TWO-STAGE EMPIRICAL LIKELIHOOD FOR LONGITUDINAL NEUROIMAGING DATA

BY XIAOYAN SHI, JOSEPH G. IBRAHIM[1], JEFFREY LIEBERMAN,
MARTIN STYNER[2], YIMEI LI AND HONGTU ZHU[1]

University of North Carolina at Chapel Hill, University of North Carolina at Chapel Hill, Columbia University, University of North Carolina at Chapel Hill, St. Jude Children's Research Hospital and University of North Carolina at Chapel Hill

Longitudinal imaging studies are essential to understanding the neural development of neuropsychiatric disorders, substance use disorders, and the normal brain. The main objective of this paper is to develop a two-stage adjusted exponentially tilted empirical likelihood (TETEL) for the spatial analysis of neuroimaging data from longitudinal studies. The TETEL method as a frequentist approach allows us to efficiently analyze longitudinal data without modeling temporal correlation and to classify different time-dependent covariate types. To account for spatial dependence, the TETEL method developed here specifically combines all the data in the closest neighborhood of each voxel (or pixel) on a 3-dimensional (3D) volume (or 2D surface) with appropriate weights to calculate adaptive parameter estimates and adaptive test statistics. Simulation studies are used to examine the finite sample performance of the adjusted exponential tilted likelihood ratio statistic and TETEL. We demonstrate the application of our statistical methods to the detection of the difference in the morphological changes of the hippocampus across time between schizophrenia patients and healthy subjects in a longitudinal schizophrenia study.

1. Introduction. Neuroimaging data, including both anatomical and functional magnetic resonance imaging (MRI), have been/are being widely collected to understand the neural development of neuropsychiatric disorders, substance use disorders, and the normal brain in various longitudinal studies [Almli et al. (2007)]. For instance, various morphometrical measures of the morphology of the cortical and subcortical structures (e.g., hippocampus) are extracted from anatomical MRIs for understanding neuroanatomical differences in brain structure across

different populations and across time. Studies of brain morphology have been conducted widely to characterize differences in brain structure across groups of healthy individuals and persons with various diseases, and across time [Thompson and Toga (2002), Thompson, Cannon and Toga (2002), Styner et al. (2005), Zhu et al. (2008a)]. Moreover, functional MRI (fMRI) is a valuable tool for understanding functional integration of different brain regions in response to specific stimuli and behavioral tasks and detecting the association between brain function and covariates of interest, such as diagnosis, behavioral task, severity of disease, age, or IQ [Friston (2007), Rogers et al. (2007), Huettel, Song and McCarthy (2004)].

Much effort has been devoted to developing frequentist and Bayesian methods for analyzing neuroimaging data using numerical simulations and theoretical reasoning. Frequentist statistical methods for analyzing neuroimaging data are often sequentially executed in two steps. The first step involves fitting a general linear model or a linear mixed model to neuroimaging data from all subjects at each voxel [Beckmann, Jenkinson and Smith (2003), Friston et al. (2005), Rowe (2005), Woolrich et al. (2004), Zhu et al. (2008a)]. The second step is to calculate adjusted p-values that account for testing the hypotheses across multiple brain regions or across many voxels of the imaging volume using various statistical methods (e.g., random field theory, false discovery rate, or permutation method) [Cao and Worsley (2001), Friston et al. (1996), Hayasaka et al. (2004), Logan and Rowe (2004), Worsley et al. (2004)]. Most of these frequentist methods have been implemented in existing neuroimaging software platforms, including statistical parametric mapping (SPM) (www.fil.ion.ucl.ac.uk/spm/) and FMRIB Software Library (FSL) (www.fmrib.ox.ac.uk/fsl/), among many others. In the recent literature, a number of papers have been published on the development of Bayesian spatial–temporal models for functional imaging data [Penny, Flandin and Trujillo-Barreto (2007), Bowman et al. (2008), Woolrich et al. (2004), Luo and Puthusserypady (2005)]. Most Bayesian approaches, however, are less practical due to the extensively computational burden of running a Markov chain Monte Carlo method in a large number of voxels [Bowman et al. (2008)], and, thus, they are limited to small or moderate anatomic regions and a small number of regions of interest (ROI). Moreover, as pointed out in Snook, Plewes and Beaulieu (2007), the major drawbacks of ROI analysis include the instability of statistical results obtained from ROI analysis and the partial volume effect in relative large ROIs.

Existing statistical methods in the neuroimaging literature have two major limitations for analyzing longitudinal neuroimaging data, as explained below. The respective strategies to resolve these two limitations are detailed in Section 2. The first limitation is that the parametric models, such as linear mixed models, require the correct specification of the temporal correlation structure and cannot properly distinguish between different types of time-dependent covariates (types I, II and III) [Diggle et al. (2002), Lai and Small (2007), Pepe and Anderson (1994)]. A distinctive feature of longitudinal neuroimaging data is that it is able to characterize individual change in neuroimaging measurements (e.g., volumetric and

morphometric) over time, and the time-dependent covariates of interest may influence change. Imaging measurements of the same individual usually exhibit positive correlation and the strength of the correlation decreases with the time separation [Liang and Zeger (1986)]. Moreover, longitudinal data may provide crucial information for a causal role of a time-dependent covariate (e.g., exposure) in the disease process [Diggle et al. (2002), Lai and Small (2007), Pepe and Anderson (1994)]. Improperly handling time-dependent covariates and ignoring (or incorrectly modeling) temporal correlation structure in imaging measures likely would influence subsequent statistical inference, such as increasing the false positive and negative errors, and result in misleading scientific inferences [Diggle et al. (2002), Lai and Small (2007)].

The second limitation is that most smoothing methods apply the same amount of smoothing throughout the whole image, which can be problematic near the edges of the significant regions. Although it is common to apply a smoothing step before applying a voxel-wise approach for the analysis of neuroimaging data [Poline and Mazoyer (1994), Shafie et al. (2003), Lindquist and Wager (2008)], the voxel-wise method suffers from the same amount of smoothing throughout the whole image and the arbitrary choice of smoothing extent [Hecke et al. (2009), Jones et al. (2005)]. Jones et al. (2005) have shown that the final results of voxel-based analysis can strongly depend on the amount of smoothing in the smoothed diffusion imaging data. Recently, Yue, Loh and Lindquist (2010) introduced a spatially smoothing method using nonstationary spatial Gaussian Markov random fields to spatially and adaptively smooth images. Their approach, however, can be computationally extensive for 3D imaging data.

In this paper we will develop strategies to resolve these two limitations. To resolve the first limitation, we develop an adjusted exponentially tilted empirical likelihood method, called AETEL, for the analysis of longitudinal neuroimaging data with time-dependent covariates. AETEL is a nonparametric method that is built on a set of estimating equations and the number of estimating equations can be larger than the number of parameters. Thus, it avoids parametric assumptions and this feature is very appealing for the analysis of real neuroimaging data, such as brain morphological measures, because the distribution of the univariate (or multivariate) neuroimaging measurements often deviates from the Gaussian distribution [Ashburner and Friston (2000), Salmond et al. (2002), Luo and Nichols (2003)]. Using more estimating equations than the number of parameters allows us to appropriately handle time-dependent covariates of different types and to make an efficient use of the estimating equations without the need of modeling the temporal correlation in longitudinal data [Lai and Small (2007), Qu, Lindsay and Li (2000)]. AETEL also provides a natural test statistic to test whether a specific covariate is of a certain type (types I, II and III).

To resolve the second limitation, we develop a two-stage AETEL, abbreviated as TETEL, for the analysis of longitudinal neuroimaging data. TETEL integrates

a smoothing method into our AETEL for carrying out statistical inference on neuroimaging data. The TETEL method, as an adaptive procedure, fits AETEL at each voxel in stage 1. Then, TETEL uses the information learned from stage 1 to discard the data from the neighboring voxels with dissimilar signal pattern and to incorporate the data from the neighboring voxels with similar signal pattern to adaptively calculate parameter estimates and test statistics. TETEL avoids using the same amount of smoothing throughout the whole image in most smoothing methods. In addition, theoretically, we can establish consistency and asymptotic normality of the estimators and test statistics obtained from TETEL.

Section 2 of this paper introduces the shape data of the hippocampus structure from a longitudinal schizophrenia study and presents the new statistical methods just described. In Section 3 we conduct simulation studies to examine the finite sample performance of the TETEL method. Section 4 illustrates an application of the proposed methods to the longitudinal schizophrenia study of the hippocampus. We present concluding remarks in Section 5.

2. Data and methods.

2.1. *Longitudinal schizophrenia study of hippocampus shape.* This is a longitudinal, randomized, controlled, multisite, double-blind study conducted at 14 academic medical centers in North America and western Europe, with partial funding from Lilly Research Laboratories [Lieberman et al. (2005), Styner et al. (2004)]. In this study 238 first-episode schizophrenia patients were enrolled meeting the following criteria: age 16 to 40 years; onset of psychiatric symptoms before age 35; diagnosis of schizophrenia, schizophreniform, or schizoaffective disorder according to the fourth edition of diagnostic and statistical manual of mental disorders (DSM-IV) criteria; and various treatment and substance dependence conditions. After random allocation at baseline, 123 patients were selected to receive a conventional antipsychotic, haloperidol (2–20 mg/d), and 115 were selected to receive an atypical antipsychotic, olanzapine (5–20 mg/d). Patients were treated and followed up to 47 months. Also, 56 healthy control subjects matched to the patient's demographic characteristics were enrolled. Neurocognitive and MRI assessments were performed at months 0 (baseline), 3, 6, 13, 24, 36, and 47 approximately, with different subjects having different visiting times, and some subjects dropped out during the course of the study.

The hippocampus, a gray matter structure in the limbic system, is involved in processes of motivation and emotions and has a central role in the formation of memory. The hippocampus is a paired structure with mirror-image halves in the left and right brain hemispheres and located inside the medial temporal lobe (Figure 1). Many MRI studies have reported the reduction of hippocampal volume demonstrated in schizophrenia subjects and at onset of the first episode of psychotic symptoms before effects associated with treatment and disease chronicity [Lieberman et al. (2005)].
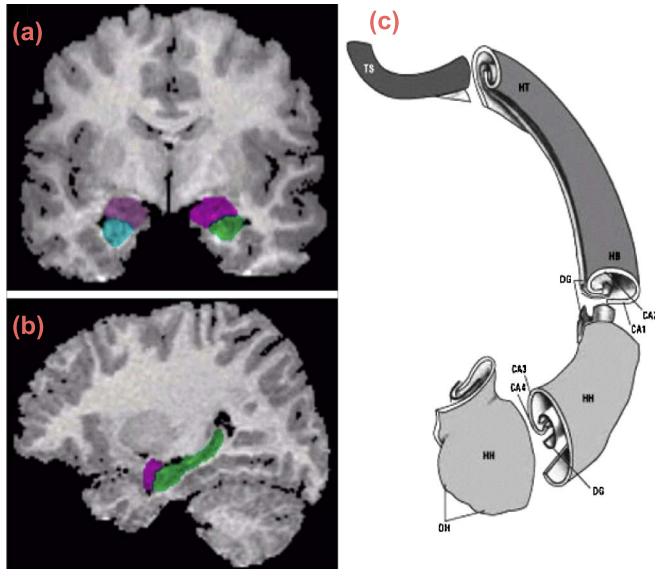
FIG. 1.   *Location of hippocampus in the context of the surrounding structures in the coronal* (a) *and sagittal* (b) *views. Subregions of the hippocampus in* (c) *showing the head of the hippocampus* (*HH*), *the digitationes hippocampi* (*DH*), *the hippocampal body* (*HB*), *the hippocampal tail* (*HT*), *the terminal segment of the HT* (*TS*), *the dentate gyrus* (*DG*), *and the fields of the cornu ammonis* (*CA*1–*CA*4). *Adapted with permission from Springer Verlag*, *Heidelberg*, *Germany* [*Duvernoy* (2005)].

