Received: 5 February 2020 | Revised: 17 April 2020 | Accepted: 13 May 2020

DOI: 10.1002/bimj.202000039

Biometrical Journal

RESEARCH PAPER

Parametric mode regression for bounded responses

Haiming Zhou[1] | **Xianzheng Huang**[2] | **for the Alzheimer's Disease Neuroimaging Initiative***

[1] Department of Statistics and Actuarial Science, Northern Illinois University, DeKalb, IL, USA

[2] Department of Statistics, University of South Carolina, Columbia, SC, USA

Correspondence
Haiming Zhou, Department of Statistics and Actuarial Science, Northern Illinois University, DeKalb, IL 60115, USA.
Email: zhouh@niu.edu

*Data used in preparation of this article were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (adni.loni.usc.edu). As such, the investigators within the ADNI contributed to the design and implementation of ADNI and/or provided data but did not participate in analysis or writing of this report. A complete listing of ADNI investigators can be found at http://adni.loni.usc.edu/wp-content/uploads/how_to_apply/ADNI_Acknowledgement_List.pdf.

Funding information
ADNI, Grant/Award Numbers: U01 AG024904, W81XWH-12-2-0012

This article has earned an open data badge "**Reproducible Research**" for making publicly available the code necessary to reproduce the reported results. The results reported in this article were reproduced partially for data confidentiality reasons and due to their computational complexity.

Abstract

We propose new parametric frameworks of regression analysis with the conditional mode of a bounded response as the focal point of interest. Covariate effects estimation and prediction based on the maximum likelihood method under two new classes of regression models are demonstrated. We also develop graphical and numerical diagnostic tools to detect various sources of model misspecification. Predictions based on different central tendency measures inferred using various regression models are compared using synthetic data in simulations. Finally, we conduct regression analysis for data from the Alzheimer's Disease Neuroimaging Initiative to demonstrate practical implementation of the proposed methods. Supporting Information that contain technical details and additional simulation and data analysis results are available online.

KEYWORDS

beta distribution, generalized biparabolic distribution, linear predictor, link function, maximum likelihood

1 | INTRODUCTION

The statistical models and methodology presented in this article are motivated by the Alzheimer's Disease Neuroimaging Initiative (ADNI) launched in 2003 and led by Principal Investigator Michael W. Weiner, MD. It is an ongoing study with a public–private partnership in the United States and Canada that gathers and analyzes thousands of subjects' brain scans, genetic profiles, and biomarkers in blood and cerebrospinal fluid. The main goal of ADNI is to understand relationships among the clinical, cognitive, imaging, genetic and biochemical biomarker characteristics of the entire spectrum

 © 2020 WILEY-VCH Verlag GmbH & Co. KGaA, Weinheim

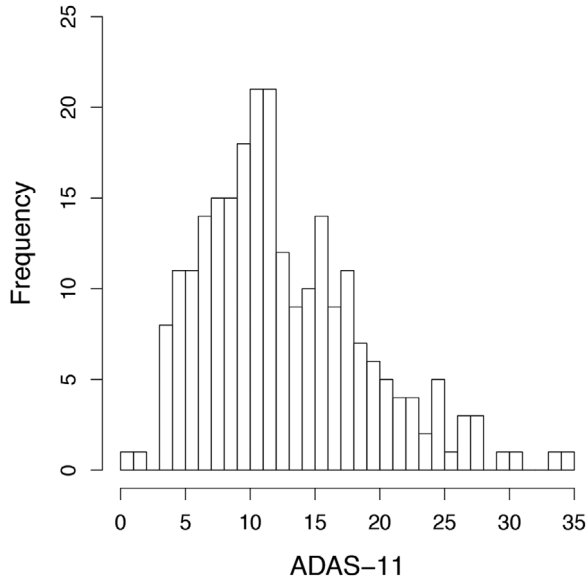

FIGURE 1 Histogram of ADAS-11 scores at month-12 of subjects in ADNI Phase 1 study

of Alzheimer's diseases (AD). Ultimately, the hope is to achieve early detection of AD in preparation for early intervention of the disease progression and also to help recruiting appropriate individuals in clinical trials. Clinical outcomes for assessing one's cognitive function in ADNI are bounded scores from well-established neuropsychological tests, such as the Alzheimer's disease assessment scale (ADAS; Kueper, Speechley, & Montero-Odasso, 2018; Rosen, Mohs, & Davis, 1984), mini-mental state examination (Tombaugh & McIntyre, 1992), and Rey auditory verbal learning test (Schmidt, 1996). Distributions of these test scores from the ADNI cohort are typically heavy-tailed and skewed. As an example, Figure 1 presents histogram of the ADAS-cognition subscale scores, also referred to as ADAS-11, of subjects at month 12 who were diagnosed with late mild cognitive impairment (LMCI) when they entered the ADNI Phase 1 study.

In order to effectively reveal the association between one's cognitive skill and potential influential biomarkers, we formulate parametric mode regression models tailored for heavy-tailed, skewed, and bounded response data, with efficient prediction as our goal of statistical inference besides identifying influential biomarkers for AD. Under the nonparametric framework, Wang, Chen, Cai, Shen, and Huang (2017) carried out mode regression analysis of ADNI data to predict cognitive impairment using neuroimaging data. They noted that mean regression analysis for studying the association between the cognitive assessment of an individual and the individual's neuroimaging features failed to yield scientifically meaningful results due to the heavy-tailed and skewed noise presented in data typically arising in this application. In addition to biomedical applications, mode regression for association study has been routinely used in econometrics (Ho, Damien, & Walker, 2017; Kemp & Silva, 2012; Lee, 1989, 1993), astronomy (Bamford et al., 2008), and traffic engineering (Einbeck & Tutz, 2006). Kemp and Silva (2012) argued that the mode is the most intuitive measure of central tendency for positively skewed data found in many econometric applications such as wages, prices, and expenditures. More generally, the conditional mode serves as a more informative summary for associations between a response Y and covariates $\mathbf{X}$ than the conditional mean or median when the distribution of Y given $\mathbf{X}$ is heavy-tailed or skewed. When comparing with predictions based on conditional means or medians, predictions based on conditional modes can provide more meaningful estimated outcomes. Yao and Li (2014) showed that, when the interval width is fixed, a mode-based prediction interval tends to have a higher coverage probability than a mean-based prediction interval.

Most existing works on mode regression involve nonparametric components (Chen, Genovese, Tibshirani, & Wasserman, 2016; Liu, Zhang, Zhao, & Lv, 2013; Yang & Yang, 2014; Yao & Li, 2014; Zhang, Zhao, & Liu, 2013; Zhao, Zhang, Liu, & Lv, 2014). These nonparametric and semiparametric approaches are developed under the frequentist framework. A few developments under the Bayesian framework include Yu and Aristodemou (2012) and Ho et al. (2017). Even though nonparametric methods and semiparametric methods can protect against misleading inference caused by inadequate parametric assumptions, often at the price of low statistical efficiency, it is not unreasonable to believe that an inference procedure may still provide reliable inference for the mode of a distribution even when certain aspects, such as tails, of the distribution are not well estimated by this procedure (Hall, 1992; Zhou & Huang, 2019). Hence, with the potential gain

in efficiency, parametric regression models can be useful in studying the association between a response and covariates via inferring the conditional mode.

Assuming a unimodal conditional distribution for the response, we formulate in Section 2 two new classes of mode regression models for a bounded response. Bounded response data are ubiquitous in practice, with the ADAS-11 score as one example. Other examples include rates or proportions, such as a disease prevalence, the fraction of household income spent on food, and the proportion of food and hygienic waste in residential solid waste. Although, technically, one can often map a bounded response to a new response whose support is the entire real line, say, via a logit transformation for a rate response, and then carry out regression analysis on the new response, it is practically more appealing to directly study the association between the original response and covariates. This practical consideration, along with the observation that many proportion responses encountered in practice are asymmetrically distributed, motivated the beta regression model proposed by Ferrari and Cribari-Neto (2004), with the mean of the beta distribution depending on covariates. Smithson and Verkuilen (2006) followed a similar strategy to formulate a regression model for a response bounded on [0, 1], where they specified a mean model and a variance model as functions of two sets of covariates separately. They later generalized this regression model by using a mixture of beta distributions (Verkuilen & Smithson, 2012). Starting from a beta regression model, Guolo and Varin (2014) incorporated the serial dependence between responses via a Gaussian copula to model time series data bounded on the unit interval. Following the construction of their regression models, all the aforementioned works carry out frequentist inference, mostly based on maximum likelihood. Under the Bayesian inferential framework, Bayes, Bazán, and García (2012) replaced the beta distribution with the beta rectangular distribution, defined as the mixture of a beta distribution and a uniform distribution, to achieve more robustness to outliers of a proportion response. Figueroa-Zúñiga, Arellano-Valle, and Ferrari (2013) introduced mixed Bayesian regression models by incorporating random effects in the linear predictor when specifying the mean function. Also considering Bayesian regression analysis for bounded data, Migliorati, Di Brisco, and Ongaro (2018) proposed a flexible beta distribution for the response given covariates based on a special mixture of two beta distributions to balance between flexibility and tractability. Unlike all the above regression models which focus on inferring the conditional mean of a bounded response, Bayes, Bazán, and De Castro (2017) developed quantile regression models for bounded responses built upon on beta distributions. Barrientos, Jara, and Quintana (2017) took on a fully nonparametric Bayesian approach to model the covariates-dependent distribution of a bounded response. One major feature of our work that distinguishes it from these existing works on regression analysis for bounded data is that the conditional mode is the focal point of inference. This very key feature motivates our construction of the regression models presented in Section 2

Following the model formulation, we outline maximum likelihood estimation of parameters in these models in Section 2. In Section 3, we propose graphical and numerical diagnostics methods for detecting various sources of model misspecification when one draws inference based on an assumed model in the two proposed families. Section 4 presents simulation studies where we carry out mode regression analysis using these assumed models based on data generated from models that may or may not belong to the two families. In these simulation experiments, we report maximum likelihood estimates (MLEs) for covariate effects, operating characteristics of the proposed diagnostics methods, and prediction intervals constructed based on the proposed mode regression models. In Section 5, we carry out mean and mode regression analysis for a dataset from ADNI. Finally, we summarize contributions of this work and discuss follow-up research in Section 6.

