}_i(\mathbf{d})\overline{\epsilon}_i(\mathbf{d}')| = O_p((N_Dh^3)^{-1}+(\log n/n)^{1/2}). \tag{15}$$

Proof of Lemma 1. Equation (12) follows directly from Theorem 1 of Einmahl and Mason (2000). It follows from (12) that for large enough N_D, there exists a constant $C_1>1$ such that

$$\sup_{\mathbf{d}\in\mathcal{D}} n^{-1}|\sum_{i=1}^n \overline{\epsilon}_i(\mathbf{d})\mathbf{x}_i| \le n^{-1/2}C_1\sup_{\mathbf{d}\in\mathcal{D}}|N_D^{-1}\pi(\mathbf{d})^{-1}\sum_{m=1}^{N_D}K_h(\mathbf{d}_m-\mathbf{d})F_n(\mathbf{d}_m)|,$$

where $F_n(\mathbf{d}_m) = n^{-1/2}\sum_{i=1}^n \mathbf{x}_i\epsilon_i(\mathbf{d}_m)$. By following the arguments in Einmahl and Mason (2000), we can show that

$$\sup_{\mathbf{d}\in\mathcal{D}}\pi(\mathbf{d})^{-1}|N_D^{-1}\sum_{m=1}^{N_D}K_h(\mathbf{d}_m-\mathbf{d})F_n(\mathbf{d}_m)| = O_p((N_Dh^3)^{-1/2}|\log h|^{1/2}) = o_p(1),$$

which yields (13).

By following Lemmas 1-4 of Li and Hsing (2010), we can prove (14) and (15). Let's consider (15) as an illustration. We define $\Delta_{n,\epsilon\epsilon}(\mathbf{d},\mathbf{d}') = \sum_{i=1}^n \overline{\epsilon}_i(\mathbf{d})\overline{\epsilon}_i(\mathbf{d}')$ and

$$\begin{aligned}\Delta^{(1)}_{n,\epsilon\epsilon}(\mathbf{d},\mathbf{d}') &= \Delta^{(1,1)}_{n,\epsilon\epsilon}(\mathbf{d},\mathbf{d}')+\Delta^{(1,2)}_{n,\epsilon\epsilon}(\mathbf{d},\mathbf{d}')\\ &= n^{-1}\sum_{i=1}^n\frac{1}{N_D^2\pi(\mathbf{d})\pi(\mathbf{d}')}\sum_{m=1}^{N_D}K_h(\mathbf{d}_m-\mathbf{d})K_h(\mathbf{d}_m-\mathbf{d}')[\epsilon_i(\mathbf{d}_m)^2-\Sigma_\epsilon(\mathbf{d}_m,\mathbf{d}_m)]\\ &+ n^{-1}\sum_{i=1}^n\frac{1}{N_D^2\pi(\mathbf{d})\pi(\mathbf{d}')}\sum_{m=1}^{N_D}K_h(\mathbf{d}_m-\mathbf{d})K_h(\mathbf{d}_m-\mathbf{d}')\Sigma_\epsilon(\mathbf{d}_m,\mathbf{d}_m),\end{aligned}$$

$$\Delta^{(2)}_{n,\epsilon\epsilon}(\mathbf{d},\mathbf{d}') = n^{-1}\sum_{i=1}^n\frac{1}{N_D^2\pi(\mathbf{d})\pi(\mathbf{d}')}\sum_{m\neq m'}K_h(\mathbf{d}_m-\mathbf{d})K_h(\mathbf{d}_{m'}-d')\epsilon_i(\mathbf{d}_m)\epsilon_i(\mathbf{d}_{m'}).$$

For large enough N_D, there exists a constant $C_1 > 1$ such that

$$\begin{aligned}
&\sup_{(\mathbf{d},\mathbf{d}')\in\mathcal{D}^2} n^{-1}|\Delta_{n,\epsilon\epsilon}(\mathbf{d},\mathbf{d}')| \\
&\le C_1 \sup_{(\mathbf{d},\mathbf{d}')\in\mathcal{D}^2} \left| n^{-1}\sum_{i=1}^{n} \frac{1}{N_D^2\pi(\mathbf{d})\pi(\mathbf{d}')} \sum_{m,m'=1}^{N_D} K_h(\mathbf{d}_m-\mathbf{d})K_h(\mathbf{d}_{m'}-d)\epsilon_i(\mathbf{d}_m)\epsilon_i(\mathbf{d}_{m'}) \right| \\
&\le C_1\{ \sup_{(\mathbf{d},\mathbf{d}')\in\mathcal{D}^2} \left|\Delta^{(1)}_{n,\epsilon\epsilon}(\mathbf{d},\mathbf{d}')\right| + \sup_{(\mathbf{d},\mathbf{d}')\in\mathcal{D}^2} \left|\Delta^{(2)}_{n,\epsilon\epsilon}(\mathbf{d},\mathbf{d}')\right|\}.
\end{aligned}$$

Similar to the arguments in Lemmas 3 and 4 of Li and Hsing (2010), we have

$$\sup_{(\mathbf{d},\mathbf{d}')\in\mathcal{D}^2} \left|\Delta^{(2)}_{n,\epsilon\epsilon}(\mathbf{d},\mathbf{d}')\right| = O(\sqrt{\log n/n}) \ \text{ a.s.}$$

Thus, we only need to consider $\sup_{(\mathbf{d},\mathbf{d}')\in\mathcal{D}^2} \left|\Delta^{(1)}_{n,\epsilon\epsilon}(\mathbf{d},\mathbf{d}')\right|$. Similar to the arguments in Lemmas 1 and 2 of Li and Hsing (2010), we can obtain

$$\sup_{(\mathbf{d},\mathbf{d}')\in\mathcal{D}^2} \left|\Delta^{(1,1)}_{n,\epsilon\epsilon}(\mathbf{d},\mathbf{d}')\right| = O_p((N_Dh^3)^{-1}(\log n/n)^{1/2}), \quad \sup_{(\mathbf{d},\mathbf{d}')\in\mathcal{D}^2} \left|\Delta^{(1,2)}_{n,\epsilon\epsilon}(\mathbf{d},\mathbf{d}')\right| = O_p((N_Dh^3)^{-1}),$$

which yield (15). This completes the proof of Lemma 1.

Lemma 2. Under Assumptions (C1)-(C7), we have the following results:

$$\sup_{\mathbf{d}_0\in\mathcal{D}_0} n^{-1}|\sum_{i=1}^{n} \mathbf{x}_i^T\Delta\boldsymbol{\beta}(\mathbf{d}_0)\epsilon_i(\mathbf{d}_0)| = O_p(n^{-1}\log(1+N_D)), \tag{16}$$

$$\sup_{\mathbf{d}_0\in\mathcal{D}_0} n^{-1}|\sum_{i=1}^{n} \Delta\eta_i(\mathbf{d}_0)\epsilon_i(\mathbf{d}_0)| = O_p(\sqrt{h^2+n^{-1/2}}), \tag{17}$$

$$\sup_{\mathbf{d}_0\in\mathcal{D}_0} n^{-1}|\sum_{i=1}^{n} \overline{\epsilon}_i(\mathbf{d}_0)\epsilon_i(\mathbf{d}_0)| = O_p(\sqrt{(N_Dh^3)^{-1}+(\log n/n)^{1/2}}). \tag{18}$$

Proof of Lemma 2. It follows from Theorem 1 and the Cauchy-Schwarz inequality that

$$\sup_{\mathbf{d}_0\in\mathcal{D}_0} n^{-1}|\sum_{i=1}^{n} \mathbf{x}_i^T\Delta\boldsymbol{\beta}(\mathbf{d}_0)\epsilon_i(\mathbf{d}_0)| \le \sup_{\mathbf{d}_0\in\mathcal{D}_0} n^{-1}||\sum_{i=1}^{n} \mathbf{x}_i\epsilon_i(\mathbf{d}_0)||_2||\Delta\boldsymbol{\beta}(\mathbf{d}_0)||_2 = O_p(n^{-1}\log(1+N_D)).$$

It follows from the Cauchy-Schwarz inequality that

$$\begin{aligned}
\sup_{\mathbf{d}_0\in\mathcal{D}_0} \{n^{-1}|\sum_{i=1}^{n} \Delta\eta_i(\mathbf{d}_0)\epsilon_i(\mathbf{d}_0)|\}^2 &\le \sup_{\mathbf{d}_0\in\mathcal{D}_0} \{n^{-1}\sum_{i=1}^{n} \Delta\eta_i(\mathbf{d}_0)^2\} \sup_{\mathbf{d}_0\in\mathcal{D}_0} \{n^{-1}\sum_{i=1}^{n} \epsilon_i(\mathbf{d}_0)^2\}, \\
\sup_{\mathbf{d}_0\in\mathcal{D}_0} \{n^{-1}|\sum_{i=1}^{n} \overline{\epsilon}_i(\mathbf{d}_0)\epsilon_i(\mathbf{d}_0)|\}^2 &\le \sup_{\mathbf{d}_0\in\mathcal{D}_0} \{n^{-1}\sum_{i=1}^{n} \overline{\epsilon}_i(\mathbf{d}_0)^2\} \sup_{\mathbf{d}_0\in\mathcal{D}_0} \{n^{-1}\sum_{i=1}^{n} \epsilon_i(\mathbf{d}_0)^2\}.
\end{aligned}$$

Let $\Delta^{(2)}_{\Sigma,\eta}(\mathbf{d}_0,\mathbf{d}_m,\mathbf{d}'_m)=\Sigma_\eta(\mathbf{d}_0,\mathbf{d}'_m)+\Sigma_\eta(\mathbf{d}_m,\mathbf{d}_0)-\Sigma_\eta(\mathbf{d}_m,\mathbf{d}'_m)-\Sigma_\eta(\mathbf{d}_0,\mathbf{d}_0)$. Based on assumption (C4), we have

$$\sup_{\mathbf{d}_0\in\mathcal{D}_0}|n^{-1}\sum_{i=1}^n\Delta\eta_i(\mathbf{d}_0)^2|\le O_p(h^2+n^{-1/2})= \tag{19}$$
$$\sup_{\mathbf{d}_0,\mathbf{d}'_0}\{|\sum_{\mathbf{d}_m,\mathbf{d}'_m}\tilde{K}^0_h(\mathbf{d}_m,\mathbf{d}_0)\tilde{K}^0_h(\mathbf{d}'_m,\mathbf{d}_0)\Delta^{(2)}_{\Sigma,\eta}(\mathbf{d}_0,\mathbf{d}_m,\mathbf{d}'_m)|+\sum_{\mathbf{d}_m,\mathbf{d}'_m}|\tilde{K}^0_h(\mathbf{d}_m,\mathbf{d}_0)\tilde{K}^0_h(\mathbf{d}'_m,\mathbf{d}'_0)|\times$$
$$\sup_{\mathbf{d}_m,\mathbf{d}'_m}||n^{-1}\sum_{i=1}^n[\eta_i(\mathbf{d}_m)-\eta_i(\mathbf{d}_0)][\eta_i(\mathbf{d}'_m)-\eta_i(\mathbf{d}_0)]-\Delta^{(2)}_{\Sigma,\eta}(\mathbf{d}_0,\mathbf{d}_m,\mathbf{d}'_m)||\}.$$

Let $\hat{\Sigma}_\epsilon(\mathbf{d}_0)=n^{-1}\sum_{i=1}^n\epsilon_i(\mathbf{d}_0)^2$ and $\lambda(K_\epsilon,n,N_D)=K_\epsilon\log(2N_D)/n+\sqrt{2\log(2N_D)/n}$. It follows assumption (C2) and Lemma 14.13 of Buhlmann and van de Geer (2011) that

$$P(\sup_{\mathbf{d}_0\in\mathcal{D}_0}|\hat{\Sigma}_\epsilon(\mathbf{d}_0)-\Sigma_\epsilon(\mathbf{d}_0,\mathbf{d}_0)|\ge 2K_\epsilon^2t+2K_\epsilon C_\epsilon\sqrt{2t}+2K_\epsilon C_\epsilon\lambda(K_\epsilon,n,N_D))\le\exp(-nt), \tag{20}$$

which yields that $\sup_{\mathbf{d}\in\mathcal{D}_0}|\hat{\Sigma}_\epsilon(\mathbf{d})-\Sigma_\epsilon(\mathbf{d}_0,\mathbf{d}_0)|=o_p(1)$. Combining (19) and (20) yields (17). Similarly, we can prove (18).

Proof of Theorem 2. We have

$$\hat{\eta}_i(\mathbf{d})-\eta_i(\mathbf{d})=\Delta_i(\mathbf{d})=\bar{\epsilon}_i(\mathbf{d})+\Delta\eta_i(\mathbf{d})+\mathbf{x}_i^T\Delta\boldsymbol{\beta}(\mathbf{d}). \tag{21}$$

Therefore, we have

$$\begin{aligned}n^{-1}\sum_{i=1}^n\hat{\eta}_i(\mathbf{d})\hat{\eta}_i(\mathbf{d}') &= n^{-1}\sum_{i=1}^n\Delta_i(\mathbf{d})\Delta_i(\mathbf{d}')+n^{-1}\sum_{i=1}^n\eta_i(\mathbf{d})\Delta_i(\mathbf{d}') \\ &\quad +n^{-1}\sum_{i=1}^n\Delta_i(\mathbf{d})\eta_i(\mathbf{d}')+n^{-1}\sum_{i=1}^n\eta_i(\mathbf{d})\eta_i(\mathbf{d}').\end{aligned} \tag{22}$$

This proof of Theorem 2 (i) consists of three steps as follows.

- Show the uniform convergence of $n^{-1}\sum_{i=1}^n\eta_i(\mathbf{d})\eta_i(\mathbf{d}')$ to $\Sigma_\eta(\mathbf{d},\mathbf{d}')$ over $(\mathbf{d},\mathbf{d}')\in\mathcal{D}^2$ in probability.
- Show that $n^{-1}\sum_{i=1}^n\eta_i(\mathbf{d})\Delta_i(\mathbf{d}')+n^{-1}\sum_{i=1}^n\Delta_i(\mathbf{d})\eta_i(\mathbf{d}')$ converges to zero uniformly for all $(\mathbf{d},\mathbf{d}')\in\mathcal{D}^2$ in probability.

- Show that $n^{-1}\sum_{i=1}^{n}\Delta_i(\mathbf{d})\Delta_i(\mathbf{d}')$ converges to zero uniformly for all $(\mathbf{d},\mathbf{d}')\in\mathcal{D}^2$ in probability.