The aim of this study is to use the boundary and medial shape of the hippocampus to examine whether hippocampal abnormalities are present in schizophrenia patients. Statistical shape modeling and analysis have emerged as important tools for understanding cortical and subcortical structures from medical images [Dryden and Mardia (1998)]. We consider two approaches for shape representation including a spherical harmonic description sampled into a triangulated surfaces (SPHARM-PDM) and a medial shape description [Pizer et al. (2003), Styner and Gerig (2003)]. The SPHARM-PDM can only represent objects of spherical topology, whereas the medial representation provides information on a rich set of features, including local thickness. These shape features are not accessible by conventional volume-based morphometry and offer us a great opportunity to address the weaknesses of conventional volumetric methods.

We consider two sets of responses of interest. The first set of responses was based on the SPHARM-PDM representation of hippocampal surfaces. We use the SPHARM-PDM [Styner et al. (2004)] shape representation to establish surface correspondence and align the surface location vectors across all subjects. The sampled SPHARM-PDM is a smooth, accurate, fine-scale shape representation (Figure 3). The hippocampal surfaces of different subjects are thus represented by the same number of location vectors (with each location vector consisting of the spatial x, y, and z coordinates of the corresponding vertex on the SPHARM-

PDM surface) and are used as the second set of responses. Covariates of interest are race (Caucasian, African American, and others), age (in years), gender, group (the schizophrenia group and the healthy control group) and time (visiting times in months).

The second set of responses was the hippocampus m-rep thickness at the 24 medial atoms of the left and the right brain (Figure 4). The m-rep is a linked set of medial primitives named medial atoms, which are formed from two equal length vectors and are composed of a position, a radius, a frame implying the tangent plane to the medial manifold, and an object angle [Styner et al. (2004)]. The m-rep thickness is the radius of each medial atom. Covariates of interest were WBV, race (Caucasian, African American, and others), age (in years), gender, diagnostic status (patient or control), and visiting times (in weeks). This WBV measure includes gray and white matter, ventricular cerebrospinal fluid, cisterns, fissures, and cortical sulci. The WBV is commonly used as a covariate in statistical analyses to control for scaling effects [Arndt et al. (1991)]. Particularly, WBV is a time-dependent covariate and may vary with the hippocampus thickness measurement.

2.2. *Estimating equations for longitudinal data.* We consider a longitudinal study of imaging data with n subjects, where a $q \times 1$ covariate $\mathbf{x}_{i,j}$ (e.g., age, gender, height, and brain volume) is obtained for the ith subject at the jth time point t_{ij} for $i = 1, \ldots, n$ and $j = 1, \ldots, m_i$. Thus, there are at least $\sum_{i=1}^{n} m_i = N$ images in the study. Based on each image, we observe or compute neuroimaging measures, denoted by $\mathbf{Y}_i = \{\mathbf{y}_{ij}(d) : d \in \mathcal{D}, j = 1, \ldots, m_i\}$, across all m_i time points from the ith subject, where d represents a voxel (or atom, or point) on $\mathcal{D}$, a specific brain region. The imaging measure $\mathbf{y}_{ij}(d)$ at each voxel d can be either univariate or multivariate. For example, the m-rep thickness is a univariate measure, whereas the location vector of SPHARM-PDM is a three-dimensional MRI measure at each point [Styner and Gerig (2003), Chung, Dalton and Davidson (2007)]. For notational simplicity, we assume that the $\mathbf{y}_{ij}(d)$ are univariate measures.

We temporarily drop voxel d from our notation. At a specific voxel d in the brain region, $\mathbf{z}_i = \{(\mathbf{y}_{ij}, \mathbf{x}_{ij}) : j = 1, \ldots, m_i\}$ is independent and satisfies a moment condition

$$(2.1) \qquad\qquad E\{g(\mathbf{z}_i, \theta)\} = 0 \qquad \text{for } i = 1, \ldots, n,$$

where θ is a $p \times 1$ vector, $g(\cdot, \cdot)$ is an $r \times 1$ vector of known functions with $r \geq p$, and E denotes the expectation with respect to the true distribution of all the $\mathbf{z}_i$'s. Equation (2.1) is often referred to as a set of unbiased estimating equations or moments model [Qin and Lawless (1994), Hansen (1982)]. The moments model (2.1) is more general than most parametric models including linear mixed model used for the analysis of neuroimaging data [Worsley et al. (2004), Qin and Lawless (1994), Hansen (1982), Schennach (2007), Owen (2001)].

For longitudinal data, although the measurements from different subjects are independent, those within the same subject may be highly correlated. The generalized estimating equations (GEE) assume a working covariance matrix for $\mathbf{y}_i =$

$(\mathbf{y}_{i1}, \ldots, \mathbf{y}_{im_i})^T$ given by V_i. Let $E(\mathbf{y}_i) = \mu_i(\beta) = (\mu_{i1}(\beta), \ldots, \mu_{im_i}(\beta))^T$ and $D_i(\beta) = \partial \mu_i(\beta)/\partial \beta$. Under the assumption that $E\{D_i(\beta)^T V_i^{-1}[\mathbf{y}_i - \mu_i(\beta)]\} = 0$, Liang and Zeger (1986) proposed to use an estimator, denoted by $\hat{\beta}_{\text{gee}}$, which solves a set of GEEs as follows:

$$(2.2) \qquad G(\beta) = \sum_{i=1}^n D_i(\beta)^T V_i^{-1}[\mathbf{y}_i - \mu_i(\beta)] = \mathbf{0}.$$

For longitudinal data with time-dependent covariates, whether $E[g(\mathbf{z}_i, \theta)] = E\{D_i(\beta)^T V_i^{-1}[\mathbf{y}_i - \mu_i(\beta)]\}$ equals zero or not depends on the type of time-dependent covariates and the structure of V_i [Lai and Small (2007)]. The time-dependent covariate $\mathbf{x}_{ij}$ is of type I if

$$(2.3) \qquad E\{\partial_\beta \mu_{is}(\beta)[\mathbf{y}_{ij} - \mu_{ij}(\beta)]\} = 0 \qquad \text{for all } s, j = 1, \ldots, m_i,$$

where $\partial_\beta = \partial/\partial\beta$. A sufficient condition for type I covariates is $E[y_{ij}|\mathbf{x}_{ij}] = E[y_{ij}|\mathbf{x}_{i1}, \ldots, \mathbf{x}_{im_i}]$. For type I covariates, we can set $g(\mathbf{z}_i, \theta) = D_i(\beta)^T V_i^{-1}[\mathbf{y}_i - \mu_i(\beta)]$ and show that $E[g(\mathbf{z}_i, \theta)] = 0$. If V_i is the covariance matrix of $\mathbf{y}_i$, then the estimator $\hat{\beta}_{\text{gee}}$ is an efficient estimator. However, $\hat{\beta}_{\text{gee}}$ is inefficient under a mis-specified V_i. To increase the efficiency, we may choose several candidate working covariance matrices $M_i^{(1)}, \ldots, M_i^{(s_0)}$ and assume $V_i^{-1} = \sum_{k=1}^{s_0} \alpha_k M_i^{(k)}$ for some unknown constants α_k [Qu, Lindsay and Li (2000)]. Then, following Qu, Lindsay and Li (2000), we consider a set of estimating equations given by

$$(2.4) \qquad g(\mathbf{z}_i, \theta) = \begin{pmatrix} D_i(\beta)^T M_i^{(1)}[\mathbf{y}_i - \mu_i(\beta)] \\ \vdots \\ D_i(\beta)^T M_i^{(s_0)}[\mathbf{y}_i - \mu_i(\beta)] \end{pmatrix} \qquad \text{for } i = 1, \ldots, n.$$

In this case, the number of functions in $g(\mathbf{z}_i, \theta)$ is $s_0 q > q$, when $s_0 > 1$.

The time-dependent covariate $\mathbf{x}_{ij}$ is of type II if

$$(2.5) \qquad E\{\partial_\beta \mu_{is}(\beta)[\mathbf{y}_{ij} - \mu_{ij}(\beta)]\} = 0 \qquad \text{for all } s \geq j, j = 1, \ldots, m_i.$$

A sufficient condition for type II covariates is

$$(2.6) \qquad p(\mathbf{x}_{i,t+1}, \ldots, \mathbf{x}_{im_i}|\mathbf{y}_{it}, \mathbf{x}_{it}) = p(\mathbf{x}_{i,t+1}, \ldots, \mathbf{x}_{im_i}|\mathbf{x}_{it}).$$

For type II covariates, we can set $g(\mathbf{z}_i, \theta) = D_i(\beta)^T[\mathbf{y}_i - \mu_i(\beta)]$, in which an independent working covariance matrix is used. However, the estimator $\hat{\beta}_{\text{gee}}$ based on the independent working correlation matrix is inefficient, since we do not use the information contained in $E\{\partial_\beta \mu_{is}(\beta)[\mathbf{y}_{ij} - \mu_{ij}(\beta)]\} = 0$ for all $s > j$. To increase the efficiency of the estimate, we choose a set of lower triangular $m_i \times m_i$ matrices $L_i^{(1)}, \ldots, L_i^{(s_0)}$, and then we consider estimating equations given by

$$(2.7) \qquad g(\mathbf{z}_i, \theta) = \begin{pmatrix} D_i(\beta)^T L_i^{(1)}[\mathbf{y}_i - \mu_i(\beta)] \\ \vdots \\ D_i(\beta)^T L_i^{(s_0)}[\mathbf{y}_i - \mu_i(\beta)] \end{pmatrix} \qquad \text{for } i = 1, \ldots, n.$$

In this case, the number of functions in $g(\mathbf{z}_i, \theta)$ is $s_0 q > q$, when $s_0 > 1$. Supposing that $m_1 = \cdots = m_n$, we can set $s_0 = m_1(m_1 + 1)/2$ and $L_i^{(b)} = \mathbf{e}_s \mathbf{e}_j^T$ for $s \geq j$ and $b = 1, \ldots, s_0$, where $\mathbf{e}_s$ is a $q \times 1$ vector with the sth component 1 and 0 otherwise. Thus, similar to Lai and Small (2007), we are able to pick $\partial_\beta \mu_{is}(\beta)[\mathbf{y}_{ij} - \mu_{ij}(\beta)]$ for all $s \geq j$.