2 | TWO FAMILIES OF REGRESSION MODELS AND MAXIMUM LIKELIHOOD ESTIMATION

2.1 | Regression models

Without loss of generality, we assume from now on that the response variable has support on [0,1], since any other bounded support can be rescaled to the unit interval. Inspired by the existing mean and quantile regression models for bounded data originating from beta regression, we first formulate a beta mode regression model for a bounded response. Recall that, for a random variable V that follows a beta distribution, its probability density function (pdf) is given by

$$f_{\text{beta}}(v;\alpha_1,\alpha_2) = \frac{\Gamma(\alpha_1+\alpha_2)}{\Gamma(\alpha_1)\Gamma(\alpha_2)} v^{\alpha_1-1}(1-v)^{\alpha_2-1}, \quad \text{for } v \in [0,1], \tag{2.1}$$

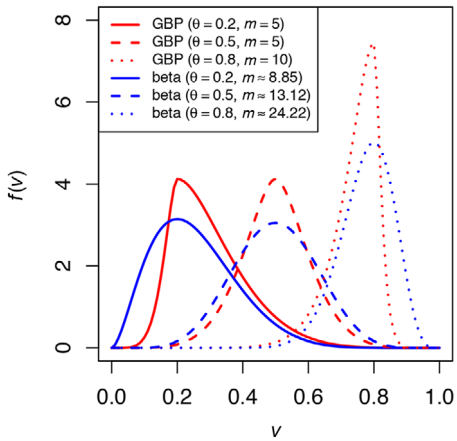

FIGURE 2 Probability density functions of GBP distributions (red lines) with $(\theta, m) = (0.2, 5)$ (solid line), $(0.5, 5)$ (dashed line), and $(0.8, 10)$ (dotted line), respectively, and beta density curves (blue lines) with the same mode and variance as those of the GBP densities depicted in the same line type

where $\Gamma(t)$ is the gamma function and α_1 and α_2 are two positive shape parameters. When $\alpha_1, \alpha_2 > 1$, there is a unique mode for the beta distribution given by $\theta = (\alpha_1 - 1)/(\alpha_1 + \alpha_2 - 2)$. Directly including the mode in the parameterization makes it more convenient to draw inference for the mode. For this reason, we set $\alpha_1 = 1 + m\theta$ and $\alpha_2 = 1 + m(1 - \theta)$, where $m > 0$. This parameterization not only signifies the parameter of central interest, θ, but also makes α_1 and α_2 larger than one, ensuring the existence of a unique mode. The variance of the beta distribution under this parameterization is $(1 + m\theta)\{1 + m(1 - \theta)\}/\{(2 + m)^2(3 + m)\}$, suggesting a smaller variance as m increases.

With the mode as our choice of central tendency measure for the bounded response, it is desirable to include the mode as one of the canonical parameters in the response distribution without additional reparameterization as we do above for a beta distribution. To the best of our knowledge, the family of generalized biparabolic distributions (GBP; García, Pérez, & Rambaud, 2009) is the only named distribution family that, first, is defined on a bounded support, second, includes symmetric and asymmetric distributions, and third, has the mode as the sole location parameter appearing in the pdf. More specifically, if V follows a GBP distribution on the support $[0, 1]$, its pdf is given by

$$f_{\mathrm{GBP}}(v; \theta, m) = \frac{(2m + 1)(m + 1)}{(3m + 1)} d^m (2 - d^m), \tag{2.2}$$

where m is a positive shape parameter, and $d = I(0 < v \leq \theta)v/\theta + I(\theta < v \leq 1)(1 - v)/(1 - \theta)$, in which $\theta \in (0, 1)$ is the mode of the distribution, and $I(\cdot)$ is the indicator function. A larger m leads to a GBP distribution more concentrated around the mode with a smaller variance. Figure 2 depicts three GBP density functions, in comparison with three beta density functions that share the same mode and variance as the corresponding depicted GBP distributions. This figure shows the general pattern that, with the mode and variance fixed, a GBP density displays a sharper drop toward zero on both sides of the mode than that for a beta density.

To complete the formulation of a regression model, we assume that, given $\mathbf{X}$, the mode of Y relates to a linear predictor $\eta(\mathbf{X}) = \beta_0 + \boldsymbol{\beta}_1^{\mathrm{T}}\mathbf{X}$ via a link function $g(t)$, that is,

$$\mathrm{Mode}(Y|\mathbf{X}) = \theta(\mathbf{X}) = g\{\eta(\mathbf{X})\}. \tag{2.3}$$

Commonly employed link functions include logit, probit, log–log, and complementary log–log. To this end, we have two regression models for Y once a link function is chosen, written succinctly as

$$Y|\mathbf{X} \sim \mathrm{beta}(1 + m\theta(\mathbf{X}),\ 1 + m\{1 - \theta(\mathbf{X})\}),$$

$$Y|\mathbf{X} \sim \mathrm{GBP}(\theta(\mathbf{X}),\ m),$$

which are henceforth referred to as the beta mode model and the GBP mode model, respectively. Note that neither regression model is a generalized linear model (McCullagh, 2019) because neither (2.1) nor (2.2) is in the form of a

density corresponding to an exponential dispersion distribution (Jørgensen, 1987) when they are parameterized via θ and m as described above. Both regression models allow heteroscedasticity in the sense that the conditional variance of Y depends on covariates. For example, under the beta mode model, $\mathrm{Var}(Y|\mathbf{X}) = \{1 + m\theta(\mathbf{X})\}[1 + m\{1 - \theta(\mathbf{X})\}]/\{(2 + m)^2(3 + m)\}$.

2.2 | Maximum likelihood estimation

Given a random sample of size n, $\mathcal{D} = \{(Y_i, \mathbf{X}_i), i = 1, \dots, n\}$, the log-likelihood function associated with a beta mode model is

$$\ell_{\text{beta}}(\boldsymbol{\Omega}; \mathcal{D}) = n \log \Gamma(2 + m) - \sum_{i=1}^{n} \log \left(\Gamma\{1 + m\theta(\mathbf{X}_i)\}\Gamma[1 + m\{1 - \theta(\mathbf{X}_i)\}]\right)$$
$$+ m \sum_{i=1}^{n} [\theta(\mathbf{X}_i) \log Y_i + \{1 - \theta(\mathbf{X}_i)\} \log(1 - Y_i)].$$

Maximizing $\ell_{\text{beta}}(\boldsymbol{\Omega}; \mathcal{D})$ with respect to $\boldsymbol{\Omega} = (\boldsymbol{\beta}^{\mathrm{T}}, m)^{\mathrm{T}}$ yields the MLE for $\boldsymbol{\Omega}$ under this model, where $\boldsymbol{\beta} = (\beta_0, \boldsymbol{\beta}_1^{\mathrm{T}})^{\mathrm{T}}$. Because the beta family is an exponential family, the corresponding likelihood function is concave, suggesting the existence of a unique MLE for $\boldsymbol{\Omega}$. Furthermore, regularity conditions required for the MLE to be consistent and asymptotically normal can also be easily verified for this regression model.

When a GBP mode model is assumed, by (2.2), the log-likelihood function is given by

$$\ell_{\text{GBP}}(\boldsymbol{\Omega}; \mathcal{D}) = n \log \left\{ \frac{(2m + 1)(m + 1)}{3m + 1} \right\} + m \sum_{i=1}^{n} \log d_i + \sum_{i=1}^{n} \log \left(2 - d_i^m\right), \tag{2.4}$$

where $d_i = I\{0 < Y_i \le \theta(\mathbf{X}_i)\}\{Y_i/\theta(\mathbf{X}_i)\} + I\{\theta(\mathbf{X}_i) < Y_i < 1\}\{(1 - Y_i)/\{1 - \theta(\mathbf{X}_i)\}$. Maximizing $\ell_{\text{GBP}}(\boldsymbol{\Omega}; \mathcal{D})$ with respect to $\boldsymbol{\Omega}$ yields the MLE for $\boldsymbol{\Omega}$ under the GBP mode regression model. Unlike the beta family, the GBP family is not an exponential family. For simplicity, let us assume m known in (2.2) and focus on the density as a function of θ for now. It can be shown that $\lim_{\theta \to \upsilon^+} (\partial^2/\partial\theta^2) \log f_{\text{GBP}}(\upsilon; \theta, m) = -2m^2/\upsilon^2$, whereas $\lim_{\theta \to \upsilon^-} (\partial^2/\partial\theta^2) \log f_{\text{GBP}}(\upsilon; \theta, m) = -2m^2/(1 - \upsilon)^2$, indicating that the Hessian function is discontinuous at any realization of the distribution except for $\upsilon = 0.5$. It can also be shown that the GBP log-likelihood is concave in a neighborhood of the truth almost surely. Moreover, regularity conditions (Cox & Hinkley, 1979, p. 281) for the consistency of MLE as the maximizer of (2.4) are satisfied for the GBP regression model, but additional conditions needed to establish asymptotic normality for MLE are not.

3 | MODEL DIAGNOSTICS

Basing statistical inference on a specific parametric model raises the concern of model misspecification that can lead to misleading inference results. This concern motivates the diagnosis tools we develop in this section.

3.1 | Graphical diagnosis

Half-normal residual plots with simulated envelopes (Atkinson, 1987) are useful graphical tools for checking the goodness-of-fit of a model with complex response distributions. Let $\hat{\mu}(\boldsymbol{x})$ and $\hat{\sigma}^2(\boldsymbol{x})$ denote the MLEs for the mean and variance of Y given $\mathbf{X} = \boldsymbol{x}$, respectively, resulting from an assumed regression model. Define the absolute standardized residual as $r_i = |Y_i - \hat{\mu}(\mathbf{X}_i)|/\hat{\sigma}(\mathbf{X}_i)$. Given data $\mathcal{D} = \{(Y_i, \mathbf{X}_i), i = 1, \dots, n\}$, the algorithm below describes how to obtain a half-normal residual plot with a simulated envelope.

Step 1 Fit the assumed regression model to data $\mathcal{D}$, calculate the absolute standardized residuals, then order the residuals from smallest to largest, denoted as $\{r_{(i)}, i = 1, \dots, n\}$. Plot $r_{(i)}$ against the half-normal quantile $q_i = \Phi^{-1}\{(i + n - 0.125)/(2n + 0.5)\}$, for $i = 1, \dots, n$, where $\Phi(\cdot)$ is the cumulative distribution function of $N(0, 1)$.

Step 2 For $k = 1, \dots, K$, conditioning on $\mathbf{X}_i$, generate a new response $Y_i^{*(k)}$ from the estimated regression model resulting from *Step 1*, for $i = 1, \dots, n$. This produces new data $\mathcal{D}^{*(k)} = \{(Y_i^{*(k)}, \mathbf{X}_i), i = 1, \dots, n\}$, for $k = 1, \dots, K$.

Step 3 For $k = 1, \dots, K$, fit the assumed regression model to data $\mathcal{D}^{*(k)}$ and calculate the ordered absolute standardized residuals $\{r_{(i)}^{*(k)}, i = 1, \dots, n\}$.