In the first step, it follows from assumption (C4) that

$$\sup_{(\mathbf{d},\mathbf{d}')\in\mathcal{D}^2}|n^{-1}\sum_{i=1}^{n}[\eta_i(\mathbf{d})\eta_i(\mathbf{d}')-\Sigma_\eta(\mathbf{d},\mathbf{d}')]|=O_p(n^{-1/2}). \tag{23}$$

In the second step, we can show that

$$\sup_{(\mathbf{d},\mathbf{d}')\in\mathcal{D}^2} n^{-1}|\sum_{i=1}^{n}\Delta_i(\mathbf{d})\eta_i(\mathbf{d}')|=o_p(h^2+(\log n/n)^{1/2}+n^{-1}\sqrt{\log(1+N_{\mathcal{D}})}). \tag{24}$$

With some simple calculations, we have

$$\sum_{i=1}^{n}\Delta_i(\mathbf{d})\eta_i(\mathbf{d}')\le\{|\sum_{i=1}^{n}\overline{\epsilon}_i(\mathbf{d})\eta_i(\mathbf{d}')|+|\sum_{i=1}^{n}\mathbf{x}_i^T\Delta\boldsymbol{\beta}(\mathbf{d})\eta_i(\mathbf{d}')|+|\sum_{i=1}^{n}\Delta\eta_i(\mathbf{d})\eta_i(\mathbf{d}')|\}. \tag{25}$$

Thus, it is sufficient to focus on the three terms on the right-hand side of (25). First, it follows from Lemma 1 that $\sup_{(\mathbf{d},\mathbf{d}')}n^{-1}\{|\sum_{i=1}^{n}\overline{\epsilon}_i(\mathbf{d})\eta_i(\mathbf{d}')|=O((\log n/n)^{1/2})$. Secondly, since $\{\mathbf{x}\eta(\mathbf{d}):\mathbf{d}\in\mathcal{D}\}$ is a Donsker class and $\sup_d||\Delta\boldsymbol{\beta}(\mathbf{d})||_2=O_p(n^{-1/2}\sqrt{\log(1+N_{\mathcal{D}})})$, we have

$$n^{-1}|\sum_{i=1}^{n}\mathbf{x}_i^T\Delta\boldsymbol{\beta}(\mathbf{d})\eta_i(\mathbf{d}')|\le\sup_d||\Delta\boldsymbol{\beta}(\mathbf{d})||_2||n^{-1}\sum_{i=1}^{n}\mathbf{x}_i^T\eta_i(\mathbf{d}')||_2=O_p(n^{-1}\sqrt{\log(1+N_{\mathcal{D}})}).$$

Thirdly, based on the definition of $\Delta\eta_i(\mathbf{d})$, we have

$$\begin{aligned}n^{-1}\sum_{i=1}^{n}\Delta\eta_i(\mathbf{d})\eta_i(\mathbf{d}')&=\{\sum_{\mathbf{d}_m}\tilde{K}_h^0(\mathbf{d}_m,\mathbf{d})\Sigma_\eta(\mathbf{d}_m,\mathbf{d}')-\Sigma_\eta(\mathbf{d},\mathbf{d}')\} \\ &+\sum_{\mathbf{d}_m}\tilde{K}_h^0(\mathbf{d}_m,d)n^{-1}\{\sum_{i=1}^{n}[\eta_i(\mathbf{d}_m)\eta_i(\mathbf{d}')-\Sigma_\eta(\mathbf{d}_m,\mathbf{d}')-\eta_i(\mathbf{d})\eta_i(\mathbf{d}')+\Sigma_\eta(\mathbf{d},\mathbf{d}')]\}.\end{aligned} \tag{26}$$

It follows from assumption (C4) that the first term on the right hand side of (26) is $O_p(h^2)$ and the second one is $O_p(n^{-1/2})$.

The third step is to show that

$$\sup_{(\mathbf{d},\mathbf{d}')}n^{-1}|\sum_{i=1}^{n}\Delta_i(\mathbf{d})\Delta_i(\mathbf{d}')|=O_p((N_Dh^3)^{-1}+(\log n/n)^{1/2}+h^2+n^{-1}\log(1+N_{\mathcal{D}})). \tag{27}$$

With some calculations, we have

$$
\begin{aligned}
|\sum_{i=1}^{n}\Delta_i(\mathbf{d})\Delta_i(\mathbf{d}')| &\leq C_1 \sup_{(d,d')}[|\sum_{i=1}^{n}\overline{\epsilon}_i(\mathbf{d})\overline{\epsilon}_i(\mathbf{d}')| + |\sum_{i=1}^{n}\overline{\epsilon}_i(\mathbf{d})\Delta\eta_i(\mathbf{d}')| \\
&+ |\sum_{i=1}^{n}\overline{\epsilon}_i(\mathbf{d})\mathbf{x}_i^T\Delta\boldsymbol{\beta}(\mathbf{d}')| + |\sum_{i=1}^{n}\Delta\eta_i(\mathbf{d}')\mathbf{x}_i^T\Delta\boldsymbol{\beta}(\mathbf{d})| \qquad (28) \\
&+ |\sum_{i=1}^{n}\Delta\eta_i(\mathbf{d})\Delta\eta_i(\mathbf{d}')| + |\sum_{i=1}^{n}\mathbf{x}_i^T\Delta\boldsymbol{\beta}(\mathbf{d})\Delta\boldsymbol{\beta}(\mathbf{d}')\mathbf{x}_i|],
\end{aligned}
$$

for a positive constant C_1. It follows from Lemma 1 that the first three terms on the right hand side of (28) uniformly converge to zero. We only need to consider the last three terms on the right hand side of (28) as follows:

$$
\begin{aligned}
n^{-1}|\sum_{i=1}^{n}\Delta\eta_i(\mathbf{d}')\mathbf{x}_i^T\Delta\boldsymbol{\beta}(\mathbf{d})| &\leq ||\Delta\boldsymbol{\beta}(\mathbf{d})||_2 n^{-1}||\sum_{i=1}^{n}\Delta\eta_i(\mathbf{d}')\mathbf{x}_i^T||_2 \\
&\leq ||\Delta\boldsymbol{\beta}(\mathbf{d})||_2\sum_{\mathbf{d}_m}|\tilde{K}_h^0(\mathbf{d}_m-\mathbf{d}')||n^{-1}\sum_{i=1}^{n}\mathbf{x}_i\eta_i(\mathbf{d}_m) - n^{-1}\sum_{i=1}^{n}\mathbf{x}_i\eta_i(\mathbf{d})||_2 \\
&= O_p(n^{-1}\sqrt{\log(1+N_{\mathcal{D}})}), \\
n^{-1}\sum_{i=1}^{n}\Delta\eta_i(\mathbf{d})\Delta\eta_i(\mathbf{d}') &= \sum_{\mathbf{d}_m,\mathbf{d}_m'}\tilde{K}_h^0(\mathbf{d}_m,\mathbf{d})\tilde{K}_h^0(\mathbf{d}_m',\mathbf{d}')n^{-1}\sum_{i=1}^{n}[\eta_i(\mathbf{d}_m)-\eta_i(\mathbf{d})][\eta_i(\mathbf{d}_m')-\eta_i(\mathbf{d}')] \\
&= O_p(h^2+n^{-1/2}), \\
n^{-1}\sum_{i=1}^{n}\mathbf{x}_i^T\Delta\boldsymbol{\beta}(\mathbf{d})\Delta\boldsymbol{\beta}(\mathbf{d}')^T\mathbf{x}_i &= \mathrm{tr}\{\Delta\boldsymbol{\beta}(\mathbf{d})\Delta\boldsymbol{\beta}(\mathbf{d}')^T n^{-1}\sum_{i=1}^{n}\mathbf{x}_i^{\otimes 2}\} = O_p(n^{-1}\log(1+N_{\mathcal{D}})).
\end{aligned}
$$

To prove Theorem 2 (ii), we note that $\hat{\epsilon}_i(\mathbf{d}_m) = \epsilon_i(\mathbf{d}_m) - \mathbf{x}_i^T[\hat{\boldsymbol{\beta}}(\mathbf{d}_m) - \boldsymbol{\beta}(\mathbf{d}_m)] - \Delta_i(\mathbf{d}_m)$ holds for all $\mathbf{d}_m \in \mathcal{D}_0$. It yields that

$$
\begin{aligned}
&\sup_{\mathbf{d}_m\in\mathcal{D}_0}|n^{-1}\sum_{i=1}^{n}\hat{\epsilon}_i(\mathbf{d}_m)^2 - n^{-1}\sum_{i=1}^{n}\epsilon_i(\mathbf{d}_m)^2| \leq \\
&2\sup_{\mathbf{d}_m\in\mathcal{D}_0} n^{-1}|\sum_{i=1}^{n}\Delta_i(\mathbf{d}_m)\epsilon_i(\mathbf{d}_m)| + 2\sup_{\mathbf{d}_m\in\mathcal{D}_0}|n^{-1}\sum_{i=1}^{n}\epsilon_i(\mathbf{d}_m)\mathbf{x}_i^T[\hat{\boldsymbol{\beta}}(\mathbf{d}_m)-\boldsymbol{\beta}(\mathbf{d}_m)]| \qquad (29) \\
&+n^{-1}\sup_{\mathbf{d}_m\in\mathcal{D}_0}\sum_{i=1}^{n}\Delta_i(\mathbf{d}_m)^2 + \sup_{\mathbf{d}_m\in\mathcal{D}_0}[\hat{\boldsymbol{\beta}}(\mathbf{d}_m)-\boldsymbol{\beta}(\mathbf{d}_m)]^T n^{-1}\sum_{i=1}^{n}\mathbf{x}_i^{\otimes 2}[\hat{\boldsymbol{\beta}}(\mathbf{d}_m)-\boldsymbol{\beta}(\mathbf{d}_m)].
\end{aligned}
$$

It follows from Lemma 2 that the first two terms on the right hand side of (29) are at the order of $O_p(n^{-1}\log(1+N_D) + \sqrt{h^2+n^{-1/2}} + \sqrt{(N_D h^3)^{-1} + (\log n/n)^{1/2}})$, while it

follows from (28) and Theorem 1 that the last two terms on the right hand side of above inequality converge to zero uniformly for all $\mathbf{d}_m \in \mathcal{D}_0$ in probability. This completes the proof of Theorem 2 (ii).

Theorem 2 (iii) directly follows from the same arguments in Lemma 6 of Li and Hsing (2010). So, we omit the details.

Proof of Theorem 3. For $s \geq 1$, we define

$$F_1(\mathbf{d}_0, h_s) = \frac{\sum_{\mathbf{d}_m \in B(\mathbf{d}_0, h_s) \cap \mathcal{D}_0} [\omega_j^{(0)}(\mathbf{d}_0, \mathbf{d}_m; h_s) - \omega_j(\mathbf{d}_0, \mathbf{d}_m; h_s)]}{\sum_{\mathbf{d}_m \in B(\mathbf{d}_0, h_s) \cap \mathcal{D}_0} \omega_j(\mathbf{d}_0, \mathbf{d}_m; h_s)}, \tag{30}$$

$$F_2(\mathbf{d}_0, h_s) = \frac{\sum_{\mathbf{d}_m \in B(\mathbf{d}_0, h_s) \cap \mathcal{D}_0} [\omega_j(\mathbf{d}_0, \mathbf{d}_m; h_1) - \omega_j^{(0)}(\mathbf{d}_0, \mathbf{d}_m; h_s)] \hat{\Delta}_j(\mathbf{d}_m)}{\sum_{\mathbf{d}_m \in B(\mathbf{d}_0, h_s) \cap \mathcal{D}_0} \omega_j(\mathbf{d}_0, \mathbf{d}_m; h_s)}.$$

For $0 \leq s \leq S$, we have the following results:

(R.1) $F_1(\mathbf{d}_0, h_s) = o_p(1), \quad F_2(\mathbf{d}_0, h_s) = o_p(\sqrt{\log(1+N_D)/n})$,

(R.2) $\tilde{\Delta}_{j*}(\mathbf{d}_0; h_s) = O_p(1) N_D h_s^3 K_{st}(0.5 C_n^{-1} n \boldsymbol{u}^{(j)}(h_s)^2 / \hat{\Sigma}_n(\sqrt{n}\hat{\beta}_j(\mathbf{d}_0; h_{s-1}))) = o_p(\sqrt{\log(1+N_D)/n})$,

(R.3) $\hat{\beta}_j(\mathbf{d}_0; h_s) - \beta_{j*}(\mathbf{d}_0) = F_0(\mathbf{d}_0, h_s)[1 + o_p(1)] = O_p(\sqrt{\log(1+N_D)/n})$,

(R.4) $\sup_{\mathbf{d}_0 \in \mathcal{D}_0} |\hat{\Sigma}_n(\sqrt{n}\hat{\beta}_j(\mathbf{d}_0; h_s)) - \Sigma_j^{(0)}(\mathbf{d}_0; h_s)| = o_p(1)$.

It follows from Theorem 1 that (R.1)-(R.4) hold for $s = 0$. For $s = 1$, we consider two different cases including (i) $\mathbf{d}_0 \in \mathcal{D} \setminus \partial\mathcal{D}^{(j)}(h_S)$ and (ii) $\mathbf{d}_0 \in \partial\mathcal{D}^{(j)}(h_S)$. Since $\Delta_{j*}(\mathbf{d}_0, \mathbf{d}_0') = 0$ for $\mathbf{d}_0 \in \mathcal{D}_{j,l}^o$ and $\mathbf{d}_0' \in B(\mathbf{d}_0; h_1)$, $D_{\beta_j}(\mathbf{d}_0, \mathbf{d}_0'; h_0)/C_n$ can be written as

$$\begin{aligned} D_{\beta_j}(\mathbf{d}_0, \mathbf{d}_0'; h_0)/C_n &= C_n^{-1} n \{\hat{\beta}_j(\mathbf{d}_0) - \hat{\beta}_j(\mathbf{d}_0')\}^2 / \hat{\Sigma}_n(\sqrt{n}\hat{\beta}_j(\mathbf{d}_0)) \\ &= C_n^{-1} n \{\hat{\Delta}_j(\mathbf{d}_0) - \hat{\Delta}_j(\mathbf{d}_0')\}^2 / \hat{\Sigma}_n(\sqrt{n}\hat{\beta}_j(\mathbf{d}_0)) = O_p(\log(1+N_D)/C_n). \end{aligned} \tag{31}$$

Note that $O_p(1)$ in above inequality is independent of $\mathbf{d}_0$ and $\mathbf{d}_0'$. Therefore, we have

$$|K_{st}(D_{\beta_j}(\mathbf{d}_0, \mathbf{d}_0'; h_0)/C_n) - K_{st}(0)| \leq O_p(1) \log(1+N_D) C_n^{-1}, \tag{32}$$

which yields

$$F_1(\mathbf{d}_0, h_s) = O_p(1) \log(1+N_D) C_n^{-1} = o_p(1). \tag{33}$$

It follows from (32) and (33) that

$$\begin{aligned} F_2(\mathbf{d}_0, h_1) &= \frac{\sum_{\mathbf{d}_m \in B(\mathbf{d}_0, h_s) \cap \mathcal{D}_0} [\omega_j(\mathbf{d}_0, \mathbf{d}_m; h_1) - \omega_j^{(0)}(\mathbf{d}_0, \mathbf{d}_m; h_s)] \hat{\Delta}_j(\mathbf{d}_m)}{\sum_{\mathbf{d}_m \in B(\mathbf{d}_0, h_s) \cap \mathcal{D}_0} \omega_j^{(0)}(\mathbf{d}_0, \mathbf{d}_m; h_s)} [1 + F_1(\mathbf{d}_0, h_1)]^{-1} \\ &\leq \{\log(1 + N_D)\}^{3/2} / (C_n \sqrt{n}) O_p(1). \end{aligned} \tag{34}$$

Since $\Delta_{j*}(\mathbf{d}_0, \mathbf{d}_0') = 0$ for $\mathbf{d}_0 \in \mathcal{D}_{j,l}^o \cap \mathcal{D}_0$ and $\mathbf{d}_0' \in B(\mathbf{d}_0; h_1) \cap \mathcal{D}_0$, we have $\tilde{\beta}_{j*}(\mathbf{d}_0; h_s) = \beta_{j*}(\mathbf{d}_0)$ for all $s = 1, \ldots, S$, which yields (R.2). It follows from (32)-(34) that

$$\hat{\beta}_j(\mathbf{d}_0; h_1) - \beta_{j*}(\mathbf{d}_0) = F_2(\mathbf{d}_0, h_1) + F_0(\mathbf{d}_0, h_1)[1 + F_1(\mathbf{d}_0, h_1)]^{-1} = F_0(\mathbf{d}_0, h_1)[1 + o_p(1)], \tag{35}$$

which yields (R.3).