The time-dependent covariate $\mathbf{x}_{ij}$ is of type III if

$$(2.8) \qquad E\{\partial_\beta \mu_{is}(\beta)[\mathbf{y}_{ij} - \mu_{ij}(\beta)]\} \neq 0 \qquad \text{for some } s > j.$$

For type III covariates, we need to choose V_i as a diagonal matrix. For instance, if $V_i = \mathbf{I}_i$, where $\mathbf{I}_i$ is an $m_i \times m_i$ identity matrix, then $g(\mathbf{z}_i, \theta) = D_i(\beta)^T[\mathbf{y}_i - \mu_i(\beta)]$. Furthermore, if we assume the specific form for the variances of all $\mathbf{y}_{ij}$, then we may set $V_i = \text{diag}(\text{Cov}(\mathbf{y}_i))$.

An overall strategy to analyze models with time-dependent covariates is first to assume that the time-dependent covariates are of type III. Then we test whether the time-dependent covariates are of type II, and if the test is not rejected, we can go on to test if they are of type I. Once the type of all the time-dependent covariates is decided, we use the corresponding estimating equations. See Section 4 for more details.

2.3. *Adjusted exponentially tilted empirical likelihood.* We consider a nonparametric method, called an exponentially tilted empirical likelihood, to carry out statistical inference about θ based on a set of estimating equations $\{g(\mathbf{z}_i, \theta) : i = 1, \ldots, n\}$ [Schennach (2007)]. The exponentially tilted empirical likelihood (ETEL) method is a combination of the exponentially tilted (ET) method and the empirical likelihood (EL) method. Both EL [Owen (2001), Qin and Lawless (1994)] and ET [Imbens, Spady and Johnson (1998)] methods combine the reliability of nonparametric methods with the effectiveness of the likelihood approach. The EL estimator exhibits desirable higher-order asymptotic properties, whereas the EL estimator may fail to be $\sqrt{n}$-convergent in the presence of model misspcification. In contrast, the ETEL estimator maintains $\sqrt{n}$-convergence under model misspecification [Schennach (2007)].

However, most empirical likelihood type methods including ETEL suffer from two pitfalls: relatively low precision of the chi-square approximation and nonexistence of solutions to the estimating equations [Chen, Variyath and Abraham (2008), Liu and Chen (2010)]. Chen, Variyath and Abraham (2008) introduce a novel adjustment to these empirical likelihood methods and develop an iterative algorithm that converges very fast. Simulation studies have shown that the adjusted empirical likelihood methods perform as well as the linear regression model with Gaussian noise when data are symmetrically distributed, while the adjusted empirical likelihood methods are superior when data have skewed distribution [Zhu et al. (2009), Chen, Variyath and Abraham (2008), Liu and Chen (2010)].

Following Chen, Variyath and Abraham (2008), we consider an adjustment of ETEL, abbreviated as AETEL, by introducing an adjustment

$$(2.9) \qquad g_{n+1}(\theta) = -\frac{a_n}{n} \sum_{i=1}^{n} g(\mathbf{z}_i, \theta),$$

where $a_n = \max(1, \log(n)/2)$. Then, AETEL is defined as

$$(2.10) \qquad \ell_{\mathrm{Aetel}}(\theta) = -(n+1)^{-1} \sum_{i=1}^{n+1} \log\big((n+1)\hat{p}_i(\theta)\big),$$

where $\hat{p}_i(\theta)$ is the solution to

$$\min_{p_1,\ldots,p_{n+1}} (n+1)^{-1} \sum_{i=1}^{n+1} [(n+1)p_i] \log[(n+1)p_i]$$

subject to

$$\sum_{i=1}^{n+1} p_i = 1,\ p_i \geq 0 \quad \text{and} \quad \sum_{i=1}^{n} p_i g(\mathbf{z}_i, \theta) + p_{n+1} g_{n+1}(\theta) = 0.$$

The maximum AETEL estimator, denoted by $\hat{\theta}_{\mathrm{Aetel}}$, minimizes a criterion given by

$$(2.11) \qquad \hat{\theta}_{\mathrm{Aetel}} = \operatorname*{argmin}_{\theta} \ell_{\mathrm{Aetel}}(\theta).$$

According to a duality theorem in convex analysis [Newey and Smith (2004)],

$$\hat{p}_{n+1}(\theta) = \frac{\exp(\hat{t}(\theta)^T g_{n+1}(\theta))}{T_g(\theta)} \quad \text{and} \quad \hat{p}_i(\theta) = \frac{\exp(\hat{t}(\theta)^T g(\mathbf{z}_i, \theta))}{T_g(\theta)}$$

for $i = 1, \ldots, n$, in which

$$T_g(\theta) = \sum_{j=1}^{n} \exp(\hat{t}(\theta)^T g(\mathbf{z}_j, \theta)) + \exp(\hat{t}(\theta)^T g_{n+1}(\theta)),$$

$$\hat{t}(\theta) = \operatorname*{argmax}_{t} \left\{ -\sum_{i=1}^{n} \exp(-t^T g(\mathbf{z}_i, \theta)) - \exp(-t^T g_{n+1}(\theta)) \right\}.$$

We use the numerical algorithm proposed by Chen, Variyath and Abraham (2008) to compute $\hat{\theta}_{\mathrm{Aetel}}$, which combines the modified Newton–Raphson algorithm and the simplex method. Compared with that of computing ETEL, this numerical algorithm of Chen, Variyath and Abraham (2008) converges very fast and the solution to AETEL is guaranteed.

We consider testing the linear hypotheses:

$$(2.12) \qquad H_0 : R\theta = \mathbf{b}_0 \quad \text{vs.} \quad H_1 : R\theta \neq \mathbf{b}_0,$$

where R is a $c_0 \times p$ matrix of full row rank and $\mathbf{b}_0$ is a $c_0 \times 1$ specified vector. Most scientific questions in neuroimaging studies can be formulated into linear hypotheses, such as a comparison of brain regions across diagnostic groups and a detection of changes in brain regions across time. The AETEL ratio statistic for testing $R\theta = \mathbf{b}_0$ can be constructed as follows:

$$(2.13) \qquad LR_{\text{Aetel}} = -2(n+1)\left\{ \sup_{\theta\,:\,R\theta=\mathbf{b}_0} \ell_{\text{Aetel}}(\theta) - \sup_{\theta} \ell_{\text{Aetel}}(\theta) \right\}.$$

Thus, to compute LR_{Aetel}, we also need to compute the maximum AETEL estimator, denoted by $\hat{\theta}_{\text{Aetel},0}$, subject to an additional constraint $R\theta = \mathbf{b}_0$.

Under some conditions on $g(\mathbf{z}_i, \theta)$, we have the following theorem, whose detailed proof can be found in a supplementary document [Shi et al. (2011)].

THEOREM 2.1.   *If assumptions* (A1)–(A4) *in the supplementary document are true, then we have the following*:

(a) $\sqrt{n}(\hat{\theta}_{\text{Aetel}} - \theta_0)$ *converges to* $v_0 = N(0, \Sigma)$ *in distribution, where* θ_0 *denotes the true value of* θ *and* $\Sigma = (DV^{-1}D^T)^{-1}$, *in which*

$$D = \lim_{n\to\infty} n^{-1} \sum_{i=1}^{n} \partial_{\theta} g(\mathbf{z}_i, \theta) \quad and \quad V = \lim_{n\to\infty} n^{-1} \sum_{i=1}^{n} g(\mathbf{z}_i, \theta)^{\otimes 2};$$

(b) *under the null hypothesis* H_0, LR_{Aetel} *converges to a* $\chi^2(c_0)$ *distribution*;

(c) *if* $E[g(\mathbf{z}_i, \theta)] = 0$ *for all* i *and* $r > p$, *then* $LR_{\text{GF}} = -2(n+1)\sup_{\theta} \ell_{\text{Aetel}}(\theta)$ *is asymptotically* $\chi^2(r - p)$.

We have established consistency and asymptotic normality of $\hat{\theta}_{\text{Aetel}}$ and the asymptotic χ^2 distribution of LR_{Aetel}. Theorem 2.1 also shows that AETEL has the same first-order asymptotic properties as ETEL [Schennach (2007)]. High-order precision of AETEL can be explored by following the arguments in Liu and Chen (2010). It will be shown that the chi-square approximation of the AETEL likelihood ratio statistics is found precise, compared with the existing ETEL [Owen (2001), Liu and Chen (2010), Chen, Variyath and Abraham (2008)]. Providing a reliable p-value at each voxel is crucial for controlling the family-wise error rate and false discovery rate (FDR) across the entire brain region [Benjamini and Hochberg (1995), Worsley et al. (2004)].

2.4. *Two-stage adaptive estimation procedure.*   We propose a two-stage adaptive estimation procedure for computing parameter estimates and likelihood ratio statistics for the spatial and adaptive analysis of neuroimaging data in 3D volumes (or 2D surfaces). To distinguish data and parameter in different voxels, we introduce voxel d into our notation. For instance, $\mathbf{z}_i(d)$ and $\theta(d)$, respectively, denote the ith observation and the parameter at voxel d.

Stage 1 is to calculate the AETEL estimator of the parameter $\theta(d)$, denoted by $\hat{\theta}_{\text{Aetel}}(d)$, based on a set of estimating equations $\{g(\mathbf{z}_i(d), \theta(d)) : i = 1, \ldots, n\}$ at each voxel $d \in \mathcal{D}$.

One chooses a set of estimating equations $\{g(\mathbf{z}_i(d), \theta(d)) : i = 1, \ldots, n\}$ according to a specific type of time-dependent covariate and then substitutes them into (2.10) to build $\ell_{\text{Aetel}}(\theta(d); d)$. Subsequently, we solve $\hat{\theta}_{\text{Aetel}}(d)$ according to (2.11) by minimizing $\ell_{\text{Aetel}}(\theta(d); d)$, and then we obtain a set of parameter estimates $\{\hat{\theta}_{\text{Aetel}}(d) : d \in \mathcal{D}\}$.

Stage 2 is to calculate the TETEL estimator of $\theta(d)$, denoted by $\hat{\theta}_{\text{Tetel}}(d)$, by utilizing the information contained in $\{\hat{\theta}_{\text{Aetel}}(d) : d \in \mathcal{D}\}$. Then, we calculate a TETEL ratio statistic, denoted by $LR_{\text{Tetel}}(d)$, for testing $H_0(d) : R\theta(d) = \mathbf{b}_0$.