Step 4 Compute $r_i^L = \min_{1\le k\le K}\{r_{(i)}^{*(k)}\}$ and $r_i^U = \max_{1\le k\le K}\{r_{(i)}^{*(k)}\}$, for $i = 1, \dots, n$. Plot points $\{(x, y) : (q_i, r_i^L), i = 1, \dots, n\}$ and $\{(x, y) : (q_i, r_i^U), i = 1, \dots, n\}$ on the same plot obtained in *Step 1* to form the envelope.

Here we set $K = 19$ as suggested by Atkinson (1987). This way, the resulting envelope has a probability approximately equal to 0.95 to cover the ordered residuals $r_{(i)}$ obtained from the original data if the assumed model agrees with the true model. A substantially larger proportion of residuals falling outside the envelope indicates a lack-of-fit of the assumed model.

3.2 | Score tests for model diagnosis

To assess the adequacy of an assumed regression model quantitatively, we develop tests using score functions constructed based on matching moments. The proposed score tests exploit certain moments of the response variable or functions of it that are special in some way so that they are difficult to be estimated well via maximizing a misspecified likelihood function.

When the assumed model is a beta mode model, we construct a bivariate score function based on the following results relating to a beta random variable V:

$$E(\log V) = \psi(\alpha_1) - \psi(\alpha_1 + \alpha_2),$$

$$E(V \log V) = \frac{\alpha_1\{\psi(\alpha_1 + 1) - \psi(\alpha_1 + \alpha_2 + 1)\}}{\alpha_1 + \alpha_2},$$

where $\psi(t) = \{(d/dt)\Gamma(t)\}/\Gamma(t)$ is the digamma function. Matching these two expectations with their sample counterparts, we formulate the following score function evaluated at $(Y_i, \mathbf{X}_i)$ for model diagnosis when the assumed regression model is a beta mode model:

$$\mathbf{S}_{i,\text{beta}}(\boldsymbol{\Omega}) = \begin{bmatrix} \log Y_i - \psi\{1 + m\theta(\mathbf{X}_i)\} + \psi(2 + m) \\ Y_i \log(Y_i) - \dfrac{\{1 + m\theta(\mathbf{X}_i)\}[\psi\{2 + m\theta(\mathbf{X}_i)\} - \psi(3 + m)]}{2 + m} \end{bmatrix}. \tag{3.1}$$

If the assumed model is a GBP mode model, we formulate a bivariate score function based on matching the first two moments of $Y|\mathbf{X} \sim \text{GBP}(\theta(\mathbf{X}), m)$ (García et al., 2009),

$$E(Y|\mathbf{X}) = \frac{6m^2\theta(\mathbf{X}) + 7m + 2}{6m^2 + 14m + 4},$$

$$\begin{aligned}\text{Var}(Y|\mathbf{X}) = &\left\{4(3m + 1)^2(m + 2)^2(2m + 3)(m + 3)\right\}^{-1}\big[4m^2(37m^2 + 61m \\ &+10)\theta(\mathbf{X})\{\theta(\mathbf{X}) - 1\} + 82m^4 + 247m^3 + 247m^2 + 96m + 12\big].\end{aligned}$$

That is, the score function evaluated at $(Y_i, \mathbf{X}_i)$ for assessing the adequacy of an assumed GBP mode model is

$$\mathbf{S}_{i,\text{GBP}}(\boldsymbol{\Omega}) = \begin{bmatrix} Y_i - E(Y_i|\mathbf{X}_i) \\ Y_i^2 - \text{Var}(Y_i|\mathbf{X}_i) - \{E(Y_i|\mathbf{X}_i)\}^2 \end{bmatrix}. \tag{3.2}$$

Generically denote by $\mathbf{S}_i(\boldsymbol{\Omega})$ the score function in (3.1) or (3.2), depending on whether one assumes a beta mode model or a GBP mode model. We mimic the Hotelling's T^2 statistic (Hotelling, 1931) to define the following test statistic:

$$Q(\hat{\boldsymbol{\Omega}}; \mathcal{D}) = \frac{n-2}{2(n-1)} \bar{\mathbf{S}}^{\mathrm{T}} \hat{\boldsymbol{\Sigma}}^{-1} \bar{\mathbf{S}}, \tag{3.3}$$

where $\hat{\boldsymbol{\Omega}}$ is the MLE of $\boldsymbol{\Omega}$ under the assumed model, $\bar{\mathbf{S}} = n^{-1} \sum_{i=1}^{n} \mathbf{S}_i(\hat{\boldsymbol{\Omega}})$, and $\hat{\boldsymbol{\Sigma}} = \{n(n-1)\}^{-1} \sum_{i=1}^{n} \{\mathbf{S}_i(\hat{\boldsymbol{\Omega}}) - \bar{\mathbf{S}}\}\{\mathbf{S}_i(\hat{\boldsymbol{\Omega}}) - \bar{\mathbf{S}}\}^{\mathrm{T}}$ is an estimator for the variance-covariace of $\bar{\mathbf{S}}$. Under the null hypothesis that the assumed model is the true model, one has $E(\bar{\mathbf{S}}) = 0$ when evaluating $\hat{\boldsymbol{\Omega}}$ at the truth, and thus a small value for $Q(\hat{\boldsymbol{\Omega}}; \mathcal{D})$ is expected under the null. In contrast, when the assumed model differs from the true model to the extent that $E(\bar{\mathbf{S}})$ substantially deviates from zero, a large realization of $Q(\hat{\boldsymbol{\Omega}}; \mathcal{D})$ is expected. According to Hotelling (1931), if $\mathbf{S}_i(\boldsymbol{\Omega})$ is a bivariate normal random variable, then $Q(\boldsymbol{\Omega}; \mathcal{D}) \sim F_{2,n-2}$ under the null. With a response supported on $[0, 1]$, a bivariate normal is not likely to approximate well the distributions of the scores in (3.1) and (3.2), although a large $Q(\hat{\boldsymbol{\Omega}}; \mathcal{D})$ still implies poor fit for relevant moments and thus casts doubt on the assumed model. To accurately approximate certain percentiles of the null distribution of $Q(\hat{\boldsymbol{\Omega}}; \mathcal{D})$, we use a parametric bootstrap procedure that leads to an estimated p-value associated with the test statistic. The algorithm in supplementary Section S1 outlines the bootstrap procedure under the null stating that the true model is a GBP mode model. A similar bootstrap procedure is used to estimate the p-value of the test statistic when one assumes a beta mode model.

Empirical evidence from simulation studies (supplementary Figure S3) suggest that this bootstrap procedure can estimate the tail of the null distribution of $Q(\hat{\boldsymbol{\Omega}}; \mathcal{D})$ well enough to preserve the right size of the proposed score tests. Besides how well one can estimate certain percentiles of a null distribution, operating characteristics of the score tests also depend on the extent of distortion on moment estimation when an inadequate model is assumed. More empirical evidence on this aspect are presented next, along with the performance of maximum likelihood estimation and predictions based on synthetic data generated from various regression models.

4 | SIMULATION STUDY

Source code to reproduce the results in this section is available as Supporting Information on the journal's web page (http://onlinelibrary.wiley.com/doi/xxx/suppinfo).

4.1 | Design of simulation experiments

In all experiments, we simulate a bivariate covariate, $\mathbf{X} = (X_1, X_2)^{\mathrm{T}}$, as the predictor in a regression model. When carrying out regression analysis, we assume a linear predictor, $\eta(\mathbf{X}) = \beta_0 + \beta_1 X_1 + \beta_2 X_2$, and the logit link $g(t) = 1/(1 + e^{-t})$ in (2.3), despite the true data generating process.

When conducting regression analysis assuming a beta mode model, we first simulate X_2 from Bernoulli(0.5), and then generate data for X_1 according to $N(I(X_2 = 1) - I(X_2 = 0), 1)$. Given covariates data, responses are generated from each of the following four conditional distributions:

(B1) $Y|\mathbf{X} \sim \text{beta}(1 + m\theta(\mathbf{X}), 1 + m\{1 - \theta(\mathbf{X})\})$, where $\theta(\mathbf{X}) = 1/[1 + \exp\{-\eta(\mathbf{X})\}]$, with $\eta(\mathbf{X}) = \beta_0 + \beta_1 X_1 + \beta_2 X_2$;
(B2) $Y|\mathbf{X} \sim \text{beta}(1 + m\theta(\mathbf{X}), 1 + m\{1 - \theta(\mathbf{X})\})$, where $\theta(\mathbf{X}) = 1/[1 + \exp\{-\eta(\mathbf{X})\}]$, with $\eta(\mathbf{X}) = \beta_0 + \beta_1 X_1 + \beta_2 X_1^2 + \beta_3 X_2$;
(B3) $Y|\mathbf{X} \sim \text{beta}(1 + m\theta(\mathbf{X}), 1 + m\{1 - \theta(\mathbf{X})\})$, where $\theta(\mathbf{X}) = 0.5\Phi[2\{\eta(\mathbf{X}) + 2\}] + 0.5\Phi[2\{\eta(\mathbf{X}) - 2\}]$, with $\eta(\mathbf{X}) = \beta_0 + \beta_1 X_1 + \beta_2 X_2$;
(B4) $Y|\mathbf{X} \sim \text{GBP}(\theta(\mathbf{X}), m)$, where $\theta(\mathbf{X}) = 1/[1 + \exp\{-\eta(\mathbf{X})\}]$, with $\eta(\mathbf{X}) = \beta_0 + \beta_1 X_1 + \beta_2 X_2$.

When a GBP mode model is assumed for regression analysis, we consider the following four regression models according to which responses are generated after data for X_1 are simulated from $N(0, 1)$ and data for X_2 are simulated from Bernoulli(0.5):

TABLE 1 Averages of MLEs for parameters in a beta mode model and a GBP mode model across 300 Monte Carlo replicates generated according to (B1) and (G1), respectively, and averages of the corresponding estimated standard deviations ($\widehat{\text{s.d.}}$) in comparison with the empirical standard deviations (s.d.)