To prove (R.4), we only need some notation as follows:

$$T_1(h_s) = \sup_{\mathbf{d}_0 \in \mathcal{D}_0} \left| \sum_{\mathbf{d}_m, \mathbf{d}_m' \in B(\mathbf{d}_0, h_s) \cap \mathcal{D}_0} \tilde{\omega}_j(\mathbf{d}_0, \mathbf{d}_m; h_s) \tilde{\omega}_j(\mathbf{d}_0, \mathbf{d}_m'; h_s) \{\hat{\Sigma}_y(\mathbf{d}_m, \mathbf{d}_m') - \Sigma_y(\mathbf{d}_m, \mathbf{d}_m')\} \right|,$$

$$T_2(h_s) = \sup_{\mathbf{d}_0 \in \mathcal{D}_0} \left| \sum_{\mathbf{d}_m, \mathbf{d}_m' \in B(\mathbf{d}_0, h_s) \cap \mathcal{D}_0} \{\tilde{\omega}_j(\mathbf{d}_0, \mathbf{d}_m; h_s) - \tilde{\omega}_j^{(0)}(\mathbf{d}_0, \mathbf{d}_m; h_s)\} \tilde{\omega}_j(\mathbf{d}_0, \mathbf{d}_m'; h_s) \Sigma_y(\mathbf{d}_m, \mathbf{d}_m') \right|,$$

$$T_3(h_s) = \sup_{\mathbf{d}_0 \in \mathcal{D}_0} \left| \sum_{\mathbf{d}_m, \mathbf{d}_m' \in B(\mathbf{d}_0, h_s) \cap \mathcal{D}_0} \{\tilde{\omega}_j(\mathbf{d}_0, \mathbf{d}_m; h_s) - \tilde{\omega}_j^{(0)}(\mathbf{d}_0, \mathbf{d}_m; h_s)\} \tilde{\omega}_j^{(0)}(\mathbf{d}_0, \mathbf{d}_m'; h_s) \Sigma_y(\mathbf{d}_m, \mathbf{d}_m') \right|.$$

A sufficient condition of (R.4) is $|T_1(h_1)| + |T_2(h_1)| + |T_3(h_1)| = o_p(1)$. It follows from Theorem 1 that $T_1(h_1) \leq \sup_{\mathbf{d}_m, \mathbf{d}_m' \in \mathcal{D}_0} |\hat{\Sigma}_y(\mathbf{d}_m, \mathbf{d}_m') - \Sigma_y(\mathbf{d}_m, \mathbf{d}_m')| = o_p(1)$. Moreover, $\tilde{\omega}_j(\mathbf{d}_0, \mathbf{d}_m; h_s) - \tilde{\omega}_j^{(0)}(\mathbf{d}_0, \mathbf{d}_m; h_s)$ equals

$$\frac{\omega_j(\mathbf{d}_0, \mathbf{d}_m; h_s) - \omega_j^{(0)}(\mathbf{d}_0, \mathbf{d}_m; h_s)}{\sum_{\mathbf{d}_m \in B(\mathbf{d}_0, h_s) \cap \mathcal{D}_0} \omega_j(\mathbf{d}_0, \mathbf{d}_m; h_s)} + F_1(\mathbf{d}_0, h_1) \tilde{\omega}_j^{(0)}(\mathbf{d}_0, \mathbf{d}_m; h_s). \tag{36}$$

Substituting (36) into $T_2(h_1)$, we have

$$T_2(h_1) \leq C_1 \{|F_1(\mathbf{d}_0, h_1)| + \frac{\sum_{\mathbf{d}_m \in B(\mathbf{d}_0, h_s) \cap \mathcal{D}_0} |\omega_j^{(0)}(\mathbf{d}_0, \mathbf{d}_m; h_1) - \omega_j(\mathbf{d}_0, \mathbf{d}_m; h_1)|}{\sum_{\mathbf{d}_m \in B(\mathbf{d}_0, h_1) \cap \mathcal{D}_0} \omega_j(\mathbf{d}_0, \mathbf{d}_m; h_1)}\} = o_p(1).$$

Similar to the derivation of $T_2(h_1)$, we can prove $T_3(h_1) = o_p(1)$.

For $\mathbf{d}_0 \in \partial\mathcal{D}^{(j)}(h_S)$, we assume $\mathbf{d}_0 \in \partial\mathcal{D}^{(j)}(h_1)$ without loss of generality. It follows from assumption (C10) that $B(\mathbf{d}_0, h_1)$ is the union of $B(\mathbf{d}_0, h_1) \cap \{\mathbf{d}_0' : \mathbf{d}_0' \in \mathcal{D}_0, \Delta_{j*}(\mathbf{d}_0, \mathbf{d}_0') = 0\}$ and $B(\mathbf{d}_0, h_1) \cap \{\mathbf{d}_0' : \mathbf{d}_0' \in \mathcal{D}_0, |\Delta_{j*}(\mathbf{d}_0, \mathbf{d}_0')| \geq \boldsymbol{u}^{(j)}(h_1)\}$. For $\mathbf{d}_0' \in B(\mathbf{d}_0, h_1) \cap \{\mathbf{d}_0' : \mathbf{d}_0' \in \mathcal{D}_0, \Delta_{j*}(\mathbf{d}_0, \mathbf{d}_0') = 0\}$, it is easy to see that (32) is true. For $\mathbf{d}_0' \in B(\mathbf{d}_0, h_1) \cap I_j(\mathbf{d}_0, \boldsymbol{u}^{(j)}(h_1), \infty)$, it follows from the inequality $2(a-b)^2 + 2b^2 \geq a^2$ for any a, b that

$$\begin{aligned} D_{\beta_j}(\mathbf{d}_0, \mathbf{d}_0'; h_0)/C_n &\geq C_n^{-1} n[0.5\Delta_{j*}(\mathbf{d}_0, \mathbf{d}_0')^2 - \{\hat{\Delta}_j(\mathbf{d}_0) - \hat{\Delta}_j(\mathbf{d}_0')\}^2]/\hat{\Sigma}_n(\sqrt{n}\hat{\beta}_j(\mathbf{d}_0)) \\ &\geq [0.5C_n^{-1} n\boldsymbol{u}^{(j)}(h_1)^2 - 4C_n^{-1} n \sup_{\mathbf{d}_0} \hat{\Delta}_j(\mathbf{d}_0)^2]/\hat{\Sigma}_n(\sqrt{n}\hat{\beta}_j(\mathbf{d}_0)). \quad (37) \end{aligned}$$

Thus, we have

$$K_{st}(D_{\beta_j}(\mathbf{d}_0, \mathbf{d}_0'; h_0)/C_n) \leq O_p(1) K_{st}(0.5C_n^{-1} n\boldsymbol{u}^{(j)2}/\hat{\Sigma}_n(\sqrt{n}\hat{\beta}_j(\mathbf{d}_0))), \quad (38)$$

which yields that

$$\begin{aligned} &\frac{\sum_{\mathbf{d}_m \in B(\mathbf{d}_0, h_1) \cap \mathcal{D}_0 \cap I_j(\mathbf{d}_0, \boldsymbol{u}^{(j)}(h_1), \infty)} K_{loc}(||\mathbf{d}_0 - \mathbf{d}_m||_2/h_1) K_{st}(D_{\beta_j}(\mathbf{d}_0, \mathbf{d}_m; h_0)/C_n)}{\sum_{\mathbf{d}_m \in B(d, h_1) \cap \mathcal{D}_0} K_{loc}(||\mathbf{d}_0 - \mathbf{d}_m||_2/h_1) K_{st}(D_{\beta_j}(\mathbf{d}_0, \mathbf{d}_m; h_0)/C_n)} \\ \leq\ & O_p(1) K_{st}(0.5C_n^{-1} n\boldsymbol{u}^{(j)}(h_1)^2/\hat{\Sigma}_n(\sqrt{n}\hat{\beta}_j(\mathbf{d}_0))) \times \\ & \frac{\sum_{\mathbf{d}_m \in B(\mathbf{d}_0, h_1) \cap I_j(\mathbf{d}_0, \boldsymbol{u}^{(j)}(h_1)^2, \infty) \cap \mathcal{D}_0} K_{loc}(||\mathbf{d}_0 - \mathbf{d}_m||_2/h_1)}{K_{loc}(0) K_{st}(0)} \\ \leq\ & O_p(1) N_D h_1^3 K_{st}(0.5C_n^{-1} n\boldsymbol{u}^{(j)}(h_1)^2/\hat{\Sigma}_n(\sqrt{n}\hat{\beta}_j(\mathbf{d}_0))). \quad (39) \end{aligned}$$

Therefore, it follows from (38) and (39) that

$$\begin{aligned} F_1(\mathbf{d}_0, h_s) &= O_p(1)\{\log(1+N_D) C_n^{-1} + N_D h_1^3 K_{st}(0.5C_n^{-1} n\boldsymbol{u}^{(j)}(h_1)^2/\hat{\Sigma}_n(\sqrt{n}\hat{\beta}_j(\mathbf{d}_0)))\} = o_p(1), \\ F_2(\mathbf{d}_0, h_1) &\leq \{\log(1+N_D)\}^{3/2}/(C_n\sqrt{n}) O_p(1) + \\ &\quad O_p(N_D h_1^3 K_{st}(0.5C_n^{-1} n\boldsymbol{u}^{(j)}(h_1)^2/\hat{\Sigma}_n(\sqrt{n}\hat{\beta}_j(\mathbf{d}_0))))\sqrt{\log(1+N_D)/n} \\ &= o_p(\sqrt{\log(1+N_D)/n}), \quad (40) \end{aligned}$$

which yield (R.1). Furthermore, it follows from (39) that

$$\begin{aligned} |\tilde{\Delta}_{j*}(\mathbf{d}_0; h_s)| &\leq \sum_{\mathbf{d}_m \in B(\mathbf{d}_0, h_1) \cap \mathcal{D}_0 \cap I_j(\mathbf{d}_0, \sqrt{C_n/n}\boldsymbol{u}^{(j)}, \infty)} \tilde{\omega}_j(\mathbf{d}_0, \mathbf{d}_m; h_s) |\beta_{j*}(\mathbf{d}_m) - \beta_{j*}(\mathbf{d}_0)| \\ &\leq O_p(1) N_D h_1^3 K_{st}(0.5C_n^{-1} n\boldsymbol{u}^{(j)}(h_1)^2/\hat{\Sigma}_n(\sqrt{n}\hat{\beta}_j(\mathbf{d}_0))). \quad (41) \end{aligned}$$

Furthermore, it follows from (32)-(34) that

$$\begin{aligned}\hat{\beta}_j(\mathbf{d}_0;h_1) &= \tilde{\beta}_{j*}(\mathbf{d}_0;h_1)+F_2(\mathbf{d}_0,h_1)+F_0(\mathbf{d}_0,h_1)[1+F_1(\mathbf{d}_0,h_1)]^{-1} \quad (42)\\ &= \beta_{j*}(\mathbf{d}_0)+\tilde{\Delta}_{j*}(\mathbf{d}_0;h_1)+F_2(\mathbf{d}_0,h_1)+F_0(\mathbf{d}_0,h_1)[1+F_1(\mathbf{d}_0,h_1)]^{-1}\\ &= \beta_{j*}(\mathbf{d}_0)+F_0(\mathbf{d}_0,h_1)[1+o_p(1)]+o_p(\sqrt{\log(1+N_D)/n}),\end{aligned}$$

which yields (R.3). Similar to the arguments near (36), we can easily prove (R.4) for $\mathbf{d}_0 \in \partial\mathcal{D}^{(j)}(h_S)\cap\mathcal{D}_0$.

Note that (R.2) and (R.3) are the key results used in deriving (31)-(42). Based on (R.1)-(R.4) for $s=1$, we can use the same arguments from (31) to (42) to prove (R.1)-(R.4) for $s=2$. Generally, if (R.1)-(R.4) are true for any s, we can use the same arguments in (31)-(42) to prove (R.1)-(R.4) for $s+1$. This finishes the proof of Theorem 3.

Proof of Theorem 4. For $s\geq 1$, we define

$$\hat{\Delta}_j(\mathbf{d}_0;h_s)=\hat{\beta}_j(\mathbf{d}_0;h_s)-\tilde{\beta}_{j*}(\mathbf{d}_0;h_s),\ \tilde{\Delta}_{j*}(\mathbf{d}_0,\mathbf{d}_0';h_s)=\tilde{\beta}_{j*}(\mathbf{d}_0;h_s)-\tilde{\beta}_{j*}(\mathbf{d}_0';h_s). \quad (43)$$

For $0\leq s\leq S$, we want to prove the following results by introduction:

$$\begin{aligned}&\text{(R.1)}\quad F_1(\mathbf{d}_0,h_s)=o_p(1),\quad F_2(\mathbf{d}_0,h_s)=o_p(\sqrt{\log(1+N_D)/n}),\\ &\text{(R.2)}\quad \tilde{\Delta}_{j*}(\mathbf{d}_0;h_s)=L_jh_s+\delta_L+O_p(1)N_Dh_s^3K_{st}(0.5M_n^2/\hat{\Sigma}_n(\sqrt{n}\hat{\beta}_j(\mathbf{d}_0))),\\ &\text{(R.3)}\quad \hat{\beta}_j(\mathbf{d}_0;h_s)-\tilde{\beta}_{j*}(\mathbf{d}_0;h_s)=F_0(\mathbf{d}_0,h_s)[1+o_p(1)]=O_p(\sqrt{\log(1+N_D)/n}),\\ &\text{(R.4)}\quad \sup_{\mathbf{d}_0\in\mathcal{D}_0}|\hat{\Sigma}_n(\sqrt{n}\hat{\beta}_j(\mathbf{d}_0;h_s))-\Sigma_j^{(1)}(\mathbf{d}_0;h_s)|=o_p(1).\end{aligned}$$

It follows from Theorem 1 that (R.1)-(R.4) hold for $s=0$. For $s=1$, we consider two different cases including (i) $\mathbf{d}_0\in\mathcal{D}\setminus\partial\mathcal{D}^{(j)}(h_S)$ and (ii) $\mathbf{d}_0\in\partial\mathcal{D}^{(j)}(h_S)$. For $\mathbf{d}_0\in\mathcal{D}\setminus\partial\mathcal{D}^{(j)}(h_S)$ and $\mathbf{d}_0'\in B(\mathbf{d}_0,h_1)$, it follows from assumption (C9) that $D_{\beta_j}(\mathbf{d}_0,\mathbf{d}_0';h_0)/C_n$ can be written as

$$\begin{aligned}D_{\beta_j}(\mathbf{d}_0,\mathbf{d}_0';h_0)/C_n &= C_n^{-1}n\{\hat{\Delta}_j(\mathbf{d}_0)-\hat{\Delta}_j(\mathbf{d}_0')+\Delta_{j*}(\mathbf{d}_0,\mathbf{d}_0')\}^2/\hat{\Sigma}_n(\sqrt{n}\hat{\beta}_j(\mathbf{d}_0))\\ &\leq 2\hat{\Sigma}_n(\sqrt{n}\hat{\beta}_j(\mathbf{d}_0))^{-1}\{\log(1+N_D)C_n^{-1}+K_j^2nC_n^{-1}||\mathbf{d}_0-\mathbf{d}_0'||_2^2\}O_p(1).\end{aligned}$$