Specifically, one combines all data in the voxel d and the set of the closest neighboring voxels of d, denoted by $N(d)$, to form a new set of estimating equations $\{\tilde{g}(\mathbf{z}_i(d), \theta(d); d) : i = 1, \ldots, n\}$ as follows:

$$(2.14) \qquad \tilde{g}(\mathbf{z}_i(d), \theta(d); d) = \sum_{d' \in N(d) \cup \{d\}} \omega(d'; d) g(\mathbf{z}_i(d'), \theta(d)),$$

where $\omega(d'; d)$ is a weight describing the similarity between voxel d and any $d' \in N(d)$. The weights $\omega(d'; d)$ at each d depend on the parameters $\{\hat{\theta}_{\text{Aetel}}(d') : d' \in N(d) \cup \{d\}\}$ calculated in Stage 1. From now on, we assume that $\omega(d'; d)$ takes the form

$$(2.15) \qquad\qquad \omega(d'; d) = \exp(-LR_{\text{Aetel}}(d'; d)/C_n),$$

where $C_n = \chi_{1-\alpha}^2(p) \log(n)/5$ and $\chi_{1-\alpha}^2(p)$ is the upper α-percentile of the $\chi^2(p)$ distribution. In addition,

$$(2.16) \qquad LR_{\text{Aetel}}(d'; d) = -2(n+1)\Big\{\ell_{\text{Aetel}}(\hat{\theta}_{\text{Aetel}}(d'); d) - \sup_\theta \ell_{\text{Aetel}}(\theta; d)\Big\},$$

in which $\ell_{\text{Aetel}}(\theta; d)$ is defined in (2.10) based on the estimating equations $\{g(\mathbf{z}_i(d), \theta(d)) : i = 1, \ldots, n\}$. Statistically, $LR_{\text{Aetel}}(d'; d)$ denotes the AETEL ratio statistic for testing the hypothesis $H_0 : \theta(d) = \hat{\theta}_{\text{Aetel}}(d')$. Note that $LR_{\text{Aetel}}(d'; d) \geq 0$ and $LR_{\text{Aetel}}(d; d) = 0$, which yields $\omega(d; d) = 1$. If $\hat{\theta}_{\text{Aetel}}(d')$ is close to $\hat{\theta}_{\text{Aetel}}(d)$, then $LR_{\text{Aetel}}(d'; d)$ is close to zero and $\omega(d'; d)$ will be close to 1. However, if the difference between $\hat{\theta}_{\text{Aetel}}(d')$ and $\hat{\theta}_{\text{Aetel}}(d)$ is large, then $LR_{\text{Aetel}}(d'; d)$ is large and $\omega(d'; d)$ will be small. Thus, $\omega(d'; d)$ defined in (2.15) truly characterizes the similarity between voxels d and d'.

One substitutes $\tilde{g}(\mathbf{z}_i(d), \theta(d); d)$ in (2.14) into (2.10) to build a new function, denoted by $\ell_{\text{Tetel}}(\theta(d); d)$, and then solves $\hat{\theta}_{\text{Tetel}}(d)$ according to (2.11) by minimizing $\ell_{\text{Tetel}}(\theta(d); d)$. Finally, to test $H_0(d) : R\theta(d) = \mathbf{b}_0$, one uses $\tilde{g}(\mathbf{z}_i(d), \theta(d); d)$ in (2.14) to calculate the TETEL ratio statistic $LR_{\text{Tetel}}(d)$ according to (2.13). Note that the key difference between $LR_{\text{Tetel}}(d)$ and $LR_{\text{Aetel}}(d)$ lies in their different sets of estimating equations.

Although the two-stage procedure only combines the data in the voxels of $N(d)$ with the data in voxel d, they may preserve the long-range correlation structure in the imaging data, because the neighborhoods of all voxels are consecutively connected. Thus, the two-stage procedure captures a substantial amount of spatial information in the imaging data. For the sake of space, we only present the asymptotic properties of $\hat{\theta}_{\text{Tetel}}(d)$ and $LR_{\text{Tetel}}(d)$ below.

THEOREM 2.2. *If assumptions* (A1)–(A3) *and* (A5)–(A7) *in the supplementary document are true, then we have the following*:

(a) $\sqrt{n}(\hat{\theta}_{\text{Tetel}}(d) - \theta_0(d))$ *converges to* $v(d) = N(0, \Sigma(d))$ *in distribution, where* $\theta_0(d)$ *is the true value of* $\theta(d)$ *in the voxel* d *and* $\Sigma(d) = [D(d)V(d)^{-1} \times D(d)^T]^{-1}$, *in which*

$$D(d) = \lim_{n \to \infty} n^{-1} \sum_{i=1}^{n} \partial_\theta \tilde{g}(\mathbf{z}_i(d), \theta_0(d); d)$$

and

$$V(d) = \lim_{n \to \infty} n^{-1} \sum_{i=1}^{n} \tilde{g}(\mathbf{z}_i(d), \theta_0(d); d)^{\otimes 2};$$

(b) *under the null hypothesis* $H_0(d)$, $LR_{\text{Tetel}}(d)$ *converges in distribution to a* $\chi^2(c_0)$ *distribution.*

Theorem 2.2 establishes the asymptotic consistency and normality of $\hat{\theta}_{\text{Tetel}}(d)$ and the asymptotic χ^2 distribution of $LR_{\text{Tetel}}(d)$. Theorem 2.2 also shows that the asymptotic variance of $\hat{\theta}_{\text{Tetel}}(d)$ depends on all the data in $N(d) \cup \{d\}$ for all subjects. Since the weights $\omega(d'; d)$ automatically put large weights on the neighboring voxels with similar pattern and small weights on the neighboring voxels with dissimilar pattern, it follows that the TETEL procedure produces more accurate parameter estimates and more powerful test statistics.

TETEL has three features. TETEL not only downweights the data from the neighboring voxels with dissimilar signal pattern, but also incorporates the data from the neighboring voxels with similar signal pattern to adaptively calculate parameter estimates and test statistics. TETEL avoids using the same amount of smoothness throughout the whole image in most smoothing methods. Our theoretical results ensure the asymptotic consistency and normality of $\hat{\theta}_{\text{Tetel}}(d)$ and the asymptotic χ^2 distribution of $LR_{\text{Tetel}}(d)$. Then, we can approximate the p-value of $LR_{\text{Tetel}}(d)$ at each voxel d. Finally, we correct for multiple comparisons by using either the family-wise error rate or false discovery rate (FDR) across the entire brain region [Benjamini and Hochberg (1995), Worsley et al. (2004)]. Since the smoothing stage in TETEL usually introduces the positive dependency among all $LR_{\text{Tetel}}(d)$, it allows us to apply FDR in Benjamini and Yekutieli (2001) to control the false discovery rate.

3. Simulation studies. Three sets of simulation studies were conducted to examine the finite sample performance of our AETEL and TETEL methods.

3.1. *Study I*: *Longitudinal data.* We considered the following model:

$$(3.1) \qquad \mathbf{y}_{ij} = \beta_0 + \beta_1 t_{ij} + \beta_2 x_i + \beta_3 t_{ij} x_i + b_i + \varepsilon_{ij}$$

for $i = 1, \ldots, n$, where t_{ij} denotes time taking values in $(1, 2, 3, 4, 5)$, x_i was independently generated from a $N(0, 1)$, and b_i was independently generated from a $N(0, 1)$. Errors ε_{ij} were independently generated from $N(0, 1)$ and $\chi^2(3) - 3$, respectively, where $\chi^2(3)$ represents a chi-squared random variable with three degrees of freedom. The $\chi^2(3) - 3$ distribution is very skewed and differs substantially from any symmetric distribution, such as a Gaussian distribution. The true value of $(\beta_0, \beta_1, \beta_2)^T$ was set at $(1, 1, 1)^T$ and β_3 was varied as 0, 0.05, 0.10, 0.15, and 0.20. We tested the hypothesis $H_0 : \beta_3 = 0$ vs. $H_1 : \beta_3 \neq 0$ using LR_{Aetel}. To assess both Type I and II error rates of LR_{Aetel}, we used generalized estimating equations assuming an exchangeable working correlation matrix to construct LR_{Aetel} and then compared it with the ETEL likelihood ratio statistic, denoted by LR_{Etel}, and the Wald statistic, denoted by W_n, obtained from the "true" linear mixed model (3.1) representing an ideal scenario. We considered $n = 40$, 60, and 80. The 1,000 replications were used to calculate the estimates of rejection rates with significance level $\alpha = 5\%$.

The type I error rates of LR_{Aetel} and W_n are reasonably accurate for all sample sizes ($n = 40$, 60, or 80) considered and for all different distributions of error terms at the 5% significant level (Table 1). In contrast, the type I error rates of LR_{Etel} are slightly inflated for $n = 40$. The type II error rates for LR_{Aetel} and W_n are similar under both error distributions and for all sample sizes (Table 1). However, the power of the three test statistics to reject the null hypothesis increases modestly when the distribution of the error terms follows the skewed distribution $\chi^2(3) - 3$ (Table 1). This decline in the type II error rate was caused by the fact that the variance of $\chi^2(3) - 3$ is larger than that of $N(0, 1)$. Compared with LR_{Aetel} and W_n, LR_{Etel} has slightly larger power, which may be due to its inflated type I error rates. Consistent with our expectation, the statistical power for rejecting the null hypothesis increases with the sample size n.

3.2. *Study II*: *Testing the type of time-dependent covariates.* We used the simulation study for a type II time-dependent covariate in Section 4.1 of Lai and Small (2007) to examine the finite sample performance of our AETEL method. The data were simulated under the mechanism

$$y_{it} = \beta_0 + \beta_1 x_{it} + \beta_2 x_{i,t-1} + b_i + e_{it} \quad \text{and} \quad x_{it} = \beta_3 x_{i,t-1} + \varepsilon_{it},$$

where b_i, e_{it}, and ε_{it} are mutually independent and normally distributed with mean 0 and variances 4, 1, and 1, respectively; the x_{it}-process is stationary, that is,

TABLE 1
Simulation study for comparing LR_{Aetel}, LR_{Etel}, and W_n for testing $H_0 : \beta_3 = 0$ against $H_1 : \beta_3 \neq 0$

		$\chi^2(3) - 3$			$N(0, 1)$		
		$n = 40$	$n = 60$	$n = 80$	$n = 40$	$n = 60$	$n = 80$
$\beta_3 = 0.0$	LR_{Etel}	0.078	0.066	0.059	0.082	0.070	0.058
	LR_{Aetel}	0.066	0.054	0.055	0.068	0.064	0.058
	W_n	0.062	0.064	0.068	0.078	0.064	0.054
$\beta_3 = 0.05$	LR_{Etel}	0.112	0.118	0.118	0.186	0.280	0.300
	LR_{Aetel}	0.088	0.104	0.102	0.156	0.254	0.278
	W_n	0.094	0.102	0.100	0.164	0.244	0.264
$\beta_3 = 0.10$	LR_{Etel}	0.198	0.182	0.286	0.548	0.866	0.804
	LR_{Aetel}	0.176	0.160	0.268	0.506	0.848	0.792
	W_n	0.168	0.164	0.270	0.474	0.786	0.728
$\beta_3 = 0.15$	LR_{Etel}	0.280	0.340	0.394	0.986	0.930	0.978
	LR_{Aetel}	0.250	0.316	0.374	0.986	0.920	0.974
	W_n	0.268	0.324	0.356	0.978	0.892	0.970
$\beta_3 = 0.20$	LR_{Etel}	0.560	0.520	0.720	0.996	1	1
	LR_{Aetel}	0.532	0.488	0.702	0.990	1	1
	W_n	0.512	0.482	0.682	0.982	0.998	1

Estimates of rejection rates were reported for $N(0, 1)$ and $\chi^2(3) - 3$ distributed data at 3 different sample sizes ($n = 40, 60, 80$) at significance level $\alpha = 5\%$. For each case, 1,000 simulated data sets were used.