	MLE	$\widehat{\text{s.d.}}$	s.d.	MLE	$\widehat{\text{s.d.}}$	s.d.
(B1)	$n = 50$			$n = 100$		
β_0	0.973 (1.40)	0.225 (0.34)	0.243	1.011 (0.93)	0.162 (0.15)	0.161
β_1	0.983 (1.01)	0.173 (0.27)	0.174	1.000 (0.80)	0.123 (0.13)	0.138
β_2	1.005 (2.49)	0.411 (0.77)	0.431	0.982 (1.80)	0.290 (0.34)	0.311
$\log m$	2.413 (1.50)	0.224 (0.04)	0.260	2.347 (0.83)	0.159 (0.02)	0.144
(G1)	$n = 50$			$n = 100$		
β_0	0.994 (0.49)	0.069 (0.12)	0.085	0.994 (0.31)	0.050 (0.06)	0.054
β_1	0.984 (0.44)	0.063 (0.13)	0.076	0.993 (0.29)	0.045 (0.07)	0.049
β_2	0.986 (0.74)	0.109 (0.19)	0.128	1.004 (0.47)	0.079 (0.08)	0.081
$\log m$	2.356 (0.81)	0.142 (0.01)	0.141	2.322 (0.61)	0.100 (0.01)	0.105

Numbers in parentheses are $100\times$(Monte Carlo standard errors) associated with the averages. The true parameter values are $\boldsymbol{\beta} = (\beta_0, \beta_1, \beta_2)^{\mathrm{T}} = (1, 1, 1)^{\mathrm{T}}$, and $\log m = \log 10 \approx 2.303$.

(G1) $Y|\mathbf{X} \sim \text{GBP}(\theta(\mathbf{X}),\ m)$, where $\theta(\mathbf{X}) = 1/[1 + \exp\{-\eta(\mathbf{X})\}]$, with $\eta(\mathbf{X}) = \beta_0 + \beta_1 X_1 + \beta_2 X_2$;
(G2) $Y|\mathbf{X} \sim \text{GBP}(\theta(\mathbf{X}),\ m)$, where $\theta(\mathbf{X}) = 1/[1 + \exp\{-\eta(\mathbf{X})\}]$, with $\eta(\mathbf{X}) = \beta_0 + \beta_1 X_1 + \beta_2 X_1^2 + \beta_3 X_2$;
(G3) $Y|\mathbf{X} \sim \text{GBP}(\theta(\mathbf{X}),\ m)$, where $\theta(\mathbf{X}) = 0.5\Phi[2\{\eta(\mathbf{X}) + 2\}] + 0.5\Phi[2\{\eta(\mathbf{X}) - 2\}]$, in which $\eta(\mathbf{X}) = \beta_0 + \beta_1 X_1 + \beta_2 X_2$;
(G4) $Y|\mathbf{X} \sim \text{beta}(1 + m\theta(\mathbf{X}),\ 1 + m\{1 - \theta(\mathbf{X})\})$, where $\theta(\mathbf{X}) = 1/[1 + \exp\{-\eta(\mathbf{X})\}]$, with $\eta(\mathbf{X}) = \beta_0 + \beta_1 X_1 + \beta_2 X_2$.

These simulation settings are designed to cover a wide range of scenarios that are of theoretical and practical interest. For instance, cases (B1)–(B4) allow correlated covariates, whereas covariates in (G1)–(G4) are independent. More importantly, we include three sources of model misspecification in this experiment that are frequently discussed in the literature on parametric regression models. In particular, (B1) and (G1) create scenarios where the assumed model coincides with the true model, and the other cases give rise to scenarios where we implement maximum likelihood estimation under a misspecified model. Under (B2) and (G2), the assumed models misspecify the linear predictor; under (B3) and (G3), the assumed models involve a misspecified link function for the mode; and under (B4) and (G4), the assumed conditional distribution of Y given covariates is not from the same family that the true conditional distribution belongs to.

4.2 | Covariate effects estimation

Given each of the above data generating processes, we generate datasets $\{(Y_i, X_{1i}, X_{2i})\}_{i=1}^n$ with $n = 50, 100$. Under each simulation setting, we repeat maximum likelihood estimation using 300 Monte Carlo datasets.

When the assumed model matches the true model, the MLE for $\boldsymbol{\Omega}$ is expected to be consistent estimator. Table 1 provides summary statistics for these MLEs and the estimated standard deviations associated with these estimates based on data generated according to (B1) and (G1), respectively, with $m = 10$ and $\boldsymbol{\beta} = (1, 1, 1)^{\mathrm{T}}$, where the estimated standard deviations result from sandwich variance estimation for M-estimators (Boos & Stefanski, 2013, Section 7.2.1). The close agreement between the MLE of $\boldsymbol{\Omega}$ and the truth and the resemblance between the estimated standard deviations and the empirical standard deviations suggest that the first two moments of the asymptotic distribution of the MLEs are estimated reasonably well.

4.3 | Performance of model diagnosis methods

Besides covariate effects estimation, we also monitor operating characteristics of the model diagnostics tools proposed in Section 3. Assuming a beta mode model, Figure 3 demonstrates the half-normal residual plots obtained based on one dataset of size $n = 100$ generated from each of (B1)–(B4), where $m = 80$ in (B1)–(B3) and $m = 10$ in (B4). Here setting $m = 80$ for (B1)–(B3) is to make sure that the conditional variance under the beta model is similar to that under the GBP

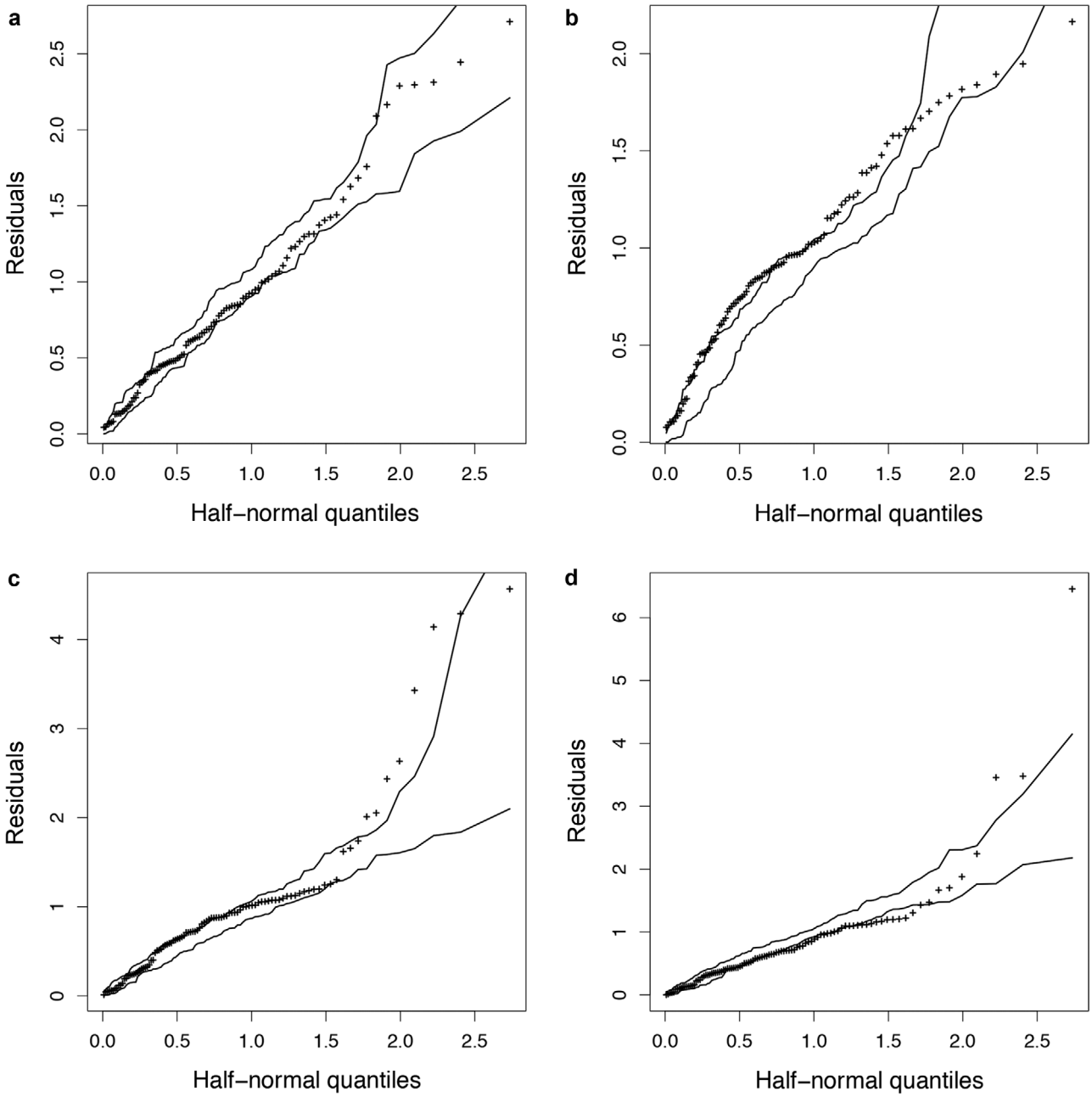

FIGURE 3 Half-normal residual plots with simulated envelopes based on one random sample of size $n = 100$ when one assumes a beta mode model for data generated from (B1)–(B4), shown in (a)–(d), respectively

model with $m = 10$. Under (B1), where the assumed model matches the true data generating process, very few residuals fall outside of the envelope. In contrast, a large proportion of the residuals are outside of the envelope under (B2), where the linear predictor is misspecified. The plots under (B3) and (B4) also witness rather high proportions of residuals outside of the envelope. These empirical evidence indicate that the half-normal residual plot is an effective graphical indicator of various sources of model misspecification.