Therefore, we have

$$\begin{aligned}|K_{st}(D_{\beta_j}(\mathbf{d}_0,\mathbf{d}_0';h_0)/C_n)-K_{st}(0)| &\leq O_p(1)\hat{\Sigma}_n(\sqrt{n}\hat{\beta}_j(\mathbf{d}_0))^{-1}|\log(1+N_D)C_n^{-1}+K_j^2nC_n^{-1}h_1^2| \\ &= O_p(1)\log(1+N_D)C_n^{-1}. \end{aligned} \tag{44}$$

For $\mathbf{d}_0 \in \partial\mathcal{D}^{(j)}(h_S)$, we assume $\mathbf{d}_0 \in \partial\mathcal{D}^{(j)}(h_1)$ without loss of generality. It follows from assumption (C10b) that $B(\mathbf{d}_0,h_1)$ is the union of $P_j(\mathbf{d}_0,h_1)$, $P_j(\mathbf{d}_0,h_1)^c \cap I_j(\mathbf{d}_0,0,\delta_L)$ and $P_j(\mathbf{d}_0,h_1)^c \cap I_j(\mathbf{d}_0,\delta_U,\infty)$, and $P_j(\mathbf{d}_0,h_1)^c \cap I_j(\mathbf{d}_0,\delta_L,\delta_U) = \emptyset$. For $\mathbf{d}_0' \in P_j(\mathbf{d}_0,h_1) \cup [P_j(\mathbf{d}_0,h_1)^c \cap I_j(\mathbf{d}_0,0,\delta_L)]$, it is easy to see that (44) is true. For $\mathbf{d}_0' \in B(\mathbf{d}_0,h_1) \cap I_j(\mathbf{d}_0,\delta_U,\infty)$, by using the same arguments in (34)-(37), we have

$$\begin{aligned}&\frac{\sum_{\mathbf{d}_m \in B(\mathbf{d}_0,h_1)\cap I_j(\mathbf{d}_0,\delta_U,\infty)} K_{loc}(||\mathbf{d}_0-\mathbf{d}_m||_2/h_1)K_{st}(D_{\beta_j}(\mathbf{d}_0,\mathbf{d}_m;h_0)/C_n)}{\sum_{\mathbf{d}_m \in B(\mathbf{d}_0,h_1)} K_{loc}(||\mathbf{d}_0-\mathbf{d}_m||_2/h_1)K_{st}(D_{\beta_j}(\mathbf{d}_0,\mathbf{d}_m;h_0)/C_n)} \\ \leq\ & O_p(1)N_Dh_1^3K_{st}(0.5M_n^2/\hat{\Sigma}_n(\sqrt{n}\hat{\beta}_j(\mathbf{d}))). \end{aligned} \tag{45}$$

Therefore, similar to (34) and (40), we have

$$\begin{aligned}F_1(\mathbf{d}_0,h_1) &= O_p(1)\{\log(1+N_D)C_n^{-1}+N_Dh_1^3K_{st}(0.5M_n^2/\hat{\Sigma}_n(\sqrt{n}\hat{\beta}_j(\mathbf{d}_0)))\} = o_p(1), \\ F_2(\mathbf{d}_0,h_1) &= o_p(\sqrt{\log(1+N_D)/n}), \end{aligned} \tag{46}$$

which yield (R.1).

We prove (R.2) as follows. For $\mathbf{d}_0 \in \partial\mathcal{D}^{(j)}(h_S) \cap \mathcal{D}_0$, it follows from (44) and (45) that

$$\begin{aligned}|\tilde{\Delta}_{j*}(\mathbf{d}_0;h_1)| &\leq |\sum_{\mathbf{d}_m \in P_j(\mathbf{d}_0,h_1)} \tilde{\omega}_j(\mathbf{d}_0,\mathbf{d}_m;h_s)[\beta_{j*}(\mathbf{d}_m)-\beta_{j*}(\mathbf{d}_0)]| \\ &+ \sum_{\mathbf{d}_m \in P_j(\mathbf{d}_0,h_1)^c\cap I_j(\mathbf{d}_0,0,\delta_L)} \tilde{\omega}_j(\mathbf{d}_0,\mathbf{d}_m;h_s)|\beta_{j*}(\mathbf{d}_m)-\beta_{j*}(\mathbf{d}_0)| \\ &+ \sum_{\mathbf{d}_m \in P_j(\mathbf{d}_0,h_1)^c\cap I_j(\mathbf{d}_0,\delta_U,\infty)} \tilde{\omega}_j(\mathbf{d}_0,\mathbf{d}_m;h_s)|\beta_{j*}(\mathbf{d}_m)-\beta_{j*}(\mathbf{d}_0)| \\ &\leq L_jh_1+\delta_L+N_Dh_1^3K_{st}(0.5M_n^2/\hat{\Sigma}_n(\sqrt{n}\hat{\beta}_j(\mathbf{d}_0)))O_p(1). \end{aligned} \tag{47}$$

However, for $\mathbf{d}_0 \in \mathcal{D} \setminus \partial\mathcal{D}^{(j)}(h_S)$, by using Taylor series expansion, we have

$$|\tilde{\Delta}_{j*}(\mathbf{d}_0;h_1)| = |\sum_{\mathbf{d}_m \in B(d,h_1)} \tilde{\omega}_j(\mathbf{d}_0,\mathbf{d}_m;h_s)[\beta_{j*}(\mathbf{d}_m)-\beta_{j*}(\mathbf{d}_0)]| \leq L_jh_1. \tag{48}$$

This yields (R.2). Similar to the arguments in Theorem 3, we can easily prove (R.3) and (R.4) for $s = 1$. So we omit the details.

For $s = 2$, we only prove the result (R.1). $D_{\beta_j}(\mathbf{d}_0, \mathbf{d}_0'; h_1)$ can be written as

$$\begin{aligned} D_{\beta_j}(\mathbf{d}_0, \mathbf{d}_0'; h_1) &= n\{\hat{\beta}_j(\mathbf{d}_0; h_1) - \hat{\beta}_j(\mathbf{d}_0'; h_1)\}^2 / \hat{\Sigma}_n(\sqrt{n}\hat{\beta}_j(\mathbf{d}_0; h_1)) \\ &= n\{\hat{\Delta}_j(\mathbf{d}_0; h_1) - \hat{\Delta}_j(\mathbf{d}_0'; h_1) + \tilde{\Delta}_{j*}(\mathbf{d}_0, \mathbf{d}_0'; h_1)\}^2 / \hat{\Sigma}_n(\sqrt{n}\hat{\beta}_j(\mathbf{d}_0; h_1)). \end{aligned} \tag{49}$$

We first consider the cases with $\mathbf{d}_0' \in P_j(d, h_2)$ for $\mathbf{d}_0 \in \mathcal{D} \setminus \partial\mathcal{D}^{(j)}(h_S)$ and $\mathbf{d}_0' \in B(\mathbf{d}_0, h_2)$ for $\mathbf{d}_0 \in \partial\mathcal{D}^{(j)}(h_S)$. It follows from (R.2) and (R.3) that

$$\begin{aligned} & nC_n^{-1}\{\hat{\Delta}_j(\mathbf{d}_0; h_1) - \hat{\Delta}_j(\mathbf{d}_0'; h_1) + \tilde{\Delta}_j(\mathbf{d}_0, \mathbf{d}_0'; h_1)\}^2 \\ & \le 2nC_n^{-1}\{\hat{\Delta}_j(\mathbf{d}_0; h_1) - \hat{\Delta}_j(\mathbf{d}_0'; h_1)\}^2 + 2nC_n^{-1}\{\tilde{\Delta}_{j*}(\mathbf{d}_0; h_1) - \tilde{\Delta}_{j*}(\mathbf{d}_0'; h_1) + \Delta_{j*}(\mathbf{d}_0, \mathbf{d}_0')\}^2 \\ & \le O_p(1)\{\log(1 + N_D)C_n^{-1} + nC_n^{-1}(h_1^2 + h_2^2)\}, \end{aligned}$$

which yields $F_1(\mathbf{d}_0, h_2) \le O_p(1)|\log(1 + N_D)C_n^{-1} + nC_n^{-1}(h_1^2 + h_2^2)|$.

For $\mathbf{d}_0' \in P_j(\mathbf{d}_0, h_1)^c \cap I_j(\mathbf{d}_0, \delta_U, \infty)$, by using the same arguments in (34)-(37), we have

$$\begin{aligned} & \{\hat{\beta}_j(\mathbf{d}_0; h_1) - \hat{\beta}_j(\mathbf{d}_0'; h_1)\}^2 \\ \ge\ & 0.5\Delta_{j*}(\mathbf{d}_0, \mathbf{d}_0')^2 - \{\tilde{\Delta}_{j*}(\mathbf{d}_0; h_1) - \tilde{\Delta}_{j*}(\mathbf{d}_0'; h_1) + \hat{\Delta}_j(\mathbf{d}_0; h_1) - \hat{\Delta}_j(\mathbf{d}_0'; h_1)\}^2 \\ \ge\ & 0.5n^{-1}C_nM_n^2 - 2\{\tilde{\Delta}_{j*}(\mathbf{d}_0; h_1) - \tilde{\Delta}_{j*}(\mathbf{d}_0'; h_1)\}^2 - 2\{\hat{\Delta}_j(\mathbf{d}_0; h_1) - \hat{\Delta}_j(\mathbf{d}_0'; h_1)\}^2 \\ \ge\ & 0.5n^{-1}C_nM_n^2 - O_p(h_1^2 + n^{-1}\log(1 + N_D)). \end{aligned}$$

Thus, we have

$$\begin{aligned} & \frac{\sum_{\mathbf{d}_m \in P_j(\mathbf{d}_0, h_2)^c \cap I_j(\mathbf{d}_0, \delta_U, \infty)} K_{loc}(||\mathbf{d}_0 - \mathbf{d}_m||_2/h_2) K_{st}(D_{\beta_j}(\mathbf{d}_0, \mathbf{d}_m; h_1)/C_n)}{\sum_{\mathbf{d}_m \in B(\mathbf{d}_0, h_2)} K_{loc}(||\mathbf{d}_0 - \mathbf{d}_m||_2/h_2) K_{st}(0)} \\ \le\ & O_p(1)N_Dh_2^3 K_{st}(0.5M_n^2/\hat{\Sigma}_n(\sqrt{n}\hat{\beta}_j(\mathbf{d}_0; h_1))) \end{aligned} \tag{50}$$

Therefore, by using the similar arguments in (34) and (40), we can get

$$\begin{aligned} F_1(\mathbf{d}_0, h_2) &\le O_p(1)|\log(1 + N_D)C_n^{-1} + nC_n^{-1}(h_1^2 + h_2^2)| \\ &+ O_p(1)N_Dh_2^3 K_{st}(0.5M_n^2/\hat{\Sigma}_n(\sqrt{n}\hat{\beta}_j(\mathbf{d}_0; h_1))), \\ F_2(\mathbf{d}_0, h_2) &= o_p(\sqrt{\log(1 + N_D)/n}). \end{aligned} \tag{51}$$

Generally, if (R.1)-(R.4) are true for any s, we can use the same arguments in (43)-(51) to prove (R.1)-(R.4) for $s+1$. This finishes the proof of Theorem 4.

Simulation Studies

Additional Simulation Results

We present some additional results obtained from the simulation studies in the main paper. Figure S1 shows some selected results based on $\hat{\beta}_3(\mathbf{d}_0, h_0)$ and $\hat{\beta}_3(\mathbf{d}_0; h_{10})$ with $N(0,1)$ distributed data and $n = 60$ from the 200 simulated data sets. The biases slightly increase from h_0 to h_{10} (Figure S1 (b) and (g)), whereas the root-mean-square errors (RMSs) and standard deviations (SDs) at h_{10} are much smaller than those at h_0 (Figure S1 (c), (d), (h), and (i)). In addition, the RMSs and their corresponding SDs are relatively close to each other at all scales for both the normal (Figure S1 (e) and (j)) and Chi-square distributed data (not shown here). Moreover, SDs in these voxels of regions of interest (ROIs) with $\beta_3(\mathbf{d}_0) > 0$ are larger than SDs in those voxels of ROI with $\beta_3(\mathbf{d}_0) = 0$ (the last column in the lower row of Figure S1), because the interior of ROI with $\beta_3(\mathbf{d}_0) = 0$ contains more pixels (Figure 3 (c)). Moreover, both the SDs at steps h_0 and h_{10} show clear spatial patterns caused by spatial correlations (Figure S1 (d) and (i)). The RMSs also show some evidence of spatial patterns (Figure S1 (c) and (h)). All these results confirm the conclusions that we make based on Table 1 in the main paper.

We test the hypotheses $H_0(\mathbf{d}_0): \beta_j(\mathbf{d}_0) = 0$ versus $H_1(\mathbf{d}_0): \beta_j(\mathbf{d}_0) \neq 0$ for $j = 1, 2, 3$ across all $\mathbf{d}_0 \in \mathcal{D}_0$ using the MASS procedure at scales h_0 and h_{10}. The $-\log_{10}(p)$ values on some selected slices are shown in Figure S2. The values that are greater than 1.3 indicate a significant effect at 5% significance level and a highly significant effect at 1% significance level if they are greater than 3. The results are consistent with that from Table 2. In the lower panels of Figure S2 at scale h_{10}, all the nonzero regions of $\beta_j(\mathbf{d}_0)$ are detected as significant at 5% significance level, while most of them are even identified

as highly significant and the boundaries between different regions are fairly identifiable. In contrast, in the upper panels of Figure S2, at scale h_0, many voxels in ROIs with $\beta_2(\mathbf{d}_0) \neq 0$ are significant at $\alpha = 5\%$ significance level, while the boundaries of ROIs are blurred.

Local Constant Estimation

As suggested by one of the referees, we compare SVCM with another estimation method, called local constant estimation (LCE). Specifically, we calculate the least squares estimate $\hat{\beta}_j(\mathbf{d}_0)$ and then use local constant method based on the Epanechnikov kernel function, $K(u) = 3/4(1-u^2)I(|u| \leq 1)$, to directly smooth the initial estimate image, which leads to a new estimate, denoted as $\tilde{\beta}_j(\mathbf{d}_0; h)$, at each voxel. Subsequently, we use the method in Stage (III) of SVCM to compute the standard errors of $\tilde{\beta}_j(\mathbf{d}_0; h)$ and construct a Wald type test. We consider small ($h_s = 1.1$), moderate ($h_m = 2$), and large bandwidths ($h_l = 4$).