$x_{i0} \sim N(0, \sigma_\varepsilon^2 / (1 - \beta_3^2))$. We refer the reader to Lai and Small (2007) for more details. Note that x_{it} is a type II covariate. We used our AETEL method with the following estimating equations: (a) the type II estimating equations according to (2.5), labeled type II; (b) the type III estimating equations according to (2.8), labeled type III; (c) GEE using the independent working correlation, labeled GEE independence; (d) GEE using the exchangeable working correlation, labeled GEE exchangeable; (e) GEE using the autoregressive AR-1 working correlation, labeled GEE AR-1. We compared the bias, root-mean-square error, and the efficiency of each case for the parameter β_1 to the GEE independence case (the efficiency is the ratio of the mean-square error of the GEE independence case to that of the case).

As we can see from Table 2, GEE independence and GEE AR-1 are biased, because they use some invalid estimating equations. The other three are all unbiased, with type II being more efficient than the other two. Combining all available valid estimating equations does improve efficiency. With the same type II estimating equations, our method has slightly less RMSE (0.0401 vs. 0.0407) than Lai and Small (2007)'s method.

TABLE 2
Results of AETEL with various estimating equations for a type II time-dependent covariate

Estimating equations	Bias	RMSE	Efficiency
Type II	0.00	0.040	1.82
Type III	0.00	0.053	1.04
GEE independence	0.00	0.054	1.00
GEE exchangeable	−0.12	0.104	–
GEE AR-1	−0.79	0.661	–

3.3. *Study III*: *Spatial data.* We simulated data at all $m = 900$ pixels on a 30×30 phantom image (Figure 2). At a given voxel d,

$$(3.2) \quad \mathbf{y}_{ij}(d) = \beta_0(d) + \beta_1(d)t_{ij} + \beta_2(d)x_i + \beta_3(d)t_{ij}x_i + b_i(d) + \varepsilon_{ij}(d)$$

for $i = 1, \ldots, n$ and $j = 1, \ldots, m_i$, where t_{ij} is the time taking values in $(1, 2, 3, 4, 5)$, x_i was independently generated from a $N(0, 1)$, and $b_i(d)$ was independently generated from a $N(0, 1)$. Errors $\varepsilon_{ij}(d)$ were independently generated from $N(0, 1)$ and $\chi^2(3) - 3$, respectively. We tested the hypotheses $H_0 : \beta_3(d) = 0$ and $H_1 : \beta_3(d) \neq 0$ across all pixels. To assess the Type I and II error rates at the pixel level, we set $\beta_0(d) = \beta_1(d) = \beta_2(d) = 0$ across all pixels d and varied $\beta_3(d)$ as 0.0, 0.05, 0.10, 0.15, and 0.20. Specifically, we created two regions of interest (ROI) by setting $\beta_3(d)$ as 0.05, 0.10, 0.15, and 0.20, and setting $\beta_3(d) = 0$ outside of the two ROIs in order to assess the finite sample performance of our method at different signal-to-noise ratios (SNRs). We considered $n = 40$ and 80.

We used generalized estimation equations with an exchangeable working correlation matrix to calculate $\hat{\theta}(d)$ and $LR_{\text{Aetel}}(d)$ in Stage 1. In Stage 2 we used the four first-order neighbors of pixel d to form $N(d)$ and then calculated $LR_{\text{Tetel}}(d)$. As a comparison with the conventional analysis on image data, we first smoothed
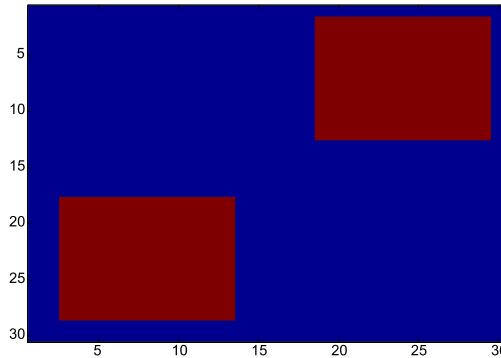

FIG. 2. *Two red regions of interest (ROIs) on a 30×30 image. The ROIs are indicated by the red area.*

TABLE 3
*Comparison of the two stages of TETEL for unsmoothed spatial data and the Wald test statistic for smoothed spatial data: true average rejection rates for voxels inside the ROI and false average rejection rates for voxels outside of the ROI were reported for $N(0, 1)$ and $\chi^2(3) - 3$ distributed data, and 2 different sample sizes ($n = 40$ and 80) at $\alpha = 5\%$. For each case,
100 simulated data sets were used*

| | | LR_{Tetal} | | | | Wald | | | |
| | | $n = 40$ | | $n = 80$ | | $n = 40$ | | $n = 80$ | |
β_3	Stage	True	False	True	False	True	False	True	False
					$N(0, 1)$				
0.05	Stage 1	0.223	0.088	0.329	0.068	0.711	0.101	0.891	0.105
	Stage 2	0.302	0.089	0.426	0.069				
0.10	Stage 1	0.571	0.087	0.820	0.069	0.964	0.15	0.991	0.158
	Stage 2	0.690	0.088	0.910	0.070				
0.15	Stage 1	0.863	0.089	0.984	0.069	0.996	0.184	0.998	0.177
	Stage 2	0.954	0.090	0.998	0.069				
0.20	Stage 1	0.987	0.089	0.999	0.069	0.999	0.193	0.999	0.192
	Stage 2	0.992	0.090	1.000	0.069				
					$\chi^2(3) - 3$				
0.05	Stage 1	0.117	0.085	0.122	0.070	0.313	0.089	0.331	0.073
	Stage 2	0.212	0.090	0.232	0.070				
0.10	Stage 1	0.193	0.087	0.259	0.069	0.567	0.099	0.858	0.099
	Stage 2	0.278	0.089	0.411	0.070				
0.15	Stage 1	0.313	0.090	0.447	0.068	0.847	0.113	0.948	0.123
	Stage 2	0.486	0.091	0.649	0.070				
0.20	Stage 1	0.463	0.090	0.660	0.069	0.947	0.130	0.979	0.145
	Stage 2	0.653	0.090	0.859	0.069				

image data by using the heat kernel smoothing method with 16 iterations, which gave an effective smoothness of about 4 pixels [Chung, Dalton and Davidson (2007)], and then calculated the Wald statistic based on GEE with an exchangeable working correlation matrix at each pixel. The 100 replications were used to approximate rejection rate with significance level $\alpha = 5\%$.

As shown in Table 3, the Type I rejection rates outside of ROIs for both LR_{Aetel} and LR_{Tetel} are relatively accurate for all cases, while the statistical power for rejecting the null hypothesis in ROIs significantly increases with the absolute value of $\beta_3(d)$. Compared with LR_{Aetel}, LR_{Tetel} has higher statistical power for rejecting the null hypothesis in ROIs with $\beta_3(d) \neq 0$. In contrast, compared with LR_{Aetel} and LR_{Tetel} based on the unsmoothed imaging data, although the Wald statistic for the smoothed imaging data has higher statistical power for rejecting the null hypothesis in ROIs, its Type I error rate is inflated and increases with the absolute value of $\beta_3(d)$. The decline in the type I and II error rates is caused by the fact that the variance of $\chi^2(3) - 3$ is larger than that of $N(0, 1)$. We also tried different

degrees of smoothness and ROIs with different sizes and found that the degree of smoothness and ROI size can have profound effect on the Type I and II error rates of the Wald statistic (not presented here).

4. Hippocampus shape.

4.1. *Hippocampus SPHARM-PDM representation.* Let $\mathbf{y}_{ij}(d)$ be the 3×1 coordinate vector at voxel d on the left and right hippocampus SPHARM-PDMs and $\mathbf{x}_{ij} = (1, \text{gender}_i, \text{age}_i, \text{SC1}_i, \text{SC2}_i, race1_i, race2_i, time_{ij})^T$, where SC1 and SC2 were, respectively, dummy variables for haloperidol-treated SC patient and olanzapine-treated SC patient versus healthy controls, and $race1$ and $race2$ were, respectively, dummy variables for Caucasian and African American versus other race. Let $\mathbf{y}_i(d) = (\mathbf{y}_{i1}(d)^T, \dots, \mathbf{y}_{im_i}(d)^T)^T$ and $A \otimes B$ denote the Kronecker product of matrices A and B. We assume that the mean and covariance matrix of $\mathbf{y}_i(d)$ are, respectively, given by

$$E(\mathbf{y}_i(d)) = \begin{pmatrix} \mathbf{x}_{i1}^T \otimes \mathbf{I}_3 \\ \cdots \\ \mathbf{x}_{im_i}^T \otimes \mathbf{I}_3 \end{pmatrix} \beta(d) \quad \text{and} \quad \text{Cov}(\mathbf{y}_{ij}(d)) = V_i(d) = R_i(\alpha(d)) \otimes \Sigma(d),$$

where $\beta(d)$ is a 24×1 vector, $R_i(\alpha(d)) = (\alpha(d)^{|j-k|})$ is the standard autoregressive AR-1 correlation matrix and $\Sigma(d)$ is a 3×3 covariance matrix of $\mathbf{y}_{ij}(d)$. We estimated $\alpha(d)$ and $\Sigma(d)$ by using Pearson residuals, which were calculated by solving GEEs with an independent working correlation matrix. For now on, $V_i(d)$ [or $\alpha(d)$ and $\Sigma(d)$] are assumed to be known. For the data analysis, we used the moment model based on GEE in (2.2) since there is no time-dependent covariate except time itself. The $g(\mathbf{z}_i(d), \theta(d); d)$ which is used in TETEL is given by

$$g(\mathbf{z}_i(d), \theta(d); d) = \sum_{i=1}^{n} \begin{pmatrix} \mathbf{x}_{i1}^T \otimes \mathbf{I}_3 \\ \cdots \\ \mathbf{x}_{im_i}^T \otimes \mathbf{I}_3 \end{pmatrix}^T V_i(d)^{-1} \left[\mathbf{y}_i(d) - \begin{pmatrix} \mathbf{x}_{i1}^T \otimes \mathbf{I}_3 \\ \cdots \\ \mathbf{x}_{im_i}^T \otimes \mathbf{I}_3 \end{pmatrix} \beta(d) \right].$$

Existing statistical methods of image data in SPM require that the error distribution is Gaussian and the variance is constant. The Shapiro–Wilk test rejects the normality assumption at many voxels of both the left and right hippocampus structures, and, thus, our nonparametric TETEL method is preferred for the analysis of this data set.