Assuming a GBP mode model, Figure 4 demonstrates the half-normal residual plots using one dataset of size $n = 100$ generated from each of (G1)–(G4), where $m = 10$ in (G1)–(G3), and $m = 80$ in (G4). Again setting $m = 80$ for (G4) is to make sure that the conditional variance under the beta model is similar to that under the GBP model with $m = 10$. Similar to what one sees in the previous figure, in the absence of model misspecification as in (G1), most residuals are within the envelope. A much larger proportion of residuals fall outside of the envelope in the presence of linear predictor

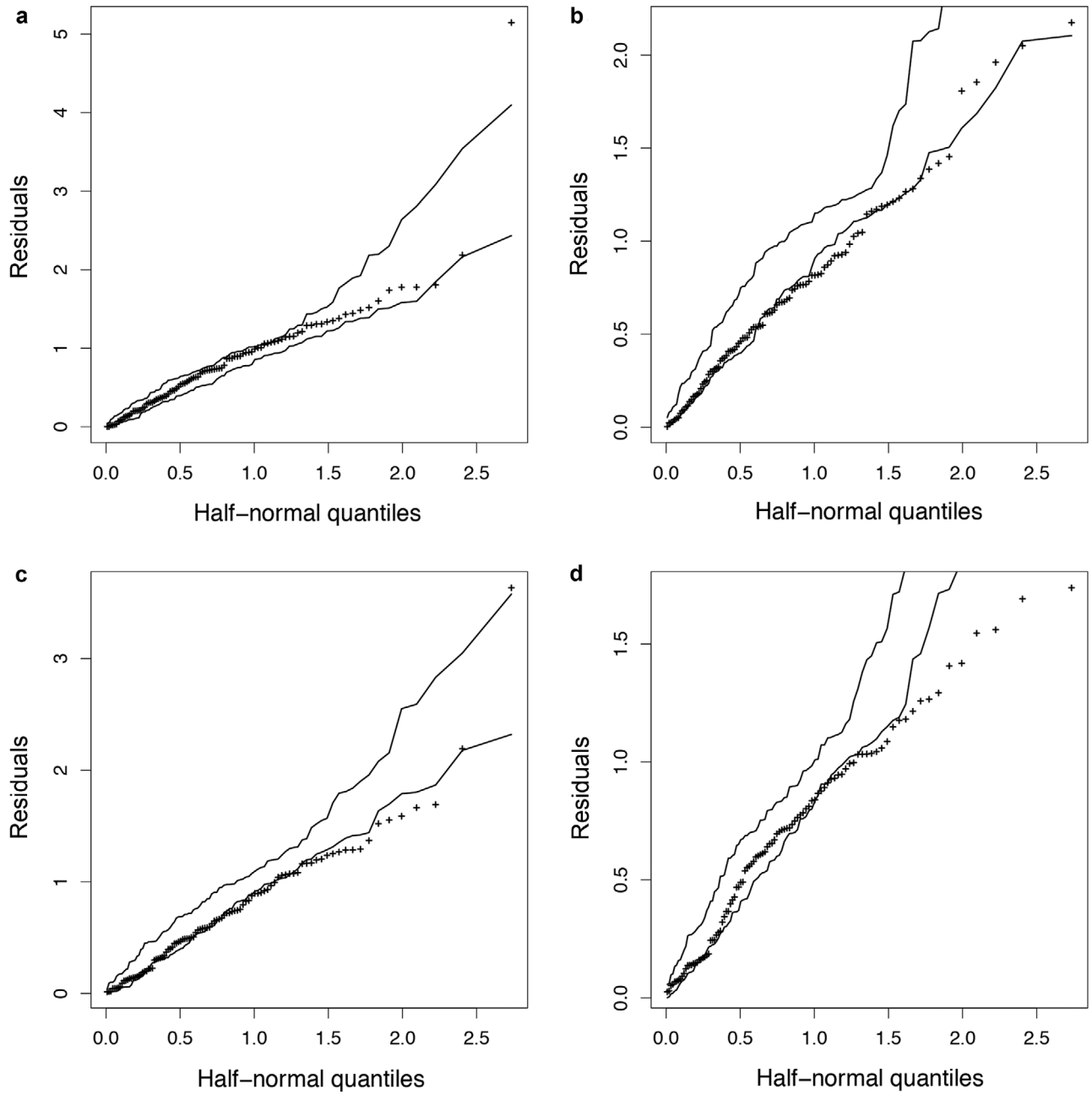

FIGURE 4 Half-normal residual plots with simulated envelopes based on one random sample of size $n = 100$ when one assumes a GBP mode model for data generated from (G1)–(G4), shown in (a)–(d), respectively

misspecification as in (G2). The plots also exhibit a moderate to high proportion of residuals outside of the envelopes under (G3) and (G4). Hence, the effectiveness of the half-normal residual plot for detecting various sources of model misspecification is also evident when one assumes a GBP mode model.

Figure 5 presents the empirical power of the score tests proposed in Section 3.2 when a beta mode model or a GBP mode model is assumed, where the empirical power of a test is defined as the rejection rate of the test at significance level 0.05 across 300 Monte Carlo replicates under each true model specification. Given one simulated dataset, we use 300 bootstrap samples to estimate the p-value associated with a score test. These empirical power indicate that the size of the score tests remain close to the nominal level in the absence of model misspecification, and, as the sample size grows, their power to detect any one of the three sources of model misspecification increases. Under each assumed mode model, the proposed score test has the highest power to detect a link misspecification, moderate power in response to a misspecified linear predictor, and the lowest power when the conditional distribution family is misspecified. The low power to detect the

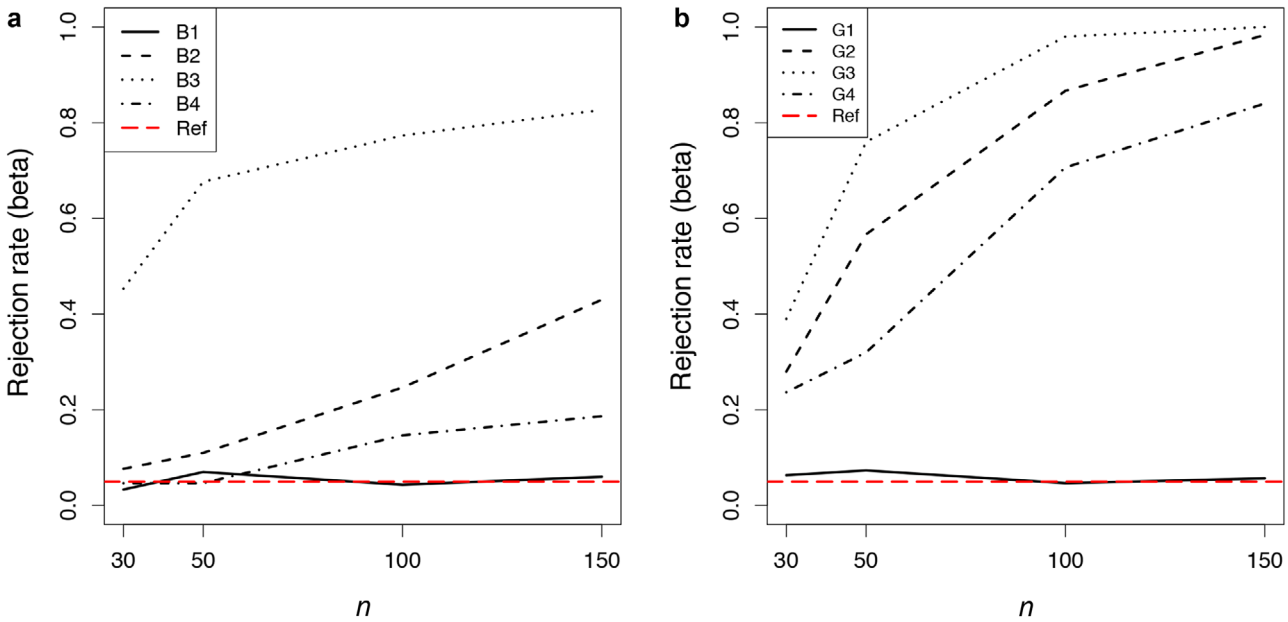

FIGURE 5 Rejection rates across 300 Monte Carlo replicates associated with the score tests under (B1)–(B4) when a beta mode model is assumed (in (a)), and under (G1)–(G4) when a GBP mode model is assumed (in (b)). True model settings are: solid lines for (B1) and (G1), dashed lines for (B2) and (G2), dotted lines for (B3) and (G3), and dashed-dotted lines for (B4) and (G4). Red horizontal long-dashed lines are reference lines at nominal level 0.05

last type of model misspecification, especially when a beta mode model is assumed, may suggest that, given a GBP mode model, there exists a member in the family of beta mode models that can approximate the GBP mode model well enough to produce reasonable estimates for the first two moments. Lastly, these numerical evidence of model misspecification match nicely with the graphical evidence from half-normal residual plots in that a higher rejection rate observed for the score test under one scenario usually goes with a higher proportion of residuals outside of the envelope in the half-normal residual plot in the same scenario.

4.4 | Predictions

Predicting an outcome is often one of the ultimate goals in regression analysis, such as in ADNI where accurate prediction of AD progression is a major goal. Suppose $\boldsymbol{x}$ is the covariate value at which one wishes to predict a bounded outcome, such as one's ADAS-11 score. In what follows, at a nominal coverage probability $q \in (0, 1)$, we first construct a prediction interval based on an estimated mode, denoted by $\mathcal{PI}_\theta(\boldsymbol{x}, q)$, then we formulate a prediction interval based on an estimated mean, denoted by $\mathcal{PI}_\mu(\boldsymbol{x}, q)$, and a prediction interval based on an estimated median, denoted by $\mathcal{PI}_\nu(\boldsymbol{x}, q)$. Define $e = Y - \theta(\boldsymbol{x})$ as the mode residual, and denote by $f_e(e|\boldsymbol{x})$ the pdf of e given $\mathbf{X} = \boldsymbol{x}$.

Under an assumed mode regression model, following maximum likelihood estimation of $\boldsymbol{\Omega}$, one obtains the MLEs for $\theta(\boldsymbol{x})$ and $\mu(\boldsymbol{x})$, as well as an estimated pdf of e given $\mathbf{X} = \boldsymbol{x}$. Denote these MLEs by $\hat{\theta}(\boldsymbol{x})$ and $\hat{\mu}(\boldsymbol{x})$, respectively, and denote by $\hat{f}_e(e|\boldsymbol{x})$ the estimated pdf. Then, based on these estimates, the narrowest $\mathcal{PI}_\theta(\boldsymbol{x}, q)$ is $[\hat{\theta}(\boldsymbol{x}) + e_1, \hat{\theta}(\boldsymbol{x}) + e_2]$, where $e_1 < 0 < e_2$ satisfy $\hat{f}_e(e_1|\boldsymbol{x}) = \hat{f}_e(e_2|\boldsymbol{x})$ and $\int_{e_1}^{e_2} \hat{f}_e(e|\boldsymbol{x})\, de = q$.