Figure S3 presents the LCE estimates obtained from the three different bandwidths based on one selected simulated data set. For the small bandwidth, effect ROIs cannot be clearly detected. As bandwidth increases, the $-\log_{10}(p)$ plots in Figure S5 reveal that the coefficients near the boundaries of all ROIs are easily oversmoothed and the edges of all ROIs are blurred at moderate and large bandwidths. In addition, Figure S5 shows that false positive rates are high for moderate and large bandwidths, as confirmed in Table S2. The failure of detecting edges by LCE is also observed from the bias plots in Figure S4 (panels (b), (g), and (l)). As shown in Figure S4 (panels (e), (j), and (o)), the ratios of RMS over SD are uniformly greater than 1, which indicates that the SDs are underestimated.

We repeated the simulation 200 times at the three different bandwidths with $N(0,1)$ and $\chi^2(3) - 3$ distributed data for two different sample sizes ($n = 60$ and 80) as we did in the main paper. For the sake of space, we only report the results for $\beta_2(\mathbf{d}_0)$ in Table S1. The bias in ROIs with no or weak signals ($\beta_2(\mathbf{d}_0) = 0$ or 0.2) is positive, whereas

the bias is negative for ROIs with median or strong signals ($\beta_2(\mathbf{d}_0) = 0.4,\ 0.6$ or 0.8). It indicates that weak signals are overestimated, whereas mediate and strong signals are underestimated mainly due to the burring edges of LCE. Table S1 reveals that the SDs are underestimated. We also calculated the rejection rates for testing $H_0 : \beta_2(\mathbf{d}_0) = 0$ in all voxels and include them in Table S2. The effect sizes (or false positive rates) are much larger than the preselected significant level $\alpha = 5\%$, and thus the Wald test is invalid even though it is very powerful for detecting relatively weak signals. Such large false positive rates may be due to positive bias in ROIs with none or weak signals and underestimated SDs.

Gaussian Kernel Smoothing

As suggested by one of the referees, we compare SVCM with a standard voxel-wise method, called Gaussian Kernel Smoothing (GKS) hereafter. The GKS consists of two steps including a smoothing step to smooth the simulated raw imaging data and an inference step to calculate the least squares estimate of $\boldsymbol{\beta}(\mathbf{d}_0)$, denoted as $\tilde{\boldsymbol{\beta}}^o(\mathbf{d}_0; h)$, and test hypothesis of interest at each voxel. In the smoothing step, we use the Gaussian kernel smoothing function and consider three different bandwidths including a small bandwidth ($h_s = 1.1$), a moderate bandwidth ($h_m = 2$), and a large bandwidth ($h_l = 4$).

Figure S6 presents the GSK estimates obtained from the three different bandwidths based on one selected simulated data set. Similar to LCE, small bandwidth does not increase signal detection especially near the boundaries of ROIs (Figure S6 (a)-(c)), while moderate and large bandwidths oversmooth the coefficient images and blur the boundaries of ROIs (Figure S6 (d)-(i)). It indicates that GKS is not capable of effectively estimating functions with potential jumps and edges. Figure S8 shows that the false positive rates are high for moderate and large bandwidths. See also Table S4. The bias plots in Figure S7 (panels (b), (g), and (l)) show strong blurred edges. It further confirms the limitation of GKS in preserving boundaries.

We repeated the simulation 200 times at the three different bandwidths with $N(0, 1)$

and $\chi^2(3) - 3$ distributed data for two different sample sizes ($n = 60, 80$) as we did in the main paper. For the sake of space, we only report the results for $\beta_2(\mathbf{d}_0)$ in Table S3. Inspecting Table S3 reveals that bias in ROIs with weak signals ($\beta_2(\mathbf{d}_0) = 0$ or 0.2 is positive, whereas bias in ROIs with mediate and strong signals ($\beta_2(\mathbf{d}_0) = 0.4,\ 0.6$ or 0.8) are negative. The rejection rate results for testing $H_0(\mathbf{d}_0) : \beta_2(\mathbf{d}_0) = 0$ are shown in Table S4. The effect sizes (false positive rates) are much larger than the preselected significant level $\alpha = 0.05$.

ADHD 200

Image Processing

The image processing is performed as follows. First, we do an AC-PC (anterior commissure - posterior commissure) correction on all images using MIPAV software (Medical Image Processing and software package, 2013), and then resampled the MRI images to 256×256×256. To correct the intensity inhomogeneity, we use N3 algorithm (Sled et al., 1998). An accurate and robust skull stripping method (Wang et al., 2011) was performed, and the skull stripping results were further manually reviewed to ensure clean skull and dura removal. After the skull-stripping, we used N3 algorithm again to correct for intensity inhomogeneity. Then the cerebellum is removed based on registration, in which we use a manually labeled cerebellum as a template. After intensity inhomogeneity correction, we use FAST in FSL (Zhang et al., 2001) to segment the human brain into three different tissues: grey matter (GM), white matter (WM), and Cerebrospinal fluid (CSF). We use HAMMER (Shen and Davatzikos, 2002) to do the registration. After registration, we get the subject-labeled image based on the Jacob template (Kabani et al., 1998), which is manually labeled into 93 ROIs. For each of the 93 ROIs in the labeled image of one subject, we compute the GM/WM/CSF tissue volumes in this ROI region combining the segmentation result of this subject.

To quantify the local volumetric group differences, we generate the RAVENS maps

(Goldszal et al., 1998; Davatzikos et al., 2001) for whole brain and for GM, WM and CSF, respectively, by using the deformation field that we get in registration. RAVENS methodology is based on a volume-preserving spatial transformation, which ensures that no volumetric information is lost during the process of spatial normalization, since this process changes an individual's brain morphology to conform it to the morphology of a template. A physical analog is the squeezing of a rubber object, which changes the density of the rubber, to maintain the same total mass in the object. Regional volumetric measurements are then performed via the resulting tissue density maps. We also do automatic subject labeling by transferring the labels of the template after deformable registration with the subjects. We have 93 ROIs in total. After labeling, we can get the ROI volumes of all subjects.

Additional Results

We are also interested in assessing the gender and diagnostic interaction. Specifically, we tested $H_0(\mathbf{d}_0): \beta_7(\mathbf{d}_0) = 0$ against $H_1(\mathbf{d}_0): \beta_7(\mathbf{d}_0) \neq 0$ for the gender×diagnosis interaction across all voxels. As s increases from 0 to 10, MASS shows an advantage in smoothing effective signals within relatively homogeneous ROIs, while preserving the edges of these ROIs (Figure S9 (a)-(b)). Inspecting Figure S9 (c) and (d) reveals that it is much easy to identify significant ROIs in the $-\log_{10}(p)$ images at scale h_{10}, which are much smoother than those at scale h_0. Thus, MASS shows a clear advantage in detecting more significant and smoothed activation regions. Furthermore, as shown in Figure S10 , the largest estimated eigenvalue is much larger than the rest estimated eigenvalues, which decrease very slowly to zero, and explains 22% of variation in data after accounting for $\mathbf{x}_i$. This is quite common in neuroimaging data (Caffo et al., 2010).

To formally detect significant ROIs, we used a cluster-form of threshold of 5% with a minimum voxel clustering value of 50 voxels. We were able to detect 26 and 10 significant clusters for testing $H_0(\mathbf{d}_0): \beta_6(\mathbf{d}_0) = 0$ (Figure S11 (a)) and $H_0(\mathbf{d}_0): \beta_7(\mathbf{d}_0) = 0$ (Figure S11 (b)), respectively, across all voxels. Table S5 lists the first

two largest predefined regions (ROIs) within the first six largest significant blocks for testing $H_0(\mathbf{d}_0) : \beta_6(\mathbf{d}_0) = 0$ and $H_0(\mathbf{d}_0) : \beta_7(\mathbf{d}_0) = 0$, respectively, along with their voxel sizes. Left and right frontal lobe white matter ROIs are the largest ROIs with significant Age×Diagnosis interaction effect while the first largest ROI with significant Gender×Diagnosis interaction effect is temporal lobe. We can also observe that size of the significant blocks for Age×Diagnosis interaction effect becomes much small starting from the fifth largest block while size of the significant blocks for Gender×Diagnosis interaction effect becomes much small starting from the third largest block.

References

Buhlmann, P. and van de Geer, S. (2011), *Statistics for High-Dimensional Data: Methods, Theory and Applications*, New York, N.Y.: Springer.

Caffo, B., Crainiceanu, C., Verduzco, G., Joel, S., Mostofsky, S., Bassett, S., and Pekar, J. (2010), "Two-stage decompositions for the analysis of functional connectivity for fMRI with application to Alzheimer's disease risk," *NeuroImage*, 51, 1140–1149.

Davatzikos, C., Genc, A., Xu, D., and Resnick, S. (2001), "Voxel-based morphometry using the RAVENS maps: methods and validation using simulated longitudinal atrophy." *NeuroImage*, 14, 1361–1369.

Einmahl, U. and Mason, D. M. (2000), "An empirical process approach to the uniform consistency of kernel-type function estimators," *Journal of Theoretical Probability*, 13, 1–37.

Goldszal, A. F., Davatzikos, C., Pham, D. L., Yan, M. X. H., Bryan, R. N., and Resnick, S. M. (1998), "An image-processing system for qualitative and quantitative volumetric analysis of brain images," *J. Comput. Assist. Tomogr.*, 22, 827–837.

Kabani, N., MacDonald, D., Holmes, C., and Evans, A. (1998), "A 3D atlas of the human brain," *Neuroimage*, 7, S717.

Li, Y. and Hsing, T. (2010), "Uniform convergence rates for nonparametric regression and principal component analysis in functional/longitudinal data," *The Annals of Statistics*, 38, 3321–3351.

Medical Image Processing, A. and software package, V. M. (2013), "Version 7.0.1," http://mipav.cit.nih.gov/clickwrap.php.

Shen, D. and Davatzikos, C. (2002), "HAMMER: hierarchical attribute matching mechanism for elastic registration,," *IEEE Trans. on Medical Imaging*, 21, 1421–1439.

Sled, J. G., Zijdenbos, A. P., and Evans., A. C. (1998), "A nonparametric method for automatic correction of intensity nonuniformity in MRI data," *IEEE Transactions on Medical Imaging*, 17, 87–97.

van der Vaar, A. W. and Wellner, J. A. (1996), *Weak Convergence and Empirical Processes*, Springer-Verlag Inc.

Wang, Y., Nie, J., Yap, P., Shi, F., Guo, L., and Shen, D. (2011), "Robust deformable-surface-based skull-stripping for large-scale studies," in *Medical Image Computing and Computer-Assisted Intervention*, eds. Fichtinger, G., Martel, A., and Peters, T., Toronto, Canada: Springer Berlin / Heidelberg, vol. 6893, pp. 635–642.

Zhang, Y., Brady, M., and Smith, S. (2001), "Segmentation of brain MR images through a hidden Markov random field model and the expectation maximization algorithm," *IEEE Trans. on Medical Imaging*, 20, 45–57.

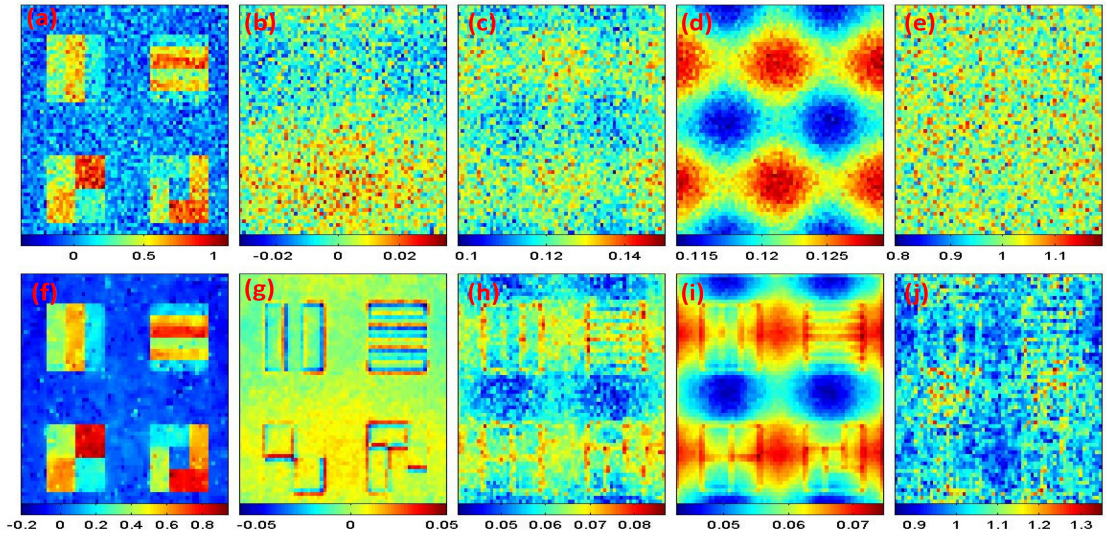

Figure S1: Simulation results: a selected slice of (a) and (f) $\hat{\beta}_3(\mathbf{d}_0; h_s)$; (b) and (g) the biases of $\hat{\beta}_3(\mathbf{d}_0; h_s)$; (c) and (h) the root-mean-square errors (RMSs) of $\hat{\beta}_3(\mathbf{d}_0; h_s)$; (d) and (i) the standard deviation estimates (SDs) of $\hat{\beta}_3(\mathbf{d}_0; h_s)$; and (e) and (f) the ratios of RMS over SD. Upper panels and lower panels correspond to h_0 and h_{10}, respectively.

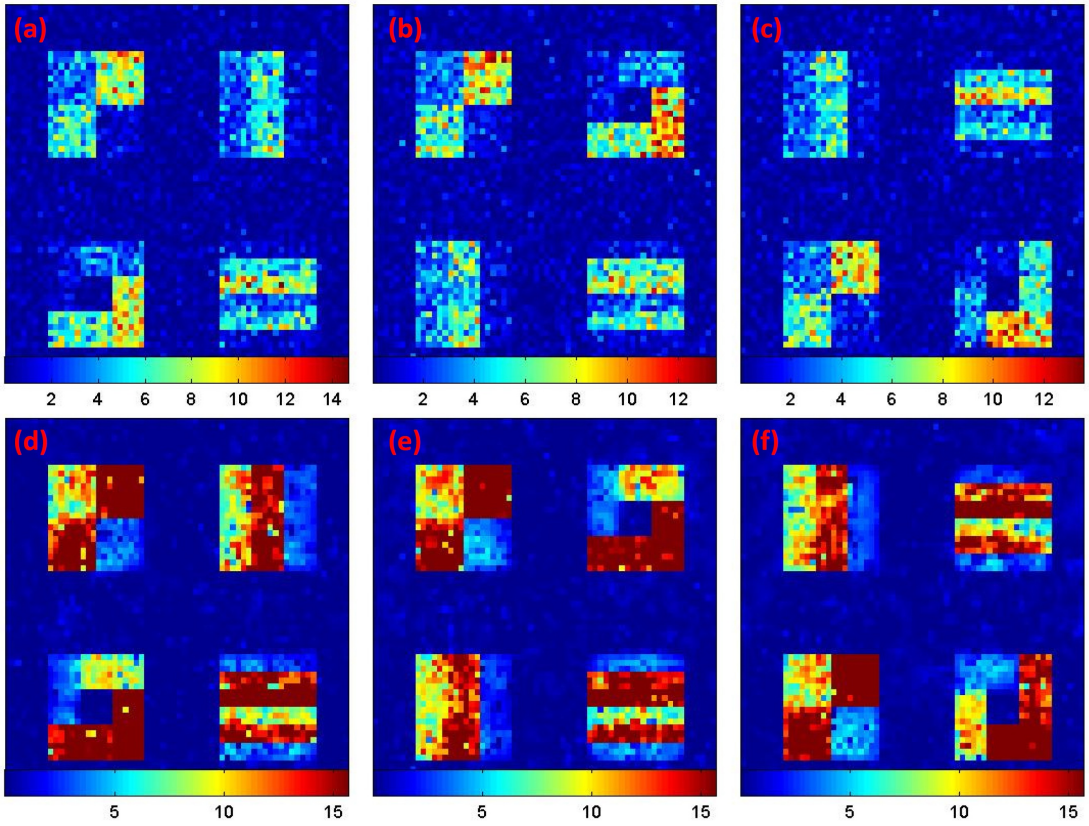

Figure S2: Simulation results: a selected slice of the $-\log_{10}(p)$ images for testing (a) and (**d**) $H_0(\mathbf{d}_0) : \beta_1(\mathbf{d}_0) = 0$; (b) and (e) $H_0(\mathbf{d}_0) : \beta_2(\mathbf{d}_0) = 0$; and (c) and (f) $H_0(\mathbf{d}_0) : \beta_3(\mathbf{d}_0) = 0$. Upper panels and lower panels correspond to h_0 and h_{10}, respectively.

