



Published in final edited form as:

IEEE Trans Med Imaging. 2012 May ; 31(5): 1100–1112. doi:10.1109/TMI.2012.2185830.

TwinMARM: Two-stage Multiscale Adaptive Regression Methods for Twin Neuroimaging Data*

Yimei Li^a, John H. Gilmore^c, Jiaping Wang^b, Martin Styner^c, Weili Lin^{d,e}, and Hongtu Zhu^{b,e}

^aDepartments of Biostatistics, St. Jude Children's Research Hospital, 262 Danny Thomas Place Memphis, TN 38105-3678

^bDepartment of Biostatistics, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599, USA

^cDepartment of Psychiatry, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599, USA

^dDepartment of Radiology, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599, USA

^eBiomedical Research Imaging Center, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599, USA

Abstract

Twin imaging studies have been valuable for understanding the relative contribution of the environment and genes on brain structures and their functions. Conventional analyses of twin imaging data include three sequential steps: spatially smoothing imaging data, independently fitting a structural equation model at each voxel, and finally correcting for multiple comparisons. However, conventional analyses are limited due to the same amount of smoothing throughout the whole image, the arbitrary choice of smoothing extent, and the decreased power in detecting environmental and genetic effects introduced by smoothing raw images. The goal of this article is to develop a two-stage multiscale adaptive regression method (TwinMARM) for spatial and adaptive analysis of twin neuroimaging and behavioral data. The first stage is to establish the relationship between twin imaging data and a set of covariates of interest, such as age and gender. The second stage is to disentangle the environmental and genetic influences on brain structures and their functions. In each stage, TwinMARM employs hierarchically nested spheres with increasing radii at each location and then captures spatial dependence among imaging observations via consecutively connected spheres across all voxels. Simulation studies show that our TwinMARM significantly outperforms conventional analyses of twin imaging data. Finally, we use our method to detect statistically significant effects of genetic and environmental variations on white matter structures in a neonatal twin study.

Keywords

Heritability; Multiscale adaptive regression model; Smooth; Structural equation model; Twin study

*This work was supported in part by NIH grants RR025747-01, P01CA142538-01, MH086633, and AG033387 to Dr. Zhu, NIH grants MH064065, HD053000, and MH070890 to Dr. Gilmore, NIH grants R01NS055754 and R01EB5-34816 to Dr. Lin, Lilly Research Laboratories, the UNC NDRC HD 03110, Eli Lilly grant F1D-MC-X252, and NIH Roadmap Grant U54 EB005149-01, NAMIC to Dr. Styner. The content is solely the responsibility of the authors and does not necessarily represent the official views of the NIH. The readers are welcome to request reprints from Dr. Hongtu Zhu., hzhu@bios.unc.edu; phone: 919-966-7272.

1 Introduction

The typical design of a twin study compares the similarity of monozygotic twins (MZ), who share the same genetic polymorphisms, to that of dizygotic twins (DZ), who share an average of 50% of their genetic polymorphisms. Under the equal environment assumption across zygosity, the known differences in genetic similarity allows us to dissect the effects of genes and environment on a known phenotype, such as total brain volume and volumes of subcortical structures. Thus, twin neuroimaging studies have been a valuable source of information for evaluating the inheritance of brain structure and function by disentangling genetic factors from environment [1, 2, 3, 4, 5, 6, 7, 8]. For instance, high heritability was found in different brain volumes including intracranial volume, total brain volume, and global gray and white matter volumes [9, 3, 10]. Moreover, cortical thickness in the sensorimotor cortex, middle frontal cortex, and anterior temporal cortex were found to be influenced by genetic factors [5, 11].

The voxel-wise methods for analyzing twin neuroimaging data on a two-dimensional (2D) surface (or in a three-dimensional (3D) volume) are sequentially executed in three steps. The first step is to use standard smoothing methods to spatially smooth the imaging data [5, 6, 7]. These smoothing methods are commonly independent of the imaging data and apply the same amount of smoothing throughout the whole image. See, for example, [12] for overviews of some smoothing methods in the neuroimaging literature. The second step involves fitting statistical models to imaging measures from all twin pairs at each voxel or surface location separately to generate a parametric map of test statistics (or p -values) [13, 14]. The most commonly used models are structural equation models, which primarily decompose the observed phenotypic variance into additive genetic, dominant genetic, common environmental components, and random noise [15, 16]. The third step is to compute adjusted p -values in order to correct for testing multiple hypotheses across thousands to millions of locations using multiple comparison correction methods (e.g., random field theory (RFT), false discovery rate (FDR), or permutation methods)[17, 18, 19, 20, 21, 22].

The existing voxel-wise methods based on Gaussian kernel smoothing and structural equation models have two major limitations for analyzing twin neuroimaging data, which underscore the great need for further methodological development. It is well known that the commonly used Gaussian kernel for smoothing imaging data usually blurs the image data near the edges of the significant regions, which can dramatically increase the numbers of false positives and negatives. As shown in [23, 24], the existing voxel-wise methods can suffer from the arbitrary choice of smoothing extent in the initial smoothing step, which leads to the large number of false positive and false negative results. An overview of different smoothing methods that are used in the neuroimaging literature can be found in [12]. Generally, most methods for smoothing raw twin imaging data take a weighted average of the imaging data at voxels in the neighborhood of each voxel, thus are primarily associated with the mean structure of imaging data. Such a way of smoothing raw twin imaging data can change the variance structure of imaging data, which is primarily associated with genetic and environmental factors [25, 16, 15, 13, 14]. Thus, directly smoothing twin imaging data can introduce substantial bias in estimating these factors and lead to a dramatic increase of the numbers of false positives and negatives. We will investigate the consequence of smoothing twin imaging data by theoretical arguments in the Appendix and using simulated data in Section 3. Recently, [26] examined how the overall significance of voxel-wise effects varied with respect to full width at half maximum (FWHM) of the Gaussian smoothing applied to the fractional anisotropy (FA) images. Their results have shown that raw FAmaps with no smoothing yielded the greatest sensitivity to detect gene effects. Our results give a theoretical justification for their results.

In general, directly smoothing raw twin imaging data by using Gaussian kernel should be avoided in practice.

As pointed out by [20] and many others, these voxel-wise methods essentially treat all voxels as independent units in the modeling stage, and they explicitly model the spatial coherence and incorporate the information that spatially contiguous regions of effect with rather sharp edges may exist in neuroimaging data. At least two groups of statistical methods have been developed to characterize spatial dependence in neuroimaging data. The first group of methods is to explicitly incorporate the spatial non-independence in the p -value or testing statistic images. These methods usually combine strength from both the magnitude of testing statistic and cluster extent, where the cluster is defined as the number of contiguous significant voxels above a specific threshold [20, 21, 22].

Instead of solely utilizing information in the p -value or testing statistic images, the other group of methods explicitly incorporates the spatial dependence in their models. It is common to use conditional autoregressive (CAR), Markov random field (MRF), and other spatial correlation priors to characterize spatial dependence among spatially connected voxels [27, 28, 29, 30, 31, 32]. In addition, a wealth of image denoising and restoration methods has been developed for pre-processing a single image in the literature. An overview of image denoising and restoration methods has been given in [33, 34] and references therein. For instance, Polzehl et al.[35, 36] developed a propagation separation(PS) approach to adaptively and spatially smooth images from a single image without detecting edge curves/surfaces explicitly. In Qiu et al.[37, 38], novel three-dimensional (3D) denoising procedures were proposed to nonparametrically estimate a 3-D jump surface from noisy data. In Tabelow et al. [39, 40, 41], the original PS idea was adapted to develop a multiscale adaptive linear model to adaptively and spatially denoise functional magnetic resonance imaging (fMRI) and diffusion tensor images (DTI) from a single subject. Under the linear model for fMRI, In Polzehl et al.[42], a structural adaptive segmentation algorithm was developed to integrate the signal detection with noise detection in one procedure. In Li et al. [43], a multiscale adaptive regression model(MARM) was proposed to integrate the PS approach [35, 36] with general statistical modeling at each voxel for spatial and adaptive analysis of neuroimaging data from multiple subjects in various study designs.

The aim of this article is to develop a pipeline, called TwinMARM, for the spatial and adaptive analysis of twin neuroimaging data. TwinMARM integrates adaptive estimation and testing in a single framework. TwinMARM consists of two stages of multiscale adaptive regression models (MARM). Each stage of TwinMARM constructs hierarchical nested neighborhoods with increasing radii at all voxels, adaptively generates weighted quasi-likelihood functions, and efficiently utilizes available information to estimate parameters. Particularly, in the first stage, weighted quasi-likelihood functions are constructed to capture the mean structure of imaging data in order to establish the relationship between twin imaging data and a set of covariates of interest, such as age and gender. Subsequently, in the second stage, a set of new weighted quasi-likelihood functions is built that utilizes the variance structure of imaging data to disentangle relative contributions of environment and genes on brain structures and their functions.

Compared to all existing methods including MARM and PS, we make several major contributions. TwinMARM represents a novel generalization of MARM and PS. Specifically, PS has been primarily used to solely average imaging intensity, which improves the accuracy in estimating the mean structure of imaging data [42, 39, 40, 41], but we are more interested in variance structure of imaging data in Twin studies. One may directly apply MARM to the structural equation model (SEM) described in Section 2, but we have found that it is computationally difficult to directly optimize the weight likelihood

function for SEM. Moreover, the existing MARM and PS methods can only smooth all unknown parameters (or a subset of parameters) simultaneously, since the weights used in MARM and PS depend the weighted distance of the whole parameter vectors. However, we cannot estimate the mean and variance structure of twin imaging data simultaneously. At the same time, extending MARM and PS to smoothing individual parameters is an unsolved problem, since it is very difficult to approximate the covariance matrix of the smoothed parameter estimates obtained from MARM and PS. Fortunately, due to the special model structure for twin data, we are able to find a novel approach, which allows us to apply MARM twice to independently smooth the mean and variance structure of twin imaging data. Theoretically, we have shown in the Appendix that our approach is valid asymptotically. TwinMARM slightly increases the amount of computational time in computing parameter estimates and test statistics, whereas it substantially outperforms the voxel-wise approach in increasing the accuracy of parameter estimates and the power of test statistics. Practically, TwinMARM developed here is applicable to different imaging modalities including DTI, fMRI, and MRI among many others. Particularly, TwinMARM gives a solution to the issue raised in [26]. Due to the importance and popularity of twin imaging studies, TwinMARM will have wide and important applications in the analysis of imaging, genetic, and clinical data collected from twin studies.

2 Methods

2.1 Structural Equation Model

Suppose we observe imaging measures and clinical variables from n_1 MZ twin pairs and $n_2 = n - n_1$ DZ twin pairs. Imaging measures can be volumes of anatomical regions, the shape representation of the surfaces of cortical or subcortical structures, fMRI signals, diffusion tensors, and tensor-derived measures [44, 45, 46, 47, 48, 49, 50, 5, 6, 7, 8]. Clinical variables may include demographic and environmental variables. Specifically, for the j -th subject in the i -th twin pair, we observe an $N_V \times 1$ vector of imaging measures, denoted by $Y_{ij} = \{y_{ij}(v) : v \in \mathcal{V}\}$, and a $k \times 1$ vector of clinical variables $x_{ij} = (x_{ij1}, \dots, x_{ijk})^T$ for $i = 1, \dots, n$ and $j = 1, 2$, where x_{ij1} is commonly set as 1 and \mathcal{V} and v , respectively, represent a specific brain region and a voxel in \mathcal{V} . For notational simplicity, we only consider univariate measure.

