Multiscale Adaptive Marginal Analysis of Longitudinal Neuroimaging Data with Time-Varying Covariates

Martha Skup[1], **Hongtu Zhu**[2,3], and **Heping Zhang**[1] **for the Alzheimer's Disease Neuroimaging Initiative**

Martha Skup: martha.skup@yale.edu; Hongtu Zhu: hzhu@bios.unc.edu; Heping Zhang: heping.zhang@yale.edu

[1]Division of Biostatistics, Yale University, New Haven, Connecticut

[2]Department of Biostatistics, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina

[3]Biomedical Research Imaging Center, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina

Summary

Neuroimaging data collected at repeated occasions are gaining increasing attention in the neuroimaging community due to their potential in answering questions regarding brain development, aging, and neurodegeneration. These datasets are large and complicated, characterized by the intricate spatial dependence structure of each response image, multiple response images per subject, and covariates that may vary with time. We propose a multiscale adaptive generalized method of moments (MA-GMM) approach to estimate marginal regression models for imaging datasets that contain time-varying, spatially-related responses and some time-varying covariates. Our method categorizes covariates into types to determine the valid moment conditions to combine during estimation. Further, instead of assuming independence of voxels (the components that make up each subject's response image at each time point) as many current neuroimaging analysis techniques do, this method "adaptively smoothes" neuroimaging response data, computing parameter estimates by iteratively building spheres around each voxel and combining observations within the spheres with weights. MA-GMM's development adds to the few available modeling approaches intended for longitudinal imaging data analysis. Simulation studies and an analysis of a real longitudinal imaging dataset from the Alzheimer's Disease Neuroimaging Initiative are used to assess the performance of MA-GMM.

Keywords

Generalized method of moments (GMM); Longitudinal neuroimaging data; Marginal modeling; Multiscale adaptive regression model (MARM); Smoothing; Time-varying covariates; Voxel-wise method

1. Introduction

Longitudinal neuroimaging investigations, where subjects are scanned at several occasions and imaging data are obtained at more than one time point, are vital to furthering our

understanding of structural and functional development in healthy and pathological brains. The response data consist of 3-dimensional (3D) images of the brain and each image is made up of hundreds of thousands of cuboid elements, called "voxels," that each reflect a measure of the dependent variable (e.g., brain volume or neural activation in response to a stimulus). The images exhibit a complicated spatial correlation structure since voxels in close proximity within the image are often related. When obtained at repeated occasions, such as in a longitudinal investigation, these datasets are more complicated still, because collected response images and possibly some covariates vary with time. This work develops a spatial method to analyze datasets collected at repeated occasions, which are characterized by time-varying, spatially-correlated imaging responses and some time-varying covariates.

1.1 Spatial Modeling of Neuroimaging Data

Traditional analyses of neuroimaging data, most often carried out in a cross-sectional setting, are typically conducted in two steps. First, the raw images undergo a series of preprocessing procedures (e.g., registration of images to a common template) to prepare the data for analysis (Lazar, 2008). A main component of preprocessing is spatial smoothing, which increases images' signal-to-noise ratio and blurs out residual anatomical differences between individual brain images. An overview of smoothing techniques can be found in Yue et al. (2010), who highlight the common problematic theme of most smoothing methods: not only is the choice of smoothing extent made independently of the data *a priori*, but the extent is assumed equal across the image, often leading to increases in false positive and false negative rates due to under-smoothing in certain regions and over-smoothing in others. Second, independent statistical models are fit to the response measures at each voxel of the template and inference is conducted voxel-by-voxel ("voxel-wise") based on each voxel's estimated model. While the non-independence of voxels is typically dealt with through multiple comparison threshold adjustment (i.e., reduction of the -level to 0.001, random field theory, or false discovery rate), the limitation to the voxel-wise approach lies in its treatment of voxels as independent units in model fitting because it fails to explicitly model the spatial properties of the images (Benjamini and Hochberg, 1995; Worsley et al., 2004).

To overcome both the spatial smoothing and the voxel-wise analysis critiques, a limited number of spatial models have been proposed for neuroimaging data (e.g., Hartvig and Jensen (2000), Banerjee et al. (2004), Besag (1986), Bowman et al. (2008), Shi et al. (2011)). Development and implementation of these more sophisticated models has been slow in coming not only because deciphering the structure of spatial dependence in brain images can be difficult but also because spatial models are complex, which often translates to the nontrivial concern of high computation costs (Lazar, 2008).

Recently, Li et al. (2011) proposed a computationally efficient approach, called the multiscale adaptive regression method (MARM), to spatially and adaptively smooth cross-sectional neuroimaging data. By skipping the smoothing step during preprocessing, this technique avoids setting an arbitrary smoothing extent and does not force the same amount of smoothing throughout the whole image, thus improving upon conventional voxel-wise approaches. Instead, MARM incorporates spatial information in model estimation by relying on an iteratively increasing neighborhood concept and does so in a way that evades the difficult task of specifying the full spatial correlation structure of voxels. Another contribution of MARM is in its ability to calculate the standard deviation of images. The method is a novel generalization to multiple subjects of the power propagation-separation (PS) approach, which was originally developed for nonparametric estimation of regression curves or surfaces (Polzehl and Spokoiny, 2000). Tabelow and colleagues used the PS idea to develop a multiscale adaptive linear model to denoise imaging data from a single subject adaptively and spatially (Tabelow et al., 2006). However, the MARM method has not been applied to analysis of longitudinal imaging datasets that include time-varying covariates. In

the present work, we extend MARM to such a setting where covariates may be time-dependent by integrating the MARM approach for cross-sectional neuroimaging data with a marginal regression model intended for longitudinal data.

1.2 Longitudinal Modeling of Neuroimaging Data

Use of marginal models, where the target of inference is the population, has become widespread for analyzing longitudinal data in many fields. In neuroimaging studies, where scientific questions of interest generally concern the population mean rather than random subject effects, development of marginal models is very relevant. The most commonly-used method of fitting marginal models in non-high-dimensional response settings is the generalized estimating equation (GEE) approach (Liang and Zeger, 1986), which combines estimating equations with a "working" correlation structure to solve for estimates of the parameters. Often, an independence model is chosen (where the working correlation structure is a diagonal matrix of ones) because it has been shown to maintain a high efficiency of estimated parameters. However, when a dataset contains time-varying covariates, either using or not using the GEE independence model can lead to bad estimators (Pan and Connett, 2002), signifying that it may not be the optimal choice for marginal analysis. Fitzmaurice (1995) showed that use of the independence working correlation structure may result in an inefficient estimator in such situations, while Pepe and Anderson (1994) reported that specification of a working correlation structure other than an independent one may lead to highly biased estimates.