Since our goal is to detect the difference in the SPHARM-PDM surface shape between the schizophrenia and control groups, we used LR_{Aetel} and LR_{Tetel} to carry out the test. Moreover, in Stage 2, we used the closest neighbors of each voxel d to form $N(d)$. The color-coded p-values of the LR_{Aetel} and LR_{Tetel} and their corrected p-values using FDR across the voxels of both the left and right reference hippocampi are shown in Figure 3 [Benjamini and Yekutieli (2001)], in which the top row is for the first stage (LR_{Aetel}) and the bottom row is for the second stage (LR_{Tetel}).
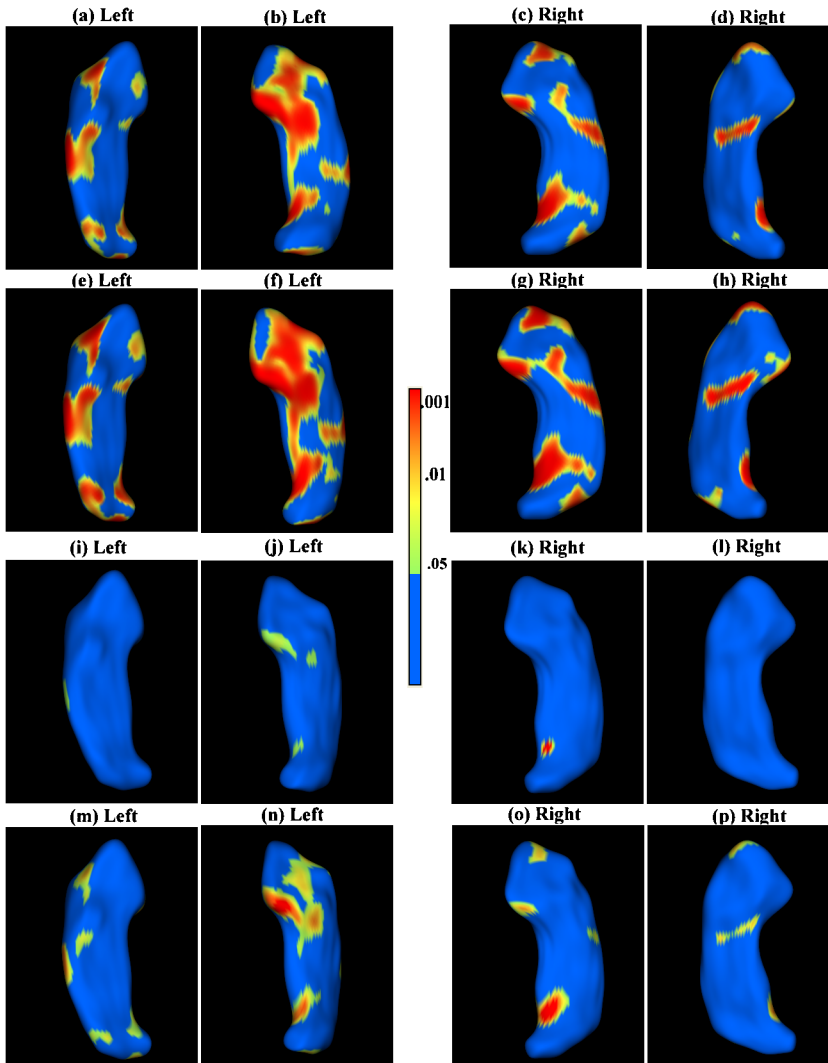
FIG. 3.   *Results from the longitudinal schizophrenia study. The first and third rows are for the first stage* (LR_{Aetel}): *the color-coded raw p-value maps of group effect for the left hippocampus* (a, b) *and the right hippocampus* (c, d), *and the corresponding color-coded corrected p-value maps of group effect for the left hippocampus* (i, j) *and the right hippocampus* (k, l). *The second and fourth rows are for the second stage* (LR_{Tetel}): *the color-coded p-value maps of group effect for the left hippocampus* (e, f) *and the right hippocampus* (g, h), *and the corresponding color-coded corrected p-value maps of group effect for the left hippocampus* (m, n) *and the right hippocampus* (o, p).

The analyses show strong shape differences in the superior, anterior parts of the left hippocampus, at the intersection of cornu ammonis 1 and cornu ammonis 2, previously not shown. Posterior shape changes at the hippocampal tail shown

in chronic schizophrenics [Styner et al. (2004)] are detected here already in first episode patients. Furthermore, the results also confirm those reported in Narr et al. (2004) by indicating a strong medial shape difference in the central, left hippocampal body in first episode patients. Comparing the first and second rows, it is clear that TETEL shows advantages in detecting more significant and smoother activation areas.

4.2. *Hippocampus m-rep thickness.* First, we considered the baseline analysis. We used the moment model based on the estimating equations $\mathbf{x}_{i1}(\mathbf{y}_{i1} - \mathbf{x}_{i1}^T \beta)$, where $\mathbf{y}_{i1}$ is the m-rep thickness measured at baseline for the ith subject at each medial atom of the left and right hippocampi; $\mathbf{x}_{i1}$ is an 8×1 vector given by $\mathbf{x}_{i1} = (1, \text{gender}_i, \text{age}_i, \text{SC1}_i, \text{SC2}_i, race1_i, race2_i, WBV_{i1})^T$ and $\beta = (\beta_0, \beta_1, \ldots, \beta_7)^T$. Existing statistical methods of image data in SPM require that the error distribution is Gaussian and the variance is constant. The Shapiro–Wilk normality test was applied to check this parametric assumption of the general linear model at each atom for the left hippocampus and right hippocampus using the residuals. Figure 4(c) and (e) show that the Shapiro–Wilk test rejects the normality assumption at many atoms of both the left and right hippocampus structures, therefore, our nonparametric AETEL method is preferred for the analysis of this data set.

Here our goal is to detect differences in thickness of the hippocampus across the three groups. Hence, we set the null hypotheses $H_0: \beta_3 = \beta_4 = 0$ at all 24 atoms for both the left and right hippocampi. Accordingly, we have

$$R = \begin{pmatrix} 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \end{pmatrix}$$

and $\mathbf{b}_0 = (0, 0)^T$. We used LR_{Aetel} to carry out the test. The color-coded p-values of the LR_{Aetel} across the atoms of both the left and right reference hippocampi are shown in Figure 5(a) and (b). The false discovery rate approach was used to correct for multiple comparisons, and the resulting adjusted p-values were shown in Figure 5(c) and (d). Before correcting for multiple comparisons, there was a significant group difference in m-rep thickness at the upper central atoms in the left hippocampus and some area in the right hippocampus. However, there is no significant group effect at any atom after correcting for multiple comparisons.

Second, we did a longitudinal data analysis. The advantage of a longitudinal study over a baseline study is that it allows us to determine (i) whether the change patterns of the response are similar or not across the three groups; (ii) whether, on average over time, there is a difference in the response across the three groups. We considered the moment model with $\mathbf{x}_{ij} = (1, \text{gender}_i, \text{age}_i, \text{SC1}_i, \text{SC2}_i, race1_i, race2_i, WBV_{ij}, time_{ij}, \text{SC1}_i * time_{ij}, \text{SC2}_i * time_{ij})^T$.

Since the WBV is a time-dependent covariate, we need to verify its appropriate type. Moreover, from a neuroscience point of view, the m-rep thickness at each atom serves as a local volumetric measure and covaries with WBV. We started with
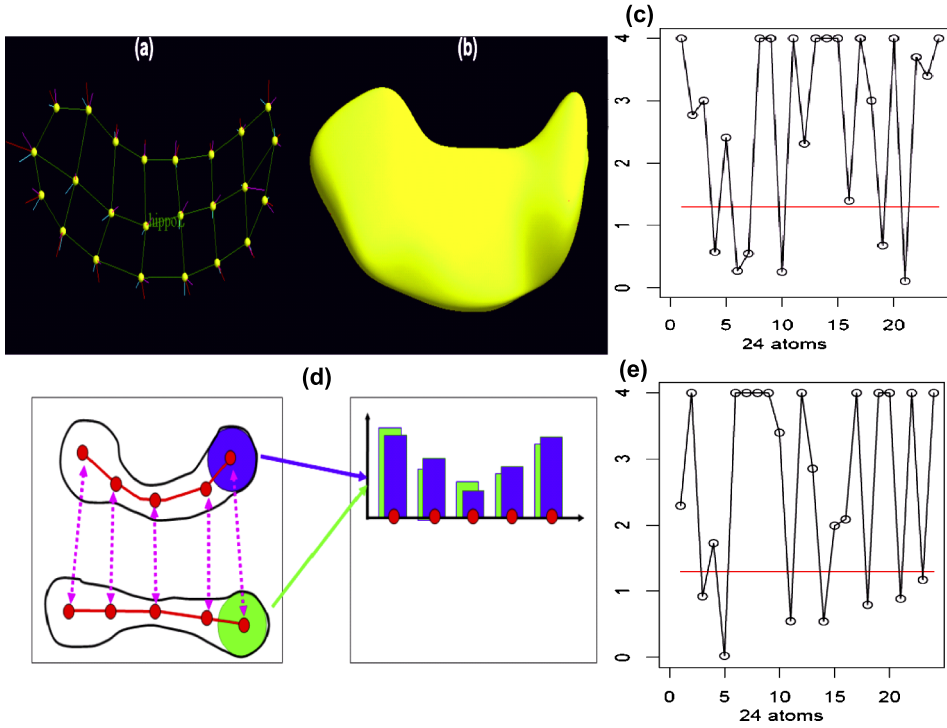
FIG. 4.   *An m-rep model of a hippocampus*: (a) *an m-rep model of the hippocampus*; (b) *the boundary surface of the m-rep model of hippocampus*; (d) *m-rep radius (or thickness) measures at the five atoms from two m-rep objects*; (c) *shows the* $-\log_{10}(p)$-*values for the Shapiro–Wilk test for the residuals at each atom on the left hippocampus*; (e) *shows the* $-\log_{10}(p)$-*values for the Shapiro–Wilk test for the residuals at each atom on the right hippocampus. The red horizontal line is the* 0.05 *cutoff line.*

type III and used GEE in (2.2) with $V_i = I_i$. Then we used the type II equations specified in (2.5) and tested whether WBV is type II against type III. The LR_{GF} did not reject for almost all 24 atoms, suggesting WBV is a type II covariate for most atoms. Furthermore, we used the type I equations specified in (2.3) and tested whether WBV is type I against type II. The LR_{GF} rejected that WBV was of type I for most atoms (Figure 5). This indicates the invalidity of some type I equations. We used goodness-of-fit statistics in Zhu et al. (2008b) to test whether some of the extra equations added for type I, such as

$$E\{\partial_{\beta_l}\mu_{is}(\beta)[\mathbf{y}_{ij} - \mu_{ij}(\beta)]\} = 0 \qquad \text{for all } s < j, j = 1, \ldots, m_i,$$

