Statistica Sinica **29** (2019), 1977-2005
doi:https://doi.org/10.5705/ss.202017.0153

TENSOR GENERALIZED ESTIMATING EQUATIONS FOR LONGITUDINAL IMAGING ANALYSIS

Xiang Zhang, Lexin Li, Hua Zhou, Yeqing Zhou,
Dinggang Shen and ADNI

*North Carolina State University, University of California, Berkeley,
University of California, Los Angeles, Shanghai University of Finance
and Economics, University of North Carolina, Chapel Hill
and the Alzheimer's Disease Neuroimaging Initiative*

Abstract: Longitudinal neuroimaging studies are becoming increasingly prevalent, where brain images are collected on multiple subjects at multiple time points. Analyses of such data are scientifically important, but also challenging. Brain images are in the form of multidimensional arrays, or tensors, which are characterized by both ultrahigh dimensionality and a complex structure. Longitudinally repeated images and induced temporal correlations add a further layer of complexity. Despite some recent efforts, there exist very few solutions for longitudinal imaging analyses. In response to the increasing need to analyze longitudinal imaging data, we propose several tensor generalized estimating equations (GEEs). The proposed GEE approach accounts for intra-subject correlation, and an imposed low-rank structure on the coefficient tensor effectively reduces the dimensionality. We also propose a scalable estimation algorithm, establish the asymptotic properties of the solution to the tensor GEEs, and investigate sparsity regularization for the purpose of region selection. We demonstrate the proposed method using simulations and by analyzing a real data set from the Alzheimer's Disease Neuroimaging Initiative.

Key words and phrases: Generalized estimating equations, longitudinal imaging, low rank tensor decomposition, magnetic resonance imaging, multidimensional array, tensor regression.

1. Introduction

Longitudinal neuroimaging studies are becoming increasingly prevalent, in which brain images are collected for multiple subjects, each at multiple time points (Zhang, Shen and Alzheimer's Disease Neuroimaging Initiative (2012)). Analyses of such images help us to understand the progression of a disease, predict the onset of disorders, and identify those regions of the brain relevant to a disease. Our motivating example is a study from the Alzheimer's Disease Neuroimaging Initiative (ADNI). Alzheimer's disease (AD) is a progressive and irreversible neurodegenerative disorder and the leading form of dementia in elderly subjects. The data set contains data on 88 subjects with mild cognitive impairment (MCI), a prodromal stage of AD. Each subject underwent a magnetic resonance imaging (MRI) scan at the following five time points: baseline, 6-month, 12-month, 18-month, and 24-month. After preprocessing, each MRI image is represented by a $32 \times 32 \times 32$ three-dimensional array. For each subject at each visit, researchers also recorded a cognitive score based on a mini-mental state examination (MMSE), which measures the disease progression. Here, researchers are interested in the association between MCI/AD and structural brain atrophy, as reflected by MRI. MRI images are equally important in terms of predicting AD/MCI, because an accurate diagnosis is critical for timely therapy and potentially delaying the disease (Zhang et al. (2011)).

Longitudinal imaging analyses are particularly challenging. Each image is in the form of a multidimensional array, or *tensor*, which is characterized by both ultrahigh dimensionality and a complex structure. For instance, a $32 \times 32 \times 32$ MRI image involves $32^3 = 32,768$ parameters, and there are rarely more than a few hundred subjects. A single image includes complex spatial correlations between its voxels. Thus, naively converting an array into a vector results in extremely high dimensionality and destroys all inherent spatial information. Moreover, repeated images of the same subject are temporally correlated. Despite the increasing availability of longitudinal imaging data, there is a relative paucity of effective solutions, and thus, a substantial demand for the systematic development of new longitudinal imaging analysis methods.

Therefore, we propose *tensor generalized estimating equations* (GEEs) for the analysis of longitudinal imaging data. Our proposed approach consists of two key components: a low-rank tensor decomposition and the GEEs. We impose a low-rank structure on the coefficient array in a GEE that implicitly utilizes the spatial structure of the image predictor. At the same time, it substantially reduces the number of free parameters, making subsequent estimations and inferences feasible. We incorporate this structure into the estimating equations to accommodate the longitudinal correlations in the data. Within this framework, we develop a scalable algorithm for solving the complicated tensor GEEs. We also examine the L_1 and smoothly clipped absolute deviation (SCAD) type penalized tensor GEEs to identify brain subregions that are highly relevant to the clinical outcome. This region-selection process is itself of vital scientific interest, and corresponds to the extensively studied variable-selection problem in classical regressions with vector-valued predictors. Furthermore, we establish the asymptotic properties of the solution to the tensor GEEs. In particular, we show that the tensor GEE estimator inherits the robustness feature of the classical GEE estimator in the sense that the estimate is consistent, even if the working correlation structure is misspecified.

Our proposed approach is related to, but also clearly distinct from existing works on longitudinal data and tensor data analyses. We briefly review the literature here, and point out the differences and our contributions. First, there is a long list of studies on longitudinal data analyses (Liang and Zeger (1986); Prentice and Zhao (1991); Li (1997); Qu, Lindsay and Li (2000); Xie and Yang (2003); Balan and Schiopu-Kratina (2005); Song et al. (2009); Wang (2011)) and variable selection for longitudinal models (Pan (2001); Fu (2003); Fan and Li (2004); Ni, Zhang and Zhang (2010); Xue, Qu and Zhou (2010); Wang, Zhou and Qu (2012)). However, these methods employ a vector of covariates, whereas in our problem, covariates take the form of a multidimensional array. Second, most existing neuroimaging studies utilize only baseline imaging data, ignoring information from the follow-up time points. However, recent studies have begun using longitudinal images for individual-based classification (Misra, Fan and Davatzikos (2009); Davatzikos et al. (2009); McEvoy et al. (2011); Hinrichs et al. (2011)) and cognitive score predictions (Zhang, Shen and Alzheimer's Disease Neuroimaging Initiative (2012)). These solutions extract a vector of summary features from the longitudinal images. In contrast, we jointly model all voxels of an image and include a tensor predictor. Other studies regress longitudinal images on a vector of predictors (Skup, Zhu and Zhang (2012); Li et al. (2013)). However, these works differ from ours in that they treat an image as a response rather than as a predictor. Third, tensor decompositions have been applied in statistical models (Zhou, Li and Zhu (2013); Zhou and Li (2014); Aston, Pigoli and Tavakoli (2017); Sun et al. (2017); Raskutti and Yuan (2016)). Our proposed approach is similar in that we impose a low-rank structure on the tensor GEE coefficient for effective dimension reduction. In that sense, our work generalizes the classical GEE from a vector to a tensor predictor. Furthermore, we generalize the tensor predictor regression (Zhou, Li and Zhu (2013); Raskutti and Yuan (2016)) from independent imaging data to longitudinal image data. Such a generalization may seem straightforward conceptually, but is far from trivial technically. To the best of our knowledge, our work is the first to systematically address a longitudinal imaging predictor in a regression context. As such, it offers both a timely response to the increasing demand for longitudinal neuroimaging, as well as a useful addition to the methodologies used in longitudinal data analyses.

The rest of the article is organized as follows. Section 2 proposes the tensor GEE, along with its estimation and regularization. Section 3 discusses the asymptotic properties of the solution to the tensor GEEs. Sections 4 and 5 present the simulations and real-data analysis, respectively. Section 6 concludes with a discussion. The Supplementary Material contains all the technical proofs.

2. Methodology

2.1. Tensor GEEs

Suppose there are n training subjects, and for the i-th subject, there are observations over m_i time points. For simplicity, we assume $m_i = m$ and that the time points are the same for all subjects. The observed data consist of $\{(Y_{ij}, \boldsymbol{X}_{ij}), i = 1, \ldots, n, j = 1, \ldots, m\}$, where Y_{ij} denotes the target response and $\boldsymbol{X}_{ij} \in \mathbb{R}^{p_1 \times \cdots \times p_D}$ is a D-dimensional array representing the image. Note that our model naturally incorporates an additional vector of covariates, $\boldsymbol{Z}$. However, we choose to drop this term to simplify the presentation. Write $\boldsymbol{Y}_i = (Y_{i1}, \ldots, Y_{im})^{\intercal}$. A key attribute of longitudinal data is that the observations from different subjects are commonly viewed as independent, whereas those from the same subject are *correlated*. That is, the intra-subject covariance matrix $\mathrm{Var}(\boldsymbol{Y}_i) \in \mathbb{R}^{m \times m}$ is not diagonal, but does have some structure.

The GEE method has been widely employed for analyzing correlated longitudinal data since the pioneering work of Liang and Zeger (1986). The method requires the specification of the first two moments of the conditional distribution of the response, given the covariates $\mu_{ij} = E(Y_{ij}|\boldsymbol{X}_{ij})$ and $\sigma_{ij}^2 = \mathrm{Var}(Y_{ij}|\boldsymbol{X}_{ij})$. Following Liang and Zeger (1986), we assume Y_{ij} is from an exponential family with a canonical link. Then, $\mu_{ij}(\boldsymbol{B}) = \mu(\theta_{ij})$ and $\sigma_{ij}^2(\boldsymbol{B}) = \phi\mu^{(1)}(\theta_{ij})$, for $i = 1, \ldots, n,\ j = 1, \ldots, m$, where $\mu(\cdot)$ is a differentiable canonical link function, $\mu^{(1)}(\cdot)$ is its first derivative, θ_{ij} is the linear systematic part, and ϕ is an overdispersion parameter. Here, we simply set $\phi = 1$. The extension to a general ϕ is straightforward. The systematic part θ_{ij} is associated with the covariates via the equation,

$$\theta_{ij} = \langle \boldsymbol{B}, \boldsymbol{X}_{ij} \rangle, \tag{2.1}$$

where $\boldsymbol{B}$ is the coefficient tensor of the same size as $\boldsymbol{X}$ that captures the effects of every array element of $\boldsymbol{X}$ on $\boldsymbol{Y}$. The inner product $\langle \boldsymbol{B}, \boldsymbol{X}_{ij} \rangle = \langle \mathrm{vec}\boldsymbol{B}, \mathrm{vec}\boldsymbol{X}_{ij} \rangle$, where the $\mathrm{vec}(\boldsymbol{B})$ operator stacks the entries of a tensor $\boldsymbol{B}$ into a column vector. The GEE estimator of $\boldsymbol{B}$ is then defined as the solution to

$$\sum_{i=1}^{n} \frac{\partial \boldsymbol{\mu}_i(\boldsymbol{B})}{\partial \mathrm{vec}(\boldsymbol{B})} \boldsymbol{V}_i^{-1} \big\{ \boldsymbol{Y}_i - \boldsymbol{\mu}_i(\boldsymbol{B}) \big\} = \mathbf{0}, \tag{2.2}$$

where $\boldsymbol{Y}_i = (Y_{i1}, \ldots, Y_{im})^{\intercal}$, $\boldsymbol{\mu}_i(\boldsymbol{B}) = [\mu_{i1}(\boldsymbol{B}), \ldots, \mu_{im}(\boldsymbol{B})]^{\intercal}$, and $\boldsymbol{V}_i = \mathrm{cov}(\boldsymbol{Y}_i)$ is the response covariance matrix of the i-th subject. The first component in (2.2) is the derivative of $\boldsymbol{\mu}_i(\boldsymbol{B})$ with respect to the vector $\mathrm{vec}(\boldsymbol{B}) \in \mathbb{R}^{\prod_d p_d}$. As such, in total, there are $\prod_d p_d$ estimating equations to solve in (2.2). For a regression with image covariates, this dimension is prohibitively high, and usually far exceeds the sample size. For instance, a $32 \times 32 \times 32$ MRI image predictor requires that we solve $32^3 = 327,68$ equations, resulting in no unique solution when the sample comprises only tens or hundreds of observations. Thus, it becomes crucial that we reduce the number of estimating equations.