These criticisms have lead to the consideration of alternative marginal regression approaches for longitudinal data analysis such as the generalized method of moments (GMM) (Lai and Small, 2007). Our goal in the present work is to develop a marginal model for the analysis of longitudinal imaging data that (i) incorporates the spatial structure of the response images in a way that is adaptive and non-arbitrary and (ii) provides an alternative approach to GEE for the marginal analysis of longitudinal datasets with some time-varying covariates in an attempt to address the above criticisms. As such, we develop a multiscale adaptive GMM (MA-GMM) model intended for neuroimaging data collected at repeated occasions. Our model relies on classifying time-dependent covariates into types to establish which of the available estimating functions to use in the procedure and, in order to estimate parameters at each voxel, it iteratively makes use of an increasing amount of information contained at surrounding voxels.

Few other methods have been developed for modeling longitudinal neuroimaging data. Some examples include Zhang et al. (2009), Shinohara et al. (2011), and Zipunnikov et al. (2011a). Another recent contribution is the work of Zipunnikov et al. (2011b), which decomposes the observed variability of high-dimensional data observed over time (such as longitudinal images) into subject specific, longitudinal subject specific, and subject-visit specific components. To our knowledge, only one other method for longitudinal neuroimaging analysis has addressed the important issue of proper handling of time-dependent covariates by incorporating covariate classifications in the construction of an exponentially tilted empirical likelihood (Shi et al., 2011). However, although some effort was made in their suggested technique to utilize spatial information in estimation via a two-stage procedure, it is not multiscale meaning that only data in the closest neighborhood of each voxel is considered.

Section 2 introduces our motivating dataset, real brain volume data collected by the Alzheimer's Disease Neuroimaging Initiative (ADNI). In Section 3, we review the GMM approach of Lai and Small (2007) for analyzing longitudinal data with time-varying covariates and present our strategy of applying this method to neuroimaging data. In

Sections 4 and 5, we demonstrate our procedure's usefulness using simulated datasets as well as the ADNI data. Finally, we present concluding remarks in Section 6.

2. Experimental Data

To illustrate the usefulness of our proposed model, we applied MA-GMM to brain structural volume data collected longitudinally by the ADNI, a large scale multi-site study collecting clinical, imaging, and laboratory data at multiple time points from healthy controls, individuals with amnestic mild cognitive impairment, and subjects with Alzheimer's Disease. Additional information about ADNI is provided in Web Appendix A.

The sample in our investigation included 244 subjects: 155 healthy controls and 89 individuals with probable Alzheimer's Disease. Each subject contributed four structural magnetic resonance imaging (MRI) scans over a two year period to the dataset. Image preparation and preprocessing (detailed in Skup et al., 2011) yielded an unsmoothed response image in a common stereotaxic space for each subject at each time point: a regional volumetric map called a "RAVENS" map composed of $7,425$ ventricle voxels (Davatzikos, 1998). The RAVENS value at each voxel within a single subject's image at each time point is thought to be directly proportional to the volume of the brain structure in the subject's original scan prior to preprocessing procedures. Comparison of brain tissue atrophy or expansion over time, therefore, can be quantified by patterns of RAVENS value change. Based on previously-reported findings, within the ventricle map, we focused our inference on the local volume in the right and left lateral ventricle brain regions in Alzheimer's Disease patients versus healthy controls; the regions contained $1,367$ and $1,255$ voxels respectively. A wide-range of collected covariates, some time-varying, were also available for analysis.

3. Model

3.1 Notation and Dataset Setup

Our motivation is to appropriately model a longitudinal imaging dataset, which consists of image response data and a set of covariates collected from n subjects at repeated occasions such that some covariates are time-varying. For simplicity of notation, we assume that each subject is observed at each time point and describe in Web Appendix C how to adapt our procedure when some subjects are not observed at all time points. We define $\mathbf{X}_{ij}$ to be a $p \times 1$ covariate vector for the i-th subject at the j-th occasion, where $i = 1, \ldots, n$ and $j = 1, \ldots, J$. Here, J denotes the number times that the i-th subject participates in. Each session produced one 3D image per subject. We let $\mathbf{X}_i = (\mathbf{X}_{i1}, \ldots, \mathbf{X}_{iJ})^T$.

Preprocessing of the collected image data yields an unsmoothed 3D response image of each subject's neuroimaging measures at each time point. Let $\boldsymbol{D}$ represent a subject's 3D volume of responses and d denote a voxel in $\boldsymbol{D}$ such that $N(\boldsymbol{D})$ is the total number of voxels in $\boldsymbol{D}$. A single image from a subject is represented by the $a \times b \times c$ 3D rectangular lattice consisting of voxels indexed by mapping of the image's coordinates (a, b, c) to $d = 1, \ldots, N(\boldsymbol{D})$. These measures (the response variables) are denoted by $\mathbf{Y}_i = \{ y_{ij}(d) : d \in \boldsymbol{D}, j = 1, \ldots, J \}$.

Our focus is to model the longitudinal relationship of the response measurements $\mathbf{Y}_i$ at voxel d to the p covariates, some of which are time-varying, integrating spatial information from voxels surrounding d into estimation at d. We will rely on the MARM methodology of Li et al. (2011) to analyze imaging responses and covariates collected longitudinally by combining the GMM approach of Hansen (1982) and Lai and Small (2007) with MARM.

3.2 Voxel-wise Generalized Method of Moments

The two-step GMM procedure, pioneered by Hansen (1982), is one of the most widely-used statistical tools in the field of econometrics. Lai and Small (2007) implemented this method to marginal regression of longitudinal data with time-varying covariates, showing that it may improve the efficiency of estimated parameters. Discussion of GMM and its uses can be found in Mátyás (1999) and its theoretical treatment can be found in Hansen (1982).

For simplicity, in this section, we temporarily drop voxel d from our notation. We consider the standard form of a marginal model for longitudinal data, where the conditional mean of each response $E(y_{ij}|\mathbf{X}_{ij}) = \mu_{ij}$ is assumed to depend on the covariates through some known link function $h(\mu_{ij}) = \mathbf{X}_{ij}$ (which we take to be the identity link) (McCullagh and Nelder, 1989). We let $\boldsymbol{\mu}_i = (\mu_{i1}, \ldots, \mu_{ij})^T$.