were not valid. For instance, for the 3rd atom on the left hippocampus, the p-value of the goodness-of-fit test for the newly added equation $E\{\partial_{\beta_l}\mu_{i2}(\beta)[\mathbf{y}_{i3} - \mu_{i3}(\beta)]\} = 0$ is smaller than 0.001 [Figure 5(e)]; for the 14th atom on the right hippocampus, the p-value of the goodness-of-fit test for the newly added equation $E\{\partial_{\beta_l}\mu_{i2}(\beta)[\mathbf{y}_{i3} - \mu_{i3}(\beta)]\} = 0$ is smaller than 0.001 [Figure 5(f)]. Therefore, we
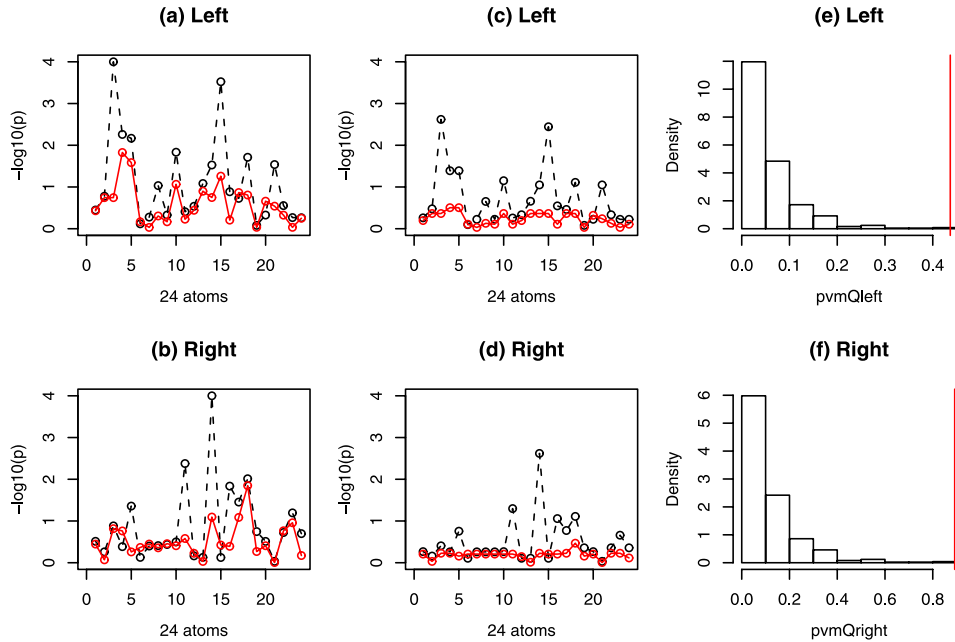
FIG. 5. *An m-rep model of a hippocampus*: *Maps of* $-\log_{10}(p)$-*values for testing WBV as a type I time-dependent covariate* (*black*) *and a type II time-dependent covariate* (*red*): (a) *uncorrected* $-\log_{10}(p)$-*values for left hippocampus*; (b) *uncorrected* $-\log_{10}(p)$-*values for right hippocampus*; (c) *corrected* $-\log_{10}(p)$-*values for left hippocampus*; (d) *corrected* $-\log_{10}(p)$-*values for right hippocampus*; (e) *the goodness-of-fit test for the equation* $E\{\partial_\beta \mu_{i2}(\beta)[\mathbf{y}_{i3} - \mu_{i3}(\beta)]\} = 0$ *for the 3rd atom on the left hippocampus*; (f) *the goodness-of-fit test for the equation* $E\{\partial_\beta \mu_{i2}(\beta)[\mathbf{y}_{i3} - \mu_{i3}(\beta)]\} = 0$ *for the 14th atom on the right hippocampus*.

treated WBV as a type II time-dependent covariate and used the corresponding estimating equations for the longitudinal data analysis.

To determine whether the changing patterns of the thickness of the hippocampus over time are similar or not across the three groups, we tested the null hypotheses $H_0 : \beta_9 = \beta_{10} = 0$ (β_9 and β_{10} are the coefficients of the interaction terms of group and time) at all 24 atoms for each of the left hippocampus and the right hippocampus, and it turned out that the interaction terms were not significant for most atoms. Next we deleted the interaction terms and tried to look at whether there are differences in the responses across the three groups on average over time with respect to the null hypotheses $H_0 : \beta_3 = \beta_4 = 0$ at all 24 atoms for each of the left hippocampus and the right hippocampus. Again we only found that there was a significant difference through time in m-rep thickness at the upper central atoms in the left hippocampus across schizophrenia patients and healthy controls groups after correcting for multiple comparisons, but the differences were not significant at other atoms, nor at any atoms on the right hippocampus. The color-coded p-values of the LR_{Aetel} across the atoms of both the left and right reference hippocampi are
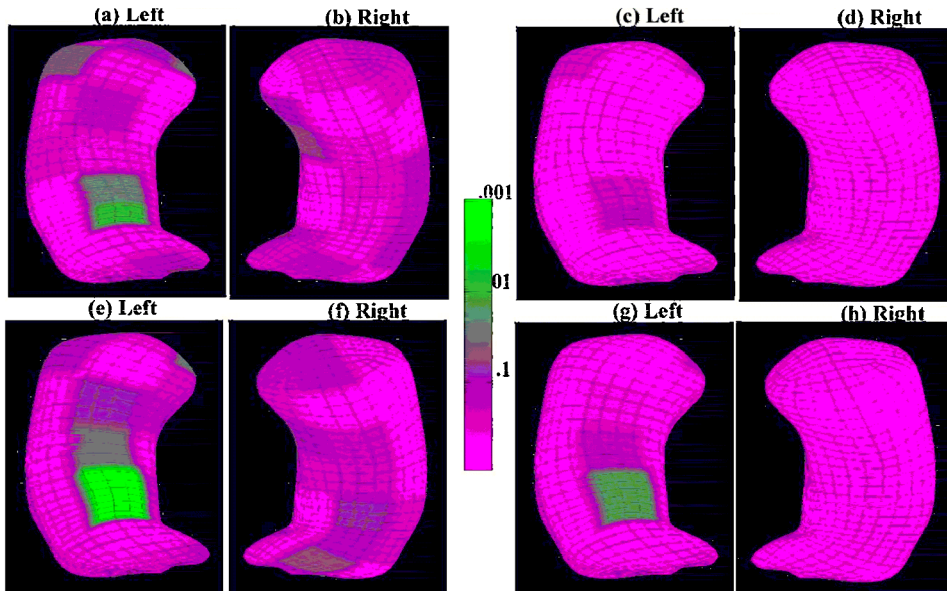
FIG. 6.   *Results from the longitudinal schizophrenia study. The top row is for the baseline analysis: the color-coded uncorrected p-value maps of group effect for* (a) *the left hippocampus and* (b) *the right hippocampus*; *the color-coded corrected p-value maps of group effect for* (c) *the left hippocampus and* (d) *the right hippocampus after correcting for multiple comparisons. The bottom row is for the longitudinal analysis: the color-coded uncorrected p-value maps of group effect for* (e) *the left hippocampus and* (f) *the right hippocampus*; *the color-coded corrected p-value maps of group effect for* (g) *the left hippocampus and* (h) *the right hippocampus after correcting for multiple comparisons.*

shown in Figure 5(e) and (f), and the corrected p-values are shown in Figure 6(g) and (h). Before correcting for multiple comparisons, there was a significant group difference in m-rep thickness at the upper central atoms in the left hippocampus, and the significance level is larger than that of the baseline analysis. Since the positive correlation is commonly observed in imaging data, we applied the false discovery rate (FDR) procedure in Benjamini and Yekutieli (2001) to correct for multiple comparisons. There is still a significant group effect at the upper central atoms in the left hippocampus.

We compared the results by making the assumption that WBV was a type II time-dependent and also a type III time-dependent covariate. Treating WBV as a type II time-dependent covariate lowered the p-values, making some nonsignificant p-values for the group effect significant. On the other hand, we found that all the standard deviations associated with the parameter estimates treating WBV as a type II time-dependent covariate were uniformly less than those treating WBV as a type III, which confirms that treating WBV as a type II gains efficiency by making use of more correct estimating equations. Table 4 compares the standard deviations of the parameter estimates between treating WBV as a type II time-

TABLE 4
Standard deviation comparison of the parameter estimates between treating WBV as a type II time-dependent covariate and a type III time-dependent covariate at atom 11 *of the left hippocampus*

	Intercept	Gender	Age	SC1	SC2	Race1	Race2	WBV	Time
Type III	0.367	0.078	0.007	0.062	0.058	0.097	0.102	0.237	0.022
Type II	0.344	0.075	0.005	0.058	0.054	0.094	0.100	0.221	0.018

dependent covariate and a type III time-dependent covariate at atom 11 of the left hippocampus.

The longitudinal analysis increased the significance level at those significant atoms for the group effect, compared to the baseline analysis. We were also able to observe the change difference across groups through time, although it is not much. Both the baseline analysis and longitudinal analysis suggest that there is an asymmetric aspect in that the left hippocampus shows larger regions of significance than the right one, and the significant positions of the group differences are around the lateral dentate gyrus and medial CA4 body regions for the left hippocampus.

5. Discussion. We have developed TETEL for spatial analysis of neuroimaging data from longitudinal studies. We have shown that AETEL allows us to efficiently analyze longitudinal data with different time-dependent covariate types. We have specifically combined all the data in the closest neighborhood of each voxel (or pixel) on a 3D volume (or 2D surface) with appropriate weights to calculate adaptive parameter estimates and adaptive test statistics. We have used simulation studies to examine the finite sample performance of AETEL and TETEL. In our longitudinal schizophrenia study, we have used the boundary and medial shape of the hippocampus to detect differences in morphological changes of the hippocampus across time between schizophrenic patients and healthy subjects. For the m-rep thickness, we have found that WBV is an important time-dependent covariate. Potential applications of our methodology include understanding normal and abnormal brain development, and identifying the neural bases of the pathophysiology and etiology of neurodegenerative and neuropsychiatric disorders.

Many issues still merit further research. One major issue is to develop a test procedure based perhaps on random field theory or resampling methods to correct for multiple comparisons in order to control the family-wise error rate under the moment model (2.1). Another major issue is to extend the test procedure to conduct cluster size inference and examine its performance in controlling the Type I error rate. The test procedure may lead to a simple cluster size test (cluster size test assesses significance for all sizes of the connected regions greater than a given

primary threshold). Models with nonparametric components using TETEL also may prove to be useful directions to consider.

SUPPLEMENTARY MATERIAL

Proofs of Theorems 2.1 and 2.2 (DOI: 10.1214/11-AOAS480SUPP; .pdf). We present assumptions and proofs of Theorems 2.1 and 2.2.

REFERENCES

ALMLI, C. R., RIVKIN, M. J., MCKINSTRY, R. C. and BRAIN DEVELOPMENT COOPERATIVE GROUP (2007). The NIH MRI study of normal brain development (objective-2): Newborns, infants, toddlers, and preschoolers. *NeuroImage* **35** 308–325.

ARNDT, S., COHEN, G., ALLIGER, R. J., SWAYZE, V. W. and ANDREASEN, N. C. (1991). Problems with ratio and proportion measures of imaged cerebral structures. *Psychiatry Res.* **40** 79–89.

ASHBURNER, J. and FRISTON, K. J. (2000). Voxel-based morphometry: The methods. *NeuroImage* **11** 805–821.

BECKMANN, C. F., JENKINSON, M. and SMITH, S. M. (2003). General multilevel linear modeling for group analysis in fMRI. *NeuroImage* **20** 1052–1063.

BENJAMINI, Y. and HOCHBERG, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *J. R. Stat. Soc. Ser. B* **57** 289–300. MR1325392

BENJAMINI, Y. and YEKUTIELI, D. (2001). The control of the false discovery rate in multiple testing under dependency. *Ann. Statist.* **29** 1165–1188. MR1869245

BOWMAN, F. D., CAFFO, B., BASSETT, S. S. and KILTS, C. (2008). A Bayesian hierarchical framework for spatial modeling of fMRI data. *NeuroImage* **39** 146–156.

CAO, J. and WORSLEY, K. J. (2001). Applications of random fields in human brain mapping. In *Spatial Statistics*: *Methodological Aspects and Applications* (M. Moore, ed.). *Lecture Notes in Statistics* **159** 170–182. Springer, New York.