Therefore, we impose a low-rank structure on the coefficient array $\boldsymbol{B}$. More specifically, we assume $\boldsymbol{B}$ in model (2.1) follows a canonical polyadic (CP) decomposition structure (Kolda and Bader (2009)),

$$\boldsymbol{B} = \sum_{r=1}^{R} \boldsymbol{\beta}_1^{(r)} \circ \cdots \circ \boldsymbol{\beta}_D^{(r)}, \tag{2.3}$$

where $\boldsymbol{\beta}_d^{(r)} \in \mathbb{R}^{p_d}$, for $d = 1, \ldots, D, r = 1, \ldots, R$, are column vectors, $\circ$ denotes the outer product, and $\boldsymbol{B}$ cannot be written as a sum of less than R outer products. The decomposition (2.3) is often represented by the shorthand $\boldsymbol{B} = [\![\boldsymbol{B}_1, \ldots, \boldsymbol{B}_D]\!]$, where $\boldsymbol{B}_d = [\boldsymbol{\beta}_d^{(1)}, \ldots, \boldsymbol{\beta}_d^{(R)}] \in \mathbb{R}^{p_d \times R}$. Under this structure, the systematic part in (2.1) becomes

$$\theta_{ij} = \left\langle \sum_{r=1}^{R} \boldsymbol{\beta}_1^{(r)} \circ \cdots \circ \boldsymbol{\beta}_D^{(r)}, \boldsymbol{X}_{ij} \right\rangle = \langle (\boldsymbol{B}_D \odot \cdots \odot \boldsymbol{B}_1) \mathbf{1}_R, \mathrm{vec}\boldsymbol{X}_{ij} \rangle.$$

We then propose the tensor GEE estimator of $\boldsymbol{B}$, which is defined as the solution to

$$\sum_{i=1}^{n} \frac{\partial \boldsymbol{\mu}_i(\boldsymbol{B})}{\partial \boldsymbol{\beta}_{\boldsymbol{B}}} \boldsymbol{V}_i^{-1} \big\{ \boldsymbol{Y}_i - \boldsymbol{\mu}_i(\boldsymbol{B}) \big\} = \mathbf{0}, \tag{2.4}$$

where $\boldsymbol{\beta}_{\boldsymbol{B}} = \mathrm{vec}(\boldsymbol{B}_1, \ldots, \boldsymbol{B}_D)$, and the subscript $\boldsymbol{B}$ indicates that $\boldsymbol{\beta}$ is constructed from the CP decomposition of a given coefficient tensor $\boldsymbol{B} = [\![\boldsymbol{B}_1, \ldots, \boldsymbol{B}_D]\!]$. Introducing the CP structure into the GEE has two important implications. First, compared with the classical GEE (2.2), the derivative in (2.4) is now with respect to $\boldsymbol{\beta}_{\boldsymbol{B}} \in \mathbb{R}^{R \sum_d p_d}$. Consequently, the number of estimating equations is reduced from the exponential order $\prod_d p_d$ to the linear order $R\sum_d p_d$. This substantial reduction in dimensionality is the key to enabling effective estimations and inferences under a limited sample size. Second, under this structure, any two elements $\beta_{i_1\ldots i_d}$ and $\beta_{j_1\ldots j_d}$ in $\boldsymbol{B}$ share common parameters if $i_d = j_d$ for any $d = 1,\ldots,D$. As a result, the coefficients are correlated if they share the same spatial locations along any one of the tensor modes. This implicitly incorporates the spatial structure of the tensor coefficient.

In (2.4), the true intra-subject covariance structure $\boldsymbol{V}_i$ is usually unknown, in practice. The classical GEE adopts a working covariance matrix, specified through a working correlation matrix $\boldsymbol{R}$. That is, $\boldsymbol{V}_i = \boldsymbol{A}_i^{1/2}(\boldsymbol{B})\boldsymbol{R}\boldsymbol{A}_i^{1/2}(\boldsymbol{B})$, where $\boldsymbol{A}_i(\boldsymbol{B})$ is an $m \times m$ diagonal matrix, with $\sigma_{ij}^2(\boldsymbol{B})$ on the diagonal, and $\boldsymbol{R}$ is the $m \times m$ working intra-subject correlation matrix. Commonly used correlation structures include independence, autocorrelation (AR), compound symmetry, and unstructured correlation, among others. The correlation matrix $\boldsymbol{R}$ may involve additional parameters, which can be estimated using a residual-based moment method.

By adopting this working correlation idea and explicitly evaluating the derivative in (2.4), we arrive at a formal definition of the tensor GEE estimator, which is the solution to $(\widehat{\boldsymbol{B}})$ of the following estimating equations:

$$\sum_{i=1}^{n}[\boldsymbol{J}_1,\ \boldsymbol{J}_2,\ldots,\boldsymbol{J}_D]^{\top}\mathrm{vec}(\boldsymbol{X}_i)\boldsymbol{A}_i^{1/2}(\boldsymbol{B})\widehat{\mathbf{R}}^{-1}\boldsymbol{A}_i^{-1/2}(\boldsymbol{B})\{\boldsymbol{Y}_i - \boldsymbol{\mu}_i(\boldsymbol{B})\} = \mathbf{0}, \quad (2.5)$$

where $\widehat{\boldsymbol{R}}$ is an estimated correlation matrix, $\mathrm{vec}(\boldsymbol{X}_i) = (\mathrm{vec}(\boldsymbol{X}_{i1}),\ldots,\mathrm{vec}(\boldsymbol{X}_{im}))$ is a $\prod_{d=1}^{D} p_d \times m$ matrix, and $\boldsymbol{J}_d$ is the $\prod_{d=1}^{D} p_d \times Rp_d$ Jacobian matrix of the form $\mathbf{\Pi}_d \times [(\boldsymbol{B}_D \odot \cdots \odot \boldsymbol{B}_{d+1} \odot \boldsymbol{B}_{d-1} \odot \cdots \odot \boldsymbol{B}_1) \otimes \boldsymbol{I}_{p_d}]$, where $\mathbf{\Pi}_d$ is the $(\prod_{d=1}^{D} p_d)$-by-$(\prod_{d=1}^{D} p_d)$ permutation matrix that reorders $\mathrm{vec}\boldsymbol{B}_{(d)}$ to obtain $\mathrm{vec}\boldsymbol{B}$; that is, $\mathrm{vec}\boldsymbol{B} = \mathbf{\Pi}_d \times \mathrm{vec}\boldsymbol{B}_{(d)}$. Note that $\mu^{(1)}(\theta_{ij})$ is canceled out by the diagonals on the matrix $\boldsymbol{A}_i^{-1}$ owing to the property of the canonical link. For ease of presentation, we denote the left-hand side of equation (2.5) as $\boldsymbol{s}(\boldsymbol{B})$, and write the tensor GEE (2.5) as $\boldsymbol{s}(\boldsymbol{B}) = \mathbf{0}$.

2.2. Estimation and rank selection

Solving the tensor GEE (2.5) with respect to $\boldsymbol{B}$ directly can be computationally intensive, because the mean of the response given the covariates is nonlinear in the parameters and the Jacobian matrices $\boldsymbol{J}_1,\ldots,\boldsymbol{J}_D$ depend on the unknown parameters. Thus, we propose a block-relaxation algorithm to solve the sub-GEE for each $\boldsymbol{B}_1,\ldots,\boldsymbol{B}_D$ iteratively, keeping all other components fixed. Specifically, when updating $\boldsymbol{B}_d \in \mathrm{I\!R}^{p_d \times R}$, the systematic part $\theta_{ij}(\boldsymbol{B})$ can be rewritten as

$$\theta_{ij}(\boldsymbol{B}) = \langle \boldsymbol{B}, \boldsymbol{X}_{ij} \rangle = \langle \boldsymbol{B}_d, \boldsymbol{X}_{ij(d)}(\boldsymbol{B}_D \odot \cdots \odot \boldsymbol{B}_{d+1} \odot \boldsymbol{B}_{d-1} \odot \cdots \odot \boldsymbol{B}_1) \rangle,$$

where $\boldsymbol{X}_{ij(d)}$ is the mode-d matricization of the tensor $\boldsymbol{X}_{ij}$, which flattens $\boldsymbol{X}_{ij}$ into a $p_d \times \prod_{d' \neq d} p_{d'}$ matrix, such that the $(k_1, \ldots, k_D)$ element of $\boldsymbol{X}_{ij}$ maps to the (k_d, l) element of the matrix $\boldsymbol{X}_{ij(d)}$, where $l = 1 + \sum_{d' \neq d}(k_{d'} - 1) \prod_{d'' < d', d'' \neq d} p_{d''}$, and $\odot$ denotes the Khatri–Rao product (Rao and Mitra (1971)). Consequently, the systematic part $\theta_{ij}(\boldsymbol{B})$ becomes linear in $\boldsymbol{B}_d$. The Jacobian matrix $\boldsymbol{J}_d$ is free of $\boldsymbol{B}_d$ and depends on the covariates and the fixed parameters only. Then, each step reduces to a standard GEE problem with Rp_d parameters, which can be solved using standard statistical software. As in the case of the classical GEE, our tensor GEE potentially has multiple roots. Our numerical simulations show that different starting values often lead to the same solution.

A problem of practical importance is choosing the rank R for the coefficient array $\boldsymbol{B}$ in its CP decomposition. This can be viewed as a *model selection* problem. Pan (2001) proposed a quasi-likelihood independence model criterion for the classical GEE model selection, which evaluates the likelihood under the independent working correlation assumption. In our tensor GEE setup, we adopt a similar criterion,

$$\mathrm{BIC}(R) = -2\ell(\widehat{\boldsymbol{B}}(R); \boldsymbol{I}_m) + \log(n) p_e, \tag{2.6}$$

where $\ell(\widehat{\boldsymbol{B}}(R); \boldsymbol{I}_m)$ is the log-likelihood evaluated at the tensor GEE estimator $\widehat{\boldsymbol{B}}(R)$, with a working rank R and the independent working correlation structure $\boldsymbol{I}_m$. For simplicity, we call this criterion the Bayesian information criterion (BIC), because the term $\log(n)$ is used. Because the CP decomposition itself is not unique, but can be made so under some minor conditions (Zhou, Li and Zhu (2013)), the actual number of estimating equations, or the effective number of parameters, is of the form $p_e = R(p_1 + p_2) - R^2$ for $D = 2$, and $p_e = R(\sum_d p_d - D + 1)$ for $D > 2$. We choose R that minimizes this criterion among a series of working ranks.