To use GMM for the estimation of requires the specification of a $q \times p$ vector of zero-mean continuous functions ("moment conditions" or "estimating functions") of , denoted $g(\mathbf{Y}_i, \mathbf{X}_i,)$, which satisfy the assumption $E_0\{g(\mathbf{Y}_i, \mathbf{X}_i,)\} = 0$. The sample version of the moment conditions is denoted $\mathbf{G}_n(\beta) = \sum_{i=1}^{n} g(\mathbf{Y}_i, \mathbf{X}_i, \beta)/n$. A standard method of moments would estimate by solving $\mathbf{G}_n( ) = 0$, but when the number of estimating functions is larger than the number of unknown parameters, e.g. $q > p$, the system of equations is over-identified and it is not possible to set the sample average of these moment functions exactly equal to zero and solve for . As one approach to deal with this issue, Hansen (1982) introduced a positive definite matrix $\mathbf{F}_n$ into the above formulation that minimizes the quadratic form in deviations between the sample moments and the population moment conditions. The GMM estimator is thus defined as

$$\widehat{\beta} = argmin_\beta \{\mathbf{G}_n(\beta)^T \mathbf{F}_n \mathbf{G}_n(\beta)\}.$$

Hansen (1982) further showed that that is efficient if the $\mathbf{F}_n$ matrix is the inverse of the covariance matrix of the moment conditions, denoted by $V^{-1} = \{\text{Cov}(g(\mathbf{Y}_i, \mathbf{X}_i, _0))\}^{-1}$ (where $_0$ is the true value of). The intuition here is that such an $\mathbf{F}_n$ gives less weight to those sample moment conditions with large variances. Thus, the GMM estimator is obtained in a two-step procedure. First, an initial consistent estimator is used to compute a consistent estimate of V^{-1}, denoted $\widehat{V}_n^{-1}$. The second step involves estimating by GMM with the matrix $\mathbf{F}_n$ set to be $\widehat{V}_n^{-1} = n\left\{\sum_{i=1}^{n} g(\mathbf{Y}_i, \mathbf{X}_i, \tilde{\beta}) \, g(\mathbf{Y}_i, \mathbf{X}_i, \tilde{\beta})^T\right\}^{-1}$.

To implement GMM in the analysis of a longitudinal dataset that contains time-varying covariates, Lai and Small (2007) suggest the following version of the two-step procedure (see Figure 1 for path diagram). Prior to estimation, their method requires the categorizing of covariates into types, which determines the valid estimating functions to include in $g(\mathbf{Y}_i, \mathbf{X}_i,)$. Definition of time-varying covariates types are discussed in Section 3.2.1.

Once the vector of valid moment functions is identified, they use GEE with an independent working correlation structure to obtain the initial consistent estimator of the parameters in the first step of the procedure, denote the resultant vector of parameter estimates as $_{\text{GEE}}$, and compute a consistent estimate of the variance of the moment conditions V^{-1} by

$\widehat{V}_n^{-1} = n\left\{\sum_{i=1}^{n} g(\mathbf{Y}_i, \mathbf{X}_i, \tilde{\beta}_{\text{GEE}}) \, g(\mathbf{Y}_i, \mathbf{X}_i, \tilde{\beta}_{\text{GEE}})^T\right\}^{-1}$. The second step of the GMM procedure solves

$$\left\{ \frac{\partial \mathbf{G}_n(\beta)}{\partial \beta} \right\}^T \widehat{V}_n^{-1} \mathbf{G}_n(\beta) = 0.$$

The asymptotic variance is estimated by

$$\left[\left\{ n^{-1} \sum_{i=1}^n \frac{\partial \boldsymbol{g}(\boldsymbol{Y}_i, \mathbf{X}_i, \beta)}{\partial \beta} \right\}^T \widehat{V}_n^{-1} \left\{ n^{-1} \sum_{i=1}^n \frac{\partial \boldsymbol{g}(\boldsymbol{Y}_i, \mathbf{X}_i, \beta)}{\partial \beta} \right\} \right]^{-1}.$$

3.2.1 Types of Covariates—Lai and Small (2007) distinguish between three types of time-dependent covariates, classifications which serve to determine which of the available moment conditions are valid for estimation. In the following definitions, x^k, $k = 1, \ldots, p$ denotes the k-th covariate (such that any covariate within $\mathbf{X}_i$ may be time-varying).

Type I Covariate: Time-varying covariate x^k is classified as being Type I if it satisfies:

$$E_{\beta_0} \left[\frac{\partial \mu_{it}(\beta_0)}{\partial \beta_k} \{ y_{ij} - \mu_{ij}(\beta_0) \} \right] = 0 \quad \text{for all } t, j, t = 1, \ldots, J, j = 1, \ldots, J. \quad (1)$$

When a covariate is of this type, there are J^2 valid moment conditions for that covariate.

Type II Covariate: Time-varying covariate x^k is classified as being Type II if it satisfies:

$$E_{\beta_0} \left[\frac{\partial \mu_{it}(\beta_0)}{\partial \beta_k} \{ y_{ij} - \mu_{ij}(\beta_0) \} \right] = 0 \text{ for all } t > j, j = 1, \ldots, J. \quad (2)$$

When a covariate is of this type, there are $J(J + 1)/2$ valid moment conditions corresponding to that covariate.

Type III Covariate: Time-varying covariate x^k is classified as being Type III if it satisfies:

$$E_{\beta_0} \left[\frac{\partial \mu_{it}(\beta_0)}{\partial \beta_k} \{ y_{ij} - \mu_{ij}(\beta_0) \} \right] \neq 0 \text{ for some } t > j. \quad (3)$$

When a covariate is of this type, there are J valid moment conditions for that covariate.

For results regarding consistently testing a covariate's type, see Lai and Small (2007).

3.3 Weighted Generalized Method of Moments

We propose a weighted GMM (wGMM) as a possible solution to incorporate into model estimation the spatial structure of imaging data collected longitudinally. To distinguish data and the parameters at different voxels, we reintroduce voxel d into our notation and consider the observed sample of data $\{y_{ij}(d), \mathbf{X}_{ij}: d \quad D, i = 1, \ldots, n, j = 1, \ldots, J\}$. For instance, (d) denotes the parameter vector at voxel d. This method (i) classifies time-varying covariates into types to determine which moment conditions are valid for estimation and (ii) then constructs and combines valid moment conditions for (d) based on data both at voxel d and in the neighborhood sphere of voxel d to make inference about (d).