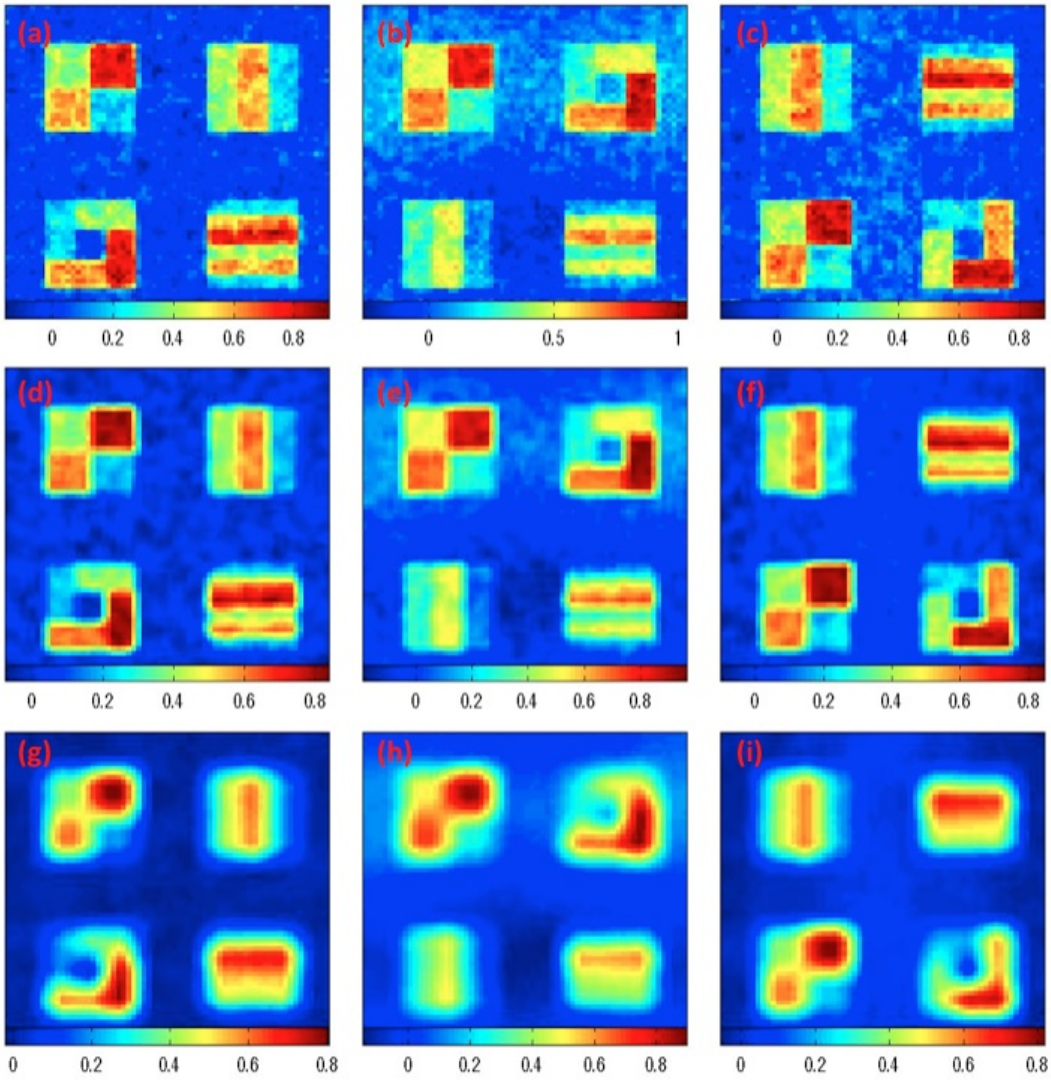

Figure S3: Simulation results from LCE: a selected slide of (a) $\tilde{\beta}_1(\mathbf{d}_0; h_s)$; (b) $\tilde{\beta}_2(\mathbf{d}_0; h_s)$; and (c) $\tilde{\beta}_3(\mathbf{d}_0; h_s)$ with small bandwidth h_s; (d) $\tilde{\beta}_1(\mathbf{d}_0; h_m)$; (e) $\tilde{\beta}_2(\mathbf{d}_0; h_m)$; and (f) $\tilde{\beta}_3(\mathbf{d}_0; h_m)$ with mediate bandwidth h_m; (g) $\tilde{\beta}_1(\mathbf{d}_0; h_l)$; (h) $\tilde{\beta}_2(\mathbf{d}_0; h_l)$; and (i) $\tilde{\beta}_3(\mathbf{d}_0; h_l)$ with large bandwidth h_l;

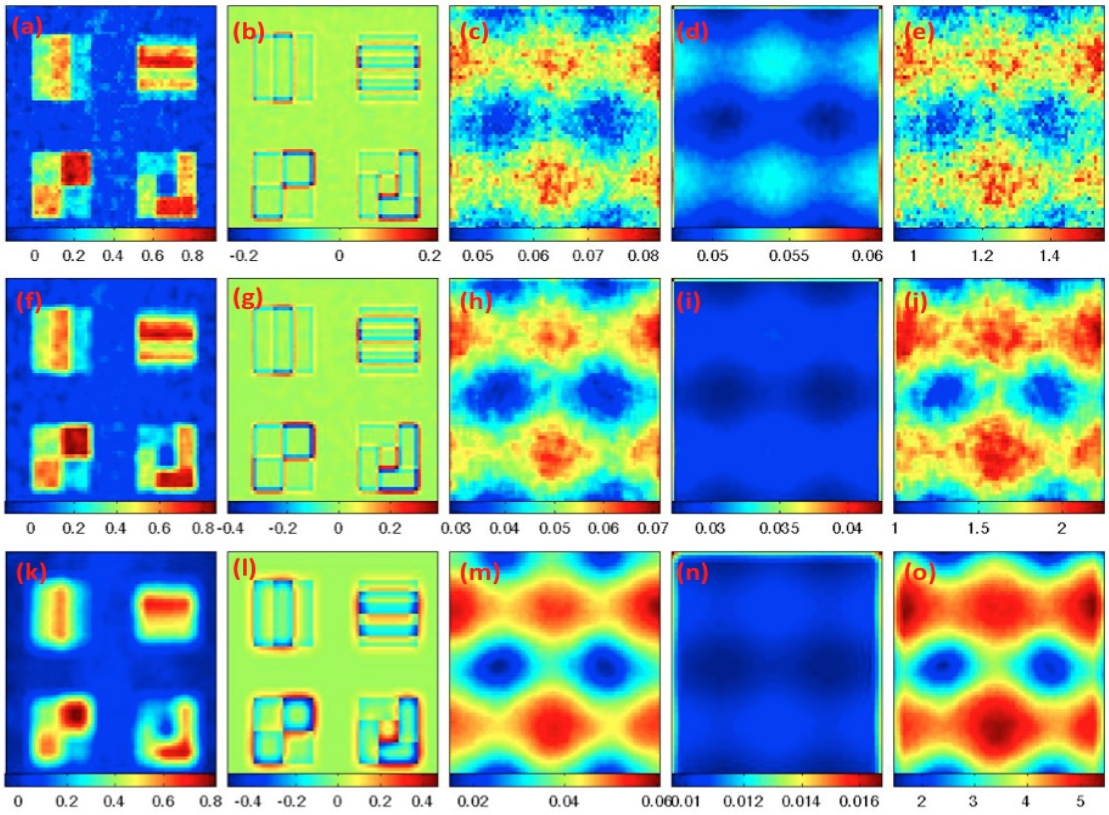

Figure S4: Simulation results from LCE: a selected slice of (a) , (f) and (k) $\tilde{\beta}_3(\mathbf{d}_0; h)$; (b), (g) and (l) the biases of $\tilde{\beta}_3(\mathbf{d}_0; h)$; (c), (h) and (m) the root-mean-square errors (RMSs) of $\tilde{\beta}_3(\mathbf{d}_0; h)$; (d), (i) and (n) the standard deviation estimates (SDs) of $\tilde{\beta}_3(\mathbf{d}_0; h)$; and (e), (j) and (o) the ratios of RMS over SD. Upper, middle and lower panels correspond to bandwidths h_s, h_m and h_l, respectively.

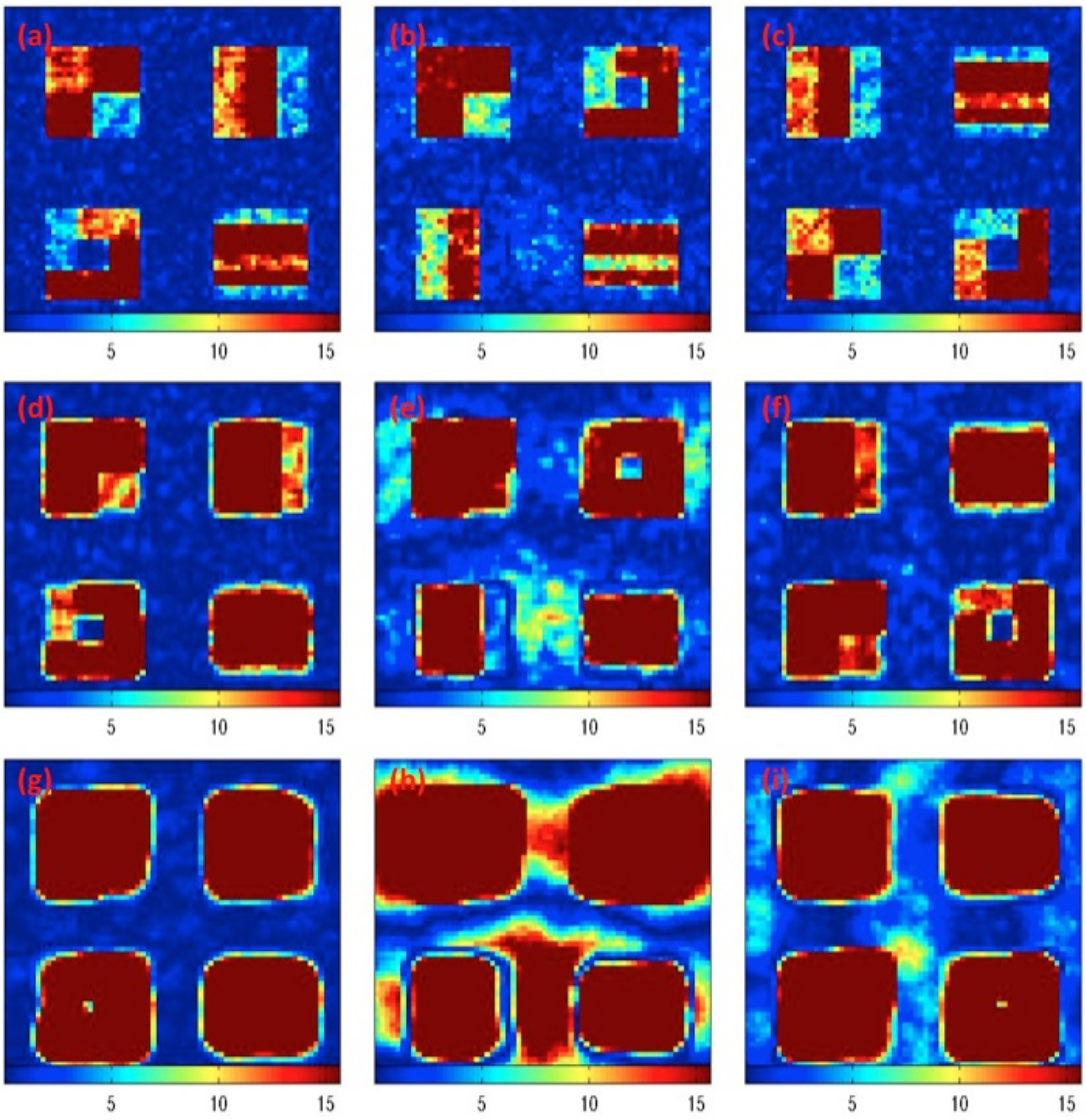

Figure S5: Simulation results from LCE: a selected slice of the $-\log_{10}(p)$ images for testing (a), (d) and (g) $H_0(\mathbf{d}_0) : \beta_1(\mathbf{d}_0) = 0$; (b), (e) and (h) $H_0(\mathbf{d}_0) : \beta_2(\mathbf{d}_0) = 0$; and (c), (f) and (i) $H_0(\mathbf{d}_0) : \beta_3(\mathbf{d}_0) = 0$. Upper, middle and lower panels correspond to bandwidths h_s, h_m and h_l, respectively.