At a specific voxel v , we consider the structural equation model

$$y_{ij}(v) = x_{ij}^T \beta(v) + a_{ij}(v) + d_{ij}(v) + c_i(v) + e_{ij}(v) \quad (1)$$

for $j = 1, 2$ and $i = 1, \dots, n = n_1 + n_2$, where $\beta(v) = (\beta_1(v), \dots, \beta_k(v))^T$ is a $k \times 1$ vector representing unknown regression coefficients, $a_{ij}(v)$, $d_{ij}(v)$, $c_i(v)$ and $e_{ij}(v)$ are, respectively, the additive genetic, dominance genetic, common environmental, and residual effects on the i -th twin pair. It is common to assume that $a_{ij}(v)$, $d_{ij}(v)$, $c_i(v)$ and $e_{ij}(v)$ are independently normally distributed with means 0 and variances $\sigma_a^2(v)$, $\sigma_d^2(v)$, $\sigma_c^2(v)$, and $\sigma_e^2(v)$, respectively [16, 15, 13, 14, 51]. Moreover, $\text{Cov}(a_{11}(v), a_{21}(v))$ equals $\sigma_a^2(v)$ for MZ twins and $\sigma_a^2(v)/2$ for DZ twins, while $\text{Cov}(d_{11}(v), d_{21}(v))$ equals $\sigma_d^2(v)$ for MZ twins and $\sigma_d^2(v)/4$ for DZ twins [52]. See a path diagram of model (1) in Figure 1. Due to an identifiability issue, we may not be able to dissect the additive genetic effect, the dominance genetic effect, and the common environmental effect from the residual environmental effect [16, 15, 13, 14]. It is common to consider two simpler models. One is the ACE model, which only includes additive genetic, common environmental, and residual effects. The other is the ADE model, which includes additive and dominance genetic, and residual effects.

Without loss of generality, we focus on the ACE model and drop the assumption of normality for $a_{ij}(v)$, $d_{ij}(v)$, $c_i(v)$ and $e_{ij}(v)$ from now on, since TwinMARM is independent of such assumption. For the ACE model, if we are able to estimate all variances

$\eta(v) = (\sigma_a^2(v), \sigma_c^2(v), \sigma_e^2(v))$, we can disentangle relative contribution of environment and genetic polymorphisms. Specifically, the genetic heritability at voxel v , denoted by $g(v)$, is defined as the ratio of the genetic variance to the total variance of $y_{ij}(v)$ given by

$$g(v) = \frac{\sigma_a^2(v)}{\sigma_a^2(v) + \sigma_c^2(v) + \sigma_e^2(v)}.$$

2.2 TwinMARM

We propose the two-stage multiscale adaptive regression method (TwinMARM) for the analysis of twin imaging data and behavioral data as follows. The first stage is to estimate $\beta = \{\beta(v) : v \in \mathcal{V}\}$, while the second stage is to estimate $\eta = \{\eta(v) : v \in \mathcal{V}\}$. In each stage, we reformulate the problem of estimating β (or η) as a regression model and then apply the multiscale adaptive regression model (MARM) in [43], which can be regarded as a generalization of the propagation-separation (PS) procedure in multiple subjects [35, 53, 36, 39, 40, 41]. The key ideas of MARM and PS are to construct hierarchical nested neighborhoods with increasing radii at all voxels, to adaptively construct weighted quasi-likelihood functions to estimate parameter estimates, and to increase the power of test statistics in detecting subtle changes of brain structures and their functions. A path diagram of TwinMARM is presented in Figure 2.

2.2.1 TwinMARM: Stage I—The first stage is to estimate $\beta(v)$ and carry out statistical inference on $\beta(v)$. Specifically, we consider a bivariate regression model given by

$$Y_i(v) = \begin{pmatrix} y_{i1}(v) \\ y_{i2}(v) \end{pmatrix} = \mathbf{x}_i^T \beta(v) + \mathbf{f}_i(v) = \begin{pmatrix} \mathbf{x}_{i1}^T \\ \mathbf{x}_{i2}^T \end{pmatrix} \beta(v) + \begin{pmatrix} f_{i1}(v) \\ f_{i2}(v) \end{pmatrix}, \quad (2)$$

where $\mathbf{x}_i = [\mathbf{x}_{i1} \quad \mathbf{x}_{i2}]$ and $\mathbf{f}_i(v) = (f_{i1}(v), f_{i2}(v))^T$, in which $f_{ij}(v) = a_{ij}(v) + c_i(v) + e_{ij}(v)$ for the ACE model. Although $\mathbf{f}_i(v)$ has different covariance structures for MZ and DZ, respectively, we assume that $\mathbf{f}_i(v)$ has mean zero and covariance $\Sigma_f(v)$ for all i to avoid estimating variance components ($\sigma_a^2(v)$, $\sigma_c^2(v)$, $\sigma_e^2(v)$) in Stage I, which leads a simple procedure for estimating β . We will appropriately account for the misspecification of covariance of $\mathbf{f}_i(v)$ in our statistical inference on $\beta(v)$.

Since all components of $\beta(v)$ are the parameters of interest and $\Sigma_f(v)$ can be regarded as nuisance parameters, we first estimate $\Sigma_f(v)$ across all voxels and then fix them at their estimated values. Specifically, let $\mathcal{L}(Y_i(v) | \mathbf{x}_i, \beta(v), \Sigma_f(v))$ be the pseudo-likelihood of $Y_i(v)$ given by

$$-0.5 \log |\Sigma_f(v)| - 0.5 \{Y_i(v) - \mathbf{x}_i^T \beta(v)\}^T \Sigma_f^{-1}(v) \{Y_i(v) - \mathbf{x}_i^T \beta(v)\}. \quad (3)$$

We calculate the maximum pseudo-likelihood estimate of $(\beta(v), \Sigma_f(v))$ given by

$$(\widehat{\beta}(v), \widehat{\Sigma}_f(v)) = \arg \max_{(\beta(v), \Sigma_f(v))} \ell(Y_i(v) | \mathbf{x}_i, \beta(v), \Sigma_f(v)). \quad (4)$$

Let $\mathbf{a}^{\otimes 2} = \mathbf{a}\mathbf{a}^T$ for any vector \mathbf{a} and $(\widehat{\beta}(v)^{(t)}, \widehat{\Sigma}_f(v)^{(t)})$ be the estimate of $(\beta(v), \Sigma_f(v))$ at the t -th iteration. Computationally, at each voxel v , by setting $\widehat{\Sigma}_f(v)^{(0)} = \mathbf{I}_2$, a 2×2 identity matrix, we iteratively update

$$\begin{aligned} \widehat{\beta}(v)^{(t+1)} &= \left\{ \sum_{i=1}^n \mathbf{x}_i \widehat{\Sigma}_f^{-1}(v)^{(t)} \mathbf{x}_i^T \right\}^{-1} \sum_{i=1}^n \mathbf{x}_i \widehat{\Sigma}_f^{-1}(v)^{(t)} Y_i(v), \\ \widehat{\Sigma}_f(v)^{(t+1)} &= (n-2)^{-1} \sum_{i=1}^n \{Y_i(v) - \mathbf{x}_i^T \widehat{\beta}(v)^{(t+1)}\}^{\otimes 2} \end{aligned} \quad (5)$$

until that the difference between $(\widehat{\beta}(v)^{(t+1)}, \widehat{\Sigma}_f(v)^{(t+1)})$ and $(\widehat{\beta}(v)^{(t)}, \widehat{\Sigma}_f(v)^{(t)})$ is smaller than a prefixed constant, say 10^{-6} . From now on, $\Sigma_f(v)$ will be fixed at $\widehat{\Sigma}_f(v)$ across all voxels v . Statistically, it can be shown that such a substitution has a negligible effect on statistical inference of β .

To estimate $\beta(v)$ at voxel v , we construct a weighted quasi-likelihood function by utilizing all imaging data in a sphere with radius h at voxel v , denoted by $B(v, h)$. Let $\omega(v, v'; h) \in [0, 1]$ be a weight function of two voxels and a radius h , which characterizes the similarity between the data in voxels v and v' such that $\sum_{v' \in B(v, h)} \omega(v, v'; h) = 1$ for all $h > 0$. If $\omega(v, v'; h) \gg 0$, it represents that the data in voxels v and v' are similar to each other, whereas $\omega(v, v'; h) \approx 0$ indicates that the data in voxel v' do not contain too much information on $\beta(v)$. The adaptive weights $\omega(v, v'; h)$ play an important role in preventing oversmoothing the estimates of $\beta(v)$ as well as preserving the edges of significant regions. We utilize all the data $\{Y_{\lambda}(v') : v' \in B(v, h)\}$ to construct the weighted log-likelihood function at voxel v at scale h , which is given by

$$\ell_{obs}(\beta(v); h) = -0.5 \sum_{i=1}^n \sum_{v' \in B(v, h)} w(v, v'; h) \{Y_i(v) - \mathbf{x}_i^T \beta(v)\}^T \widehat{\Sigma}_f^{-1}(v') \{Y_i(v) - \mathbf{x}_i^T \beta(v)\}. \quad (6)$$

By directly maximizing $\ell_{obs}(\beta(v); h)$, we can obtain the maximum pseudo-likelihood estimate of $\beta(v)$, denoted by $\widehat{\beta}(v; h)$, which is given by

$$\widehat{\beta}(v; h) = \Sigma_{n,1}^{-1}(v; h) \sum_{v' \in B(v, h)} \omega(v, v'; h) \sum_{i=1}^n \mathbf{x}_i \widehat{\Sigma}_f^{-1}(v') Y_i(v'), \quad (7)$$

where $\Sigma_{n,1}(v; h) = \sum_{v' \in B(v, h)} \omega(v, v'; h) \sum_{i=1}^n \mathbf{x}_i \widehat{\Sigma}_f^{-1}(v') \mathbf{x}_i^T$. The covariance matrix of $\widehat{\beta}(v; h)$ can be approximated by

$$\text{Cov}(\widehat{\beta}(v; h)) \approx \Sigma_n(\widehat{\beta}(v; h)) = \Sigma_{n,1}^{-1}(v; h) \Sigma_{n,2}(v; h) \Sigma_{n,1}^{-1}(v; h), \quad (8)$$

where $\Sigma_{n,2}(v; h) = \sum_{i=1}^n \left[\sum_{v' \in B(v, h)} \omega(v, v'; h) \mathbf{x}_i \widehat{\Sigma}_f^{-1}(v') \{Y_i(v) - \mathbf{x}_i^T \widehat{\beta}(v; h)\}^{\otimes 2} \right]$. Although [39] have obtained the same $\widehat{\beta}(v; h)$ as in expression (7), the covariance estimate of $\widehat{\beta}(v; h)$ in expression (8) has a simple form compared to the corresponding one in [39] with an additional spatial correlation assumption. We have found that such spatial correlation

assumption is unnecessary since $\Sigma_{n,2}(v; h)$ has implicitly incorporated the spatial correlation in the data.

Based on $\hat{\beta}(v; h)$, we can further construct test statistics to examine scientific questions about $\beta(v)$. For instance, such questions may compare brain structure across different groups (MZ versus DZ) or detect change in brain structure across time [3, 4, 5, 6, 7, 8]. These questions can be formulated as the linear hypotheses about $\beta(v)$

$$H_{0,\mu}: R_1\beta(v)=b_0 \quad \text{vs.} \quad H_{1,\mu}: R_1\beta(v) \neq b_0, \quad (9)$$

where $\mu = R_1\beta(v)$, R_1 is a $r \times k$ matrix of full row rank and b_0 is a $r \times 1$ specified vector. We test the null hypothesis $H_{0,\mu}: R_1\beta(v) = b_0$ using the Wald test statistic

$$W_{\mu}(v;h)=\{R_1\widehat{\beta}(v;h)-\mathbf{b}_0\}^T \{R_1\Sigma_n(\widehat{\beta}(v;h))R_1^T\}^{-1} \{R_1\widehat{\beta}(v;h)-\mathbf{b}_0\}. \quad (10)$$

Under $H_{0,\mu}$, $W_{\mu}(v; h)$ is asymptotically distributed as $\chi^2(r)$ [43].

2.2.2 Propagation-Seperation procedure—Following the propagation-seperation (PS) procedure proposed in [35, 53, 36], [39, 40, 41] and [43], we adaptively determine $\{\omega(v, v'; h) : v, v' \in \mathcal{V}\}$ and then calculate $\hat{\beta}(v; h)$ and $W_{\mu}(v; h)$ as h increases from 0 to a preset value r_0 . Since PS and the choice of its associated parameters have been described in details in [35, 53, 36] and [43], we briefly discuss them here for the sake of simplicity.