Relying on the idea that voxels which are not at the boundaries of significant regions often have a neighborhood in which (d) is nearly constant, we define $\mathbf{B}(d, h)$ to be the set of all

voxels in the spherical neighborhood of voxel d with radius h and denote voxels within $\mathbf{B}(d, h)$ as d'. We carry out estimation of the parameters of interest at each voxel d by iterating the previously described GMM procedure along a sequence of spheres $\mathbf{B}(d, h)$ each of which integrates increasingly more spatial information into the procedure than the previous iteration via weights, as described in Li et al. (2011). Note that throughout this section, the weights, which we denote as $\omega(d, d'; h)$, characterize the similarity between voxels d and d' within each sphere $\mathbf{B}(d, h)$ and are formally defined in Web Appendix D. We let $\omega(d, d'; h) \in [0, 1]$ and $\sum_{d' \in \mathbf{B}(d, h)} \omega(d, d'; h) = 1$. If $\omega(d, d'; h) \approx 0$, this signifies that data at voxel d' does not contain much information on $\beta(d)$, whereas if $\omega(d, d'; h) \gg 0$, this indicates that data at voxels d and d' are similar to each other.

As before, we first classify the p covariates into types to choose a set of estimating equations $g(\mathbf{Y}_i(d), \mathbf{X}_i, \beta(d))$: either time-invarying or time-varying (I, II, III) according to the definitions in Section 3.2.1. With this, we form a new set of weighted estimating functions, denoted $g(\mathbf{Y}_i(d), \mathbf{X}_i, \beta(d); \omega(d, d'; h))$, using all data in the sphere of voxel d defined by radius h. Specifically,

$$g(\mathbf{Y}_i(d), \mathbf{X}_i, \beta(d); \omega(d, d'; h)) = \sum_{d' \in \mathbf{B}(d,h)} \omega(d, d'; h)\, g(\mathbf{Y}_i(d'), \mathbf{X}_i, \beta(d)).$$

The weighted sample version of this vector of continuous functions can be expressed as:

$$\mathbf{G}_n\,(\beta(d); \omega(d, d'; h)) = n^{-1} \sum_{i=1}^{n} g(\mathbf{Y}_i(d), \mathbf{X}_i, \beta(d); \omega(d, d'; h)). \quad (4)$$

At each iteration, we carry out the following two-step procedure. First, using data within spheres defined by radius h, we obtain an initial weighted estimator of the parameter vector $\beta(d)$ using weighted generalized estimating equations with an independent working correlation structure (wGEE-Ind) following the methodology of Li et al. (2011) and denote the resulting initial estimator as $\tilde{\beta}^{\omega}_{\text{GEE}}(d, h)$. Web Appendix B contains details regarding the functions necessary for estimation. This wGEE-Ind estimator allows us to set $\mathbf{F}^{\omega}_n(d, h) = \widehat{V}^{-1}_{n,\omega}(d, h)$, the inverse of the covariance matrix of the moment conditions, such that

$$\mathbf{F}^{\omega}_n(d, h) = n \left\{ \sum_{i=1}^{n} g\,(\mathbf{Y}_i(d), \mathbf{X}_i,\, \beta(d); \omega(d, d'; h))\, g\,(\mathbf{Y}_i(d), \mathbf{X}_i, \beta(d); \omega(d, d'; h))^T \right\}^{-1}$$

where $g(\mathbf{Y}_i(d), \mathbf{X}_i, \beta(d); \omega(d, d'; h))$ is evaluated at $\tilde{\beta}^{\omega}_{\text{GEE}}(d, h)$.

In the second step of our procedure, at each iteration, we perform estimation of the wGMM estimator, $\widehat{\beta}^{\omega}_{\text{GMM}}(d, h)$, by solving

$$\left\{ \frac{\partial \mathbf{G}_n(\beta(d), \omega(d, d'; h))}{\partial \beta(d)} \right\}^T \mathbf{F}^{\omega}_n(d, h)\, \mathbf{G}_n(\beta(d), \omega(d, d'; h)) = 0. \quad (5)$$

The asymptotic covariance matrix of $\widehat{\beta}^{\omega}_{\text{GMM}}(d, h)$ can be consistently estimated by

$$\sum(\widehat{\beta}^{\omega}_{\mathrm{GMM}}(d, h)) = \left[\left\{ n^{-1} \sum_{i=1}^{n} \frac{\partial g(Y_i(d), \mathbf{X}_i, \beta(d); \omega(d, d'; h))}{\partial \beta(d)} \right\}^{T} \mathbf{F}^{\omega}_n(d, h) \right.$$
$$\left. \times \left\{ n^{-1} \sum_{i=1}^{n} \frac{\partial g(Y_i(d), \mathbf{X}_i, \beta(d); \omega(d, d'; h))}{\partial \beta(d)} \right\} \right]^{-1} \quad (6)$$

where $g(\mathbf{Y}_i(d), \mathbf{X}_i, \beta(d); \omega(d, d'; h)) / \partial\beta(d)$ is evaluated at $\beta(d) = \widehat{\beta}^{\omega}_{\mathrm{GMM}}(d, h)$.

For an extension of our methodology to the case when not all subjects are observed at each time point, we refer to reader to Web Appendix C.

3.4 Multiscale Adaptive Estimation and Testing Procedure

Important questions to consider in the above described method are (i) how to determine the weights {$\omega(d, d'; h): d, d' \in \boldsymbol{D}$} used in the two-step wGMM procedure, (ii) how to iteratively carry out estimation and inference, and (iii) how to decide on covariate type for those time-varying covariates for which classification is unknown. The third of these questions is addressed in Section 3.5. To address the first two questions, we use the multiscale adaptive estimation and testing (MAET) procedure to adaptively compute weights and to iteratively update parameter estimates in S iterations at each voxel d by evolving the estimate of $\beta(d)$ along a sequence of spheres $\mathbf{B}(d, h_s)$ with increasing radii $h_1 = 0 < h_2 < \ldots < h_S$. See Web Figure 1 for an illustration depicting the increasing spherical neighborhood concept of this procedure. Combining the MAET procedure with the wGMM approach described in Section 3.3 gives us our proposed MA-GMM model.

MAET is conducted in five main stages which we outline in a flowchart by Figure 2. Important preset ingredients of the procedure are: the kernels $K_{loc}(u)$ and $K_{st}(u)$ and the parameters c_h, S_{sc}, S, C_n, and C. In Web Appendix D, we describe each of these stages in detail and include definitions for finding weights and parameters. Briefly, weights at $s = 1$ (the first iteration and thus the smallest radius $h_s = h_1 = 0$) are set to 1 such that $\omega(d, d'; h_1) = 1$. The two-step wGMM procedure described in Section 3.3 is carried out at each voxel d: (i) a consistent estimate of the parameter vector $\beta(d)$ is obtained via wGEE-Ind, denoted $\tilde{\beta}^{\omega}_{\mathrm{GEE}}(d, h_1)$, then used to find $\mathbf{F}^{\omega}_n(d, h_1)$ and (ii) the parameter vector $\beta(d)$ is estimated via wGMM and denoted $\widehat{\beta}^{\omega}_{\mathrm{GMM}}(d, h_1)$. Effectively, the estimates at the first iteration can be considered naive GMM estimates since $h_1 = 0$ and no neighborhood information is used in the estimation of parameters at d. Next, at iteration $s = 2$, based on the information contained in $\{\widehat{\beta}^{\omega}_{\mathrm{GMM}}(d, h_1): d \in \boldsymbol{D}\}$, new weights $\omega(d, d'; h_2)$ are computed for all the voxels within spheres $\mathbf{B}(d, h_2)$ defined by radius h_2. The parameters at each voxel d are updated with two-step wGMM using weighted information from voxels d' within the neighborhood of voxel d (i.e., the sphere defined by the radius h_2). The procedure continues from h_3 to h_S as we adaptively determine weights and update the parameters for each of the S iterations.