To formulate a $(100 \times q)\%$ mean-based prediction interval, we first make sure that $\hat{\mu}(\boldsymbol{x}) \in \mathcal{PI}_\mu(\boldsymbol{x}, q)$, then we construct an interval with the desired coverage probability that is close to $\hat{\theta}(\boldsymbol{x})$ as much as possible in order to achieve the narrowest $\mathcal{PI}_\mu(\boldsymbol{x}, q)$. Clearly, if $\hat{\mu}(\boldsymbol{x})$ already falls in $\mathcal{PI}_\theta(\boldsymbol{x}, q)$ constructed above, which is the narrowest by construction, then one may also use this interval as $\mathcal{PI}_\mu(\boldsymbol{x}, q)$. Otherwise, we construct $\mathcal{PI}_\mu(\boldsymbol{x}, q)$ with $\hat{\mu}(\boldsymbol{x})$ on one of the boundaries depending on how $\hat{\mu}(\boldsymbol{x})$ compares with $\hat{\theta}(\boldsymbol{x})$. In particular, if $\hat{\mu}(\boldsymbol{x}) \geq \hat{\theta}(\boldsymbol{x})$, then we set $\mathcal{PI}_\mu(\boldsymbol{x}, q) = [\hat{\mu}(\boldsymbol{x}) - c,\ \hat{\mu}(\boldsymbol{x})]$, where $c > 0$ is chosen such that $\int_{\hat{\mu}(\boldsymbol{x}) - c}^{\hat{\mu}(\boldsymbol{x})} \hat{f}_{Y|\mathbf{X}}(y|\boldsymbol{x})dy = q$, in which $\hat{f}_{Y|\mathbf{X}}(y|\boldsymbol{x})$ is pdf of the assumed distribution of Y given $\mathbf{X} = \boldsymbol{x}$ evaluated at $\hat{\boldsymbol{\Omega}}$. If $\hat{\mu}(\boldsymbol{x}) < \hat{\theta}(\boldsymbol{x})$, then we let $\mathcal{PI}_\mu(\boldsymbol{x}, q) = [\hat{\mu}(\boldsymbol{x}),\ \hat{\mu}(\boldsymbol{x}) + c]$, where $c > 0$ satisfies $\int_{\hat{\mu}(\boldsymbol{x})}^{\hat{\mu}(\boldsymbol{x}) + c} \hat{f}_{Y|\mathbf{X}}(y|\boldsymbol{x})dy =$

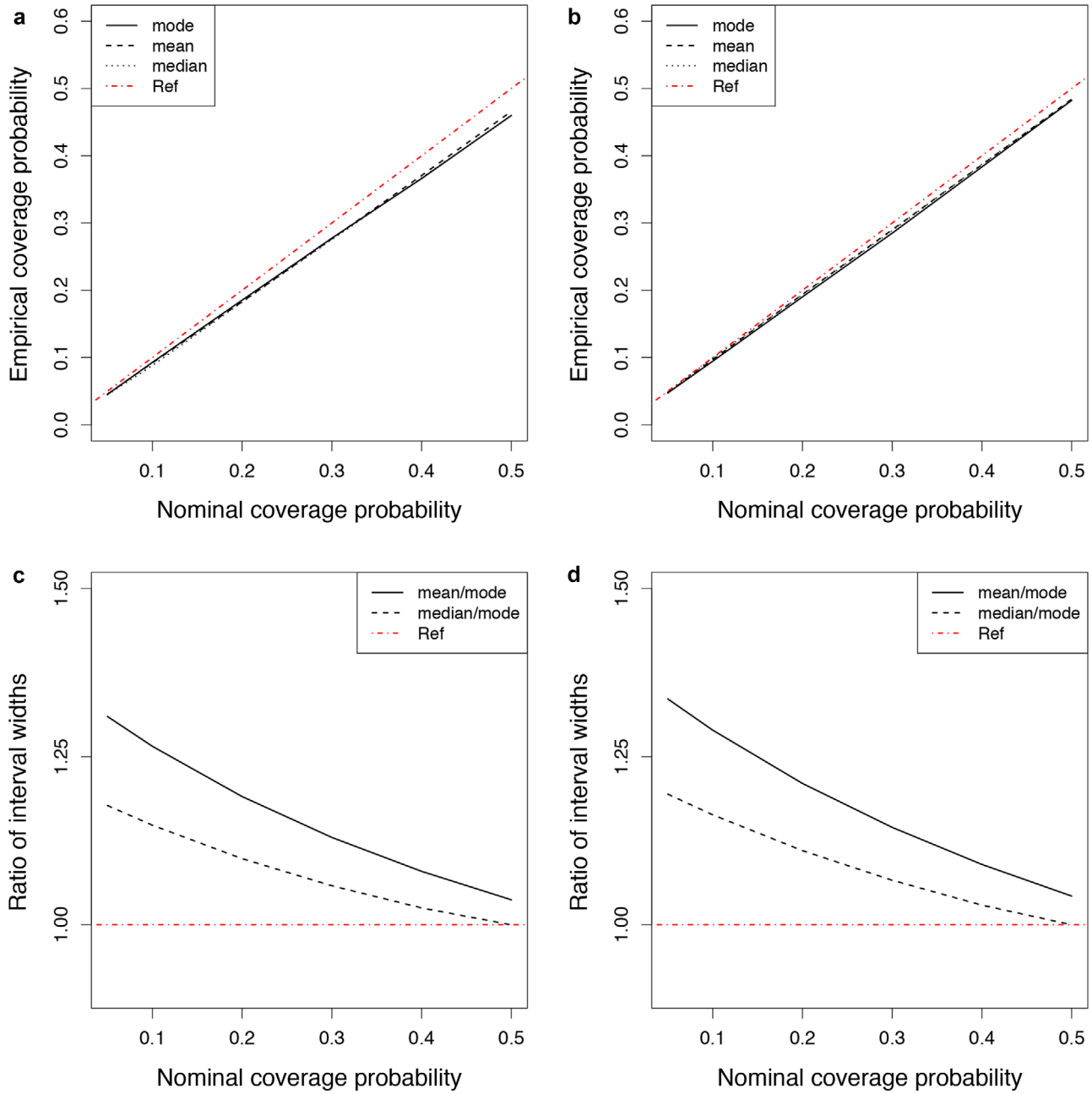

FIGURE 6 Prediction intervals based on data from the beta mode model in (B1). Top panels (a) and (b) depict average empirical coverage probabilities (across 300 Monte Carlo replicates) of mode-based prediction intervals (solid lines), those of mean-based prediction intervals (dashed lines), and those of median-based prediction intervals (dotted lines) versus nominal coverage probabilities. Red dash-dotted lines are 45° reference lines. Lower panels (c) and (d) depict ratios of the average width of mean-based prediction intervals over that of mode-based prediction intervals (solid lines) and ratios of the average width of median-based prediction intervals over that of mode-based prediction intervals (dashed lines) versus nominal coverage probabilities. Red dash-dotted horizontal lines are reference lines at value one. Panels (a) and (c) are for $n = 50$. Panels (b) and (d) are for $n = 100$

q. A $(100 \times q)\%$ median-based prediction interval can be similarly constructed once an estimated median, denoted by $\hat{\nu}(\boldsymbol{x})$, is obtained.

With the assumed model being a beta mode model, Figure 6 depicts in upper panels averages empirical coverage probabilities of $\mathcal{PI}_\theta(\cdot, q)$, $\mathcal{PI}_\mu(\cdot, q)$, and $\mathcal{PI}_\nu(\cdot, q)$ versus nominal coverage probabilities q when 300 Monte Carlo replicate datasets are generated from (B1) with $\boldsymbol{\beta} = (3, 1, 1)$, $m = 10$ and $n = 50, 100$, where q ranges from 0.05 to 0.5. The empirical coverage probability of each type of prediction intervals is obtained via fivefold cross-validation. Take $\mathcal{PI}_\theta(\cdot, q)$ as

an example, its empirical coverage probability based on one dataset is defined as $n^{-1}\sum_{k=1}^{5}\sum_{i\in\mathcal{I}_k} I\{Y_i \in \mathcal{PI}_\theta^{(-k)}(\mathbf{X}_i, q)\}$, where $\mathcal{PI}_\theta^{(-k)}(\mathbf{X}_i, q)$ is the $(100\times q)\%$ mode-based prediction interval constructed using data excluding the kth testing dataset corresponding to the index set $\mathcal{I}_k$, for $k = 1, \dots, 5$. The lower panels of Figure 6 compare the average width of the three types of prediction intervals.

According to these figures, mode-based prediction intervals, mean-based prediction intervals, and media-based prediction intervals achieve similar empirical coverage probabilities that become closer to the nominal coverage probability as n increases. More importantly, the mode-based prediction interval tends to be narrowest among the three, and the mean-based prediction interval is the widest. By construction, it is expected that $\mathcal{PI}_\theta(\boldsymbol{x}, q) = \mathcal{PI}_\mu(\boldsymbol{x}, q) = \mathcal{PI}_\nu(\boldsymbol{x}, q)$ when q is not low since $\mathcal{PI}_\theta(\boldsymbol{x}, q)$ with a moderate or high coverage probability is very likely to include the estimated mean and the estimated median. Certainly, these prediction intervals are also expected to be more similar when the conditional distribution for the response is less skewed.

To demonstrate the impact of outliers on the aforementioned prediction intervals, we contaminate each of 300 Monte Carlo replicate datasets generated from the beta mode model by replacing 5% of randomly chosen responses with random numbers simulated from uniform$(0, t)$, where t is the 0.001th quantile of the true conditional distribution of the response. This contamination produces data with a heavier (left) tail than the distribution specified in (B1). Despite the model misspecification, we fit the resultant data assuming a beta mode regression model and construct prediction intervals based on the three central tendency measures. Figure 7 shows the comparison between different types of prediction intervals in regard to empirical coverage probability and width. From there, one can see that a direct consequence of fitting (and making predictions based on) a beta mode model to data from an underlying distribution with a heavier (than assumed) tail is inflated coverage probabilities, despite the choice of central tendency measure for prediction. Interestingly, even though the empirical coverage probability of the mode-based prediction interval is higher than those of the other two types of prediction intervals, the mode-based prediction interval remains the narrowest among the three. In conclusion, even in the presence of severe outliers, the conditional mode still yields more reliable and precise predictions than the conditional mean or median does. Parallel pictures when data are generated from (G1), with or without outliers contamination, and one assumes a GBP mode model are provided in supplementary Figures S4 and S5.

5 | APPLICATION TO ADNI DATA

There has been a consensus among medical researchers that regional brain atrophy in the medial temporal lobe structures, such as the entorhinal cortex (ERC) and hippocampus (HPC), are correlated with clinical alterations in the predementia phase of mild cognitive impairment (MCI) and various dementia phases of AD (Devanand et al., 2007; Jauhiainen et al., 2009). While early detection and intervention in MCI subjects has been actively pursued by many researchers, there are mixed opinions among them regarding the roles volumetric measures of ERC and HPC play in predicting an MCI subject's risk of developing AD (deToledo Morrell et al., 2004; Hämäläinen et al., 2007; Jack et al., 1999; Killiany et al., 2002; Whitwell et al., 2008).

Using one's ADAS-11 score as a surrogate for one's severity of cognitive impairment, we apply the proposed mode regression models to data from the ADNI database (http://adni.loni.usc.edu) to study the association between one's ADAS-11 score at month 12 and the volumetric changes in ERC and HPC at month 6 compared to their baseline measures. In particular, the dataset we consider consists of a cohort of 245 subjects who were diagnosed with LMCI when they entered the ADNI Phase 1 study and were followed up at least at both months 6 and 12. The original response variable is a subject's ADAS-11 score at month 12, which has a bounded support on $[0, 70]$. We rescale the support to the unit interval by dividing ADAS-11 scores by 70.