The key idea of PS is to build a sequence of nested spheres with increasing radii $h_0 = 0 < h_1 < \dots < h_S = r_0$ at each voxel $v \in \mathcal{V}$. When $h = h_1$, we extract features from the imaging data as well as $\{\hat{\beta}(v) : v \in \mathcal{V}\}$ and compute weights $\omega(v, v'; h_1)$ at scale h_1 for all $v, v' \in \mathcal{V}$. Then, we adaptively determine $\omega(v, v'; h_s)$ and update $\hat{\beta}(v; h_s)$ from h_1 to $h_S = r_0$. A path diagram of the multiscale strategy is given below:

$$\begin{array}{ccccc} \{\widehat{\beta}(v):v \in \mathcal{V}\} & \Rightarrow & \omega(v, v'; h_1) & \cdots & \omega(v, v'; h_S=r_0) \\ & & \downarrow & \nearrow & \downarrow \\ & & \{\widehat{\beta}(v; h_1):v \in \mathcal{V}\} & \cdots & \{\widehat{\beta}(v; h_S), W_{\mu}(v; h_S):v \in \mathcal{V}\}. \end{array}$$

The four key steps of PS including weights adaptation, estimation, stop checking, and inference are presented as follows.

- In the weights adaptation step (i), we prefix a series $\{h_s=c_h^s:s=1, \dots, S\}$ of radii with $c_h \in (1, 2)$, say $c_h = 1.15$. We then set $s = 1$ and $h_1 = c_h$. The adaptive weights are given by

$$\omega(v, v'; h_s)=K_{loc}(\|v-v'\|_2/h_s)K_{st}(D_{\beta}(v, v'; h_{s-1})/C_n), \quad (11)$$

where $K_{loc}(u) = (1-u)_+$ and $K_{st}(u) = \min(1, 2(1-u^2))_+$ according to our experience and previous literature [39, 43], and $\|\cdot\|_2$ denotes the Euclidean norm of a vector. Moreover, $D_{\beta}(v, v'; h_{s-1})$ is set as

$$D_{\beta}(v, v'; h_{s-1})=\{\widehat{\beta}(v; h_{s-1})-\widehat{\beta}(v'; h_{s-1})\}^T \Sigma_n^{-1}(\widehat{\beta}(v; h_{s-1}))\{\widehat{\beta}(v; h_{s-1})-\widehat{\beta}(v'; h_{s-1})\}. \quad (12)$$

Following Li et al (2010), we choose $C_n = n^{1/3} \chi^2(k)^{0.95}$ for $D_{\beta}(v, v'; h_{s-1})$ defined in (12), where $\chi^2(k)^{0.95}$ is the upper 0.05-percentile of the $\chi^2(k)$ distribution. The adaptive weight $K_{st}(D_{\beta}(v, v'; h_{s-1})/C_n)$ downweights the role of a voxel $v' \in B(v; h_s)$ in $\ell_{obs}(\beta(v); h_s)$ if the value of $D_{\beta}(v, v'; h_{s-1})$ is large. The weight $K_{loc}(\|v - v'\|_2/h_s)$ gives less weight to a voxel $v' \in B(v; h_s)$ whose location is far from the voxel v . Since the theoretical justification for choosing various parameters in $\omega(v, v'; h_s)$ has been discussed in [35, 53, 36] and [43], we omit them here for the sake of simplicity.

- In the estimation step (ii), for the radius h_s , we calculate $\hat{\beta}(v; h_s)$ given in (7) at each voxel $v \in \mathcal{V}$.
- In the stop checking step (iii), after the S_0 -th iteration, we calculate a stopping criterion based on a normalized distance between $\hat{\beta}(v; h_{S_0})$ and $\hat{\beta}(v; h_s)$ given by

$$D(\hat{\beta}(v; h_{S_0}), \hat{\beta}(v; h_s)) = \{\hat{\beta}(v; h_{S_0}) - \hat{\beta}(v; h_s)\}^T \Sigma_n^{-1} \{\hat{\beta}(v; h_{S_0}) - \hat{\beta}(v; h_s)\}. \quad (13)$$

Then, we compare $D(\hat{\beta}(v; h_{S_0}), \hat{\beta}(v; h_s))$ with a benchmark, denoted by $\tilde{C}(s)$, for $s > S_0$. If $D(\hat{\beta}(v; h_{S_0}), \hat{\beta}(v; h_s))$ is greater than $\tilde{C}(s)$, then we set $\hat{\beta}(v, h_s) = \hat{\beta}(v; h_{s-1})$, and $s = S$. If $s = S$, we go to the inference step (iv). If $s \leq S_0$ or $D(\hat{\beta}(v; h_{S_0}), \hat{\beta}(v; h_s)) < \tilde{C}(s)$ for $S-1 \leq s > S_0$, then we set $h_{s+1} = c_h h_s$, increase s by 1 and continue with the weights adaptation step (i). Throughout the paper, we set $S_0 = 3$ and $\tilde{C}(s) = \chi^2(p)^{0.8/(s-2)^{0.9}}$.

- In the inference step (iv), when $s = S$, we compute $W_{\mu}(v; h_S)$ for all $v \in \mathcal{V}$ and then apply either FDR or RFT correction to obtain the corrected p -values of $W_{\mu}(v; h_S)$ in order to detect significant voxels [19, 20, 21, 22]. Otherwise, we set $h_{s+1} = c_h h_s$ and continue with the weights adaptation step. The maximal step S can be taken to be relatively small, say 10, and thus the largest spherical neighborhood of each voxel only contains a relatively small number of voxels compared to the whole volume. Throughout the paper, we have used the false discovery rate method under the positive dependency in [17], since the test statistics obtained from MARM satisfy such dependency condition.

2.2.3 TwinMARM: Stage II—The second stage is to estimate $\eta(v) = (\sigma_a^2(v), \sigma_c^2(v), \sigma_e^2(v))^T$ and carry out statistical inference on $\eta(v)$. Given $\{\hat{\beta}(v; h_s) : v \in \mathcal{V}\}$ obtained from Stage I, we compute residuals $\widehat{r}_{ij}(v) = y_{ij}(v) - \mathbf{x}_{ij}^T \widehat{\beta}(v; h_s)$ for all i, j across all $v \in \mathcal{V}$. Then, we consider a trivariate regression model given by

$$\mathbf{r}_i(v) = \begin{pmatrix} \widehat{r}_{i1}^2(v) \\ \widehat{r}_{i2}^2(v) \\ \widehat{r}_{i1}(v)\widehat{r}_{i2}(v) \end{pmatrix} = \mathbf{z}_i^T \eta(v) + \boldsymbol{\varepsilon}_i(v) = \begin{pmatrix} \mathbf{z}_{i1}^T \\ \mathbf{z}_{i2}^T \\ \mathbf{z}_{i3}^T \end{pmatrix} \eta(v) + \begin{pmatrix} \varepsilon_{i1}(v) \\ \varepsilon_{i2}(v) \\ \varepsilon_{i3}(v) \end{pmatrix}, \quad (14)$$

where $\boldsymbol{\varepsilon}_i(v) = (\varepsilon_{i1}(v), \varepsilon_{i2}(v), \varepsilon_{i3}(v))^T$ and $\mathbf{z}_i = [\mathbf{z}_{i1} \ \mathbf{z}_{i2} \ \mathbf{z}_{i3}]$ in which $\mathbf{z}_{i1} = \mathbf{z}_{i2} = (1, 1, 1)^T$ and \mathbf{z}_{i3} equals $(0.5, 1, 0)^T$ for DZ and $(1, 1, 0)^T$ for MZ. We have given the detailed derivation of model (14) in the Appendix. Moreover, $\boldsymbol{\varepsilon}_i(v)$ is assumed to have mean zero and covariance $\Sigma_{\boldsymbol{\varepsilon}_i}(v)$ for all i . Although the homogeneous variance of $\boldsymbol{\varepsilon}_i(v)$ may not be realistic, we can appropriately account for such a misspecified assumption as we did for $\Sigma_f(v)$ in Stage I. Since models (2) and (14) are similar in nature, most developments for model (2) can be directly applied to model (14) and then we omit them for the sake of simplicity.

There are several minor differences between models (2) and (14). In model (14), all regression coefficients in $\boldsymbol{\eta}(v)$ are constrained to be non-negative. Without loss of generality, $\Sigma_\varepsilon(v)$ is assumed to be known since it can be replaced by a consistent estimate. The quasi-likelihood function for $\boldsymbol{\eta}(v)$ is given by

$$\ell_{obs}(\boldsymbol{\eta}(v); h_0) = -0.5 \sum_{i=1}^n \{\mathbf{r}_i(v) - \mathbf{z}_i^T \boldsymbol{\eta}(v)\}^T \Sigma_\varepsilon^{-1}(v) \{\mathbf{r}_i(v) - \mathbf{z}_i^T \boldsymbol{\eta}(v)\}. \quad (15)$$

Let $\tilde{\boldsymbol{\eta}}(v) = \left\{ \sum_{i=1}^n \mathbf{z}_i \Sigma_\varepsilon^{-1}(v) \mathbf{z}_i^T \right\}^{-1} \sum_{i=1}^n \mathbf{z}_i \Sigma_\varepsilon^{-1}(v) \mathbf{r}_i(v)$ be an unconstrained estimate of $\boldsymbol{\eta}(v)$ in model (15). By discarding these terms independent of $\boldsymbol{\eta}(v)$, we can show that $\ell_{obs}(\boldsymbol{\eta}(v); h_0)$ can be simplified into

$$-0.5 \{\boldsymbol{\eta}(v) - \tilde{\boldsymbol{\eta}}(v)\}^T \sum_{i=1}^n \mathbf{z}_i \Sigma_\varepsilon^{-1}(v) \mathbf{z}_i^T \{\boldsymbol{\eta}(v) - \tilde{\boldsymbol{\eta}}(v)\}.$$

The constrained estimate of $\boldsymbol{\eta}(v)$, denoted by $\hat{\boldsymbol{\eta}}(v)$, is given by

$$\hat{\boldsymbol{\eta}}(v) = \arg \min_{\boldsymbol{\eta}(v) \in [0, \infty)^3} \{\boldsymbol{\eta}(v) - \tilde{\boldsymbol{\eta}}(v)\}^T \sum_{i=1}^n \mathbf{z}_i \Sigma_\varepsilon^{-1}(v) \mathbf{z}_i^T \{\boldsymbol{\eta}(v) - \tilde{\boldsymbol{\eta}}(v)\}. \quad (16)$$

Since the set $[0, \infty)^3 = [0, \infty) \times [0, \infty) \times [0, \infty)$ is a convex cone, $\hat{\boldsymbol{\eta}}(v)$ is unique [54]. To compute $\hat{\boldsymbol{\eta}}(v)$, we employ a hinge algorithm for cone projection [54].

Furthermore, similar to (6), we can utilize weights $\omega_e(v, v'; h)$ to construct a weighted quasi-likelihood function as follows:

$$\ell_{obs}(\boldsymbol{\eta}(v); h) = -0.5 \sum_{i=1}^n \sum_{v' \in B(v, h)} \omega_e(v, v'; h) \{\mathbf{r}_i(v) - \mathbf{z}_i^T \boldsymbol{\eta}(v)\}^T \Sigma_\varepsilon^{-1}(v) \{\mathbf{r}_i(v) - \mathbf{z}_i^T \boldsymbol{\eta}(v)\}. \quad (17)$$

Similarly, $\hat{\boldsymbol{\eta}}(v; h)$ is given by

$$\arg \min_{\boldsymbol{\eta}(v) \in [0, \infty)^3} \{\boldsymbol{\eta}(v) - \tilde{\boldsymbol{\eta}}(v; h)\}^T \sum_{v' \in B(v, h)} \omega_e(v, v'; h) \sum_{i=1}^n \mathbf{z}_i \Sigma_\varepsilon^{-1}(v') \mathbf{z}_i^T \{\boldsymbol{\eta}(v) - \tilde{\boldsymbol{\eta}}(v; h)\}, \quad (18)$$

where $\tilde{\boldsymbol{\eta}}(v; h)$ equals

$$\left\{ \sum_{v' \in B(v, h)} \omega_e(v, v'; h) \sum_{i=1}^n \mathbf{z}_i \Sigma_\varepsilon^{-1}(v') \mathbf{z}_i^T \right\}^{-1} \sum_{v' \in B(v, h)} \omega_e(v, v'; h) \sum_{i=1}^n \mathbf{z}_i \Sigma_\varepsilon^{-1}(v') \mathbf{r}_i(v'). \quad (19)$$

The same hinge algorithm [54] is used to calculate $\hat{\boldsymbol{\eta}}(v; h)$.