3.5 Determining the Covariate Type

Our overall approach to covariate classification when analyzing neuroimaging datasets that contain time-varying covariates is similar to that of Lai and Small (2007). When the type of a time-dependent covariate is unknown, we first assume it is Type III and, in each brain region of interest (ROI), we conduct a voxel-by-voxel test to determine whether that covariate is Type II. If the test is not rejected in the majority of the voxels in the region, we can go on to test if it is Type I. For computational ease and because we would not expect a covariate's classification to change at a voxel-wise level as more spatial information is incorporated into estimation, we carry out this procedure at the first iteration of MA-GMM

when the weights are equal to one. For more test details, we refer the reader to Lai and Small (2007).

As an example, suppose we test whether a covariate is of Type III verses that it is of Type II. We can set up our null and alternative hypotheses as follows:

$$
\begin{aligned}
H_0: g_{\mathbf{a}}(Y_i(d), \mathbf{X}_i, \beta(d); \omega(d, d'; h_1)) &= 0 \quad \text{and} \quad g_{\mathbf{b}}(Y_i(d), \mathbf{X}_i, \beta(d); \omega(d, d'; h_1)) = 0 \\
H_1: g_{\mathbf{a}}(Y_i(d), \mathbf{X}_i, \beta(d); \omega(d, d'; h_1)) &= 0 \quad \text{and} \quad g_{\mathbf{b}}(Y_i(d), \mathbf{X}_i, \beta(d); \omega(d, d'; h_1)) \neq 0
\end{aligned} \quad (7)
$$

where $g_{\mathbf{a}}$ denotes the moment functions valid for Type III covariates and $g_{\mathbf{b}}$ denotes the additional moment functions that are valid for Type II covariates that are not valid for Type III covariates. The corresponding statistic for such a test can be computed with

$$
\mathbf{Q}(d) = n \times \left\{ \mathbf{M}_{\mathbf{ab}}(\widehat{\beta}_{\mathrm{GMM}}^{\mathbf{ab}}(d)) - \mathbf{M}_{\mathbf{a}}(\widehat{\beta}_{\mathrm{GMM}}^{\mathbf{a}}(d)) \right\}, \quad (8)
$$

where $\mathbf{M}_{\mathbf{ab}}(\widehat{\beta}_{\mathrm{GMM}}^{\mathbf{ab}}(d))$ is the GMM minimand which uses the full set of moment conditions in the null hypothesis and $\mathbf{M}_{\mathbf{a}}(\widehat{\beta}_{\mathrm{GMM}}^{\mathbf{a}}(d))$ is the GMM minimand which uses the moment conditions $g_{\mathbf{a}}$. The statistic map of the $Q(d)$ statistics in a particular ROI is then thresholded by comparing the statistics to a χ^2 critical value with degrees of freedom equal to the dimension of $g_{\mathbf{b}}$, or the number of moment conditions that are valid for Type II covariates but not valid for Type III covariates. If the majority of voxels within an ROI reject the null hypothesis specified by (7), we classify the time-dependent covariate as Type III. If not, we then conduct another test to compare Type II against Type I. We make our decision globally instead of voxel-by-voxel and thus an assumption we hold throughout the application of MA-GMM to neuroimaging data is that the properties of the covariate in question (e.g., its dependence on the response or on time) do not differ depending on voxel location.

4. Simulations

To examine the finite sample performance of our proposed MA-GMM approach, we conducted Monte Carlo simulation studies at three settings, each of which contained a dataset with a different type of time-varying covariate. We briefly summarize the main findings below, but defer many of the details to Web Appendix F. For the three settings, we generated the covariate data as well as 64×64 phantom response images, each made up of $d = 4096$ pixels with five known ROIs of varying shapes and activation (to investigate performance at various signal-to-noise ratios), for $n = 60$ and $n = 100$ subjects. In each of the settings, we assessed the performance of MA-GMM estimators that varied in covariate type specification as well as two baseline estimators, a multiscale adaptive GEE estimator with an independent working correlation structure (MA-GEE-Ind) and, using a smoothing extent of four *mm*, a smoothed first iteration GEE estimator with an independent working correlation structure (Sm-GEE-Ind), across iterations and ROIs.

Results for our first and second simulation studies illustrated that when there is a Type I or Type II time-dependent covariate in a longitudinal imaging dataset, the MA-GMM estimator that correctly classifies the covariate type (using moment conditions specified by (1) or (2), respectively) can be substantially more efficient than GEE with an independent working correlation structure and in general performs better (less bias and greater precision) than MA-GMM estimators that do not use as many of the available valid estimating functions. Further, while estimators become slightly more biased in the two settings with the incorporation of more spatial information, their root mean square error and power for correctly rejecting Wald tests improve. Our third simulation showed that when there is a time-dependent covariate with a Type III categorization in a dataset, the MA-GMM

estimator that uses moment conditions (3) performs in a similar fashion to the MA-GEE-Ind estimator. In all three simulation studies, incorporating spatial information into estimation served to increase the precision of the estimators and increased the power to reject hypotheses that test the significant effects of interest.

5. Application to Alzheimer's Disease Neuroimaging Initiative Data

We used longitudinal structural MRI data and a subset of available covariates collected by the ADNI over the course of a two year period from Alzheimer's Disease (AD) patients and healthy controls (HC). Associations of interest, between brain structure and the diagnostic group by age interaction, were examined by fitting a MA-GMM model to the response measure (the RAVENS value) at each voxel of the ventricle template. Our goal was to show that our method could lend support for a robust finding shown in these subjects, namely, between-group differences in increased ventricular expansion over time. While previous methods have examined this question, none have considered time-varying covariate classification in modeling and few have utilized data from more than two time points.