Besides carrying out mode regression analysis, we also adopt the beta mean regression model for a rate or proportion response proposed by Ferrari and Cribari-Neto (2004) to study the association. In their proposed regression model, the authors reparameterized the beta distribution by setting $\alpha_1 = \mu\phi$ and $\alpha_2 = (1-\mu)\phi$, where $\mu \in [0, 1]$ is the mean parameter and ϕ is a positive shape parameter, with a larger ϕ resulting in a smaller variance, and they incorporated the linear predictor $\eta(\mathbf{X})$ by letting $\mu = g\{\eta(\mathbf{X})\}$. In a preliminary analysis, we fit the beta mean model, beta mode model, and GBP mode model to the data using various link functions $g(t)$. Based on values of log-likelihood, we choose the log-log link in the regression models for further analysis.

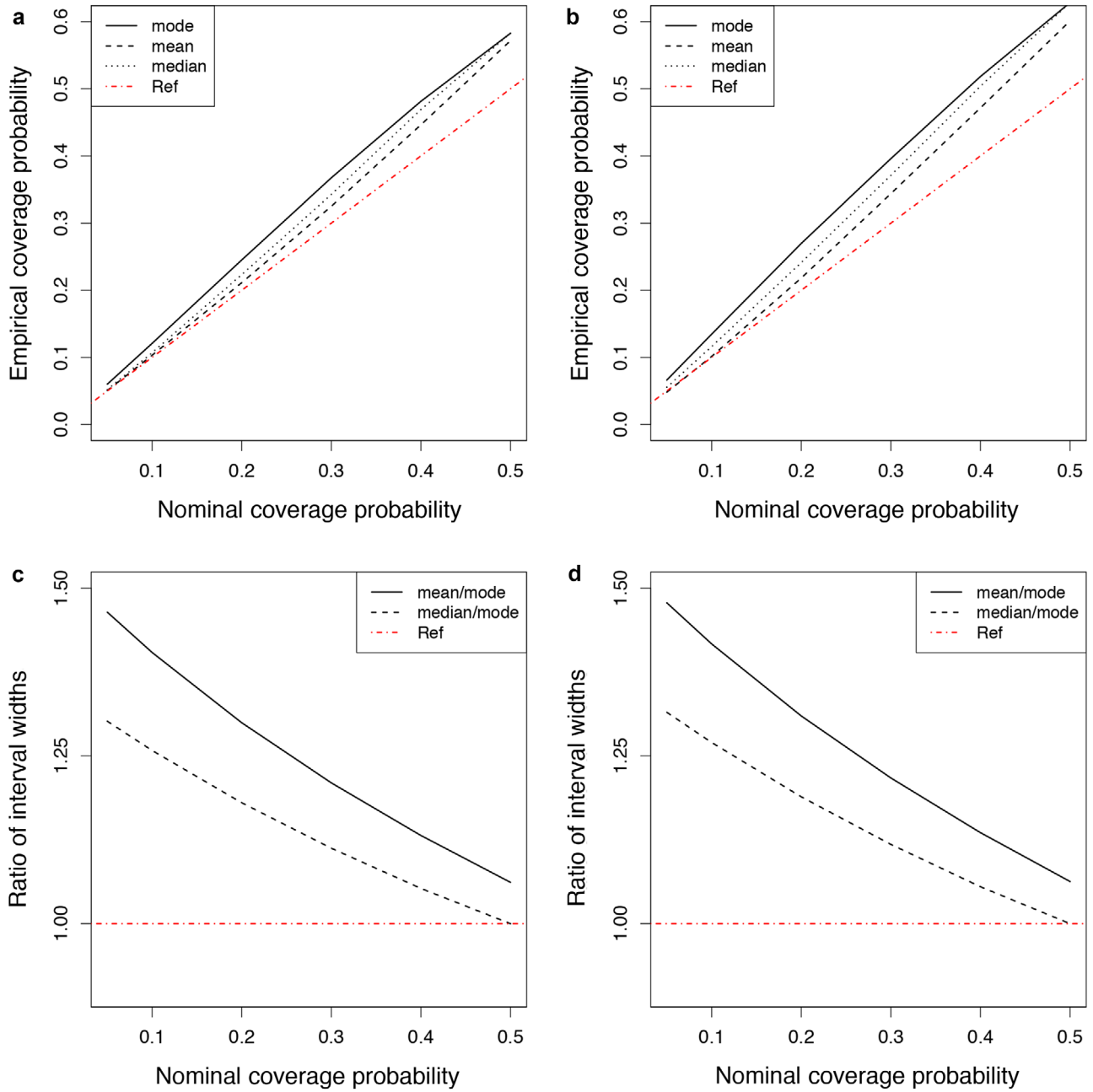

FIGURE 7 Prediction intervals based on beta mode regression using data from the beta mode model in (B1), with outliers replacing 5% of the original data. Top panels (a) and (b) depict average empirical coverage probabilities (across 300 Monte Carlo replicates) of mode-based prediction intervals (solid lines), those of mean-based prediction intervals (dashed lines), and those of median-based prediction intervals (dotted lines) versus nominal coverage probabilities. Red dash-dotted lines are 45° reference lines. Lower panels (c) and (d) depict ratios of the average width of mean-based prediction intervals over that of mode-based prediction intervals (solid lines) and ratios of the average width of median-based prediction intervals over that of mode-based prediction intervals (dashed lines) versus nominal coverage probabilities. Red dash-dotted horizontal lines are reference lines at value one. Panels (a) and (c) are for $n = 50$. Panels (b) and (d) are for $n = 100$

Panels (a) and (b) in Figure 8 provide the half-normal residual plots associated with the beta mode model and GBP mode model based on this dataset. Having majority of the residuals from GBP mode regression falling outside of the envelope suggests a poor fit of the GBP model for the data, and the beta mode model is more adequate for the current data. The score test when the null hypothesis states a beta mode model yields an estimated p-value of 0.45, while the score test when one assumes a GBP mode model gives an estimated p-value of 0. Gathering these graphical and numerical diagnosis, we conclude that the beta mode model potentially captures the underlying conditional distribution better than the GBP mode model does.

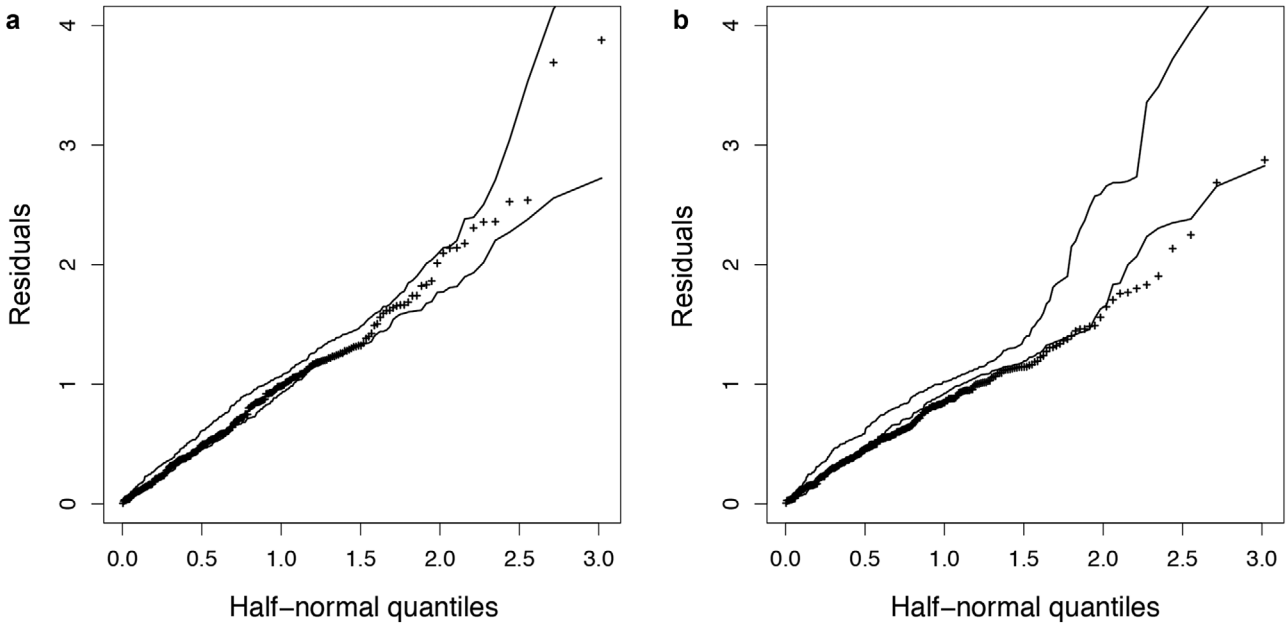

FIGURE 8 Half-normal residual plots with simulated envelope for the ADNI data associated with the beta mode model (in (a)) and the GBP mode model (in (b))

TABLE 2 Maximum likelihood estimates corresponding to each regression model for the ADNI data

Parameter	Beta mean model	Beta mode model	GBP mode model
β_0 (Intercept)	−0.555 (0.020)	−0.697 (0.026)	−0.971 (0.017)
β_1 (ERC.change)	−0.102 (0.043)	−0.125 (0.052)	−0.117 (0.028)
β_2 (HPC.change)	−0.170 (0.084)	−0.216 (0.107)	−0.112 (0.109)
ϕ or $\log m$	17.990 (1.609)	2.772 (0.099)	1.826 (0.068)

Numbers in parentheses are estimated standard errors associated with the MLEs.

Table 2 provides MLEs of unknown parameters in each of the three considered regression models. According to Table 2, inference for the covariate effects from all three regression models suggest that the volumetric change in ERC is a statistically significant predictor for one's cognitive impairment. However, results from the GBP mode model does not indicate that the volumetric change in HPC is significantly associated with the response (with a p-value of 0.303), although inference from both beta mean and beta mode model imply a significant effect of the change in hippocampal volume on the ADAS-11 score (with p-values 0.042 and 0.043, respectively).