2.3. Regularization for region selection

Selecting brain subregions that are highly relevant to the disease outcome is of vital scientific interest. This allows researchers to concentrate on brain subregions, thus improving their understanding of the disease pathology, and is useful for hypothesis generation and validation. In our setup, region selection translates to a *sparse* estimation of the elements of the coefficient tensor $\boldsymbol{B}$, and is analogous to the extensively studied variable selection problem in classical vector-valued regressions. We adopt the L_1-type regularization to achieve this goal. Specifically, we consider the following regularized tensor GEE:

$$n^{-1}\boldsymbol{s}(\boldsymbol{B}) - \begin{pmatrix} \partial_{\beta^{(1)}_{11}} P_\lambda(|\beta^{(1)}_{11}|, \rho_n) \\ \vdots \\ \partial_{\beta^{(r)}_{di}} P_\lambda(|\beta^{(r)}_{di}|, \rho_n) \\ \vdots \\ \partial_{\beta^{(R)}_{Dp_D}} P_\lambda(|\beta^{(R)}_{Dp_D}|, \rho_n) \end{pmatrix} = \mathbf{0}, \tag{2.7}$$

where $P_\lambda(|\beta|, \rho_n)$ is a scalar penalty function, ρ_n is the penalty tuning parameter, λ is an index for the penalty family, and $\partial_\beta P_\lambda(|\beta|, \rho_n)$ is the subgradient with respect to the argument β. We consider two specific penalty functions: the lasso (Tibshirani (1996)), in which $P_\lambda(|\beta|, \rho_n) = \rho_n|\beta|$ with $\lambda = 1$, and the SCAD (Fan and Li (2001)), in which $\partial/\partial|\beta| P_\lambda(|\beta|, \rho_n) = \rho_n\{1_{\{|\beta|\le\rho_n\}} + (\lambda\rho_n - |\beta|)_+/(\lambda - 1)$ $1_{\{|\beta|>\rho_n\}}\}$, for $\lambda > 2$.

Owing to the separability of the parameters in the regularization term, the alternating updating strategy still applies. When updating $\boldsymbol{B}_d$, we solve the penalized sub-GEE

$$n^{-1}\boldsymbol{s}_d(\boldsymbol{B}_d) - \begin{pmatrix} \partial_{\beta^{(1)}_{d1}} P_\lambda(|\beta^{(1)}_{d1}|, \rho_n) \\ \vdots \\ \partial_{\beta^{(r)}_{di}} P_\lambda(|\beta^{(r)}_{di}|, \rho_n) \\ \vdots \\ \partial_{\beta^{(R)}_{dp_d}} P_\lambda(|\beta^{(R)}_{dp_D}|, \rho_n) \end{pmatrix} = \mathbf{0}, \tag{2.8}$$

where $\boldsymbol{s}_d$ is the sub-estimation equation for block $\boldsymbol{B}_d$. There are Rp_d equations to solve in this step. The anti-derivative of $\boldsymbol{s}_d$ is recognized as the loss of an Aitken linear model with a block-diagonal covariance matrix. Thus, after a linear transformation of $\boldsymbol{Y}_i$ and using the working design matrix, the solution to (2.8) is the same as the minimizer of a regular penalized weighted least squares problem, for which many software packages exist. The fitting procedure reduces to alternating the penalized weighted least squares.

Note that, in addition to the region selection, regularization is useful for stabilizing the estimates, handling small-n-large-p, and incorporating prior subject knowledge. The above regularization paradigm can be extended to incorporate other forms of regularization, such as the L_2-type ridge regularization, or different penalties along different modes of the tensor coefficient.

3. Theory

Next, we study the asymptotic properties of the tensor GEE estimator. We first note that there are two specifications, or potential misspecifications, in the tensor GEE. The first is the working correlation structure. We show that the tensor GEE estimator remains consistent, even if the working correlation structure is misspecified. This is an analogous result to that of the classical GEE. Thus, we extend the work of Xie and Yang (2003), Balan and Schiopu-Kratina (2005), and Wang (2011). We achieve this by assuming the rank is fixed and known. This is similar in spirit to the classical GEE setup, where a linear model is imposed and the rank is, in effect, set to one. The second specification is the working rank of the CP decomposition in the tensor GEE. We show that for the normal linear model, the rank selected by the BIC under an independent correlation structure is consistent, even if this structure might have been misspecified. This justifies the BIC criterion (2.6) and, to some extent, the asymptotic investigation under a known rank. Note that the assumption of a known rank is common in theoretical analyses of estimators based on low-rank structures (Zhou, Li and Zhu (2013); Sun and Li (2017)). Furthermore, our asymptotic study is carried out in the classical sense that the number of parameters (dimension) is fixed and the sample size goes to infinity. We believe such a fixed-dimension asymptotic study is useful because it reveals the basic properties and offers a statistical guarantee for our tensor GEE estimator. More importantly, it establishes that both our tensor estimator and the rank estimator remain consistent under a potentially incorrect working correlation structure. In principle, we can also consider the scenario where the dimension diverges to infinity along with the sample size. In this regard, we have obtained preliminary asymptotic results, but leave a comprehensive treatment of the tensor GEE under a diverging dimension for future research.

3.1. Regularity conditions

We begin with a list of regularity conditions for the asymptotics of the tensor GEE with a fixed number of parameters. Let $||\boldsymbol{x}||$ denote the Euclidean norm of a vector $\boldsymbol{x}$ and let $||\boldsymbol{X}||_F$ be the Frobenius norm of a matrix $\boldsymbol{X}$. Denote $\boldsymbol{N}_n$ as the neighborhood of the true tensor coefficient $\{\boldsymbol{B} : ||\boldsymbol{\beta}_{\boldsymbol{B}} - \boldsymbol{\beta}_{\boldsymbol{B}_0}|| \leq \triangle n^{-1/2}\}$ for some constant $\triangle > 0$.

(A1) For some constant $c_1 > 0$, $||\boldsymbol{X}_{ij}||_F \leq c_1$, for $i = 1, \ldots, n$, $j = 1, \ldots, m$.

(A2) The true value $\boldsymbol{B}_0$ of the unknown parameter lies in the interior of a compact parameter space $\mathcal{B}$ and follows the rank-R CP structure defined in (2.3).

(A3) Let $\boldsymbol{I}(\boldsymbol{B}) = n^{-1}\sum_{i=1}^{n}[\boldsymbol{J}_1,\ \boldsymbol{J}_2, \ldots, \boldsymbol{J}_D]^{\intercal}\text{vec}\boldsymbol{X}_i\text{vec}^{\intercal}\boldsymbol{X}_i[\boldsymbol{J}_1,\ \boldsymbol{J}_2, \ldots, \boldsymbol{J}_D]$. There exist two constants $0 < c_2 < c_3$, such that $c_2 \le \lambda_{\min}(\boldsymbol{I}(\boldsymbol{B})) \le \lambda_{\max}(\boldsymbol{I}(\boldsymbol{B})) \le c_3$ over the set $\boldsymbol{N}_n$, where $\lambda_{\min}$ and $\lambda_{\max}$ are the smallest and largest eigenvalues, respectively. In addition, $\boldsymbol{I}(\boldsymbol{B})$ has a constant rank on the same set.

(A4) The true intra-subject correlation matrix $\boldsymbol{R}_0$ has eigenvalues bounded by zero and infinity. There exists a positive definite matrix $\tilde{\boldsymbol{R}}$ with eigenvalues bounded away from zero and infinity, such that $\|\widehat{\boldsymbol{R}}^{-1} - \tilde{\boldsymbol{R}}^{-1}\|_F = O_p(n^{-1/2})$, where $\widehat{\boldsymbol{R}}$ is an estimator of the correlation matrix.

(A5) For $\delta > 0$ and $c_4 > 0$, $E(\|\boldsymbol{A}_i^{-1/2}(\boldsymbol{B}_0)(\boldsymbol{Y}_i - \boldsymbol{\mu}_i(\boldsymbol{B}_0))\|)^{2+\delta} \le c_4$, for all $1 \le i \le n$.

(A6) For some constant $c_5 > 0$, $||\partial\theta_{ij}(\boldsymbol{\beta}_{\boldsymbol{B}})/\partial\boldsymbol{\beta}_{\boldsymbol{B}}|| \le c_5$, for $i = 1, \ldots, n$, $j = 1, \ldots, m$.

(A7) Denote by $\mu^{(k)}(\theta_{ij})$, for $i = 1, \ldots, n$, $j = 1, \ldots, m$, and $k = 2, 3$, the k-th derivative of $\mu(\theta_{ij})$. For some positive constants $c_6 < c_7$ and c_8, we have $c_6 < |\mu^{(1)}(\theta_{ij})| < c_7$ and $|\mu^{(k)}(\theta_{ij})| < c_8$ over the set $\boldsymbol{N}_n$.

(A8) Denote by $\boldsymbol{H}_{ij}(\boldsymbol{B}) = (\partial^2\theta_{ij}(\boldsymbol{\beta}_{\boldsymbol{B}}))/(\partial\boldsymbol{\beta}_{\boldsymbol{B}}\partial\boldsymbol{\beta}_{\boldsymbol{B}}^{\intercal})$. That is, $\boldsymbol{H}_{ij}(\boldsymbol{B})$ is the Hessian matrix of the linear systematic part θ_{ij}. There exist two positive constants $c_9 < c_{10}$, such that, for $i = 1, \ldots, n$ and $j = 1, \ldots, m$, $c_9 \le \lambda_{\min}(\boldsymbol{H}_{ij}(\boldsymbol{B})) \le \lambda_{\max}(\boldsymbol{H}_{ij}(\boldsymbol{B})) \le c_{10}$ over the set $\boldsymbol{N}_n$.

A few remarks are in order. Conditions (A2) and (A3) are required for the model identifiability of the tensor GEE (Zhou, Li and Zhu (2013)). Note that the matrix $\boldsymbol{I}(\boldsymbol{B})$ in (A3) is an $R\sum_{d=1}^{D} p_d \times R\sum_{d=1}^{D} p_d$ matrix. Thus, (A3) is much weaker than the nonsingularity condition on the design matrix if we directly vectorize the tensor covariate. Condition (A4) is commonly imposed in the GEE literature. It requires only that $\widehat{\boldsymbol{R}}$ be a consistent estimator of some $\tilde{\boldsymbol{R}}$, in the sense that $\|\widehat{\boldsymbol{R}}^{-1} - \tilde{\boldsymbol{R}}^{-1}\|_F = O_p(n^{-1/2})$. Here, $\tilde{\boldsymbol{R}}$ needs to be well behaved in that it is positive definite with eigenvalues bounded by zero and infinity, but $\tilde{\boldsymbol{R}}$ does *not* have to be the true intra-subject correlation $\boldsymbol{R}_0$. This condition essentially leads to the robust feature in Theorem 1 that the tensor GEE estimate is consistent, even if the working correlation structure is misspecified. Condition (A5) regulates the tail behavior of the residuals so that the noise does not accumulate too fast, and we employ the Lindeberg–Feller central limit theorem to control the asymptotic behavior of the residuals. Condition (A6) states that the gradients of the systematic part are well defined. Condition (A7) concerns the canonical link and holds, in general, for common exponential families, such as the binomial and Poisson distributions. Condition (A8) ensures that the Hessian matrix $\boldsymbol{H}(\boldsymbol{B})$ of the linear systematic part, which is highly sparse, is well behaved in the neighborhood of the true value.