To carry out our analysis, we included in the model the following variables as predictors: intracranial volume (ICV) at each scan, education, sex, age at each scan, and diagnostic status (AD or HC), as well as the two-way interaction between age and diagnostic status. In this set-up, the time-varying covariates included ICV, age, and the interaction between age and diagnostic status, while the time-independent covariates included education, sex, and diagnostic status. As done in previous research, the covariate of ICV was included to correct for individual subjects' varying brain sizes (e.g., Risacher et al. (2010)). By specifying ICV as a time-varying covariate, we were able to examine the longitudinal change of local volume relative to any change in ICV measurement over time. Further information regarding our classification of ICV as time-varying can be found in Web Appendix G.

The time-dependent covariate of age was clearly a Type I time-varying covariate (i.e., for age, $(x_{i1}, x_{i2}, x_{i3}, x_{i4})$ was a function of x_{ij} for each of the $j = 1, 2, 3, 4$ time points) and thus we treated the age by diagnosis interaction as a Type I time-dependent covariate as well. We investigated ICV's covariate classification using the approach described in Section 3.5 in two steps. First, at each voxel in both the right and left lateral ventricle ROIs, we tested the null hypothesis that ICV was a Type II covariate versus the alternative that it was a Type III covariate. To do so, we obtained two sets of unweighted GMM estimates of model parameters at each voxel using the estimating equations considered valid for the covariates in the two respective models under consideration (one model classified ICV as Type III and the second classified ICV as Type II; the latter utilized an additional 12 moment conditions in estimation). We computed minimands corresponding to both models and obtained a statistic map of $Q(d)$ statistics (see equation (8)), then thresholded the resulting statistic map according to a χ^2 critical value with 12 degrees of freedom. In both ROIs, the majority of hypothesis tests were not rejected, indicating that the Type II classification was more appropriate for this covariate than a Type III classification. Subsequently, we tested the null hypothesis that ICV was a Type I covariate against the alternative that it was a Type II. Again, in each of the two ROIs, we failed to reject the null at over 95% of the voxels. Thus, we treated ICV as a Type I covariate for the remainder of our analysis. Our initial assessment of our covariate categorization choice can be found in Web Appendix H.

After specifying covariate type, we applied the iterative MA-GMM procedure to each voxel within the unsmoothed RAVENS maps separately, as is typical in statistical brain mapping. We obtained estimates for the model's coefficients and their standard errors at each voxel at the last iteration of the procedure, from which we calculated multivariate Wald test statistics and associated p-values, showing the significance of covariate effects at each voxel. We

limited our focus to only those Wald statistic maps which tested for the presence of an age by diagnostic group interaction and generated *p*-value maps which corresponded to the hypothesis test that the parameter associated with the age by diagnosis interaction at that voxel was equal to zero. In a cluster-level inference approach (as done in Skup et al. (2011)), we thresholded the maps at *p*-value   0.05 with an extent threshold of   20 contiguous voxels. The primary threshold was set to a significance level of 0.05 in order find a compromise between sensitivity to cluster extent and separation of maximal voxel results. The anatomical localization of each cluster was established by comparing the coordinates of the cluster to a labeled template (Kabani et al., 1998).

Using a MA-GMM model and inference procedure where ICV was specified as a Type I time-varying covariate, we identified two clusters (a 100 voxel cluster and a 20 voxel cluster) in the right lateral ventricle ROI and one cluster (a 135 voxel cluster) in the left lateral ventricle ROI that resulted in a significant Wald test of interest. Post-hoc analyses at a representative peak voxel within the clusters that spanned the ROIs that showed evidence of an age by diagnosis interaction revealed that AD subjects demonstrated increased ventricle volume expansion compared to HC subjects, thereby confirming that our method could lend support for one of the robust findings reported by previous researchers (e.g., Weiner (2008)). Figure 3 depicts the uncorrected $-\log_{10}(p)$ maps corresponding to the two-way interaction in the two ROIs. Figure 4 shows the fitted brain volume average trajectories of AD and HC subjects at the voxel of peak difference.

To investigate the impact that covariate classification had on our final results, we conducted two additional analyses in an identical fashion, but treated ICV as a Type II and a Type III time-varying covariate, respectively. We present these results in Web Appendix I, which shows that MA-GMM with appropriately classified covariates can detect larger significant activation areas compared to methods that incorrectly classify or do not consider time-varying covariates.

6. Discussion

We propose a novel marginal regression method for the spatial analysis of longitudinal neuroimaging datasets with time-varying covariates. MA-GMM has two main features that address some challenges accompanying analysis of such data. First, our method distinguishes between covariate types to determine which moment conditions are valid for use in the procedure. Additionally, MA-GMM iteratively and adaptively builds spheres with increasing radii around voxels, efficiently incorporating information within the spheres into estimation. Using computational resources provided by Yale's High Performance Computing Center, we were able to complete our ADNI analysis of the ventricle data in less than two hours.

Our simulation studies and implementation of the method to ADNI data revealed that appropriate choice of time-varying covariate classification serves not only to lessen the bias of estimators but also increases their efficiency. Further, exploiting spatial relatedness of response images in our modeling procedure enhances the precision of estimators as more information from surrounding voxels is incorporated into the estimates at each voxel.

An interesting consideration for future research would be to investigate the impact that varying image resolution has on the robustness of conclusions obtained using MA-GMM. A question to ask is whether we can optimize the tuning of the parameters associated with the multiscale adaptive procedure to accommodate high resolution images. For instance, the parameter c_h may need to be tweeked so as to incorporate less neighboring voxels in the early iterations, compared to low resolution images. Additionally, it is likely that the

computation time needed to carry out the procedure would increase with higher resolution images.

Applications of longitudinal models are rather limited in the brain imaging literature and thus this work is an important contribution. To our knowledge, most existing neuroimaging software platforms have few if any suitable methods of analyzing neuroimaging data from longitudinal studies. Not only does MA-GMM provide an option to fit marginal regression models to neuroimaging data collected at numerous time points, but it takes special care to improve estimators when datasets are characterized by covariates that vary with time. Further, it overcomes the unsupported independence-of-voxels assumptions of standard cross-sectional neuroimaging techniques. As such, we capture some unique characteristics of longitudinal neuroimaging data in modeling in a way that is computationally tractable.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

References

Banerjee, S.; Carlin, B.; Gelfand, A. Hierarchical Modeling and Analysis for Spatial Data. Chapman: Hall/CRC; 2004.

Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. Journal of the Royal Statistical Society: Series B (Statistical Methodology). 1995; 57:289–300.

Besag J. On the statistical analysis of dirty pictures. Journal of the Royal Statistical Society: Series B (Statistical Methodology). 1986; 48:259–302.