CHEN, J., VARIYATH, A. M. and ABRAHAM, B. (2008). Adjusted empirical likelihood and its properties. *J. Comput. Graph. Statist.* **17** 426–443. MR2439967

CHUNG, M. K., DALTON, K. M. and DAVIDSON, R. J. (2007). Tensor-based cortical surface morphometry via weighted spherical harmonic representation. *IEEE Transactions on Medical Imaging* **26** 566–581.

DIGGLE, P. J., HEAGERTY, P. J., LIANG, K.-Y. and ZEGER, S. L. (2002). *Analysis of Longitudinal Data*, 2nd ed. *Oxford Statistical Science Series* **25**. Oxford Univ. Press, Oxford. MR2049007

DRYDEN, I. L. and MARDIA, K. V. (1998). *Statistical Shape Analysis*. Wiley, Chichester. MR1646114

DUVERNOY, H. (2005). *The Human Hippocampus*. Springer, New York.

FRISTON, K. J. (2007). *Statistical Parametric Mapping*: *The Analysis of Functional Brain Images*. Academic Press, London.

FRISTON, K. J., HOLMES, A. P., POLINE, J. B., PRICE, C. J. and FRITH, C. D. (1996). Detecting activations in PET and fMRI: Levels of inference and power. *NeuroImage* **4** 223–235.

FRISTON, K. J., STEPHAN, K. E., LUND, T. E., MORCOM, A. and KIEBEL, S. (2005). Mixed-effects and fMRI studies. *NeuroImage* **24** 244–252.

HANSEN, L. P. (1982). Large sample properties of generalized method of moments estimators. *Econometrica* **50** 1029–1054. MR0666123

HAYASAKA, S., PHAN, L. K., LIBERZON, I., WORSLEY, K. J. and NICHOLS, T. E. (2004). Nonstationary cluster-size inference with random field and permutation methods. *NeuroImage* **22** 676–687.

HECKE, W. V., SIJBERS, J., BACKER, S. D., POOT, D., PARIZEL, P. M. and LEEMANS, A. (2009). On the construction of a ground truth framework for evaluating voxel-based diffusion tensor MRI analysis methods. *NeuroImage* **46** 692–707.

HUETTEL, S. A., SONG, A. W. and MCCARTHY, G. (2004). *Functional Magnetic Resonance Imaging*. Sinauer, Sunderland, MA.

IMBENS, G. W., SPADY, R. H. and JOHNSON, P. (1998). Information-theoretic approaches to inference in moment condition models. *Econometrica* **66** 333–357. MR1612246

JONES, D. K., SYMMS, D. K., CERCIGNANI, M. and HOWARD, R. J. (2005). The effect of filter size on VBM analyses of DT-MRI data. *NeuroImage* **26** 546–554.

LAI, T. L. and SMALL, D. (2007). Marginal regression analysis of longitudinal data with time-dependent covariates: A generalized method-of-moments approach. *J. R. Stat. Soc. Ser. B Stat. Methodol*. **69** 79–99. MR2301501

LIANG, K. Y. and ZEGER, S. L. (1986). Longitudinal data analysis using generalized linear models. *Biometrika* **73** 13–22. MR0836430

LIEBERMAN, J. A., TOLLEFSON, G. D., CHARLES, C., ZIPURSKY, R., SHARMA, T., KAHN, R. S., KEEFE, R. S. E., GREEN, A. I., GUR, R. E., MCEVOY, J., PERKINS, D., HAMER, R. M., GU, H. and TOHEN, M. (2005). Antipsychotic drug effects on brain morphology in first-episode psychosis. *Archives of General Psychiatry* **62** 361–370.

LINDQUIST, M. A. and WAGER, T. D. (2008). Spatial smoothing in fMRI using prolate spheroidal wave functions. *Hum. Brain Mapp*. **29** 1276–1287.

LIU, Y. and CHEN, J. (2010). Adjusted empirical likelihood with high-order precision. *Ann. Statist*. **38** 1341–1362. MR2662345

LOGAN, B. R. and ROWE, D. B. (2004). An evalution of thresholding techniques in fMRI analysis. *NeuroImage* **22** 95–108.

LUO, W.-L. and NICHOLS, T. E. (2003). Diagnosis and exploration of massively univariate neuroimaging models. *NeuroImage* **19** 1014–1032.

LUO, H. and PUTHUSSERYPADY, S. (2005). A sparse Bayesian method for determination of flexible design matrix for fMRI data analysis. *IEEE Trans. Circuits Syst. I. Regul. Pap*. **52** 2699–2706.

NARR, K. L., THOMPSON, P. M., SZESZKO, P., ROBINSON, D., JANG, S., WOODS, R. P., KIM, S., HAYASHI, K. M., ASUNCTION, D., TOGA, A. W. and BILDER, R. M. (2004). Regional specificity of hippocampal volume reductions in first-episode schizophrenia. *NeuroImage* **21** 1563–1575.

NEWEY, W. K. and SMITH, R. J. (2004). Higher order properties of GMM and generalized empirical likelihood estimators. *Econometrica* **72** 219–255. MR2031017

OWEN, A. B. (2001). *Empirical Likelihood*. Chapman & Hall/CRC, New York.

PENNY, W., FLANDIN, G. and TRUJILLO-BARRETO, N. (2007). Bayesian comparison of spatially regularised general linear models. *Hum. Brain Mapp*. **28** 275–293.

PEPE, M. S. and ANDERSON, G. L. (1994). A cautionary note on inference for marginal regression models with longitudinal data and general correlated response data. *Comm. Statist. Simul. Comput*. **23** 939–951.

PIZER, S. M., FLETCHER, P. T., JOSHI, S., THALL, A., CHEN, J. Z., FRIDMAN, Y., FRITSCH, D. S., GASH, A. G., GLOTZER, J. M., JIROUTEK, M. R., LU, C., MULLER, K. E.,

TRACTON, G., YUSHKEVICH, P. and CHANEY, E. L. (2003). Deformable m-reps for 3D medical image segmentation. *Int. J. Comput. Vis.* **55** 85–106.

POLINE, J. and MAZOYER, B. (1994). Analysis of individual brain activation maps using hierarchical description and multiscale detection. *IEEE Transactions in Medical Imaging* **4** 702–710.

QIN, J. and LAWLESS, J. (1994). Empirical likelihood and general estimating equations. *Ann. Statist.* **22** 300–325. MR1272085

QU, A., LINDSAY, B. G. and LI, B. (2000). Improving generalised estimating equations using quadratic inference functions. *Biometrika* **87** 823–836. MR1813977

ROGERS, B. P., MORGAN, V. L., NEWTON, A. T. and GORE, J. C. (2007). Assessing functional connectivity in the human brain by fMRI. *Magn. Reson. Imaging* **25** 1347–1357.

ROWE, D. B. (2005). Parameter estimation in the magnitude-only and complex-valued fMRI data models. *NeuroImage* **25** 1124–1132.

SALMOND, C. H., ASHBURNER, J., VARGHA-KHADEM, F., CONNELLY, A., GADIAN, D. G. and FRISTON, K. J. (2002). Distributional assumptions in voxel-based morphometry. *NeuroImage* **17** 1027–1030.

SCHENNACH, S. M. (2007). Point estimation with exponentially tilted empirical likelihood. *Ann. Statist.* **35** 634–672. MR2336862

SHAFIE, K., SIGAL, B., SIEGMUND, D. and WORSLEY, K. J. (2003). Rotation space random fields with an application to fMRI data. *Ann. Statist.* **31** 1732–1771. MR2036389

SHI, X., IBRAHIM, J. G., LIEBERMAN, J., STYNER, M., LI, Y. and ZHU, H. (2011). Supplement to "Two-stage empirical likelihood for longitudinal neuroimaging data." DOI:10.1214/11-AOAS480SUPP.

SNOOK, L., PLEWES, C. and BEAULIEU, C. (2007). Voxel based versus region of interest analysis in diffusion tensor imaging of neurodevelopment. *NeuroImage* **34** 243–252.

STYNER, M. and GERIG, G. (2003). Automatic and robust computation of 3d medial models incorporating object variability. *Int. J. Comput. Vis.* **55** 107–122.

STYNER, M., LIEBERMAN, J. A., PANTAZIS, D. and GERIG, G. (2004). Boundary and medial shape analysis of the hippocampus in schizophrenia. *Med. Image Anal.* **8** 197–203.

STYNER, M., LIEBERMAN, J. A., MCCLURE, R. K., WEINBERGER, D. R., JONES, D. W. and GERIG, G. (2005). Morphometric analysis of lateral ventricles in schizophrenia and healthy controls regarding genetic and disease-specific factors. *Proc. Natl. Acad. Sci. USA* **102** 4872–4877.

THOMPSON, P. M., CANNON, T. D. and TOGA, A. W. (2002). Mapping genetic influences on human brain structure. *Annals of Medicine* **24** 523–536.

THOMPSON, P. M. and TOGA, A. W. (2002). A framework for computational anatomy. *Comput. Vis. Sci.* **5** 13–34.

WOOLRICH, M. W., BEHRENS, T. E. J., BECKMANN, C. F., JENKINSON, M. and SMITH, S. M. (2004). Multilevel linear modelling for fMRI group analysis using Bayesian inference. *NeuroImage* **21** 1732–1747.

WORSLEY, K. J., TAYLOR, J. E., TOMAIUOLO, F. and LERCH, J. (2004). Unified univariate and multivariate random field theory. *NeuroImage* **23** 189–195.

YUE, Y., LOH, J. M. and LINDQUIST, M. A. (2010). Adaptive spatial smoothing of fMRI images. *Stat. Interface* **3** 3–13. MR2609707

ZHU, H., LI, Y., TANG, N., BANSAL, R., HAO, X., WEISSMAN, M. M. and PETERSON, B. S. (2008a). Statistical modelling of brain morphological measures within family pedigrees. *Statist. Sinica* **18** 1569–1591. MR2469324

ZHU, H., IBRAHIM, J. G., TANG, N. and ZHANG, H. (2008b). Diagnostic measures for empirical likelihood of general estimating equations. *Biometrika* **95** 489–507. MR2521595

ZHU, H. T., ZHOU, H., CHEN, J., LI, Y., STYNER, M. and LIEBERMAN, J. (2009). Adjusted
    exponentially tilted likelihood with applications to brain morphology. *Biometrics* **65** 919–927.

X. SHI   J. LIEBERMAN
J. G. IBRAHIM   NEW YORK STATE PSYCHIATRIC INSTITUTE
M. STYNER   1051 RIVERSIDE DRIVE
H. ZHU   NEW YORK, NEW YORK 10032
DEPARTMENT OF BIOSTATISTICS AND   USA
    BIOMEDICAL RESEARCH IMAGING CENTER
UNIVERSITY OF NORTH CAROLINA AT CHAPEL HILL
CHAPEL HILL, NORTH CAROLINA 27599-7420
USA
E-MAIL: hzhu@bios.unc.edu

                        Y. LI
                        ST. JUDE CHILDREN'S RESEARCH HOSPITAL
                        262 DANNY THOMAS PLACE
                        MEMPHIS, TENNESSEE 38105-3678
                        USA