The second focus common in twin study research is the testing of genetic and environmental influences on brain structures and their functions. See for example, [2], [55, 11], [5], [6], [7],

and [8], among many others. Such questions of interest can be formulated as the linear hypotheses about $\boldsymbol{\eta}(v)$ as follows:

$$H_{0,A}:R_2\boldsymbol{\eta}(v)=0 \quad \text{vs.} \quad H_{1,A}:R_2\boldsymbol{\eta}(v)>0, \quad (20)$$

where R_2 is an $m \times 3$ matrix. For instance, in an ACE model, R_2 equals $(1, 0, 0)$ for genetic effect and $(0, 1, 0)$ for environmental effect. Similar to testing $H_{0,\mu}$, we test $H_{0,A}$ using the Wald test statistic

$$W_A(v;h)=R_2\widehat{\boldsymbol{\eta}}(v;h)^T\{R_2\Sigma_n(\widehat{\boldsymbol{\eta}}(v;h))R_2^T\}^{-1}R_2\widehat{\boldsymbol{\eta}}(v;h). \quad (21)$$

where $\Sigma_n(\widehat{\boldsymbol{\eta}}(v;h))$ is the covariance matrix of $\widehat{\boldsymbol{\eta}}(v;h)$. The statistic $W_A(v;h)$ is asymptotically distributed as a mixture of $\chi^2(1)$ and $\chi^2(0)$ distributions under the null hypothesis $H_{0,A}$, where $\chi^2(0) = 0$. However, to test the joint genetic and environmental effects, R_2 takes the form of $[(1, 0, 0)^T(0, 1, 0)^T]$ and $W_A(v;h)$ is asymptotically distributed as a mixture of $\chi^2(2)$, $\chi^2(1)$ and $\chi^2(0)$ distributions with the weights depending on the ratio of MZ over DZ [51, 56].

3 Simulation Studies

We carried out some simulation studies to examine the finite-sample performance of TwinMARM. We first simulated MRI measures from n pairs of siblings according to the ACE model, in which $\boldsymbol{\beta}(v) = (\beta_1(v), \beta_2(v), \beta_3(v))^T$ and $\mathbf{x}_{ij} = (x_{1ij}, x_{2ij}, x_{3ij})^T$ for $j = 1, 2$. Each family contains only two siblings. Among the n pairs of twins, 60% are identical twins. We set $x_{1ij} \equiv 1$, generated x_{2ij} independently from a Bernoulli distribution with probability of success 0.5 and generated x_{3ij} independently from the Gaussian distribution with zero mean and unit variance. The x_{2ij} and x_{3ij} were chosen to represent gender identity and scaled age, respectively. We set $(\beta_2(v), \beta_3(v), \sigma_c^2(v), \sigma_a^2(v))^T = (1, 1, 1, 1)^T$ across all voxels v . For $(\beta_1(v), \sigma_a^2(v))$, we divided the 64×64 phantom image into five different regions of interest (ROIs) with different shapes and then varied $(\beta_1(v), \sigma_a^2(v))$ as $(0, 0)$, $(0.3, 0.5)$, $(0.6, 1)$, $(0.9, 1.5)$ and $(1.2, 2.0)$ across these five ROIs. The true $(\beta_1(v), \sigma_a^2(v))$ was displayed for all ROIs with black, blue, red, yellow, and white colors representing $(\beta_1(v), \sigma_a^2(v)) = (0, 0)$, $(0.3, 0.5)$, $(0.6, 1)$, $(0.9, 1.5)$ and $(1.2, 2.0)$ (Fig. 3(A)). We independently generated $\mathbf{e}_i(v) = (e_{i1}(v), e_{i2}(v))^T$, $\mathbf{a}_i(v) = (a_{i1}(v), a_{i2}(v))^T$, and $c_i(v)$ from multivariate Gaussian distributions with zero means and covariance matrices as specified in Fig. 1. We set $n = 100$ and 400.

To carry out the analysis, we first fit the structural equation model

$y_{ij}(v) = \mathbf{x}_{ij}^T \boldsymbol{\beta}(v) + a_{ij}(v) + c_i(v) + e_{ij}(v)$, in which the correlation pattern of $a_{ij}(v)$, $c_i(v)$ and $e_{ij}(v)$ was specified according to Fig. 1. We calculated the maximum likelihood estimates (MLEs) of $(\beta(v), \sigma_a^2(v), \sigma_c^2(v), \sigma_e^2(v))^T$ at each pixel by using the expectation-maximization (EM) algorithm [57]. Then, we applied the TwinMARM procedure described in Section 2.2 to calculate adaptive parameter estimates across all pixels at 11 different scales. We also smoothed the subjects' images by using an isotropic Gaussian kernel with FWHM 3 pixels and then applied the EM algorithm to the smoothed data. Furthermore, for $(\beta_1(v), \sigma_a^2(v))$, we calculated the bias, the empirical standard error RMS, the mean of the standard error estimates SD, and the ratio of RMS over SD, abbreviated as RE, at each pixel of all five ROIs based on the results obtained from the 1000 $n = 100$ and 400 simulated data sets.

For simplicity, we present some selected results for $\hat{\beta}_1(v; h)$ and $\hat{\sigma}_a^2(v; h)$ and their corresponding MLEs obtained from the EM algorithm with smoothed and unsmoothed data. We first considered $\hat{\sigma}_a^2(v; h)$. The biases of $\sigma_a(v; h)^2$ are almost the same for TwinMARM at h_{10} and EM with unsmoothed data (Fig. 3(B), (C), (E) and (F)), while the bias from EM with smoothed data is greatly increased (Fig. 5(A) and (C) and Fig. 3(D) and (G)). Inspecting the results from the EM algorithm reveals that SD is consistently larger than that obtained from TwinMARM at h_{10} (Fig. 5(B) and (D)). Overall, for $\hat{\sigma}_a^2(v; h)$, TwinMARM outperforms EM in terms of smaller RMS (Fig. 3(B), (C), (E) and (F)).

Secondly, we examine $\hat{\beta}_1(v; h)$. The biases of $\hat{\beta}_1(v; h)$ are almost the same for TwinMARM at h_{10} and EM with unsmoothed data (Fig. 6(A) and (C) and Fig. 4(B), (C), (E) and (F)), while the bias from EM with smoothed data is greatly increased (Fig. 6(A) and (C) and Fig. 4(D) and (G)). Inspecting the results from the EM algorithm reveals that SD is consistently larger than that obtained from TwinMARM at h_{10} (Fig. 6(B) and (D)). For $\hat{\beta}_1(v; h)$, TwinMARM outperforms EM in terms of smaller RMS (Fig. 4(B), (C), (E) and (F)).

To assess both Type I and II error rates of $W_A(v; h)$ at the pixel level, we tested the hypotheses $H_0: \sigma_a^2(v)=0$ and $H_1: \sigma_a^2(v)>0$ across all pixels. We applied the same PS procedure and computed the p -values of $W_A(v; h)$ at each scale. The 1,000 replications were used to calculate the estimates and standard errors of rejection rates with significance level $\alpha = 5\%$. For the test statistic $W_A(v; h)$, the Type I rejection rates in ROI with $\sigma_a^2(v)=0$ were relatively small for all scales, while the statistical power for rejecting the null hypothesis in ROIs with $\sigma_a^2(v)>0$ significantly increased with larger radius h (Fig. 7). For all ROIs with $\sigma_a^2(v)>0$, $W_A(v; h_{10})$ outperforms the likelihood ratio test statistic (LRT) based on the EM algorithm without smoothing and EM algorithm with smoothing in terms of power (Fig. 7).

Finally, we note that there are at least three limitations associated with our simulation studies. First, it should be noted that the simulated phantom image is 2D, whereas most real neuroimaging data is 3D. TwinMARM should work better in 3D and the variance reduction should be larger, since more neighboring voxels are available at each voxel for calculating adaptive weights and parameter estimation. Secondly, we did not directly simulate ‘real’ functional (or structural) imaging data from twin studies, since it is not straightforward to simulate real imaging data according to model (1). Thirdly, we did not evaluate the reconstruction and registration errors in reconstructing and registering real imaging data from twin studies.

4 Real-World Example

We considered the early postnatal brain development project led by Dr. Gilmore at the University of North Carolina at Chapel Hill [9]. This study was approved by the Institutional Review Board. A total of 49 paired twins (36 males and 62 females) were selected. All 49 pairs were scanned as neonates within a few weeks after birth at term. Written consent was obtained from their parents before imaging acquisition. The mean gestational age at MR scanning was 246 ± 18.3 days (range: 192 to 270 days). All infants were fed and calmed to sleep on a warm blanket with proper ear protection and they slept comfortably inside the MR scanner. None of infants was sedated during the imaging session.

We used a 3T Allegra head-only MR system (Siemens Medical Inc., Erlangen, Germany) to acquire all the images. The system was equipped with a maximal gradient strength of 40 mT/m and a maximal slew rate of 400 mT/(m-msec). We used a single shot EPI DTI sequence (TR/TE=5400/73 msec) with eddy current compensation to acquire the DTI images. The acquisition sequence consists of five repetitions of six gradient scans ($b = 1000$

s/mm²) at 6 non-collinear directions and a single reference scan (in total $5 \times 7 = 35$ images). The voxel resolution was isotropic 2 mm, and the in-plane field of view was set at 256 mm in both directions.

We then employed a nonlinear fluid deformation based high-dimensional, unbiased atlas computation method to process all 98 DTI datasets [58]. The atlas building procedure started with an affine registration and was followed by a nonlinear registration of a set of feature images for all subjects. The feature images are the maximum eigenvalue of the Hessian of the fractional anisotropy (FA) image, which are sensitive to the geometry of white matter. Using the computed deformation fields, we warped all tensor images into the unbiased atlas space via log-euclidean based interpolation [59]. We also averaged all the warped tensor images to create a study specific DTI atlas.

FA has been widely used as a measurement to assess directional organization of the brain which is greatly influenced by the magnitude and orientation of white matter tracts. Here, FA images are employed to identify the spatial patterns of white matter maturation. We considered a linear model $y_{ij}(v) = \beta_1(v) + \beta_2(v)G_{ij} + \beta_3(v)Z_{ij} + a_{ij}(v) + c_i(v) + \varepsilon_{ij}(v)$ for $i = 1, \dots, 49$ and $j = 1, 2$ at each voxel of the template, where G_{ij} and Z_{ij} , respectively, represent the dummy variables for gender (male=1 and female=0) and zygote (MZ=1 and DZ=0), and $a_{ij}(v)$, $c_i(v)$ and $\varepsilon_{ij}(v)$ are, respectively, the additive genetic, common environmental, and residual effects on the i -th twin pair as specified in Fig. 1. We focused on the major white matter regions. We calculated the MLE of parameters by using the EM algorithm and then applied our TwinMARM method with $c_h = 1.15$ and $S = 10$ to carry out statistical analysis.

We tested $H: \sigma_a^2(v) = 0$ for additive genetic effect across all voxels v in the white matter regions. To correct for multiple comparisons, we used the raw p -value smaller than 0.05 along with a 20 voxel extent threshold. We found significant regions in the inferior frontal gyrus, triangular part and mid cingulate cortex regions. We identified more voxels by using TwinMARM compared to EM (Table 1). We also tested $H_0: \sigma_c^2(v) = 0$ for common environmental effect across all voxels v . Several interesting regions identified included the right precentral gyrus, inferior frontal gyrus, triangular part, supplementary motor area, insula, hippocampus, right fusiform, and thalamus. Fig. 8 shows some selected slices of $-\log_{10}(p)$ map for environmental effect obtained from the 10th iteration of TwinMARM. The results show that TwinMARM may have significantly improved sensitivity and reliability as areas of significance in the TwinMARM results appear larger and smoother compared to the voxel-wise analysis approach, which is close to the result obtained from the first iteration of TwinMARM. The FA measure shows strong effects of genetic factors and common environment in the frontal lobes and in the limbic lobes. The detection of a genetic effect is partially consistent with many other studies that have detected strong genetic influences on brain morphology [11, 4, 60, 55].