Bowman F, Caffo B, Bassett S, Kilts C. A Bayesian hierarchical framework for spatial modeling of fMRI data. NeuroImage. 2008; 39:146–156. [PubMed: 17936016]

Davatzikos C. Mapping image data to stereotaxic spaces: applications to brain mapping. Human Brain Mapping. 1998; 6:334–338. [PubMed: 9788069]

Fitzmaurice G. A caveat concerning independence estimating equations with multivariate binary data. Biometrics. 1995; 51:309–317. [PubMed: 7766784]

Hansen L. Large sample properties of generalized method of moments estimators. Econometrica: Journal of the Econometric Society. 1982; 50:1029–1054.

Hartvig N, Jensen J. Spatial mixture modeling of fMRI data. Human Brain Mapping. 2000; 11:233–248. [PubMed: 11144753]

Kabani N, MacDonald D, Holmes C, Evans A. A 3D atlas of the human brain. NeuroImage. 1998; 7:S717.

Lai T, Small D. Marginal regression analysis of longitudinal data with time-dependent covariates: a generalized method-of-moments approach. Journal of the Royal Statistical Society: Series B (Statistical Methodology). 2007; 69:79–99.

Lazar, N. The Statistical Analysis of Functional MRI Data. New York: Springer; 2008.

Li Y, Zhu H, Shen D, Lin W, Gilmore J, Ibrahim J. Multiscale adaptive regression models for neuroimaging data. Journal of the Royal Statistical Society: Series B (Statistical Methodology). 2011; 73:559–578.

Liang K, Zeger S. Longitudinal data analysis using generalized linear models. Biometrika. 1986; 73:13–22.

Mátyás, L. Generalized Method of Moments Estimation. Cambridge: Cambridge University Press; 1999.

McCullagh, P.; Nelder, J. Generalized linear models. Chapman: Hall, CRC; 1989.

Pan W, Connett J. Selecting the working correlation structure in generalized estimating equations with application to the lung health study. Statistica Sinica. 2002; 12:475–490.

Pepe M, Anderson G. A cautionary note on inference for marginal regression models with longitudinal data and general correlated response data. Communications in Statistics-Simulation and Computation. 1994; 23:939–951.

Polzehl J, Spokoiny V. Adaptive weights smoothing with applications to image restoration. Journal of the Royal Statistical Society: Series B (Statistical Methodology). 2000; 62:335–354.

Risacher S, Shen L, West J, Kim S, McDonald B, Beckett L, Harvey D, Jack C Jr, Weiner M, Saykin A. Longitudinal mri atrophy biomarkers: relationship to conversion in the adni cohort. Neurobiology of aging. 2010; 31:1401–1418. [PubMed: 20620664]

Shi X, Ibrahim J, Lieberman J, Styner M, Li Y, Zhu H. Two-stage empirical likelihood for longitudinal neuroimaging data. Annals of Applied Statistics. 2011; 5:1132–1158. [PubMed: 21765885]

Shinohara R, Crainiceanu C, Caffo B, Reich D. Longitudinal analysis of spatiotemporal processes: A case study of dynamic contrast-enhanced magnetic resonance imaging in multiple sclerosis. Johns Hopkins University, Department of Biostatistics Working Paper Series. 2011b Working Paper 231.

Skup M, Zhu H, Wang Y, Giovanello K, Lin J, Shen D, Shi F, Gao W, Lin W, Fan Y, Zhang H. ADNI. Sex differences in grey matter atrophy patterns among AD and aMCI patients: results from ADNI. NeuroImage. 2011; 56:890–906. [PubMed: 21356315]

Tabelow K, Polzehl J, Voss H, Spokoiny V. Analyzing fMRI experiments with structural adaptive smoothing procedures. NeuroImage. 2006; 33:55–62. [PubMed: 16891126]

Weiner M. Expanding ventricles may detect preclinical Alzheimer disease. Neurology. 2008; 70:824–825. [PubMed: 18332339]

Worsley K, Taylor J, Tomaiuolo F, Lerch J. Unified univariate and multivariate random field theory. NeuroImage. 2004; 23:189–195.

Yue Y, Loh J, Lindquist M. Adaptive spatial smoothing of fMRI images. Statistics and Its Interface. 2010; 3:3–13.

Zhang X, Johnson T, Little R, Cao Y. Longitudinal image analysis of tumor/brain change in contrast uptake induced by radiation. The University of Michigan, Department of Biostatistics Working Paper Series. 2009 Working Paper 78.

Zipunnikov V, Caffo B, Yousem D, Davatzikos C, Schwartz B, Crainiceanu C. Functional principal components model for high-dimensional brain imaging. NeuroImage. 2011a; 58:772–784. [PubMed: 21798354]