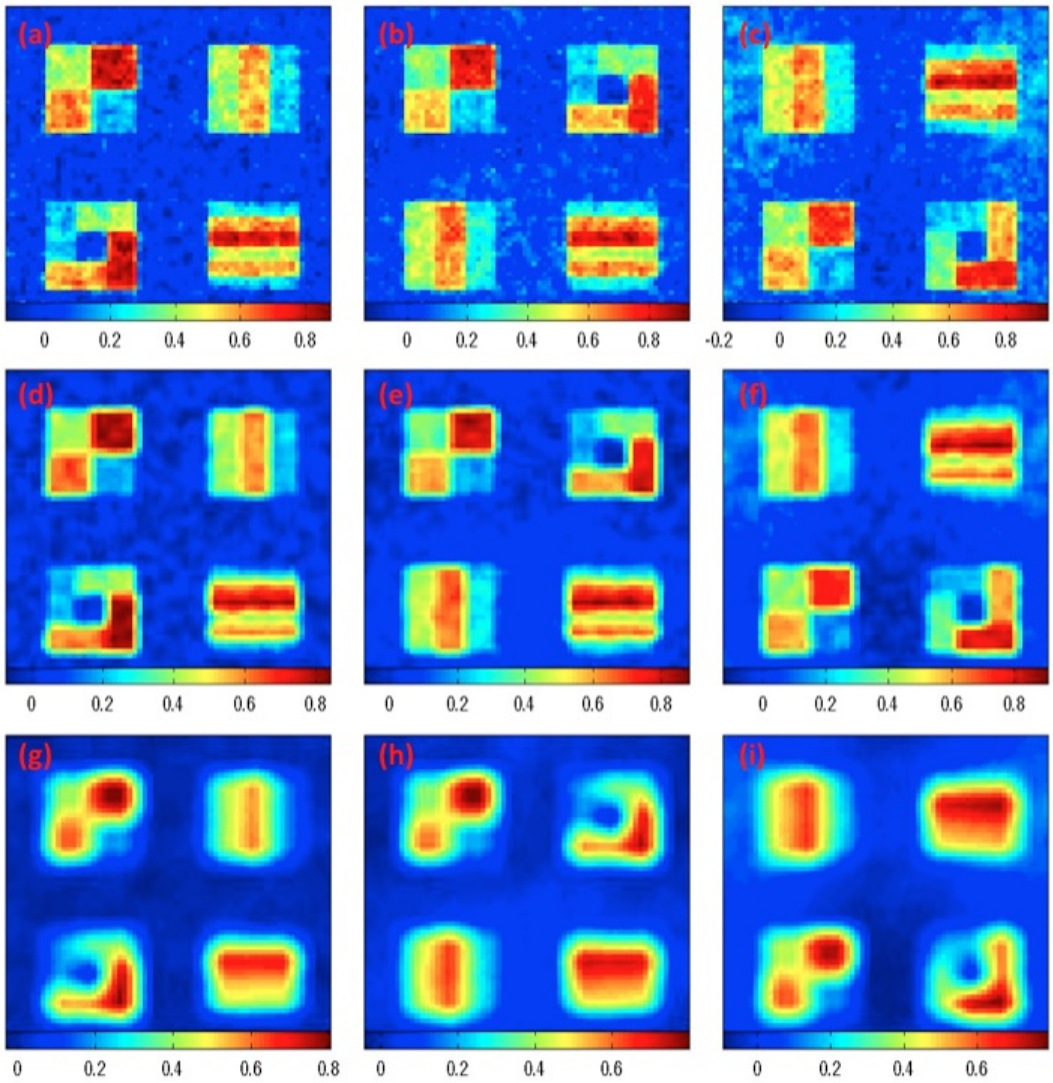

Figure S6: Simulation results from GKS: a selected slide of (a) $\tilde{\beta}_1^o(\mathbf{d}_0; h_s)$; (b) $\tilde{\beta}_2^o(\mathbf{d}_0; h_s)$; and (c) $\tilde{\beta}_3^o(\mathbf{d}_0; h_s)$ with small bandwidth h_s; (d) $\tilde{\beta}_1^o(\mathbf{d}_0; h_m)$; (e) $\tilde{\beta}_2^o(\mathbf{d}_0; h_m)$; and (f) $\tilde{\beta}_3^o(\mathbf{d}_0; h_m)$ with mediate bandwidth h_m; (g) $\tilde{\beta}_1^o(\mathbf{d}_0; h_l)$; (h) $\tilde{\beta}_2^o(\mathbf{d}_0; h_l)$; and (i) $\tilde{\beta}_3^o(\mathbf{d}_0; h_l)$ with large bandwidth h_l;

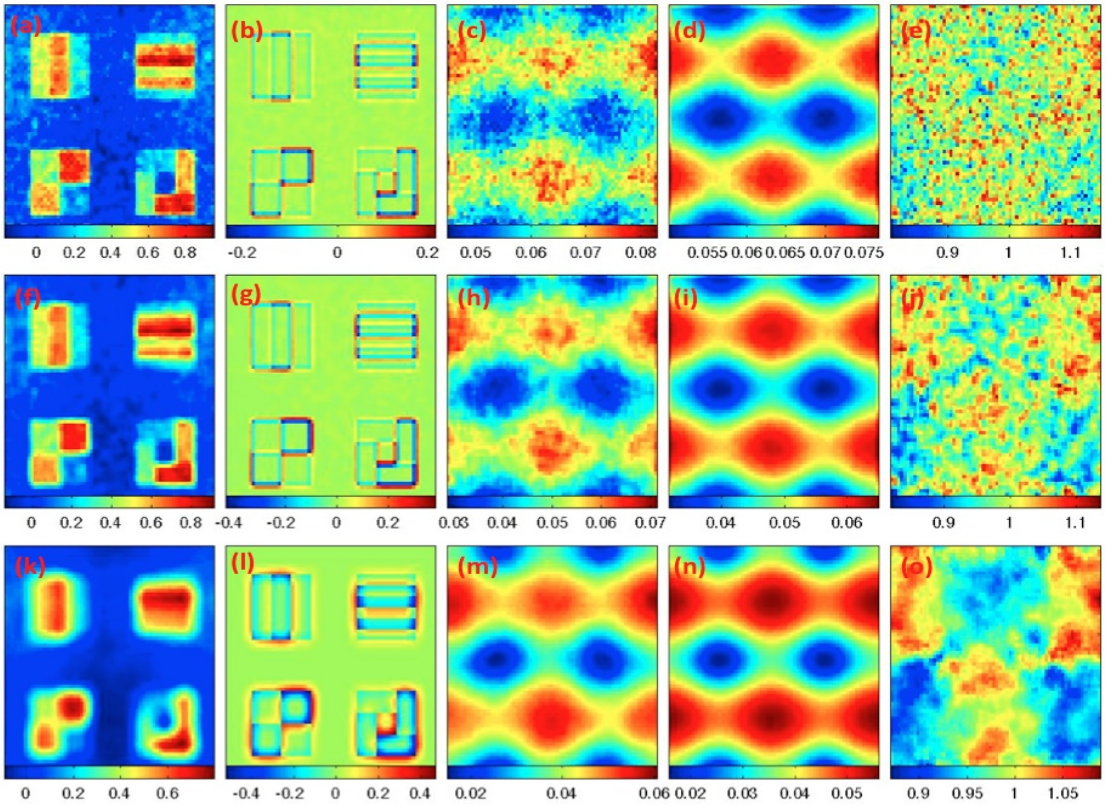

Figure S7: Simulation results from GKS: a selected slice of (a) , (f) and (k) $\tilde{\beta}_3^o(\mathbf{d}_0; h)$; (b), (g) and (l) the biases of $\tilde{\beta}_3^o(\mathbf{d}_0; h)$; (c), (h) and (m) the root-mean-square errors (RMSs) of $\tilde{\beta}_3^o(\mathbf{d}_0; h)$; (d), (i) and (n) the standard deviation estimates (SDs) of $\tilde{\beta}_3^o(\mathbf{d}_0; h)$; and (e), (j) and (o) the ratios of RMS over SD. Upper, middle and lower panels correspond to bandwidths h_s, h_m and h_l, respectively.

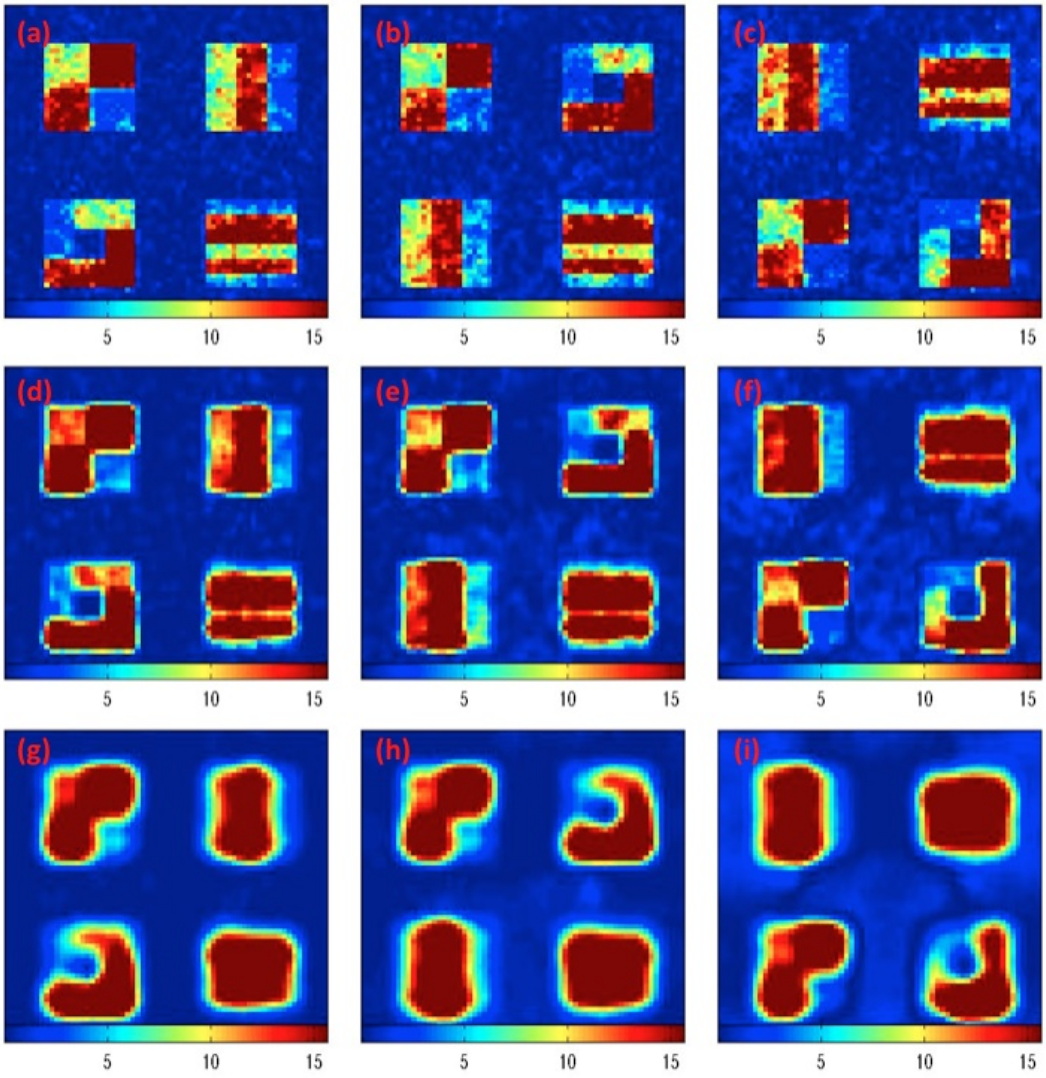

Figure S8: Simulation results from GKS: a selected slice of the $-\log_{10}(p)$ images for testing (a), (d) and (g) $H_0(\mathbf{d}_0) : \beta_1(\mathbf{d}_0) = 0$; (b), (e) and (h) $H_0(\mathbf{d}_0) : \beta_2(\mathbf{d}_0) = 0$; and (c), (f) and (i) $H_0(\mathbf{d}_0) : \beta_3(\mathbf{d}_0) = 0$. Upper, middle and lower panels correspond to bandwidths h_s, h_m and h_l, respectively.

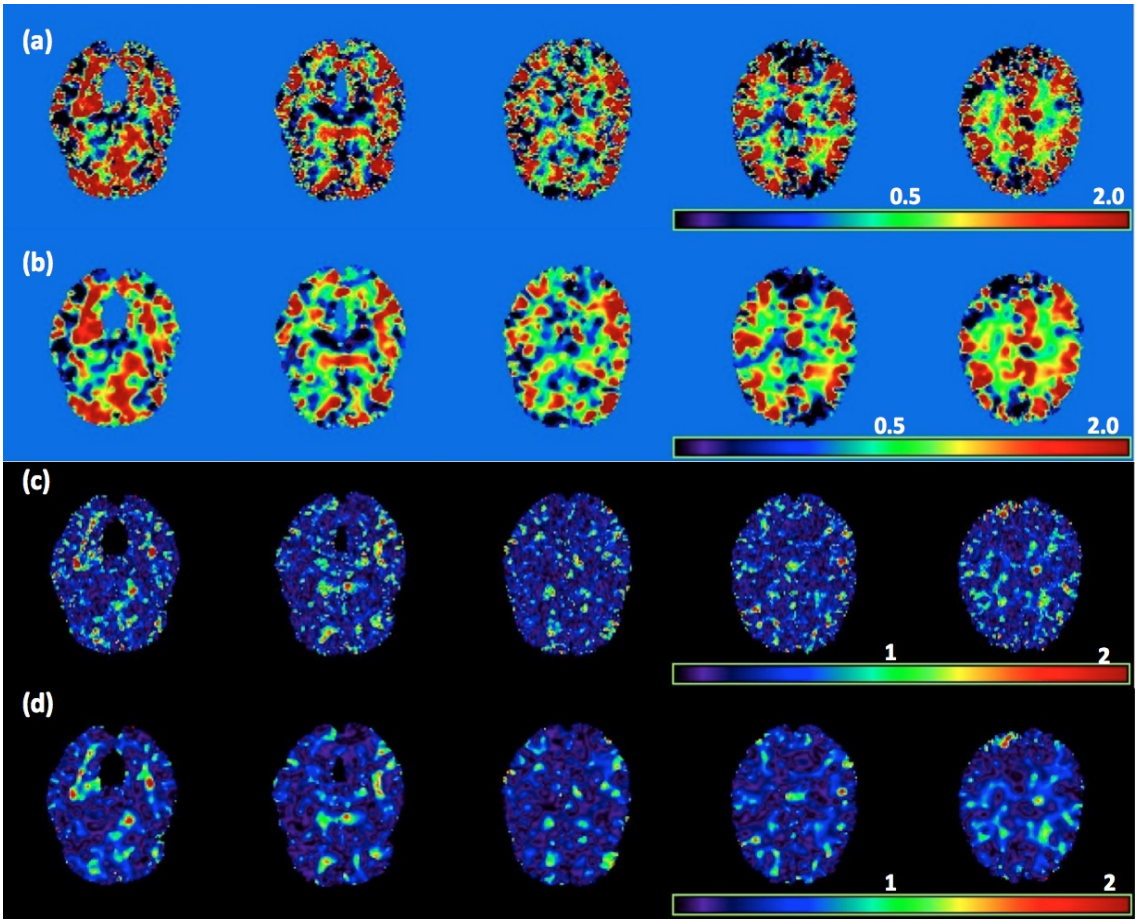

Figure S9: Results from the ADHD 200 data: five selected slices of (a) $\hat{\beta}_7(\mathbf{d}_0; h_0)$, (b) $\hat{\beta}_7(\mathbf{d}_0; h_{10})$, (c) the $-\log_{10}(p)$ images for testing $H_0(\mathbf{d}_0) : \beta_7(\mathbf{d}_0) = 0$ at scale h_0 and (d) at scale h_{10}. Moreover, $\beta_7(\mathbf{d}_0)$ is associated with the gender$\times$diagnosis interaction.

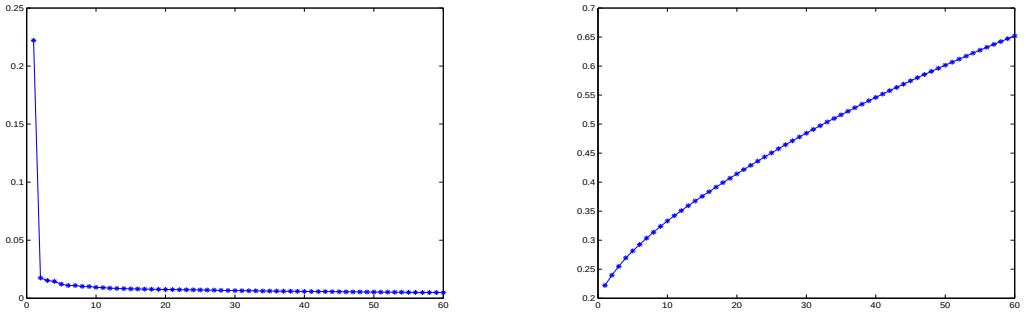

Figure S10: Results from the ADHD 200 data: the first 60 relative eigenvalues of $\hat{\Sigma}_\eta$ (left) and their cumulative variation explained (right).