We also observed the smoother and larger significant clusters of genetic factors and common environment by using the EM algorithm with smoothed data from FWHM=6mm and 9mm (Fig. 8(F), (E') and (F')). First, based on our simulation studies, we observed that the EM algorithm for unsmoothed data and TwinMARM do not introduce bias in the estimates of $\sigma_a^2(v)$, whereas we observed the substantial bias in the estimates of $\sigma_a^2(v)$ for the heavily smoothing imaging data (Figs. 3 and 5). Secondly, these smoother and larger significant clusters do not appear in the results obtained from the unsmoothed data. Thus, we feel that these smoother and larger significant clusters obtained from the heavily smoothed data may be false positive results (Fig. 8 and Table 1 and 2). However, a word of caution is needed in interpreting these significant clusters to be false positive, since two key

assumptions underlying EM and TwinMARM, including the perfect registration assumption and the structural equation model (1), may be invalid for this real data set.

5 Conclusions and Discussion

We have developed TwinMARM for the statistical analysis of twin neuroimaging data. TwinMARM consists of two stages of multiscale adaptive regression methods. TwinMARM have three features: being spatial, being hierarchical, and being adaptive. The TwinMARM procedure is a simultaneous smoothing and estimation procedure. Each stage of TwinMARM constructs hierarchical nested spheres with increasing radii at all voxels to adaptively compute weighted quasi-likelihood functions and efficiently utilizes available information to obtain parameter estimates. This leads to improved accuracy of parameter estimates and increased statistical power in detecting subtle changes of brain structure and function.

As shown in the simulation studies, the commonly used Gaussian kernel for smoothing imaging data can introduce two levels of biases for the analysis of twin imaging data, which can dramatically increase the numbers of false positives and negatives. The first level of biases originates from the mean structure, which is associated with the regression coefficients $\beta(v)$. Most smoothing methods, which are independent of imaging data, usually blur the image data near the edges of the significant regions, and consequently introduce biases in $\beta(v)$. The second level of biases comes from the variance structure, which is associated with the genetic coefficients $\sigma^2(v)$. Most smoothing methods are not only conducted independently of imaging data but also of the statistical model fitted to the imaging data. Statistically, standard smoothing methods can change the variance structure in the imaging data which are the focus of twin imaging studies. Thus, the second level of biases is much more severe than the first level of biases. In practice, we should avoid directly smoothing twin imaging data.

Several limitations need to be addressed in future research. The TwinMARM procedure is based on the two key assumptions, including the perfect registration assumption and the structural equation model (1). A first limitation of our study is that the perfect registration assumption is demonstrably false for real imaging data from twin studies. It is important and interesting to integrate the registration method, smoothing method, and TwinMARM into a single framework in order to appropriately account for registration errors in the statistical analysis of imaging data from twin studies. A second limitation of our study is that the structural equation model (1) can be invalid for real imaging data from twin studies. Heritability estimates used for most twin studies rest on restrictive assumptions in model (1), which are usually not tested, and if they are, can often found to be violated by the real data [61].

There are several current topics for our research. We are developing a Graphical User Interface (GUI) to pack our code for TwinMARM, which will be freely downloadable from our website at <http://www.bios.unc.edu/research/bias>. The development of multiscale-adaptive methods to determine multiscale neighborhoods that adapt to the pattern of imaging data at each voxel would undoubtedly be a useful resource for neuroimaging researchers. Further, we will extend TwinMARM from the univariate to multivariate and complex imaging measures, such as deformation tensors. We will develop a multiscale adaptive model for neuroimaging data from familial studies [25]. More research is needed for optimizing the choices of the parameters in TwinMARM. It is interesting to incorporate other edge-preserving local smoothing methods into TwinMARM and MARM [33, 34, 37, 38].

References

1. Baare WF, Hulshoff Pol HE, Boomsma DI, Posthuma D, de Geus EJ, Schnack HG, van Haren NE, van Oel CJ, Kahn RS. Quantitative genetic modeling of variation in human brain morphology. *Cereb Cortex*. 2001; 11:816–824. [PubMed: 11532887]
2. Wright IC, Sham P, Murray RM, Weinberger DR, Bullmore ET. Genetic contributions to regional variability in human brain structure: methods and preliminary results. *NeuroImage*. 2002; 17:256–271. [PubMed: 12482082]
3. Peper JS, Brouwer RM, Boomsma DI, Kahn RS, Hulshoff Pol HE. Genetic influences on human brain structure: a review of brain imaging studies in twins. *Human Brain Mapping*. 2007; 28:446–473.
4. Chiang MC, Barysheva M, Shattuck DW, Lee AD, Madsen SK, Avedissian C, Klunder AD, Toga AW, McMahon KL, de Zubicaray GI, Wright MJ, Srivastava A, Balov N, Thompson PM. Genetics of brain fiber architecture and intellectual performance. *Journal of Neuroscience*. 2009; 29:2212–2224. [PubMed: 19228974]
5. Yoon U, Fahim C, Perusse D, Evans AC. Lateralized genetic and environmental influences on human brain morphology of 8-year-old twins. *NeuroImage*. 2010; 53:1117–1125. [PubMed: 20074649]
6. Pell GS, Briellmann RS, Lawrence KM, Glencross D, Wellard RM, Berkovic SF, Jackson GD. Reduced variance in monozygous twins for multiple mr parameters: Implications for disease studies and the genetic basis of brain structure. *NeuroImage*. 2010; 49:1536–1544. [PubMed: 19747554]
7. Jahanshad N, Lee AD, Barysheva M, McMahon KL, de Zubicaray GI, Martin NG, Wright MJ, Toga AW, Thompson PM. Genetic influences on brain asymmetry: A dti study of 374 twins and siblings. *NeuroImage*. 2010; 52:455–469. [PubMed: 20430102]
8. Chou YY, Leporé N, Chiang MC, Avedissian C, Barysheva M, McMahon KL, de Zubicaray GI, Meredith M, Wright MJ, Toga AW, Thompson PM. Mapping genetic influences on ventricular structure in twins. *Neuroimage*. 2009; 44:1312–1323. [PubMed: 19041405]
9. Gilmore JH, Schmitt JE, Knickmeyer RA, Smith JK, Lin W, Styner M, Gerig G, Neale MC. Genetic and environmental contributions to neonatal brain structure: a twin study. *Human Brain Mapping*. 2010; 31:1174–1182. [PubMed: 20063301]
10. Geschwind DH, Miller BL, DeCarli C, Carmelli D. Heritability of lobar brain volumes in twins supports genetic models of cerebral laterality and handedness. *Proc Natl Acad Sci USA*. 2002; 99:3176–3181. [PubMed: 11867730]
11. Thompson PM, Cannon TD, Narr KL, van Erp T, Poutanen V, Huttunen M, Lonnqvist J, Standertskjold-Nordenstam CG, Kaprio J, Khaledy M, Dail R, Zoumalan CI, Toga A. Genetic influences on brain structure. *Nature Neuroscience*. 2001; 4:1253–1358.
12. Yue Y, Loh JM, Lindquist MA. Adaptive spatial smoothing of fmri images. *Statistics and its Interface*. 2010; 3:3–14.
13. Kempthorne O, Osborne RH. The interpretation of twin data. *American Journal of Human Genetics*. 1961; 13:320–339. [PubMed: 13752449]
14. Haseman JK, Elston RC. The estimation of genetic variance from twin data. *Behavior Genetics*. 1970; 1:11–19. [PubMed: 5535855]
15. Neale MC, Heath AC, Hewitt JK, Eaves LJ, Fulker DW. Fitting genetic models with lisrel: Hypothesis testing. *Behavior Genetics*. 1989; 19:37–49. [PubMed: 2712812]
16. Feng R, Zhou G, Zhang M, Zhang H. Analysis of twin data using sas. *Biometrics*. 2009; 65:584–589. [PubMed: 18647295]
17. Benjamini Y, Yekutieli D. The control of the false discovery rate in multiple testing under dependency. *The Annals of Statistics*. 2001; 29:1165–1188.
18. Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society, Ser B*. 1995; 57:289–300.
19. Nichols T, Hayasaka S. Controlling the familywise error rate in functional neuroimaging: a comparative review. *Stat Methods Med Res*. 2003; 12:419–426. [PubMed: 14599004]
20. Worsley KJ, Taylor JE, Tomaiuolo F, Lerch J. Unified univariate and multivariate random field theory. *NeuroImage*. 2004; 23:189–195.

21. Chumbley J, Worsley J, Flandin KG, Friston KJ. Topological fdr for neuroimaging. *Neuroimage*. 2010; 49(4):3057–3064. [PubMed: 19944173]
22. Salimi-Khorshidi G, Smith SM, Nichols TE. Adjusting the effect of nonstationarity in cluster-based and tfce inference. *Neuroimage*. 2011; 54(3):2006–2019. [PubMed: 20955803]
23. Hecke WV, Sijbers J, Backer SD, Poot D, Parizel PM, Leemans A. On the construction of a ground truth framework for evaluating voxel-based diffusion tensor mri analysis methods. *NeuroImage*. 2009; 46:692–707. [PubMed: 19268708]
24. Jones DK, Symms MR, Cercignani M, Howard RJ. The effect of filter size on vbm analyses of dt-mri data. *NeuroImage*. 2005; 26:546–554. [PubMed: 15907311]
25. Zhu HT, Li YM, Tang NS, Bansal R, Hao XJ, Weissman MM, Peterson BG. Statistical modelling of brain morphological measures within family pedigrees. *Statistica Sinica*. 2008; 18:1554–1569.
26. Chiang, MC.; Avedissian, C.; Barysheva, M.; Toga, AW.; McMahan, KL.; Zubicaray, GI.; Wright, MJ.; Thompson, PM. Extending genetic linkage analysis to diffusion tensor images to map single gene effects on brain fiber architecture. *MICCAI '09 Proceedings of the 12th International Conference on Medical Image Computing and Computer-Assisted Intervention: Part II*; 2009. p. 506-513.
27. Besag JE. On the statistical analysis of dirty pictures (with discussion). *Journal of the Royal Statistical Society, Ser B*. 1986; 48:259–302.
28. Banerjee, S.; Carlin, BP.; Gelfand, AE. *Hierarchical Modeling and Analysis for Spatial Data*. CRC/Chapman and Hall; 2004.
29. Penny WD, Trujillo-Barreto NJ, Friston KJ. Bayesian fmri time series analysis with spatial priors. *NeuroImage*. 2005; 24:350–362. [PubMed: 15627578]
30. Groves AR, Chappell MA, Woolrich MW. Combined spatial and non-spatial prior for inference on mri time-series. *NeuroImage*. 2009; 45(3):795–809. [PubMed: 19162204]
31. Zhu HT, Gu M, Peterson BS. Maximum likelihood from spatial random effects models via the stochastic approximation expectation maximization algorithm. *Statistics and Computing*. 2007; 15:163–177.
32. Bowman FD. Spatio-temporal models for region of interest analyses of functional mapping experiments. *Journal of American Statistical Association*. 2007; 102:442–453.
33. Qiu, P. *Image Processing and Jump Regression Analysis*. New York: John Wiley & Sons; 2005.
34. Qiu P. Jump surface estimation, edge detection, and image restoration. *Journal of American Statistical Association*. 2007; 102:745–756.
35. Polzehl J, Spokoyny VG. Adaptive weights smoothing with applications to image restoration. *J R Statist Soc B*. 2000; 62:335–354.
36. Polzehl J, Spokoyny VG. Propagation-separation approach for local likelihood estimation. *Probab Theory Relat Fields*. 2006; 135:335–362.
37. Qiu P, Mukherjee PS. Edge structure preserving image denoising. *Signal Processing*. 2010; 90:2851–2862.
38. Mukherjee PS, Qiu P. 3-d 3-d image denoising by local smoothing and nonparametric regression. *Technometrics*. 2011; 53:196–208.
39. Tabelow K, Polzehl J, Voss HU, Spokoyny V. Analyzing fmri experiments with structural adaptive smoothing procedures. *NeuroImage*. 2006; 33:55–62. [PubMed: 16891126]
40. Tabelow K, Polzehl J, Spokoyny V, Voss HU. Diffusion tensor imaging: structural adaptive smoothing. *NeuroImage*. 2008; 39:1763–1773. [PubMed: 18060811]
41. Tabelow K, Polzehl J, Ulug AM, Dyke JP, Watts R, Heier LA, Voss HU. Accurate localization of brain activity in presurgical fmri by structure adaptive smoothing. *IEEE Trans Med Imaging*. 2008; 27:531–537. [PubMed: 18390349]
42. Polzehl J, Voss HU, Tabelow K. Structural adaptive segmentation for statistical parametric mapping. *Neuroimage*. 2010; 52:515–523. [PubMed: 20420928]
43. Li YM, Zhu HT, Shen DG, Lin WL, Gilmore J, Ibrahim JG. Multiscale adaptive regression models for neuroimaging data. *Journal of the Royal Statistical Society, Series B*. 2011 in press.
44. Ashburner J, Friston KJ. Voxel-based morphometry: the methods. *Neuroimage*. 2000; 11:805–821. [PubMed: 10860804]