3.2. Consistency and asymptotic normality

Before we turn to the asymptotics of the tensor GEE estimator, we address two components involved in the estimating equations: the initial estimator and the correlation estimator. Recall that the tensor GEE estimator $\widehat{\boldsymbol{B}}$ is obtained by solving the equations

$$\sum_{i=1}^{n}[\boldsymbol{J}_1,\ldots,\boldsymbol{J}_D]^{\mathsf{T}}\mathrm{vec}\boldsymbol{X}_i\boldsymbol{A}_i^{1/2}(\boldsymbol{B})\widehat{\boldsymbol{R}}^{-1}\boldsymbol{A}_i^{-1/2}(\boldsymbol{B})\big\{\boldsymbol{Y}_i-\boldsymbol{\mu}_i(\boldsymbol{B})\big\}=\boldsymbol{0},$$

where $\widehat{\boldsymbol{R}}$ is any estimator of the intra-subject correlation matrix satisfying condition (A4). Note that $\widehat{\boldsymbol{R}}$ is often obtained using the residual-based moment method, which in turn requires an initial estimator of $\boldsymbol{B}_0$. Next, we examine several frequently used estimators of $\widehat{\boldsymbol{B}}$ and $\widehat{\boldsymbol{R}}$.

A customary initial estimator of $\widehat{\boldsymbol{B}}$ in the GEE literature assumes an independent working correlation. That is, we ignore potential intra-subject correlation, in which case, the corresponding tensor GEE becomes

$$\sum_{i=1}^{n}[\boldsymbol{J}_1,\ldots,\boldsymbol{J}_D]^{\mathsf{T}}\mathrm{vec}\boldsymbol{X}_i\big\{\boldsymbol{Y}_i-\boldsymbol{\mu}_i(\boldsymbol{B})\big\}=\boldsymbol{0}.$$

Denoting the equations as $\boldsymbol{s}_{init}(\boldsymbol{B})=\boldsymbol{0}$ and the solution as $\widehat{\boldsymbol{B}}_{init}$, the following lemma shows that this is a consistent estimator of the true $\boldsymbol{B}_0$.

Lemma 1. *Under conditions* (A1)–(A3) *and* (A5)–(A8)*, there exists a root $\widehat{\boldsymbol{B}}_{init}$ of the equations $\boldsymbol{s}_{init}(\boldsymbol{B})=\boldsymbol{0}$ satisfying*

$$\|\boldsymbol{\beta}_{\widehat{\boldsymbol{B}}_{init}}-\boldsymbol{\beta}_{\boldsymbol{B}_0}\|=O_p(n^{-1/2}).$$

Here, $\boldsymbol{\beta}_{\boldsymbol{B}}=\mathrm{vec}(\boldsymbol{B}_1,\ldots,\boldsymbol{B}_D)$, which is constructed based on the CP decomposition of a given tensor $\boldsymbol{B}=[\![\boldsymbol{B}_1,\ldots,\boldsymbol{B}_D]\!]$, as defined previously. Given a consistent initial estimator of $\boldsymbol{B}_0$, there exist multiple choices for the working correlation structure, such as autocorrelation, compound symmetry, and the nonparametric structure (Balan and Schiopu-Kratina (2005)). We investigate these choices in Sections 4 and 5.

Next, we establish the consistency and asymptotic normality of the tensor GEE estimator defined in (2.5).

Theorem 1. *Under conditions* (A1)–(A8)*, there exists a root* $\widehat{\boldsymbol{B}}$ *of the equations* $\boldsymbol{s}(\boldsymbol{B}) = \boldsymbol{0}$ *satisfying*

$$\|\boldsymbol{\beta}_{\widehat{\boldsymbol{B}}} - \boldsymbol{\beta}_{\boldsymbol{B}_0}\| = O_p(n^{-1/2}).$$

The key point in Theorem 1, as implied by condition (A4), is that the consistency of the tensor coefficient estimator $\widehat{\boldsymbol{B}}$ does *not* require the estimated working correlation $\widehat{\boldsymbol{R}}$ to be a consistent estimator of the true correlation $\boldsymbol{R}_0$. As a result, we are protected from a potential misspecification of the intra-subject correlation structure. This robustness feature is well known for GEE estimators with vector-valued covariates. Theorem 1 confirms and extends this result to the tensor GEE case with image covariates. Note that although the asymptotics of the classical GEE can, in principle, be generalized to tensor data by directly vectorizing the coefficient array, the ultrahigh dimensionality of the parameters would make the regularity conditions, such as (A3), unrealistic. In contrast, Theorem 1 ensures the consistency and robustness properties by taking into account the structural information of the tensor coefficient under the GEE framework. Under condition (A4), we define

$$\tilde{\boldsymbol{M}}_n(\boldsymbol{B})=\sum_{i=1}^{n}[\boldsymbol{J}_1,\ldots,\boldsymbol{J}_D]^{\top}\mathrm{vec}\boldsymbol{X}_i\boldsymbol{A}_i^{1/2}(\boldsymbol{B})\tilde{\boldsymbol{R}}^{-1}\boldsymbol{R}_0\tilde{\boldsymbol{R}}^{-1}\boldsymbol{A}_i^{1/2}(\boldsymbol{B})\mathrm{vec}^{\top}\boldsymbol{X}_i[\boldsymbol{J}_1,\ldots,\boldsymbol{J}_D],$$

$$\tilde{\boldsymbol{D}}_{n1}(\boldsymbol{B})=\sum_{i=1}^{n}[\boldsymbol{J}_1,\ldots,\boldsymbol{J}_D]^{\top}\mathrm{vec}\boldsymbol{X}_i\boldsymbol{A}_i^{1/2}(\boldsymbol{B})\tilde{\boldsymbol{R}}^{-1}\boldsymbol{A}_i^{1/2}(\boldsymbol{B})\mathrm{vec}^{\top}\boldsymbol{X}_i[\boldsymbol{J}_1,\ldots,\boldsymbol{J}_D].$$

As we show in the appendix, $\tilde{\boldsymbol{M}}_n(\boldsymbol{B})$ approximates the covariance matrix of $\boldsymbol{s}(\boldsymbol{B})$ in (2.5), whereas $\tilde{\boldsymbol{D}}_{n1}(\boldsymbol{B})$ approximates the leading term of the negative gradient of $\boldsymbol{s}(\boldsymbol{B})$ with respect to $\boldsymbol{\beta}_{\boldsymbol{B}}$. The following theorem establishes the asymptotic normality of the tensor GEE estimator.

Theorem 2. *Under conditions* (A1)–(A8)*, for any vector* $\boldsymbol{b} \in \mathbb{R}^{R\sum_{d=1}^{D} p_d}$*, such that* $\|\boldsymbol{b}\| = 1$*, we have*

$$\boldsymbol{b}^{\top}\tilde{\boldsymbol{M}}_n^{-1/2}(\boldsymbol{B}_0)\tilde{\boldsymbol{D}}_{n1}(\boldsymbol{B}_0)\big(\boldsymbol{\beta}_{\widehat{\boldsymbol{B}}} - \boldsymbol{\beta}_{\boldsymbol{B}_0}\big) \to \mathrm{Normal}(0,1) \textit{ in distribution.}$$

3.3. Rank selection consistency

Next, we establish that the rank selected by the BIC in (2.6) under the independent working correlation is a consistent estimator of the true rank. This result is useful in two ways. First, it justifies, to some extent, the asymptotic study in the previous section under a known rank. Second, it improves our understanding of the interaction between the working correlation and the rank specification. That is, the rank selected under a potentially misspecified correlation structure remains consistent. Note that this rank selection consistency result is not established in Zhou, Li and Zhu (2013). Therefore, to the best of our knowledge, this study is the first to find such a result. For simplicity, we only consider the Gaussian linear model case, and leave the GLM case for future research.

We employ the same regularity conditions (A1)–(A8) in Section 3.2, except that we replace (A3) with the following condition:

(A3*) There exist two positive constants $c_1^* < c_2^*$, such that $c_1^* \leq \lambda_{\min}(\boldsymbol{I}(\boldsymbol{B})) \leq \lambda_{\max}(\boldsymbol{I}(\boldsymbol{B})) \leq c_2^*$ for all parameter points $\boldsymbol{B}$ in the interior of the parameter space. In addition, the rank is constant over the set $\{\boldsymbol{B} : ||\boldsymbol{\beta}_{\boldsymbol{B}} - \boldsymbol{\beta}_{\boldsymbol{B}_0}|| \leq \triangle n^{-1/2}\}$, for some $\triangle > 0$.

The reason for requiring (A3*) is that we need to characterize the behavior of some underfitted estimators with rank smaller than the true rank. These underfitted estimators may not reside in the neighborhood of the true parameters. However, note that (A3*) is a fairly mild condition, and the difference between (A3*) and (A3) is small. This is because, when the dimension is fixed, $\boldsymbol{I}(\boldsymbol{B})$ has fixed dimensions. Then, the condition on the bounded eigenvalues essentially requires that the matrix be nonsingular. The following theorem establishes the rank selection consistency.

Theorem 3. *Let* $\widehat{R} = \arg\min BIC(R)$ *and* $R_0 = rank(\boldsymbol{B}_0)$. *For the Gaussian linear model, under conditions* (A1)–(A8) *and the modified condition* (A3*), *we have*

$$\Pr(\widehat{R} = R_0) \to 1, \ \textit{as } n \to \infty.$$

That is, with high probability, the rank selected by the BIC recovers the true rank. From a model selection perspective, this rank selection consistency implies that neither the overfitted model with a higher rank nor the underfitted model with an insufficient rank is favored by the BIC.

3.4. Region selection consistency

Recall that under the CP structure, the element in the coefficient tensor $\boldsymbol{B}$ can be written as $\beta_{i_1 \ldots i_D} = \sum_{r=1}^{R} \boldsymbol{\beta}_{1i_1}^{(r)} \times \cdots \times \boldsymbol{\beta}_{Di_D}^{(r)}$. In the imaging application, where $D = 3$, for example, the region at (i_1, i_2, i_3) is nonactive if $\beta_{i_1,i_2,i_3} = 0$. This can be induced if one of $\{\boldsymbol{\beta}_{1i_1}^{(r)}, \boldsymbol{\beta}_{2i_2}^{(r)}, \boldsymbol{\beta}_{3i_3}^{(r)}\}$ is zero for each $r = 1, \ldots, R$.

Therefore, correctly recovering the sparsity pattern of $\boldsymbol{\beta}_{\boldsymbol{B}}$ results in the selection of active regions of $\boldsymbol{B}$. Next, we establish that this selection is consistent for the SCAD regularized tensor GEE in (2.7).

Theorem 4. *Under conditions* (A1)–(A8), $\rho_n = o(1)$, *and* $n^{-1/2}\log n = o(\rho_n)$, *there exists one solution,* $\boldsymbol{\beta}_{\widehat{\boldsymbol{B}}}$, *to the SCAD regularized tensor GEE, such that*

$$\Pr(supp(\boldsymbol{\beta}_{\widehat{\boldsymbol{B}}}) = supp(\boldsymbol{\beta}_{\boldsymbol{B}_0})) \to 1, \ \ as\ n \to \infty,$$

where $supp(\boldsymbol{\beta})$ *denotes the support of the vector* $\boldsymbol{\beta}$.

This theorem states that the support of the true tensor coefficient, $\text{supp}(\boldsymbol{\beta}_{\boldsymbol{B}_0})$, can be recovered with high probability using the SCAD regularized tensor GEE. As such, it establishes the region selection consistency in the context of the tensor GEE.