Zipunnikov V, Greven S, Caffo B, Reich D, Crainiceanu C. Longitudinal high-dimensional data analysis. Johns Hopkins University, Department of Biostatistics Working Paper Series. 2011b Working Paper 234.
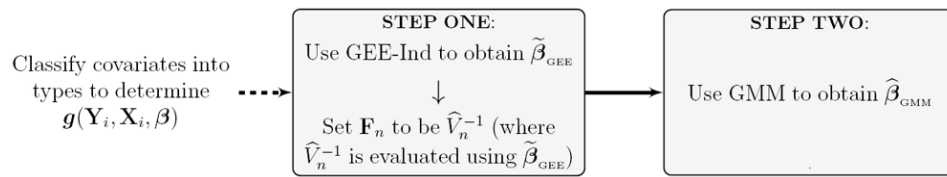
Figure 1.
Two-step GMM procedure for fitting a marginal model to longitudinal datasets that contain time-varying covariates, as proposed by Lai and Small (2007).
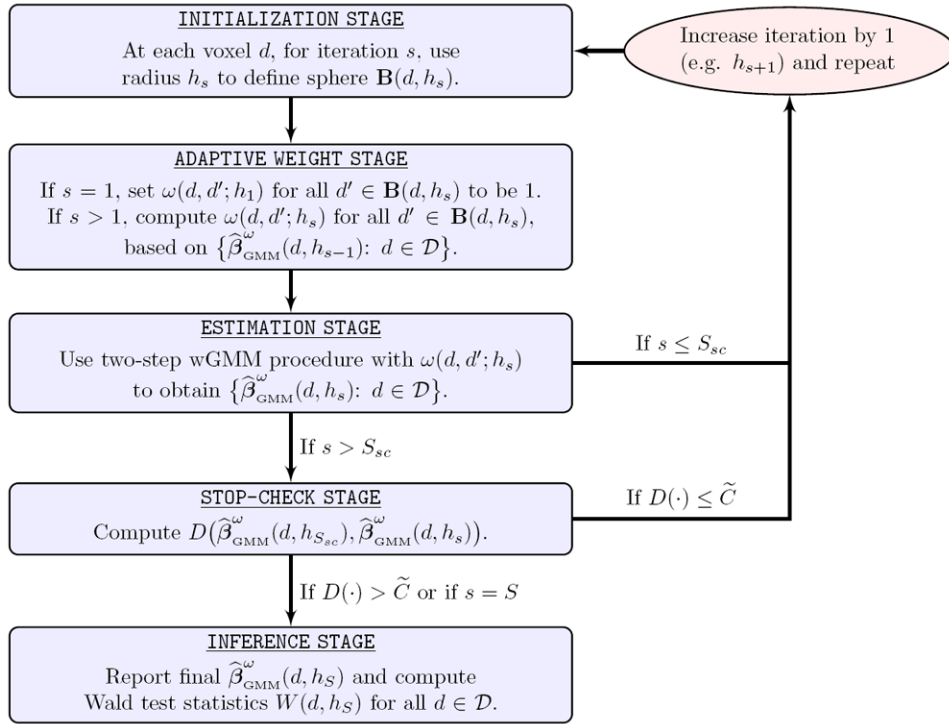
Figure 2.
A flowchart depicting the five stages of the multiscale adaptive GMM strategy. The procedure iterates stages 1 through 4 until termination of the algorithm is determined by the stop-check stage. Pre-specification of a sequence of radii $\{h_1 = 0, \ldots, h_s, \ldots, h_S\}$ as well as other predefined parameters such as C, S_{sc}, and S is required.
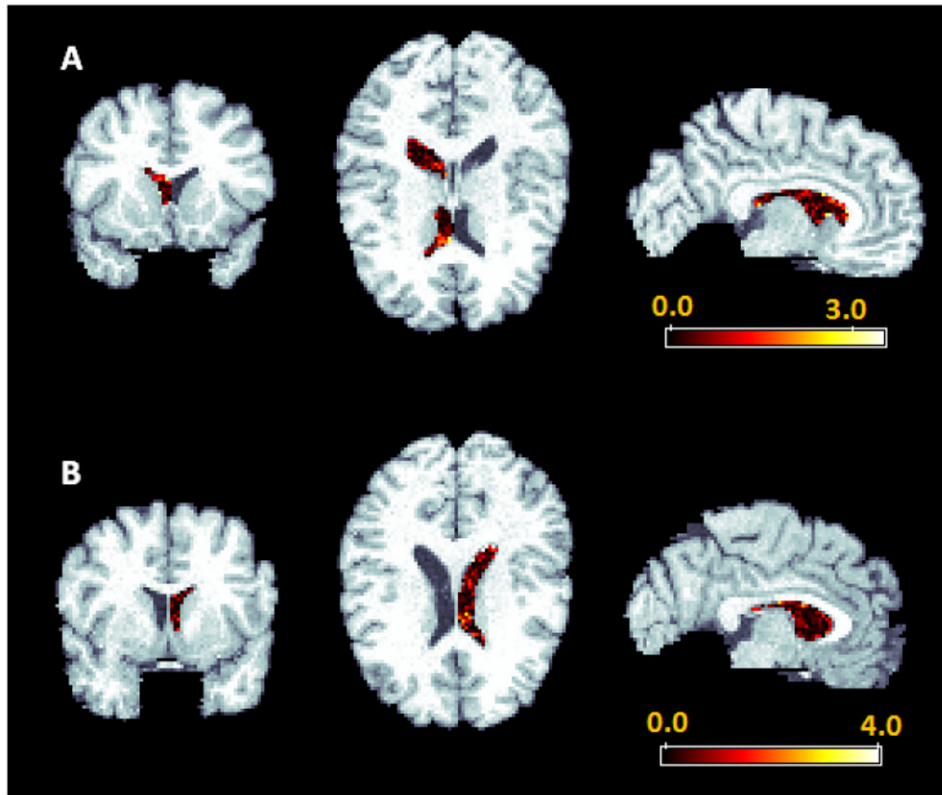
Figure 3.
Voxel-based analysis of RAVENS maps, showing $-\log_{10}(p)$ map from the MA-GMM analysis where ICV was considered a Type I time-varying covariate for various regions of interest displayed on the original intensity template. (A) Right Lateral Ventricle (B) Left Lateral Ventricle.
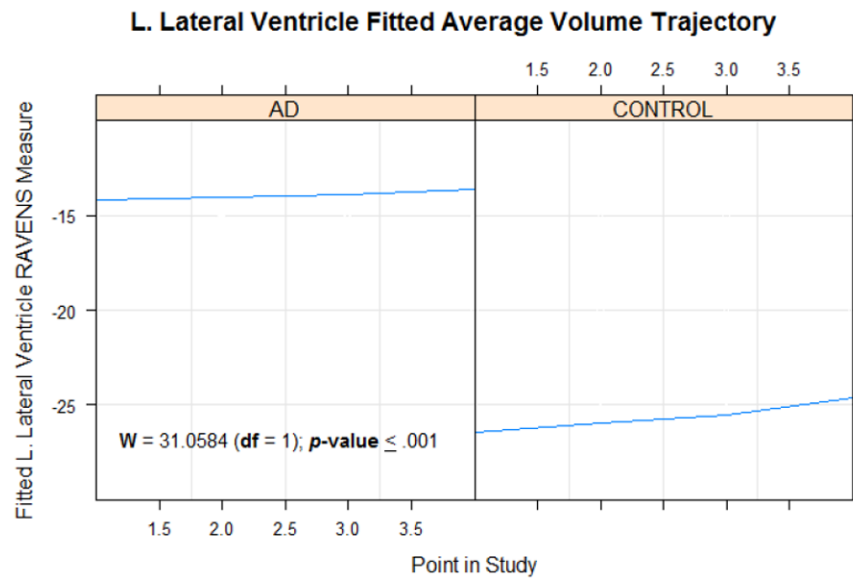
L. Lateral Ventricle Fitted Average Volume Trajectory

W = 31.0584 (**df** = 1); ***p*-value** ≤ .001

Figure 4.
Fitted mean brain volumes, corrected for ICV, for the left lateral ventricle, by diagnostic group. To correct for ICV, fitted values were adjusted as follows: corrected value = fitted value − ×ICV (where is the estimated coefficient from the estimated model corresponding to ICV). The fitted value represents the estimated RAVENS measure at the voxel based on the MA-GMM model. The x–axis represents time and the y–axis represents the corrected fitted mean brain volumes as defined by the equation. Significant pairwise differences between groups are noted.