45. Chung MK, Dalton KM, Davidson RJ. Tensor-based cortical surface morphometry via weighted spherical harmonic representation. *IEEE Trans Med Imaging*. 2008; 27:1143–1151. [PubMed: 18672431]
46. Styner M, Lieberman JA, McClure RK, Weinberger DR, Jones DW, Gerig G. Morphometric analysis of lateral ventricles in schizophrenia and healthy controls regarding genetic and disease-specific factors. *Proc Natl Acad Sci USA*. 2005; 102:4872–4677. [PubMed: 15772166]
47. Thompson PM, Toga AW. A framework for computational anatomy. *Comput Visual*. 2002; 5:13–34.
48. Plomin R, Kosslyn SM. Genes, brain and cognition. *Nature Neuroscience*. 2001; 4:1153–1154.
49. Thompson PM, Woods RP, Mega MS, Toga AW. Mathematical/computational challenges in creating population-based brain atlases. *Human Brain Mapping*. 2000; 9:81–92. [PubMed: 10680765]
50. Zhu HT, Ibrahim JG, Tang N, Rowe DB, Hao X, Bansal R, Peterson BS. A statistical analysis of brain morphology using wild bootstrapping. *IEEE Trans Med Imaging*. 2007; 26:954–966. [PubMed: 17649909]
51. Wang X, Guo X, He M, Zhang H. Statistical inference in mixed models and analysis of twin and family data. *Biometrics*. 2011 page in press.
52. Falconer, DS.; Mackay, TFC. *Introduction to Quantitative Genetics*. Longman: Harlow, Essex; 1996.
53. Polzehl J, Spokoiny VG. Image denoising: pointwise adaptive approach. *Annals of Statistics*. 2003; 31:30–57.
54. Meyer, MC. *An algorithm for quadratic programming with applications in statistics*. 2009.
55. Thompson PM, Cannon TD, Toga AW. Mapping genetic influences on human brain structure. *Ann Med*. 2002; 24:523–536. [PubMed: 12553492]
56. Shapiro A. Towards a unified theory of inequality constrained testing in multivariate analysis. *International Statistical Review*. 1988; 56:49–62.
57. Dempster AP, Laird NM, Rubin DB. Maximum likelihood from incomplete data via the em algorithm. *Journal of the Royal Statistical Society Series B*. 1977; 39:1–38.
58. Goodlett CB, Fletcher PT, Gilmore JH, Gerig G. Group analysis of dti fiber tract statistics with application to neurodevelopment. *NeuroImage*. 2009; 45:S133–S142. [PubMed: 19059345]
59. Arsigny V, Fillard P, Pennec X, Ayache N. Log-euclidean metrics for fast and simple calculus on diffusion tensors. *Magnetic Resonance in Medicine*. 2006; 56:411–421. [PubMed: 16788917]
60. Chiang MC, McMahon KL, de Zubicaray GI, Martin NG, Hickie I, Toga AW, Wright MJ, Thompson PM. Genetics of white matter development: A dti study of 705 twins and their siblings aged 12 to 29. *NeuroImage*. 2011; 54:2308–2317. [PubMed: 20950689]
61. Schonemann P. Models and muddles of heritability. *Genetica*. 1997; 99:97–108. [PubMed: 9463078]
62. van der Vaar, AW. *Asymptotic Statistics*. Cambridge University Press; 1998.

6 Appendix: Effect of smoothing twin imaging data

We examined the effect of smoothing twin imaging data on its associated statistical analysis as follows. As an illustration, we focus on an arbitrary voxel v and its spherical neighborhood, denoted by $\{v' : v' \in B(v, h)\}$. Standard smoothing methods primarily take a weighted average of all observations $\{y_{ij}(v') : v' \in B(v, h)\}$, denoted by $\tilde{y}_{ij}(v)$, such that $\tilde{y}_{ij}(v) = \sum_{v' \in B(v, h)} \omega(v, v'; h) y_{ij}(v')$, where $\sum_{v' \in B(v, h)} \omega(v, v'; h) = 1$ and $\omega(v, v'; h) > 0$. Under model (1), we can show that

$$\begin{aligned} \tilde{y}_{ij}(v) &= \sum_{v' \in B(v, h)} \omega(v, v'; h) \{x_{ij}^T \beta(v') + a_{ij}(v') + d_{ij}(v') + c_i(v') + e_{ij}(v')\} \\ &= x_{ij}^T \tilde{\beta}(v) + \tilde{a}_{ij}(v) + \tilde{d}_{ij}(v) + \tilde{c}_i(v) + \tilde{e}_{ij}(v), \end{aligned} \quad (22)$$

where $\hat{\boldsymbol{\beta}}(v) = \sum_{v' \in B(v,h)} \omega(v, v'; h) \boldsymbol{\beta}(v')$,

$$\begin{aligned} \tilde{a}_{ij}(v) &= \sum_{v' \in B(v,h)} \omega(v, v'; h) a_{ij}(v'), & \tilde{d}_{ij}(v) &= \sum_{v' \in B(v,h)} \omega(v, v'; h) d_{ij}(v'), \\ \tilde{c}_{ij}(v) &= \sum_{v' \in B(v,h)} \omega(v, v'; h) c_{ij}(v'), & \tilde{e}_{ij}(v) &= \sum_{v' \in B(v,h)} \omega(v, v'; h) e_{ij}(v'). \end{aligned}$$

It can be shown that the variance of $\tilde{a}_{ij}(v)$, denoted by $\tilde{\sigma}_a^2(v)$, depends on both the variances of $\{a_{ij}(v') : v' \in B(v, h)\}$ and their spatial correlations. Specifically, we have

$$\tilde{\sigma}_a^2(v) = \sum_{v' \in B(v,h)} \omega(v, v'; h)^2 \sigma_a^2(v') + 2 \sum_{v' \neq v'' \in B(v,h)} \omega(v, v'; h) \omega(v, v''; h) \text{Cov}(a_{ij}(v'), a_{ij}(v'')).$$

Similar arguments hold for the variances of $\tilde{d}_{ij}(v)$, $\tilde{c}_{ij}(v)$ and $\tilde{e}_{ij}(v)$ denoted by $\tilde{\sigma}_d^2(v)$, $\tilde{\sigma}_c^2(v)$, and $\tilde{\sigma}_e^2(v)$, respectively. Due to the presence of spatial correlations and other factors (e.g., the inhomogeneity of $\tilde{\sigma}_a^2(v')$ across $v' \in B(v, h)$), it is common that

$$g(v) = \frac{\sigma_a^2(v) + \sigma_d^2(v)}{\sigma_a^2(v) + \sigma_d^2(v) + \sigma_e^2(v) + \sigma_c^2(v)} \neq \tilde{g}(v) = \frac{\tilde{\sigma}_a^2(v) + \tilde{\sigma}_d^2(v)}{\tilde{\sigma}_a^2(v) + \tilde{\sigma}_d^2(v) + \tilde{\sigma}_e^2(v) + \tilde{\sigma}_c^2(v)}.$$

Generally, the Gaussian kernel smoothing method can easily modify the variance structure of twin imaging data. Because such variance structures are often the focus of twin studies, we thus suggest avoiding directly smoothing twin imaging data in practice.

We show that TwinMARM does not introduce bias in the variance structure of twin imaging data. This consists of two steps.

- Step (i) As $n \rightarrow \infty$, $\hat{\boldsymbol{\beta}}(v; h_S)$ converges $\boldsymbol{\beta}_*(v)$ in probability and $\sqrt{n}\{\hat{\boldsymbol{\beta}}(v; h_S) - \boldsymbol{\beta}_*(v)\}$ is asymptotically normal distributed for all $v \in \mathcal{V}$ under the mild conditions given in [43], where $\boldsymbol{\beta}_*(v)$ is the true value $\boldsymbol{\beta}(v)$ in voxel v . Please see Theorems 1–3 in [43] for the details.
- Step (ii) As $n \rightarrow \infty$, $\tilde{\boldsymbol{\eta}}(v; h_S)$ converges $\boldsymbol{\eta}_*(v)$ and $\sqrt{n}\{\tilde{\boldsymbol{\eta}}(v; h_S) - \boldsymbol{\eta}_*(v)\}$ converges to a random variable in distribution for all $v \in \mathcal{V}$. We can essentially use the same idea in proving Step (i) to prove Step (ii), but there is a complexity in dealing with substituting $\boldsymbol{\beta}_*(v)$ by $\hat{\boldsymbol{\beta}}(v; h_S)$. We will show that such substitution is negligible below.

We just highlight the key idea underlying the proof of Step (ii). It follows from Step (i) and the ACE model that for $j = 1, 2$,

$$\begin{aligned} \tilde{r}_{ij}(v) &= y_{ij}(v) - \mathbf{x}_{ij}^T \hat{\boldsymbol{\beta}}(v; h_S) = y_{ij}(v) - \mathbf{x}_{ij}^T \boldsymbol{\beta}_*(v) + \mathbf{x}_{ij}^T \{\hat{\boldsymbol{\beta}}(v; h_S) - \boldsymbol{\beta}_*(v)\} \\ &= n^{-1/2} \mathbf{x}_{ij}^T \sqrt{n} \{\hat{\boldsymbol{\beta}}(v; h_S) - \boldsymbol{\beta}_*(v)\} + f_{ij}(v). \end{aligned} \quad (23)$$

Let $\Delta \boldsymbol{\beta}(v; h_S) = \{\hat{\boldsymbol{\beta}}(v; h_S) - \boldsymbol{\beta}_*(v)\}$. With some simple algebra, for $j = 1, 2$, we can show that

$$\begin{aligned}\widehat{r}_{ij}^2(v) &= f_{ij}^2(v) + n^{-1}O_p(1) + 2n^{-1/2}\mathbf{x}_{ij}^T \sqrt{n}\Delta\beta(v; h_s) f_{ij}(v) \\ &= \mathbf{z}_{ij}^T \eta_*(v) + \varepsilon_{ij}(v),\end{aligned}$$

in which we have used the results in Step (i) and $O_p(1) = \{\mathbf{x}_{ij}^T \sqrt{n}\widehat{\Delta\beta}(v; h_s)\}^2$ (van der Vaart, 1998) and set

$$\varepsilon_{ij}(v) = f_{ij}^2(v) - \mathbf{z}_{ij}^T \eta_*(v) + n^{-1}O_p(1) + 2n^{-1/2}\mathbf{x}_{ij}^T \sqrt{n}\Delta\beta(v; h_s) f_{ij}(v). \quad (24)$$

Similarly, we can show that

$$\begin{aligned}\widehat{r}_{i3}(v)\widehat{r}_{i2}(v) &= \mathbf{z}_{i3}^T \eta_*(v) + \varepsilon_{i3}(v), \\ \varepsilon_{i3}(v) &= f_{i1}(v)f_{i2}(v) - \mathbf{z}_{i3}^T \eta_*(v) + n^{-1}O_p(1) + n^{-1/2}\mathbf{x}_{i2}^T \sqrt{n}\Delta\beta(v; h_s) f_{i1}(v) + n^{-1/2}\mathbf{x}_{i1}^T \sqrt{n}\Delta\beta(v; h_s) f_{i2}(v),\end{aligned} \quad (25)$$

where $O_p(1) = \{\mathbf{x}_{i1}^T \sqrt{n}\widehat{\Delta\beta}(v; h_s)\}\{\mathbf{x}_{i2}^T \sqrt{n}\widehat{\Delta\beta}(v; h_s)\}$. Based on these results, we can obtain model (14), in which $\mathbf{e}_\lambda(v)$ is almost unbiased except an $O_p(n^{-1/2})$ term. We will show that such $O_p(n^{-1/2})$ term has a negligible effect below.