4. Simulations

We carried out extensive simulations to investigate the finite-sample performance of our proposed tensor GEE approach. We adopt the following simulation setup. We generate the responses according to the normal linear model

$$\boldsymbol{Y}_i \sim \text{MVN}(\boldsymbol{\mu}_i, \sigma^2 \mathbf{R}_0), \quad i = 1, \ldots, n,$$

where $\boldsymbol{Y}_i = (Y_{i1}, \ldots, Y_{im})^{\mathsf{T}}$, $\boldsymbol{\mu}_i = (\mu_{i1}, \ldots, \mu_{im})^{\mathsf{T}}$, σ^2 is a scale parameter, and $\boldsymbol{R}_0$ is the true $m \times m$ intra-subject correlation matrix. We choose $\boldsymbol{R}_0$ such that it has an exchangeable (compound symmetric) structure with the off-diagonal coefficient $\rho_n = 0.8$. The mean function is of the form $\mu_{ij} = \boldsymbol{\gamma}^{\mathsf{T}}\boldsymbol{Z}_{ij} + \langle \boldsymbol{B}, \boldsymbol{X}_{ij}\rangle$, for $i = 1, \ldots, n$ and $j = 1, \ldots, m$, where $\boldsymbol{Z}_{ij} \in \mathbb{R}^{5}$ denotes the additional covariates, with all elements generated from a standard normal distribution. In addition, $\boldsymbol{\gamma} \in \mathbb{R}^{5}$ is the corresponding coefficient vector, with all elements equal to one, $\boldsymbol{X}_{ij} \in \mathbb{R}^{64\times 64}$ denotes the 2D matrix covariate, again with all elements from a standard normal distribution, and $\boldsymbol{B} \in \mathbb{R}^{64\times 64}$ is the matrix coefficient. The entries of $\boldsymbol{B}$ take the value zero or one, and $\boldsymbol{B}$ contains a series of shapes, as shown in Figure 1, including a “square,” “T-shape,” “disk,” “triangle,” and “butterfly.” Our goal is to recover these shapes in $\boldsymbol{B}$ by inferring the association between Y_{ij} and $\boldsymbol{X}_{ij}$.

4.1. Signal recovery

In reality, the true signal rarely has an exact low-rank structure. Therefore, the tensor GEE model essentially provides a low-rank *approximation* to the true signal. Thus, our first task is to verify whether this approximation is adequate in the sense that it can recover the true signal area and shape to a reasonable degree. We set $n = 500$ and $m = 4$ and show the tensor GEE estimates and the corresponding BIC values under three working ranks of $R = 1, 2$, and 3 in Figure 1. We first assume that the correlation structure is correctly specified. We examine a potential misspecification in the next section. In this setup, the "square" has a true rank equal to one, "T-shape" has rank two, and the remaining shapes have ranks much larger than three. It is clear from Figure 1 that the tensor GEE produces a reasonable recovery of the true signal, even for signals with a high rank (e.g., "disk" and "butterfly"). All shapes can be clearly recognized, even though the surrounding area is gray and noisy. Moreover, the BIC criterion (2.6) successfully identifies the correct, or best approximate rank for all of the signals.

4.2. Effect of correlation specification

We have shown that the tensor GEE estimator remains asymptotically consistent, even when the working correlation structure is misspecified. However, this describes only the *large*-sample behavior. In this section, we investigate the potential effect of a correlation misspecification when the sample size is *small* or *moderate*.

We choose the "butterfly" signal and fit the tensor GEE model using three working correlation structures: exchangeable, autoregressive of order one (AR-1), and independent. Table 1 reports the averages and standard errors (in parentheses) for 100 simulation replicates of the squared bias, variance, and mean squared error (MSE) of the tensor GEE estimate. We observe that the estimator based on the correct working correlation structure (i.e., the exchangeable structure) outperforms those based on misspecified correlation structures. When the sample size is moderate ($n = 100$), the estimators have comparable bias and the variation in the MSE is mostly from the variance part of the estimator. This agrees with the theory that the choice of working correlation structure affects the asymptotic variance of the estimator. When the sample size becomes relatively large ($n = 150$), the estimators perform similarly using the scaling term of $n^{-1/2}$ on the variance. When the sample size is small ($n = 50$), the estimators have relatively large bias and the independent working structure yields similar results to those of the exchangeable structure. This suggests that when the sample size is *limited*, using a simple independent working structure is preferable to using a more complex correlation structure.

Nevertheless, we should bear in mind that the above observations reflect the

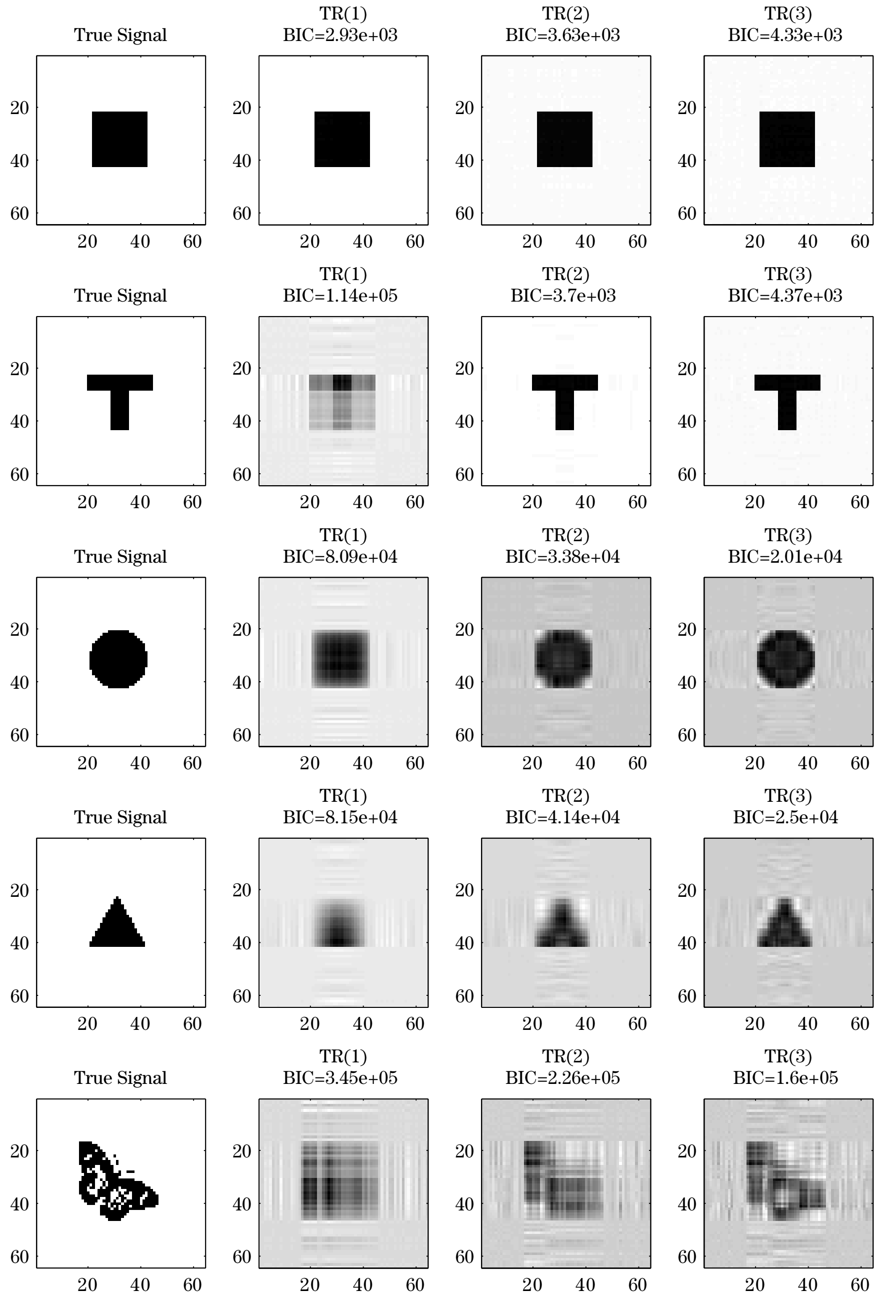

Figure 1. True and recovered image signals by the tensor GEE with varying ranks. $n = 500, m = 4$. The correlation structure is correctly specified. TR(R) is the estimate from the rank-R tensor model.

Table 1. Bias, variance, and MSE of the tensor GEE estimates under various working correlation structures. The result is based on 100 simulation replicates. The true intra-subject correlation is exchangeable with $\rho_n = 0.8$.

n	m	Working Correlation	Bias^2	Variance	MSE
50	10	Exchangeable	122.0	383.6	**505.6 (7.9)**
		AR-1	139.1	530.0	669.1(15.8)
		Independence	119.1	393.9	513.0(11.0)
100	10	Exchangeable	85.8	128.9	**214.7 (2.2)**
		AR-1	88.0	159.1	247.1 (3.0)
		Independence	93.0	141.2	234.2 (2.8)
150	10	Exchangeable	86.1	51.3	**137.2 (0.6)**
		AR-1	85.6	56.0	141.6 (0.6)
		Independence	84.9	62.3	147.2 (0.9)

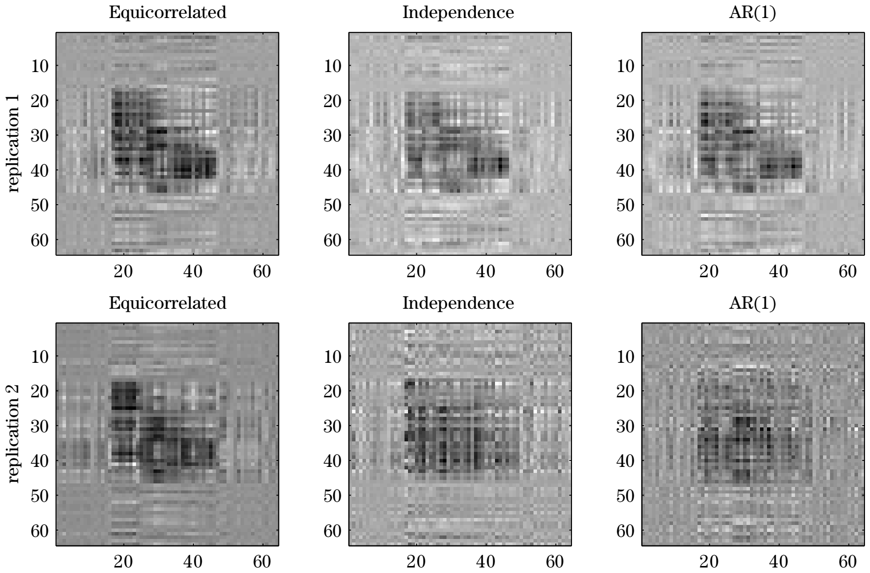

Figure 2. Snapshots of tensor GEE estimations with different working correlation structures. The true correlation is an equicorrelated structure. The comparison is row-wise. The first row shows a replicate where the estimates are "close" to the average behavior, and thus, the visual quality of the estimates under different correlations structures are similar. The second row shows a replicate where the estimates are "far away" from the average. Here, the estimate under the correct correlation structure (panel 1) is superior to those under the incorrect structures.

average behavior of the estimate. Figure 2 shows *two* snapshots of the estimated signals under the three working correlations with $n = 100$. In top top panel, the estimates are "close" to the average in the sense that the bias, variance, and MSE values for this single data realization are similar to the averages reported in Table 1. Consequently, the visual qualities of the three recovered signals are similar. However, in the bottom panel, the estimates are "far away" from the average. Here, the quality of the estimated signal under the correct working correlation structure is superior to those under the incorrect specifications. Thus, as long as the sample size of the longitudinal imaging study is moderate to large, a longitudinal model should be favored over a model that ignores potential intra-subject correlation.