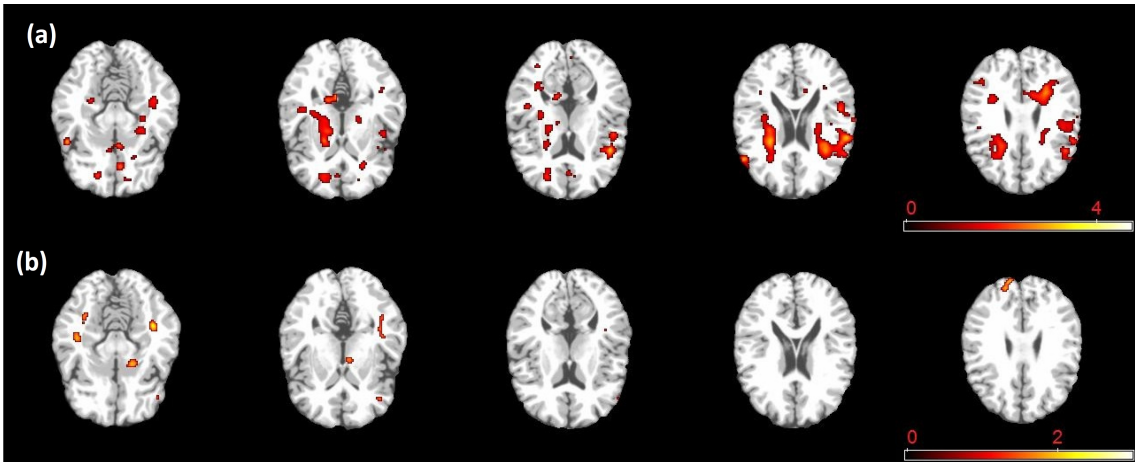

Figure S11: Results from the ADHD 200 data: The 26 and 10 significant blocks to test $H_0 : \beta_6(\mathbf{d}) = 0$ (a) and $H_0 : \beta_7(\mathbf{d}) = 0$ (b) overlaid with $-\log_{10}(p)$ values, respectively, on selected slices, where $\beta_6(\mathbf{d}_0)$ and $\beta_7(\mathbf{d}_0)$ are, respectively, associated with the age×diagnosis and gender×diagnosis interactions.

Table S1: Simulation results from LCE: Average Bias, RMS, SD, and RE of $\beta_2(\mathbf{d}_0)$ parameters in the five ROIs at three different bandwidths (h_s, h_m, h_l), $N(0,1)$ and $\chi^2(3)-3$ distributed data, and 2 different sample sizes $(n = 60, 80)$. BIAS denotes the bias of the mean of estimates; RMS denotes the root-mean-square error; SD denotes the mean of the standard deviation estimates; RE denotes the ratio of RMS over SD. For each case, 200 simulated data sets were used.

		$\chi^2(3)-3$						$N(0,1)$					
		$n=60$			$n=80$			$n=60$			$n=80$		
$\beta_2(\mathbf{d}_0)$		h_s	h_m	h_l	h_s	h_m	h_l	h_s	h_m	h_l	h_s	h_m	h_l
0.0	BIAS	0.01	0.01	0.03	0.01	0.01	0.03	0.01	0.01	0.03	0.01	0.01	0.03
	RMS	0.07	0.05	0.05	0.06	0.04	0.04	0.07	0.05	0.04	0.06	0.04	0.04
	SD	0.05	0.03	0.01	0.05	0.02	0.01	0.05	0.03	0.01	0.05	0.02	0.01
	RE	1.27	1.85	4.39	1.23	1.74	4.04	1.20	1.67	3.82	1.23	1.76	4.09
0.2	BIAS	0.01	0.01	0.02	0.00	0.01	0.02	0.00	0.01	0.01	0.00	0.01	0.02
	RMS	0.07	0.06	0.05	0.06	0.05	0.04	0.07	0.05	0.04	0.06	0.05	0.04
	SD	0.05	0.03	0.01	0.05	0.02	0.01	0.05	0.03	0.01	0.05	0.02	0.01
	RE	1.35	2.08	5.28	1.27	1.89	4.66	1.25	1.84	4.51	1.31	1.99	5.00
0.4	BIAS	-0.01	-0.01	-0.03	0.00	-0.01	-0.02	0.00	-0.01	-0.02	0.00	-0.01	-0.02
	RMS	0.08	0.06	0.06	0.06	0.05	0.04	0.07	0.05	0.05	0.06	0.05	0.04
	SD	0.06	0.03	0.01	0.05	0.02	0.01	0.06	0.03	0.01	0.05	0.02	0.01
	RE	1.38	2.15	5.53	1.28	1.93	4.80	1.26	1.88	4.65	1.33	2.05	5.20
0.6	BIAS	-0.03	-0.06	-0.13	-0.03	-0.06	-0.13	-0.03	-0.06	-0.13	-0.03	-0.06	-0.13
	RMS	0.07	0.06	0.05	0.06	0.04	0.04	0.07	0.05	0.04	0.06	0.05	0.04
	SD	0.05	0.03	0.01	0.05	0.02	0.01	0.05	0.03	0.01	0.05	0.02	0.01
	RE	1.33	2.04	5.13	1.25	1.82	4.44	1.23	1.79	4.34	1.29	1.94	4.82
0.8	BIAS	-0.04	-0.09	-0.20	-0.04	-0.09	-0.20	-0.04	-0.09	-0.20	-0.04	-0.09	-0.20
	RMS	0.08	0.06	0.06	0.06	0.05	0.04	0.07	0.06	0.05	0.06	0.05	0.05
	SD	0.06	0.03	0.01	0.05	0.02	0.01	0.06	0.03	0.01	0.05	0.02	0.01
	RE	1.40	2.22	5.73	1.31	2.00	5.04	1.28	1.92	4.82	1.35	2.11	5.39

Table S2: Simulation Study for $W_\beta(\mathbf{d}_0; h)$ from LCE: estimates (ES) and standard errors (SE) of rejection rates for pixels inside the five ROIs were reported at three different bandwidths (h_s, h_m, h_l), $N(0,1)$ and $\chi^2(3) - 3$ distributed data, and 2 different sample sizes $(n = 60, 80)$ at $\alpha = 5\%$. For each case, 200 simulated data sets were used.

		$\chi^2(3)-3$				$N(0,1)$			
		$n=60$		$n=80$		$n=60$		$n=80$	
$\beta_2(\mathbf{d}_0)$	h	ES	SE	ES	SE	ES	SE	ES	SE
0.0	h_s	0.135	0.078	0.129	0.091	0.115	0.076	0.130	0.093
	h_m	0.335	0.206	0.317	0.215	0.289	0.211	0.312	0.217
	h_l	0.688	0.191	0.688	0.198	0.658	0.205	0.670	0.203
0.2	h_s	0.883	0.070	0.950	0.045	0.892	0.071	0.944	0.047
	h_m	0.969	0.051	0.989	0.028	0.975	0.048	0.986	0.032
	h_l	0.992	0.019	0.998	0.009	0.994	0.016	0.996	0.012
0.4	h_s	1.000	0.002	1.000	0.001	1.000	0.001	1.000	0.000
	h_m	1.000	0.002	1.000	0.001	1.000	0.001	1.000	0.001
	h_l	1.000	0.001	1.000	0.000	1.000	0.001	1.000	0.000
0.6	h_s	1.000	0.000	1.000	0.000	1.000	0.000	1.000	0.000
	h_m	1.000	0.000	1.000	0.000	1.000	0.000	1.000	0.000
	h_l	1.000	0.000	1.000	0.000	1.000	0.000	1.000	0.000
0.8	h_s	1.000	0.000	1.000	0.000	1.000	0.000	1.000	0.000
	h_m	1.000	0.000	1.000	0.000	1.000	0.000	1.000	0.000
	h_l	1.000	0.000	1.000	0.000	1.000	0.000	1.000	0.000

Table S3: Simulation results from GKS: Average Bias, RMS, SD, and RE of $\beta_2(\mathbf{d}_0)$ parameters in the five ROIs at three different bandwidths (h_s, h_m, h_l), $N(0,1)$ and $\chi^2(3)-3$ distributed data, and 2 different sample sizes $(n = 60, 80)$. BIAS denotes the bias of the mean of estimates; RMS denotes the root-mean-square error; SD denotes the mean of the standard deviation estimates; RE denotes the ratio of RMS over SD. For each case, 200 simulated data sets were used.

		$\chi^2(3)-3$						$N(0,1)$					
		$n=60$			$n=80$			$n=60$			$n=80$		
$\beta_2(\mathbf{d}_0)$		h_s	h_m	h_l	h_s	h_m	h_l	h_s	h_m	h_l	h_s	h_m	h_l
0.0	BIAS	0.01	0.01	0.03	0.01	0.01	0.03	0.01	0.01	0.03	0.01	0.01	0.03
	RMS	0.07	0.05	0.05	0.06	0.04	0.04	0.07	0.05	0.04	0.06	0.04	0.04
	SD	0.07	0.05	0.04	0.06	0.04	0.04	0.07	0.05	0.04	0.06	0.04	0.04
	RE	1.03	1.05	1.06	0.99	0.99	0.99	0.97	0.95	0.93	1.00	1.00	0.99
0.2	BIAS	0.01	0.01	0.02	0.00	0.01	0.02	0.00	0.01	0.01	0.00	0.01	0.02
	RMS	0.07	0.06	0.05	0.06	0.05	0.04	0.07	0.05	0.04	0.06	0.05	0.04
	SD	0.07	0.06	0.05	0.06	0.05	0.04	0.07	0.06	0.05	0.06	0.05	0.04
	RE	1.05	1.07	1.09	0.99	0.98	0.96	0.97	0.95	0.93	1.02	1.03	1.03
0.4	BIAS	-0.01	-0.01	-0.03	0.00	-0.01	-0.02	0.00	-0.01	-0.02	0.00	-0.01	-0.02
	RMS	0.08	0.06	0.06	0.06	0.05	0.04	0.07	0.05	0.05	0.06	0.05	0.04
	SD	0.07	0.06	0.05	0.06	0.05	0.04	0.07	0.06	0.05	0.06	0.05	0.04
	RE	1.05	1.08	1.10	0.98	0.97	0.96	0.97	0.95	0.93	1.02	1.03	1.04
0.6	BIAS	-0.03	-0.06	-0.13	-0.03	-0.06	-0.13	-0.03	-0.06	-0.13	-0.03	-0.06	-0.13
	RMS	0.07	0.06	0.05	0.06	0.04	0.04	0.07	0.05	0.04	0.06	0.05	0.04
	SD	0.07	0.05	0.05	0.06	0.05	0.04	0.07	0.05	0.05	0.06	0.05	0.04
	RE	1.05	1.08	1.10	0.98	0.97	0.95	0.97	0.95	0.93	1.02	1.03	1.04
0.8	BIAS	-0.04	-0.09	-0.20	-0.04	-0.09	-0.20	-0.04	-0.09	-0.20	-0.04	-0.09	-0.20
	RMS	0.08	0.06	0.06	0.06	0.05	0.04	0.07	0.06	0.05	0.06	0.05	0.05
	SD	0.07	0.06	0.05	0.06	0.05	0.05	0.07	0.06	0.05	0.06	0.05	0.05
	RE	1.05	1.08	1.09	0.98	0.97	0.96	0.96	0.93	0.92	1.01	1.02	1.03

Table S4: Simulation Study for $W_\beta(\mathbf{d}_0; h)$ from GKS: estimates (ES) and standard errors (SE) of rejection rates for pixels inside the five ROIs were reported at three different bandwidths (h_s, h_m, h_l), $N(0,1)$ and $\chi^2(3) - 3$ distributed data, and 2 different sample sizes $(n = 60, 80)$ at $\alpha = 5\%$. For each case, 200 simulated data sets were used.

		$\chi^2(3)-3$				$N(0,1)$			
		$n=60$		$n=80$		$n=60$		$n=80$	
$\beta_2(\mathbf{d}_0)$	h	ES	SE	ES	SE	ES	SE	ES	SE
0.0	h_s	0.068	0.050	0.064	0.065	0.056	0.051	0.066	0.066
	h_m	0.118	0.194	0.115	0.218	0.100	0.203	0.117	0.216
	h_l	0.199	0.276	0.206	0.307	0.178	0.292	0.206	0.306
0.2	h_s	0.783	0.110	0.886	0.085	0.785	0.117	0.880	0.085
	h_m	0.872	0.142	0.936	0.109	0.879	0.151	0.930	0.108
	h_l	0.897	0.157	0.941	0.122	0.901	0.168	0.937	0.121
0.4	h_s	0.998	0.005	1.000	0.001	0.999	0.004	1.000	0.001
	h_m	0.998	0.014	1.000	0.005	0.999	0.009	1.000	0.004
	h_l	0.995	0.026	0.999	0.010	0.997	0.017	0.999	0.009
0.6	h_s	1.000	0.000	1.000	0.000	1.000	0.000	1.000	0.000
	h_m	1.000	0.000	1.000	0.000	1.000	0.000	1.000	0.000
	h_l	1.000	0.000	1.000	0.000	1.000	0.000	1.000	0.000
0.8	h_s	1.000	0.000	1.000	0.000	1.000	0.000	1.000	0.000
	h_m	1.000	0.000	1.000	0.000	1.000	0.000	1.000	0.000
	h_l	1.000	0.000	1.000	0.000	1.000	0.000	1.000	0.000

Table S5: Results from the ADHD 200 data: the first two largest significant regions of the first six largest significant blocks for hypothesis tests $H_0 : \beta_6(\mathbf{d}) = 0$ and $H_0 : \beta_7(\mathbf{d}) = 0$ with block and region voxel sizes. WM, L and R, respectively, represent white matter, left hemisphere, and right hemisphere. Moreover, $\beta_6(\mathbf{d}_0)$ and $\beta_7(\mathbf{d}_0)$ are, respectively, associated with the age×diagnosis ($A \times D$) and gender×diagnosis ($G \times D$) interactions.

			1st largest predefined ROI		2nd largest predefined ROI	
	block	size	ROI label	size	ROI label	size
A × D	1	3954	frontal lobe WM L	1567	frontal lobe WM R	455
	2	2065	frontal lobe WM R	900	anterior limb of internal capsule R	220
	3	1642	nucleus accumbens L	1019	frontal lobe WM L	213
	4	1143	parietal lobe WM R	688	superior parietal lobule R	132
	5	282	frontal lobe WM R	260	lateral front-orbital gyrus R	22
	6	250	temporal lobe WM L	131	frontal lobe WM L	35
G × D	1	228	temporal lobe WM L	184	middle temporal gyrus L	22
	2	216	frontal lobe WM L	163	superior frontal gyrus L	33
	3	95	temporal lobe WM R	66	lateral occipitotemporal gyrus R	21
	4	94	medial frontal gyrus R	44	frontal lobe WM R	24
	5	89	frontal lobe WM L	49	globus palladus L	21
	6	83	superior occipital gyrus R	71	occipital lobe WM R	7