Then, we need to show that $\widehat{\boldsymbol{\eta}}(v; h)$ converges $\boldsymbol{\eta}_*(v)$ in probability and $\sqrt{n}\{\widehat{\boldsymbol{\eta}}(v; h) - \boldsymbol{\eta}_*(v)\}$ asymptotically converges a random variable in distribution for all $v \in \mathcal{V}$ under some mild conditions, which are similar to those give in [43]. We focus on $h = h_0 = 0$. For $h > 0$, we can generalize the proofs given in [43]. Recall that

$\tilde{\boldsymbol{\eta}}(v) = \left\{ \sum_{i=1}^n \mathbf{z}_i \Sigma_\varepsilon^{-1}(v) \mathbf{z}_i^T \right\}^{-1} \sum_{i=1}^n \mathbf{z}_i \Sigma_\varepsilon^{-1}(v) \mathbf{r}_i(v)$. We define

$$\begin{aligned}\tilde{\boldsymbol{\varepsilon}}_i(v) &= (f_{i1}^2(v) - \mathbf{z}_{i1}^T \eta_*(v), f_{i2}^2(v) - \mathbf{z}_{i2}^T \eta_*(v), f_{i1}(v)f_{i2}(v) - \mathbf{z}_{i3}^T \eta_*(v))^T, \\ \mathbf{g}_i(v) &= (2\mathbf{x}_{i1}^T f_{i1}(v), 2\mathbf{x}_{i2}^T f_{i2}(v), \mathbf{x}_{i1}^T f_{i2}(v) + \mathbf{x}_{i2}^T f_{i1}(v))^T.\end{aligned} \quad (26)$$

By using the law of large number [62], we have

$$\frac{1}{\sqrt{n}} \sum_{i=1}^n \mathbf{z}_i \Sigma_\varepsilon^{-1}(v) n^{-1} O_p(1) = n^{-1/2} O_p(1) \text{ and } \sqrt{n} \Delta\beta(v; h_s) n^{-1} \sum_{i=1}^n \mathbf{z}_i \Sigma_\varepsilon^{-1}(v) \mathbf{g}_i(v) = o_p(1)$$

in which we have used $\sqrt{n}\Delta\beta(v; h_s) = O_p(1)$ obtained from Step (i) and $n^{-1} \sum_{i=1}^n \mathbf{z}_i \Sigma_\varepsilon^{-1}(v) \mathbf{g}_i(v)$ converges to zero in probability. Thus, it follows from (14) that

$$\begin{aligned}\sqrt{n}[\widehat{\boldsymbol{\eta}}(v) - \boldsymbol{\eta}_*(v)] &= \left\{ n^{-1} \sum_{i=1}^n \mathbf{z}_i \Sigma_\varepsilon^{-1}(v) \mathbf{z}_i^T \right\}^{-1} \frac{1}{\sqrt{n}} \sum_{i=1}^n \mathbf{z}_i \Sigma_\varepsilon^{-1}(v) \boldsymbol{\varepsilon}_i(v) \\ &= \left\{ n^{-1} \sum_{i=1}^n \mathbf{z}_i \Sigma_\varepsilon^{-1}(v) \mathbf{z}_i^T \right\}^{-1} \frac{1}{\sqrt{n}} \sum_{i=1}^n \mathbf{z}_i \Sigma_\varepsilon^{-1}(v) \{ \tilde{\boldsymbol{\varepsilon}}_i(v) + n^{-1} O_p(1) + n^{-1/2} \mathbf{g}_i(v) \sqrt{n} \Delta\beta(v; h_s) \} \\ &= \left\{ n^{-1} \sum_{i=1}^n \mathbf{z}_i \Sigma_\varepsilon^{-1}(v) \mathbf{z}_i^T \right\}^{-1} \frac{1}{\sqrt{n}} \sum_{i=1}^n \mathbf{z}_i \Sigma_\varepsilon^{-1}(v) \tilde{\boldsymbol{\varepsilon}}_i(v) + o_p(1),\end{aligned}$$

which yields that it is negligible to substitute $\boldsymbol{\beta}_*(v)$ by $\widehat{\boldsymbol{\beta}}(v; h_s)$. Furthermore, we calculate

$$\omega_e(v, v'; h) = K_{loc}(\|v - v'\|_2 / h) K_{st}(D_\eta(v, v'; h) / C_n),$$

where $D_\eta(v, v'; h) = \{\widehat{\eta}(v; h) - \widehat{\eta}(v'; h)\}^T \Sigma_n^{-1}(\widehat{\beta}(v; h)) \{\widehat{\eta}(v; h) - \widehat{\eta}(v'; h)\}$. The choice of $\omega_e(v, v'; h)$ ensures the convergence of $\widehat{\boldsymbol{\eta}}(v; h_1)$. Specifically, if $\boldsymbol{\eta}_*(v) = \boldsymbol{\eta}_*(v')$, then we have $D_\eta(v, v'; h) = O_p(n^{1/2})$ and $K_{st}(D_\eta(v, v'; h) / C_n)$ converges to infinity and zero, respectively. Therefore, $\omega_e(v, v'; h)$ can give a small weight to voxel v' . Similar to the proof of Theorems 1–3 in [43], we can show that as $n \rightarrow \infty$, $\widehat{\boldsymbol{\eta}}(v; h)$ converges $\boldsymbol{\eta}_*(v)$ and $\sqrt{n}\{\widehat{\eta}(v; h) - \eta_*(v)\} = O_p(1)$. Finally, we can prove Step (ii) for any fixed S .

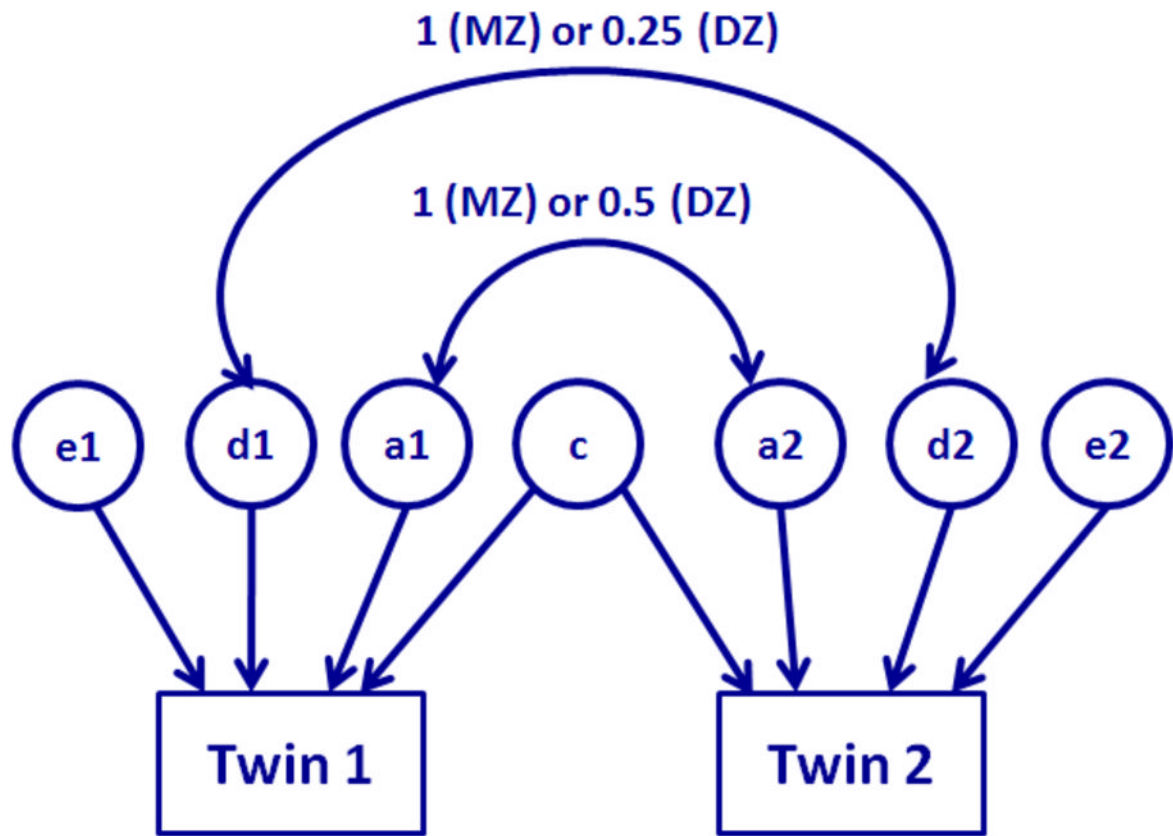


Fig. 1. Diagram for the structural equation model for twin data. The correlation of additive effects (a_1 , a_2) is 1 for MZ twin and 0.5 for DZ twin. The correlation of dominant effects (d_1 , d_2) is 1 for MZ twin and 0.25 for DZ twin. The twin share the same common environmental effect (c). Residual effects (e_1 , e_2) for twin are not correlated.

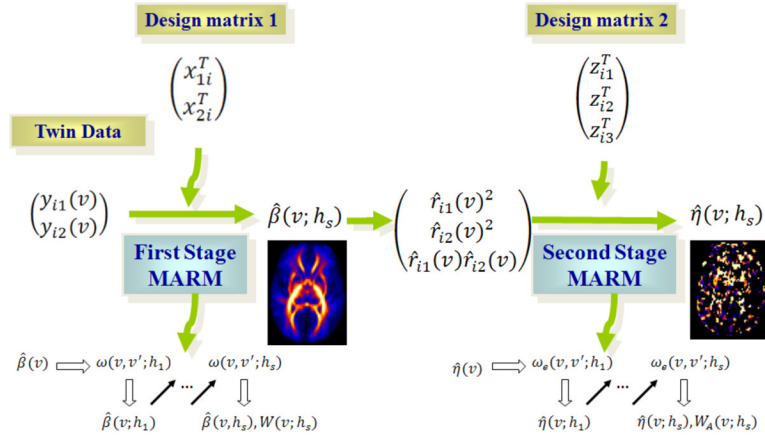


Fig. 2. Diagram for the TwinMARM method for twin data. The first stage is to estimate $\beta = \{\beta(v) : v \in \mathcal{V}\}$, while the second stage is to estimate $\eta = \{\eta(v) : v \in \mathcal{V}\}$. In each stage, we reformulate the problem of estimating β (or η) as a regression model and then apply the multiscale adaptive regression method (MARM) to spatially and adaptively calculate $\hat{\beta}(v; h_s)$ and $\hat{\eta}(v; h_s)$ and their associated test statistics. Moreover, y_{ij} denote imaging measures for twin pairs, x_{ij} is a vector of clinical variables, β is the vector of unknown regression parameters, r_{ij} are residuals obtained from first stage, η include the variances of additive genetic, common environmental and residual effects, and Z_{ij} is the design matrix for the second stage. $\omega(v, v'; h_s)$ is the weight of the first stage, whereas $\omega_d(v, v'; h_s)$ is the weight of the second stage.

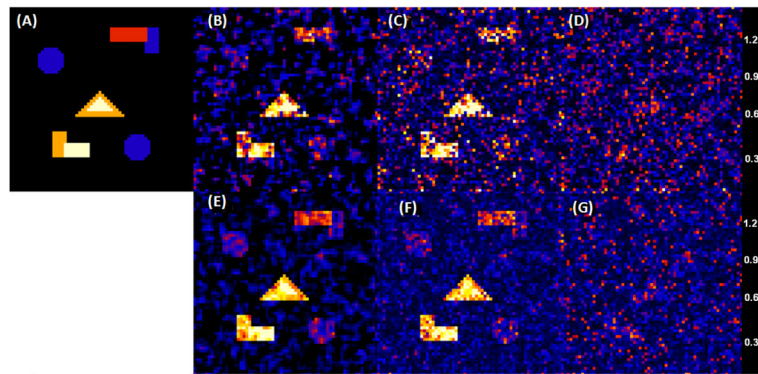


Fig. 3.