4.3. Regularized estimation and comparison

We next study the empirical performance of the regularized tensor GEE (denoted as "regularization"), comparing it with that of several alternative solutions: the tensor GEE without regularization ("no regularization"), the lasso regularized vector GEE applied to the vectorized image predictor (Fu (2003) "Fu-Lasso"), the SCAD regularized vector GEE (Wang, Zhou and Qu (2012) "Wang-SCAD"), and the sandwich estimator (Guillaume et al. (2014) "SwE"). We adopt the same simulation setup as in Section 4.1, vary the sample size n, and fix $m = 4$. For our regularized tensor GEE, we implemented both the lasso and SCAD penalty, and found their performance to be visually very similar. As such, we only report the results based on the SCAD here. The penalty parameter is tuned based on an independent validation data set. Note that the sandwich estimator of Guillaume et al. (2014) treats the image as a response, whereas we treat the image as a predictor. We used the software provided by Guillaume et al. (2014) for the calculation. We experimented with various shapes and obtained similar results. To conserve space, we report only the results of "T-shape" and "butterfly" in Figures 3 and 4. In both cases, our regularized tensor GEE outperforms the alternative solutions, especially when the sample size is limited.

4.4. Computation time

In this section, we investigate the computation time of our proposed tensor GEE. We consider the same simulation setup as in Section 4.1, but vary the sample size and the image dimension. First, we set $m = 10$ and the matrix covariate dimension to 64×64, and increase n from 50 to 500 by an increment of 50. Second, we set $n = 200$, $m = 4$, and increase the matrix covariate dimension from 32×32 to 128×128 by an increment of 16. All simulations are carried out on a laptop computer with an Intel Xeon 2.60 GHz processor. Figure 5 reports the average computation time, in seconds, along with its confidence interval,

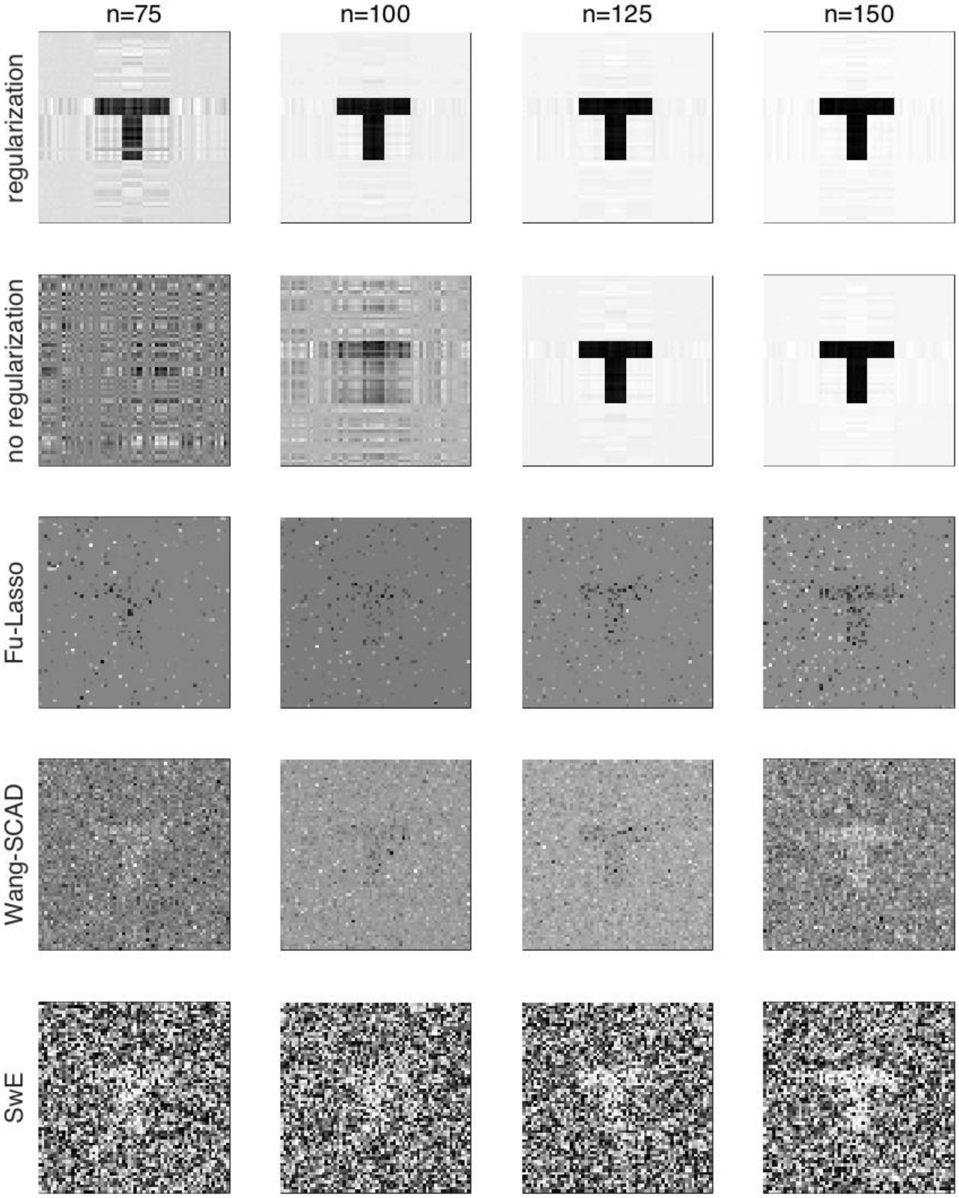

Figure 3. Comparison of the tensor GEE with and without regularization, the lasso regularized vector GEE (Fu (2003) “Fu-Lasso”), the SCAD regularized vector GEE (Wang, Zhou and Qu (2012) “Wang-SCAD”), and the sandwich estimator (Guillaume et al. (2014) “SwE”). The sample size n varies and $m = 4$. The matrix covariate is of size 64×64, and the true signal shape is “T-shape.”

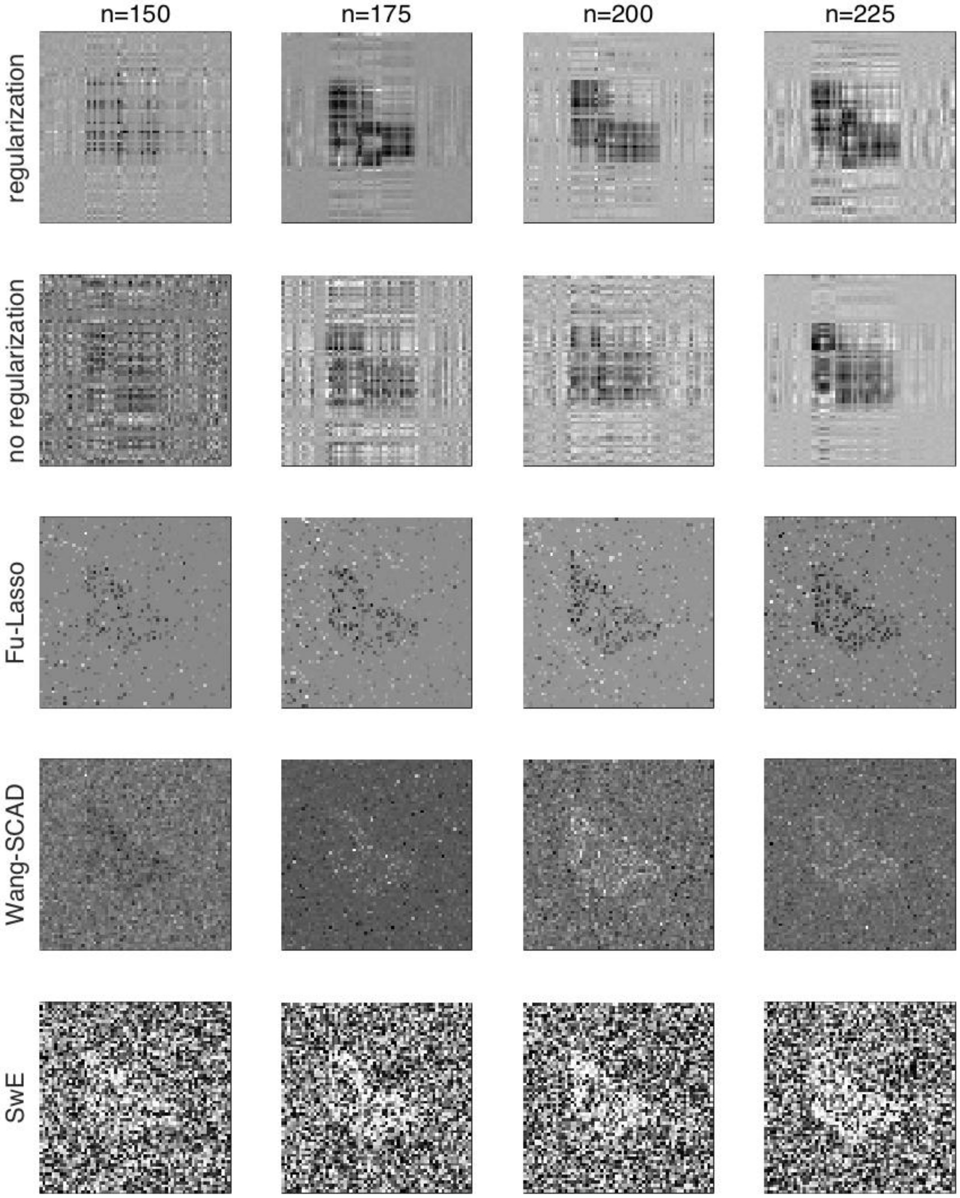

Figure 4. Comparison of the tensor GEE with and without regularization, the lasso regularized vector GEE (Fu, 2003, “Fu-Lasso”), the SCAD regularized vector GEE (Wang, Zhou and Qu, 2012, “Wang-SCAD”), and the sandwich estimator (Guillaume et al., 2014, “SwE”). The sample size n varies and $m = 4$. The matrix covariate is of size 64×64, and the true signal shape is “butterfly.”

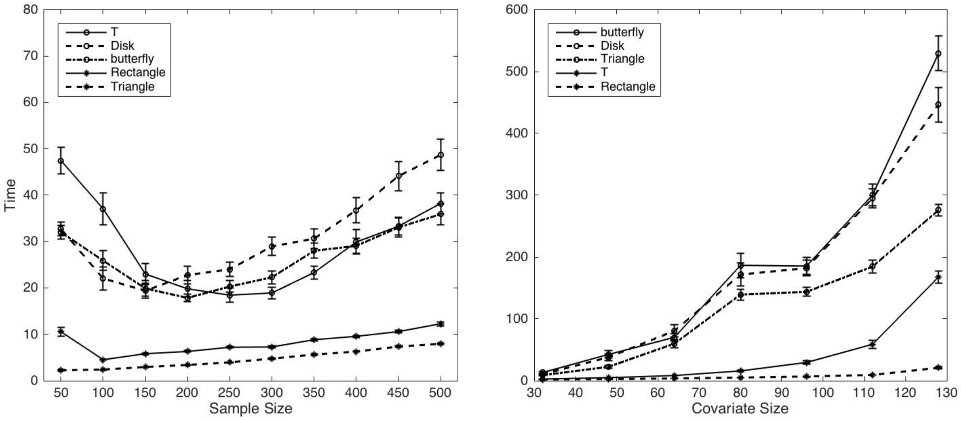

Figure 5. Computation time (in seconds) of the tensor GEE with varying sample sizes and image dimensions for various signal shapes.

based on 100 data replications for various signal shapes. Overall, we find that the computation time of our method is reasonable.

5. Real-Data Analysis

5.1. Alzheimer's disease

AD is a progressive and irreversible neurodegenerative disorder and the leading form of dementia in elderly subjects. It is characterized by gradual impairment of cognitive and memory functions, and has been projected to quadruple in terms of prevalence by the year 2050 (Brookmeyer et al. (2007)). Amnestic MCI is a prodromal stage to Alzheimer's disease, and individuals with MCI convert to AD at an annual rate that can reach as high as 15% (Petersen et al. (1999)). There is a pressing need for accurate and early diagnoses of AD and MCI, as well as for monitoring the disease progression. The data we analyze here are obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI). The data set contains observations on $n = 88$ MCI subjects with longitudinal MRI images of white matter at baseline, 6-month, 12-month, 18-month, and 24-month intervals ($m = 5$). The data set also contains recordings of participants' MMSE scores. This score measures the orientation to time and place, immediate and delayed recall of three words, attention and calculations, language, and visuoconstructional functions (Folstein, Folstein and McHugh (1975)), and it is our response variable. All MRI images have been preprocessed using the preprocessing protocol given in Zhang, Shen and Alzheimer's Disease Neuroimaging Initiative (2012). We recognize the importance of preprocessing in an imaging analysis. Thus, we conduct our analysis *after* proper preprocessing, and have two objectives. The first is to predict future clinical scores based on the data at previous time points. Here, the goal is not to use an MRI to replace a cognitive test, but instead to better understand the association between brain structure and cognition as the disease progress. The second goal is to identify brain subregions that are highly relevant to the disorder so as to better understand the disease pathology. We fit the proposed tensor GEE to these data. The rank is fixed at three, because this has been shown to provide a reasonable trade-off between dimension reduction and model flexibility (Zhou, Li and Zhu (2013)).

5.2. Prediction and disease prognosis

By averaging consecutive time points, we first downsize the original 256-dimensional MRI images to a smaller dimension (32, 64, and 128 dimensions). This downsizing step sacrifices image resolution, but facilitates the computation and reduces the dimensionality. This trade-off is the result of the limited sample size and the very high number of unknown parameters. See Li, Zhou and Li (2018) for an alternative method of image downsizing. Next, we consider two ways of evaluating the prediction accuracy.

We first use the data from the early months to predict the "future" cognitive outcome in the last month of scans. This evaluation scheme is useful to understanding the progression of the disease, and is often used in longitudinal imaging analyses; for example, see Zhang, Shen and Alzheimer's Disease Neuroimaging Initiative (2012). Specifically, we fit the tensor GEE using the data on all subjects from the baseline to the 12-month scans, and use the prediction of the MMSE at 18 months to select the tuning parameter. With the selected tuning parameter, we then refit the model using the data from the baseline to the 18-month scans, and then evaluate the prediction accuracy of all subjects using the "future" MMSE score at 24 months, based on the root mean squared error (RMSE), $\{\sum_{i=1}^{n} n^{-1}(Y_{im} - \hat{Y}_{im})^2\}^{1/2}$. Table 2 summarizes the results, which show that the MRI images of three different sizes yield similar results. The best RMSE achieved by our tensor GEE is 2.147 under an AR(1) working correlation structure, the SCAD penalty, and the downsized image dimension $32 \times 32 \times 32$. This is only slightly worse than the best reported RMSE of 2.035 in Zhang, Shen and Alzheimer's Disease Neuroimaging Initiative (2012). Note that Zhang, Shen and Alzheimer's Disease Neuroimaging Initiative (2012) used multiple imaging

Table 2. Prediction of the MMSE score at a "future" time for all subjects.

Working Correlation	Independence	Equicorrelated	AR(1)	Unstructured
	Image dimesion 32 × 32 × 32			
regularization (Lasso)	2.460	2.349	2.270	2.570
regularization (SCAD)	2.324	2.202	**2.147**	2.674
no regularization	2.526	2.427	2.429	2.628
	Image dimesion 64 × 64 × 64			
regularization (Lasso)	2.364	**2.153**	2.245	2.771
regularization (SCAD)	2.627	2.517	2.659	2.924
no regularization	4.490	4.154	4.776	3.749
	Image dimesion 128 × 128 × 128			
regularization (Lasso)	2.369	2.315	**2.293**	2.702
regularization (SCAD)	2.815	2.874	3.663	3.037
no regularization	6.805	5.008	4.036	7.979

modalities and additional biomarkers, which are supposed to improve the prediction accuracy, whereas we used just one imaging modality.

We next consider a leave-one-out cross-validation evaluation, which is useful to understanding the generalization capability across different individuals. Specifically, we use all scans of a single subject as the testing set, and fit the tensor GEE on the remaining data as the training set. We tune the regularization parameter through five-fold cross-validation on the training set. We evaluate the prediction accuracy using the RMSE of the predicted MMSE score, averaged across all months for the test subject. Table 3 reports both the mean and standard deviation (in parentheses) of the RMSE, averaged across all subjects. The best RMSE achieved by our tensor GEE is 3.172, again under an AR(1) working correlation structure, the SCAD penalty, and the downsized image dimension $32 \times 32 \times 32$. This is slightly worse than the best RMSE in Table 2, as expected. At the same time, the two tables exhibit a consistent pattern in that the tensor GEE with regularization outperforms that without regularization.

5.3. Region selection

Next, we investigate brain region selection using the regularized tensor GEE. We apply both the lasso and the SCAD penalties. Owing to the graphical similarity of the results, we report the SCAD estimate only. Figure 6 shows the estimate (marked in red) overlaid on an image of an arbitrarily chosen subject from three views (top, side, and bottom). The identified anatomical regions correspond mainly to the cerebral cortex, part of the temporal lobe, the parietal lobe, and the frontal lobe (Braak and Braak (1991); Desikan et al. (2009);

Table 3. Prediction of the MMSE score of a "future" subject at all times using leave-one-out cross-validation.

Working Correlation	Independence	Equicorrelated	AR(1)	Unstructured
	Image dimesion $32 \times 32 \times 32$			
regularization (Lasso)	3.225 (1.851)	3.404(1.710)	3.272 (1.886)	3.982(2.557)
regularization (SCAD)	3.250 (1.928)	3.392(1.624)	**3.172(1.551)**	3.790(2.634)
no regularization	4.271 (2.936)	4.063(2.571)	4.415 (3.186)	4.492(3.294)
	Image dimesion $64 \times 64 \times 64$			
regularization (Lasso)	3.381 (1.949)	3.825(1.973)	3.333 (1.877)	3.645(1.955)
regularization (SCAD)	**3.282(1.761)**	3.414(1.723)	3.592 (2.166)	3.873(1.937)
no regularization	4.670 (2.179)	5.025(2.851)	4.681 (1.870)	4.452(2.353)
	Image dimesion $128 \times 128 \times 128$			
regularization (Lasso)	3.409 (1.743)	3.968(1.850)	**3.296(1.983)**	3.301(1.574)
regularization (SCAD)	4.123 (2.326)	3.929(1.895)	3.696 (2.065)	3.780(1.862)
no regularization	5.605 (3.716)	5.532(4.713)	5.654 (2.969)	6.037(9.493)

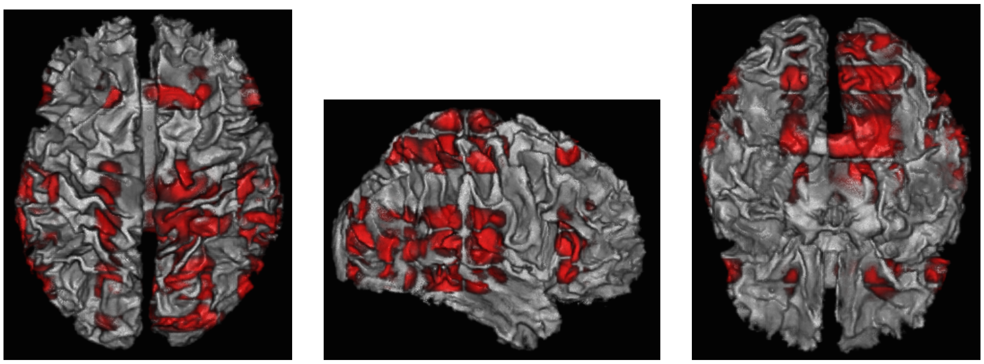

Figure 6. The ADNI data: regularized estimate overlaid on a randomly selected subject.

Yao et al. (2012)). With AD, patients experience significant widespread damage over the brain, causing shrinkage of brain volume (Yao et al. (2012); Harasty et al. (1999)) and a thinning of cortical thickness (Desikan et al. (2009); Yao et al. (2012)). The affected brain regions include those involved in controlling language (Broca's area) (Harasty et al. (1999)), reasoning (superior and inferior frontal gyri) (Harasty et al. (1999)), part of the sensory area (primary auditory cortex, olfactory cortex, insula, and operculum) (Braak and Braak (1991); Lee et al. (2013)), somatosensory association area (Yao et al. (2012); Tales et al. (2005); Mapstone, Steffenella and Duffy (2003)), memory loss (hippocampus) (den Heijer et al. (2010)), and motor function (Buchman and Bennett (2011)). However, these regions are affected at different stages of AD, indicating the capability of the proposed method to locate brain atrophy as the disease progresses. For example, damage to the hippocampus, which is highly correlated with memory loss, is commonly detected at the earliest stage of the disease. Damage to regions related to language, communication, and motor functions is normally detected at the later stages of the disease. The fact that our findings are consistent with the results reported in previous studies, particularly in longitudinal studies, demonstrates the efficacy of our proposed method in identifying correct biomarkers that are closely related to AD and MCI.

6. Discussion

We have proposed a tensor GEE for longitudinal imaging analyses. With the increasing availability of longitudinal image data and the relative paucity of effective analytical solutions, our proposed method provides a timely and useful response. Specifically, it combines the GEE approach for handling longitudinal correlations and a low-rank decomposition for significant dimension reduction and tensor structure preservation. The proposed algorithm scales with the image data size and is easy to implement using existing statistical software. Simulation studies and a real-data analysis show the potential of our method for both signal recovery and outcome prediction.