Results from a simulation study of comparing EM and TwinMARM at 2 different sample sizes ($n = 100, 400$). The first row contains the results for $\sigma_a^2(v)$ as $n = 100$: Panel (A) is the ground truth image of $\sigma_a^2(v)$, in which five ROIs with black, blue, red, yellow, and white color represent $\sigma_a^2(v) = 0, 0.3, 0.6, 0.9$ and 1.2 , respectively. Panel (B) is a selected slice of $\widehat{\sigma}_a^2(v; h_{10})$ obtained from a simulated dataset by using the TwinMARM method. Panel (C) is a selected slice of $\widehat{\sigma}_a^2(v)$ obtained from the same dataset as panel (B) by using the voxel-wise EM method. Panel (D) is a selected slice of $\widehat{\sigma}_a^2(v)$ by using EM, after smoothing the same simulated data set as panel (B). The second row contains panels (E), (F) and (G) as $n = 400$, which are the corresponding results of panels of panels (B), (C) and (D), respectively.

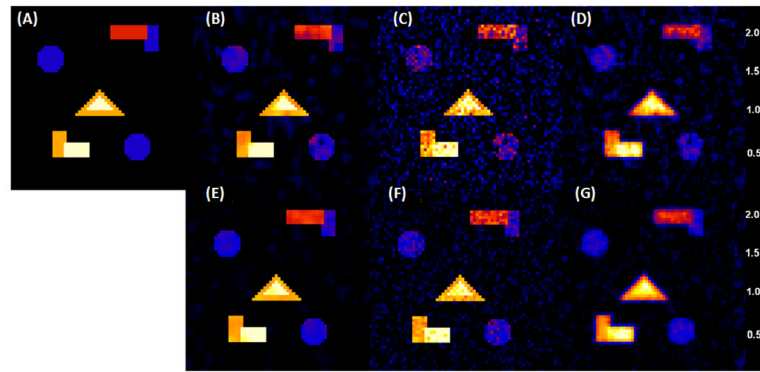


Fig. 4.

Results from a simulation study of comparing EM and TwinMARM at 2 different sample sizes ($n = 100, 400$). The first row contains the results for $\beta_1(v)$, when sample size $n = 100$: Panel (A) is the ground truth image of $\beta_1(v)$, in which five ROIs with black, blue, red, yellow, and white color represent $\beta_1(v) = 0, 0.5, 1.0, 1.5,$ and 2.0 , respectively. Panel (B) is a selected slice of $\hat{\beta}_1(v; h_{10})$ obtained from a simulated dataset by using TwinMARM. Panel (C) is a selected slice of $\hat{\beta}_1(v)$ obtained from the same simulated dataset as panel (B) by using EM. Panel (D) is a selected slice of $\hat{\beta}_1(v)$ obtained by using EM, after smoothing the same simulated data set as panel (B). The second row contains panels (E), (F) and (G) as $n = 400$, which are the corresponding results of panels of panels (B), (C) and (D), respectively.

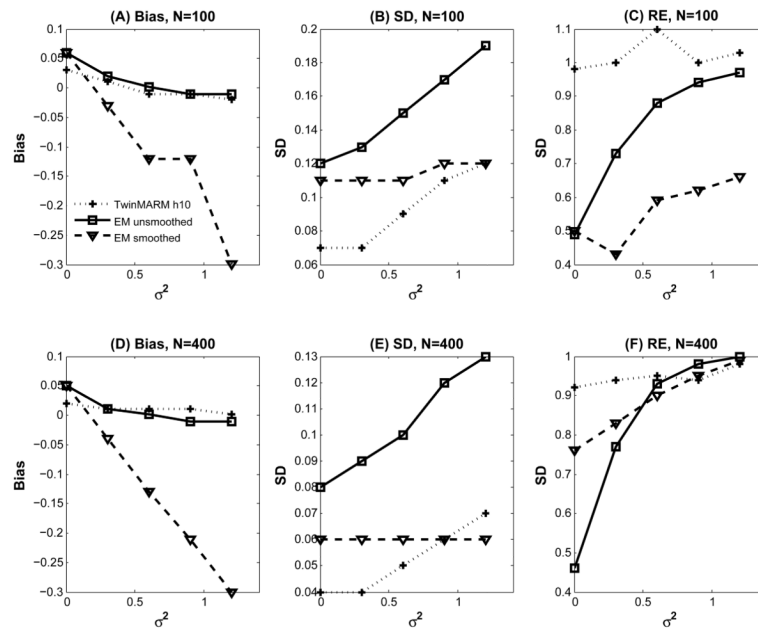


Fig. 5.

Results from a simulation study of comparing EM and TwinMARM at 2 different sample sizes ($n = 100, 400$). The first row contains the results for $\sigma_a^2(v)$ as $n = 100$: Panel (A) is the bias curve of $\sigma_a^2(v)=0, 0.3, 0.6, 0.9$ and 1.2 , respectively. Panel (B) is the SD curve of $\widehat{\sigma}_a^2(v;h_{10})$ obtained from a simulated dataset by using TwinMARM, EM, and EM with smoothed data. Panel (C) is the ratio of RMS over SD. The second row contains panels (D), (E), and (F) as $n = 400$, which are the corresponding results of panels of panels (A), (B), and (C) respectively.

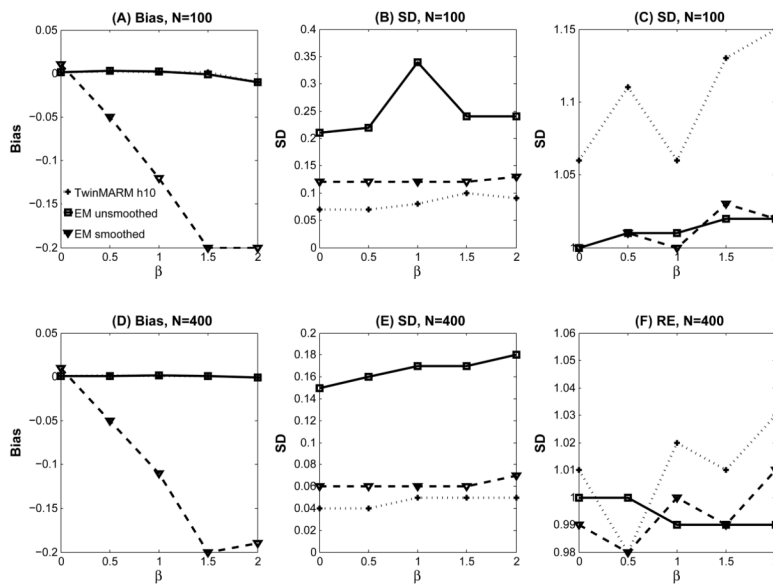


Fig. 6. Results from a simulation study of comparing EM and TwinMARM at 2 different sample sizes ($n = 100, 400$). The first row contains the results for $\beta_1(\nu)$ as $n = 100$: Panel (A) is the bias curve of $\beta_1(\nu) = 0, 0.5, 1.0, 1.5,$ and 2.0 , respectively. Panel (B) is the SD curve of $\hat{\beta}(\nu)$ obtained from a simulated dataset by using TwinMARM, EM, and EM with smoothed data. Panel (C) is the ratio of RMS over SD. The second row contains panels (D), (E), and (F) as $n = 400$, which are the corresponding results of panels of panels (A), (B), and (C) respectively.

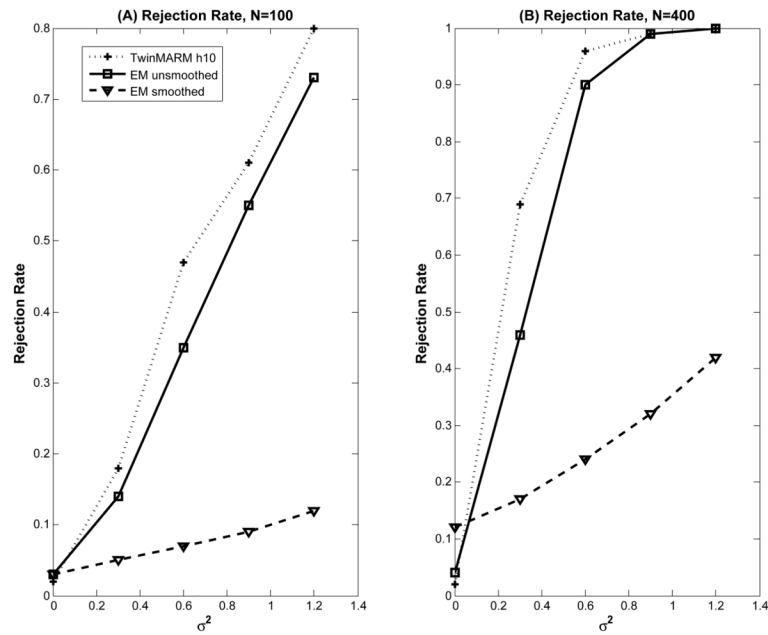


Fig. 7. Simulation results for $W_A(v; h)$: rejection rates for pixels inside the five ROIs were reported by using TwinMARM at the h_{10} scale, EM, EM after smoothing the same simulated data and 2 different sample sizes ($n = 100, 400$) at $\alpha = 5\%$. For each case, 1,000 simulated datasets were used.

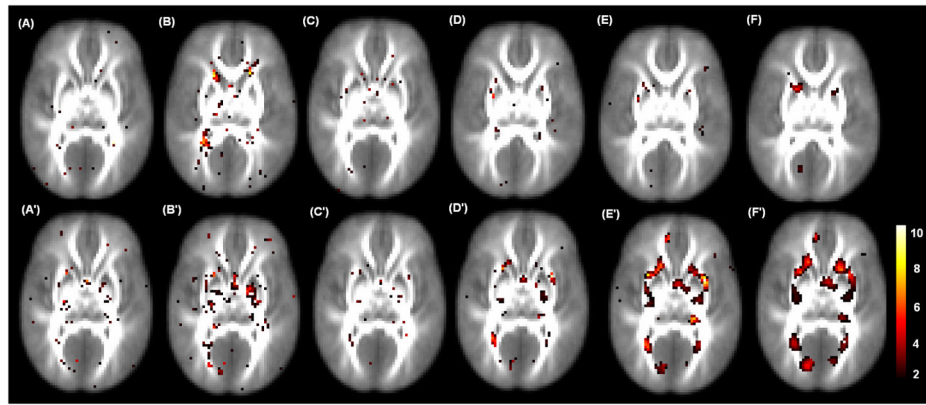


Fig. 8. Results from the 49 twin pairs in a neonatal project on brain development on the selected 27th and 30th slices. Panels (A)–(F) : the $-\log_{10}(p)$ values for testing genetic effects by using TwinMARM at the 1st and 10th iterations and EM with FWHM equal to 0mm, 3mm, 6mm, 9mm for the 27th slice; Panels (A')–(F') : the corresponding $-\log_{10}(p)$ values for testing environmental effects for the 30th slice.

Real data analysis results. # of voxels whose p -value is less than a specific threshold for testing the genetic and environmental effect for EM method without smoothing, smoothing with Gaussian kernels with FWHM equal to 3mm, 6mm, and 9mm and TwinMARM method at the 1st and 10th iteration.

Table 1

	Effect	EM smoothness (mm)				TwinMARM (round)			
		0	3	6	9	1 st	10 th		
p -value<0.01	Genetic	13	26	19	32	19	19	47	
	Environment	11	41	140	152	23	23	47	
p -value<0.05	Genetic	148	233	208	254	198	198	433	
	Environment	171	567	1318	1984	232	232	715	
p -value<0.10	Genetic	548	865	812	719	676	676	1430	
	Environment	615	1708	3769	5804	956	956	2568	

Real data analysis results. # of cluster which has over 20 contiguous voxels with p -value less than a specific threshold for testing the genetic and environmental effect by using EM method without smoothing, smoothing with Gaussian kernels (FWHM=3mm, 6mm and 9mm) and TwinMARM method at the 1^{st} and 10^{th} iteration.

Table 2

Effect	EM smoothness (mm)						TwinMARM (round)			
	0	3	6	9	1 st	10 th				
p -value<0.01	Genetic	0	0	0	1	0	0	0	0	
	Environment	0	0	1	2	0	0	0	0	
p -value<0.05	Genetic	0	0	3	3	0	0	3	3	
	Environment	0	4	17	11	0	0	4	4	
p -value<0.10	Genetic	0	10	9	8	1	1	9	9	
	Environment	4	21	18	10	3	3	20	20	