Our method involves two specifications: a working correlation structure and a working rank for the tensor coefficient. We have examined how to select these values in practice, as well as the potential consequences of their misspecification. For the working correlation structure, we have shown that, asymptotically, the tensor GEE estimator remains consistent, even if the correlation structure is misspecified. In practice, our numerical investigation suggests that a simple independent working correlation is probably preferable when the sample size is limited, whereas a data adaptive choice of a suitable working correlation is preferable for larger samples. This is useful, because many multicenter large-scale longitudinal imaging data sets, such as ADNI, becoming available. The same correlation structure selection problem is also encountered in the classical vector-valued GEE; for further discussion, see Pan and Connett (2002). For the working rank of the tensor CP decomposition, we again show that, asymptotically, the BIC criterion under the independent correlation structure selects the true rank with probability approaching one, even if this correlation structure is misspecified. In practice, the rank selection reflects a *bias–variance trade-off*. When the selected rank is smaller than the true rank, the resulting estimator is biased, but involves fewer unknown parameters, and thus is less variable. When the selection is greater than the true rank, the estimator becomes unbiased, but is also more variable with a larger number of parameters. In general, our findings suggest that the reduced-rank structure provides a reasonable *approximation* of the coefficient tensor.

Numerous problems remain open and warrant further research. The first is the rank selection, including the selection consistency for a more general family of models, its convergence rate, and its selection under a diverging dimension. Note that this problem is not yet fully solved, even in the context of tensor predictor regression on a single image observation per subject, and is particularly challenging. The second is to conduct an asymptotic study of our tensor GEE with a diverging dimension. This is important to improve our understanding of the properties of the tensor GEE. We have obtained some preliminary results, extending those of the vector GEE (Wang (2011)) to the tensor version. However, the asymptotic properties under a diverging dimension combine with the diverging rank selection, and therefore warrants further research.

Supplementary Material

The proofs of the main theorems and some technical lemmas are available in the online Supplementary Material. A Matlab software package is available upon request.

Acknowledgments

The work is partially supported by NSF grants DMS-1310319 and DMS-1613137, and NIH grants HG006139, GM105785, GM53275, and AG034570.

References

Aston, J. A. D., Pigoli, D. and Tavakoli, S. (2017). Tests for separability in nonparametric covariance operators of random surfaces. *The Annals of Statistics* **45**, 1431–1461.

Balan, R. M. and Schiopu-Kratina, I. (2005). Asymptotic results with generalized estimating equations for longitudinal data. *The Annals of Statistics* **33**, 522–541.

Braak, H. and Braak, E. (1991). Neuropathological stageing of Alzheimer-related changes. *Acta Neuropathologica* **82**, 239–259.

Brookmeyer, R., Johnson, E., Ziegler-Graham, K. and Arrighi, H. M. (2007). Forecasting the global burden of Alzheimers disease. *Alzheimer's & Dementia* **3**, 186 – 191.

Buchman, A. and Bennett, D. (2011). Loss of motor function in preclinical Alzheimer's disease. *Expert Review Neurotherapeutics* **11**, 665–676.

Davatzikos, C., Xu, F., An, Y., Fan, Y. and Resnick, S. M. (2009). Longitudinal progression of Alzheimer's-like patterns of atrophy in normal older adults: the spare-ad index. *Brain* **132**, 2026–2035.

den Heijer, T., van der Lijn, F., Koudstaal, P. J., Hofman, A., van der Lugt, A., Krestin, G. P., Niessen, W. J. and Breteler, M. M. B. (2010). A 10-year follow-up of hippocampal volume on magnetic resonance imaging in early dementia and cognitive decline. *Brain* **133**, 1163–1172.

Desikan, R., Cabral, H., Hess, C., Dillon, W., Salat, D., Buckner, R., Fischl, B. and Initiative, A. D. N. (2009). Automated MRI measures identify individuals with mild cognitive impairment and Alzheimer's disease. *Brain* **132**, 2048–2057.

Fan, J. and Li, R. (2001). Variable selection via nonconcave penalized likelihood and its oracle properties. *Journal of the American Statistical Association* **96**, 1348–1360.

Fan, J. and Li, R. (2004). New estimation and model selection procedures for semiparametric modeling in longitudinal data analysis. *Journal of the American Statistical Association* **99**, 710–723.

Folstein, M. F., Folstein, S. E. and McHugh, P. R. (1975). Mini-mental state: A practical method for grading the cognitive state of patients for the clinician. *Journal of Psychiatric Research* **12**, 189 –198.

Fu, W. J. (2003). Penalized estimating equations. *Biometrics* **59**, 126–132.

Guillaume, B., Hua, X., Thompson, P. M., Waldorp, L., Nichols, T. E., Initiative, A. D. N. and et al. (2014). Fast and accurate modelling of longitudinal and repeated measures neuroimaging data. *NeuroImage* **94**, 287–302.

Harasty, J. A., Halliday, G. M., Kril, J. J. and Code, C. (1999). Specific temporoparietal gyral atrophy reflects the pattern of language dissolution in Alzheimer's disease. *Brain* **122**, 675–686.

Hinrichs, C., Singh, V., Xu, G. and Johnson, S. C. (2011). Predictive markers for AD in a multi-modality framework: An analysis of MCI progression in the ADNI population. *NeuroImage* **55**, 574 – 589.

Kolda, T. G. and Bader, B. W. (2009). Tensor decompositions and applications. *SIAM Review* **51**, 455–500.

Lee, T. M., Sun, D., Leung, M.-K., Chu, L.-W. and Keysers, C. (2013). Neural activities during affective processing in people with Alzheimer's disease. *Neurobiology of Aging* **34**, 706–715.

Li, B. (1997). On the consistency of generalized estimating equations. In: Selected Proceedings of the Symposium on Estimating Functions (Athens, GA, 1996). Vol. 32 of IMS Lecture Notes Monograph Series. Hayward, CA: Institute of Mathematical Statistics, pp. 115–136.

Li, X., Zhou, H. and Li, L. (2018). Tucker tensor regression and neuroimaging analysis. *Statistics in Biosciences* **10**, 520–545.

Li, Y., Gilmore, J. H., Shen, D., Styner, M., Lin, W. and Zhu, H. (2013). Multiscale adaptive generalized estimating equations for longitudinal neuroimaging data. *NeuroImage* **72**, 91 – 105.

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

Mapstone, M., Steffenella, T. and Duffy, C. (2003). A visuospatial variant of mild cognitive impairment: getting lost between aging and AD. *Neurology* **60**, 802–808.

McEvoy, L. K., Holland, D., Hagler, D. J., Fennema-Notestine, C., Brewer, J. B. and Dale, A. M. (2011). Mild cognitive impairment: Baseline and longitudinal structural MR imaging measures improve predictive prognosis. *Radiology* **259**, 834–843.

Misra, C., Fan, Y. and Davatzikos, C. (2009). Baseline and longitudinal patterns of brain atrophy in MCI patients, and their use in prediction of short-term conversion to AD: Results from ADNI. *NeuroImage* **44**, 1415 – 1422.

Ni, X., Zhang, D. and Zhang, H. H. (2010). Variable selection for semiparametric mixed models in longitudinal studies. *Biometrics* **66**, 79–88.

Pan, W. (2001). Akaike's information criterion in generalized estimating equations. *Biometrics* **57**, 120–125.

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

Petersen, R., Smith, G., Waring, S., Ivnik, R., Tangalos, E. and Kokmen, E. (1999). Mild cognitive impairment: clinical characterization and outcome. *Archives of Neurology* **56**, 303–308.

Prentice, R. L. and Zhao, L. P. (1991). Estimating equations for parameters in means and covariances of multivariate discrete and continuous responses. *Biometrics* **47**, 825–839.

Qu, A., Lindsay, B. G. and Li, B. (2000). Improving generalised estimating equations using quadratic inference functions. *Biometrika* **87**, 823–836.

Rao, C. R. and Mitra, S. K. (1971). *Generalized Inverse of Matrices and its Applications.* John Wiley & Sons, Inc., New York-London-Sydney.

Raskutti, G. and Yuan, M. (2016). Convex regularization for high-dimensional tensor regression. *arXiv preprint arXiv:1512.01215,* 639.

Skup, M., Zhu, H. and Zhang, H. (2012). Multiscale adaptive marginal analysis of longitudinal neuroimaging data with time-varying covariates. *Biometrics* **68**, 1083–1092.

Song, P. X.-K., Jiang, Z., Park, E. and Qu, A. (2009). Quadratic inference functions in marginal models for longitudinal data. *Statistics in Medicine* **28**, 3683–3696.

Sun, W. and Li, L. (2017). Sparse tensor response regression and neuroimaging analysis. *The Journal of Machine Learning Research* **18**, 4908–4944.

Sun, W., Lu, J., Liu, H. and Cheng, G. (2017). Provable sparse tensor decomposition. *Journal of the Royal Statistical Society, Series B (Statistical Methodology)* **79**, 899–916.

Tales, A., Haworth, J., Nelson, S., J. Snowden, R. and Wilcock, G. (2005). Abnormal visual search in mild cognitive impairment and Alzheimer's disease. *Neurocase* **11**, 80–84.

Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society. Series B. (Statistical Methodology)* **58**, 267–288.

Wang, L. (2011). GEE analysis of clustered binary data with diverging number of covariates. *The Annals of Statistics* **39**, 389–417.

Wang, L., Zhou, J. and Qu, A. (2012). Penalized generalized estimating equations for high-dimensional longitudinal data analysis. *Biometrics* **68**, 353–360.

Xie, M. and Yang, Y. (2003). Asymptotics for generalized estimating equations with large cluster sizes. *The Annals of Statistics* **31**, 310–347.

Xue, L., Qu, A. and Zhou, J. (2010). Consistent model selection for marginal generalized additive model for correlated data. *Journal of the American Statistical Association* **105**, 1518–1530.

Yao, Z., Hu, B., Liang, C., Zhao, L., Jackson, M. and the Alzheimer's Disease Neuroimaging Initiative (2012). A longitudinal study of atrophy in amnestic mild cognitive impairment and normal aging revealed by cortical thickness. *PLoS One* **7**, e48973.

Zhang, D., Shen, D. and Alzheimer's Disease Neuroimaging Initiative (2012). Predicting future clinical changes of MCI patients using longitudinal and multimodal biomarkers. *PLoS One* **7**, e33182.

Zhang, D., Wang, Y., Zhou, L., Yuan, H., Shen, D. and the Alzheimers Disease Neuroimaging Initiative (2011). Multimodal classification of Alzheimer's disease and mild cognitive impairment. *Neuroimage* **55**, 856–867.

Zhou, H. and Li, L. (2014). Regularized matrix regression. *Journal of the Royal Statistical Society. Series B (Statistical Methodology)* **76**, 463–483.

Zhou, H., Li, L. and Zhu, H. (2013). Tensor regression with applications in neuroimaging data analysis. *Journal of the American Statistical Association* **108**, 540–552.

Department of Statistics, North Carolina State University, Raleigh, NC 27697, USA.

E-mail: xzhang23@ncsu.edu

Division of Biostatistcs, University of California, Berkeley, CA 94720, USA.

E-mail: lexinli@berkeley.edu

Department of Biostatistics, University of California, Los Angeles, CA 90095, USA.

E-mail: huazhou@ucla.edu

School of Statistics and Management, Shanghai University of Finance and Economics, Shanghai, 200433, China.

E-mail: yqzhou1991@hotmail.com

Department of Radiology, University of North Carolina, Chapel Hill, NC 27599, USA.

E-mail: dinggang_shen@med.unc.edu

(Received April 2017; accepted February 2018)