Figure 9 presents the histogram of mean residuals and the histogram of mode residuals resulting from beta mean regression and beta mode regression, respectively. Both histograms suggest a right-skewed distribution of the ADAS-11 score conditional on the two volumetric measures, and the two residual distributions are overall similar. Despite such similarity, it is worth stressing that, in general, interpretations of a covariate effect inferred by the two models are different even though we use the same notations in Table 2 for regression coefficients under different regression models. To avoid such abuse of notation, let us write the mean function under the beta mean regression model as $\mu(\mathbf{X}) = g(b_0 + \boldsymbol{b}_1^{\mathrm{T}}\mathbf{X})$, in contrast to the mode function under the beta mode regression model, $\theta(\mathbf{X}) = g(\beta_0 + \boldsymbol{\beta}_1^{\mathrm{T}}\mathbf{X})$. Under the parameterization leading to the beta mean model, the mode function is $\theta(\mathbf{X}) = \{\phi\mu(\mathbf{X}) - 1\}/(\phi - 2)$. This indicates that, when ϕ is large, in particular, much larger than one, one has $\theta(\mathbf{X}) \approx \mu(\mathbf{X})$. Similarly, under the parameterization for the beta mode model, the mean function is $\mu(\mathbf{X}) = \{m\theta(\mathbf{X}) + 1\}/(m + 2)$, suggesting that, when m is much larger than one, $\mu(\mathbf{X}) \approx \theta(\mathbf{X})$. According to Table 2, $\hat{\phi} \approx 17.99$ and $\hat{m} \approx 15.99$, both fairly large compared to one, which can be the reason for the similarity between the MLEs of the two sets of covariate effects under the two models, $\hat{\boldsymbol{b}}_1$ and $\hat{\boldsymbol{\beta}}_1$, in terms of magnitude and statistical significance. This serves as an example where beta mean regression and beta model regression perform similarly in terms of identifying influential predictors. With the rich data information for a large collection of biomarkers available in ADNI, one may consider carrying out variable selection based on beta mode regression, which is beyond the scope of the current study. In the follow-up project, we will give variable selection based on mode regression a more careful treatment, where

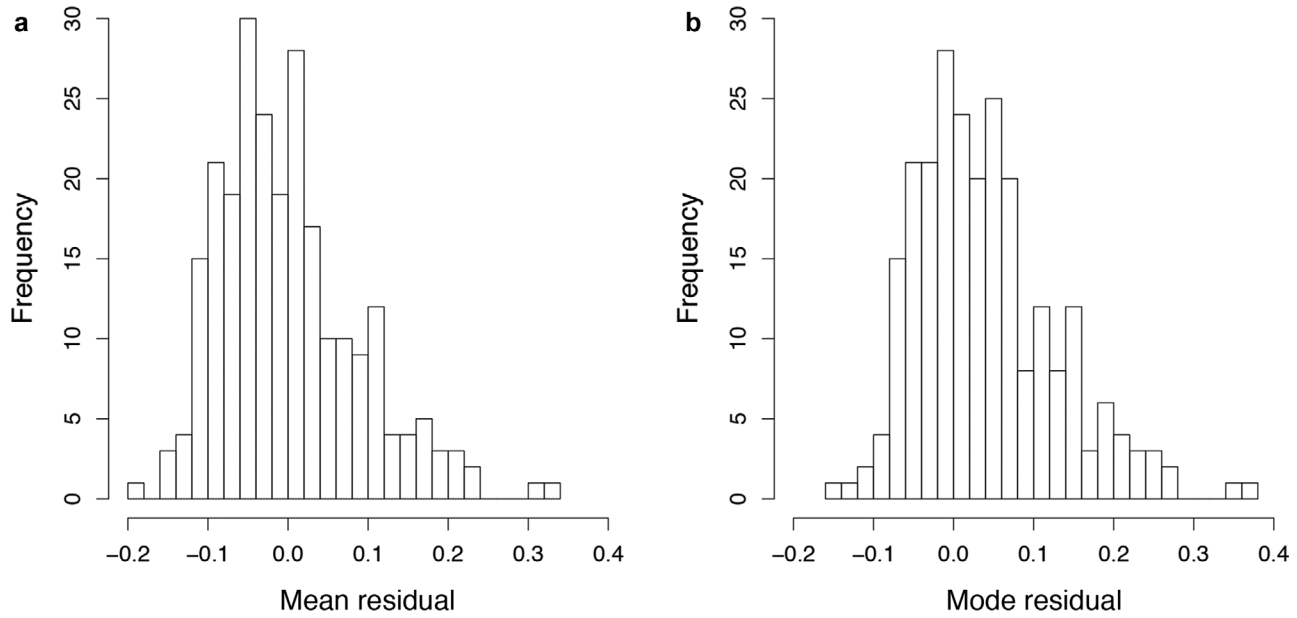

FIGURE 9 Histograms of residuals when fitting a beta mean model (in (a)) and when fitting a beta mode model (in (b)) to the ADNI data

TABLE 3 Fivefold cross-validated coverage probabilities of mode-based prediction intervals, $\mathcal{PI}_\theta(\cdot, q)$, mean-based prediction intervals, $\mathcal{PI}_\mu(\cdot, q)$, and median-based prediction intervals, $\mathcal{PI}_\nu(\cdot, q)$, under each regression model for the ADNI data

q	Beta mode model			GBP mode model		
	$\mathcal{PI}_\theta$	$\mathcal{PI}_\mu$	$\mathcal{PI}_\nu$	$\mathcal{PI}_\theta$	$\mathcal{PI}_\mu$	$\mathcal{PI}_\nu$
0.1	0.127 (0.021)	0.086 (0.022)	0.094 (0.022)	0.078 (0.020)	0.098 (0.028)	0.151 (0.025)
0.2	0.233 (0.043)	0.216 (0.044)	0.224 (0.043)	0.159 (0.041)	0.253 (0.053)	0.269 (0.048)
0.5	0.531 (0.114)	0.531 (0.114)	0.531 (0.114)	0.522 (0.114)	0.527 (0.117)	0.522 (0.114)

Numbers in parentheses are the average widths of the prediction intervals.

we formulate flexible regression models built upon the currently proposed beta or GBP mode model that allow heavier tails (to capture severe outliers) and multimodality. By then, we will have a fairer comparison with the variable selection procedure proposed by Wang et al. (2017) applying to ADNI data that is based on nonparametric mode regression.

We next construct prediction intervals based on the estimated densities from beta mode regression and GBP mode regression following the method described in Section 4.4. For each regression model, we compare the mode-based prediction interval, the mean-based prediction interval, and the median-based prediction interval, that is, $\mathcal{PI}_\theta(\cdot, q)$, $\mathcal{PI}_\mu(\cdot, q)$, and $\mathcal{PI}_\nu(\cdot, q)$, for a given q in regard to their empirical coverage probabilities and widths. A fivefold cross-validation procedure is used to obtain the empirical coverage probability of a considered type of prediction interval. Among the three types of prediction intervals based on different central tendency measures, the narrower interval that also has an empirical coverage probability close to q is more preferable. Table 3 reports summary statistics of these prediction intervals. Comparing the three types of prediction intervals with a fixed q under each mode regression model, one can see that $\mathcal{PI}_\theta(\cdot, q)$ tends to be the narrowest among the three when q is small, while all possessing empirical coverage probabilities close to the nominal level. Hence, in this application of assessing an LMCI subject's extent of cognitive impairment in the near future, using the mode tends to provide more accurate prediction than when using the mean or median.

6 | DISCUSSION

We propose two classes of regression models for studying the association between a bounded response and covariates via inferring the conditional mode of the response. Among all existing regression methodology, only a small subset of them are designed for mode regression, and an even smaller collection of them are in the parametric paradigm. The two mode regression models proposed in our study contribute new regression platforms for association studies when a bounded

response is of interest. Under each proposed mode regression model, we have developed model diagnostic tools to detect various forms of inadequate parametric assumptions.

Besides allowing the mode to depend on covariates, one may consider covariate-dependent shape parameter $m(\mathbf{X})$ to expand the class of mode regression models. A more flexible family of mode regression models can be formulated as mixtures of beta or GBP distributions, or mixing beta or GBP with a uniform distribution by mimicking the construction of beta rectangular distributions (Hahn, 2008). These mixture distributions will allow inclusion of multimodal distributions and distributions with heavier tails than those of beta or GBP distributions.

The family of GBP distributions is a rare distribution family that directly includes the mode in the parameterization, which makes it especially suitable for mode regression. If, unlike responses in our current study, the support of the response is unknown, then we have additional parameter(s) in the GBP density relating to the support, resulting in a nonregular model. In this case, maximum likelihood estimation can break down or leads to estimators that do not possess properties one usually sees in an MLE under a regular model (Cheng & Amin, 1983). Parameter estimations and properties of MLEs for parameters in a nonregular GBP regression model demand systematic investigations.

ACKNOWLEDGMENTS

Data collection and sharing for this project was funded by the Alzheimer's Disease Neuroimaging Initiative (ADNI) (National Institutes of Health Grant U01 AG024904) and DOD ADNI (Department of Defense award number W81XWH-12-2-0012). ADNI is funded by the National Institute on Aging, the National Institute of Biomedical Imaging and Bioengineering, and through generous contributions from the following: AbbVie, Alzheimers Association; Alzheimers Drug Discovery Foundation; Araclon Biotech; BioClinica, Inc.; Biogen; Bristol-Myers Squibb Company; CereSpir, Inc.; Cogstate; Eisai Inc.; Elan Pharmaceuticals, Inc.; Eli Lilly and Company; EuroImmun; F. Hoffmann-La Roche Ltd and its affiliated company Genentech, Inc.; Fujirebio; GE Healthcare; IXICO Ltd.; Janssen Alzheimer Immunotherapy Research & Development, LLC.; Johnson & Johnson Pharmaceutical Research & Development LLC.; Lumosity; Lundbeck; Merck & Co., Inc.; Meso Scale Diagnostics, LLC.; NeuroRx Research; Neurotrack Technologies; Novartis Pharmaceuticals Corporation; Pfizer Inc.; Piramal Imaging; Servier; Takeda Pharmaceutical Company; and Transition Therapeutics. The Canadian Institutes of Health Research is providing funds to support ADNI clinical sites in Canada. Private sector contributions are facilitated by the Foundation for the National Institutes of Health (www.fnih.org). The grantee organization is the Northern California Institute for Research and Education, and the study is coordinated by the Alzheimers Therapeutic Research Institute at the University of Southern California. ADNI data are disseminated by the Laboratory for Neuro Imaging at the University of Southern California.

CONFLICT OF INTEREST

The authors have declared no conflict of interest.

OPEN RESEARCH BADGES

This article has earned an Open Data badge for making publicly available the digitally-shareable data necessary to reproduce the reported results. The data is available in the Supporting Information section.

This article has earned an open data badge "**Reproducible Research**" for making publicly available the code necessary to reproduce the reported results. The results reported in this article were reproduced partially for data confidentiality reasons and due to their computational complexity.

ORCID

Haiming Zhou https://orcid.org/0000-0002-2777-5354

REFERENCES

Atkinson, A. C. (1987). *Plots, transformations, and regression: An introduction to graphical methods of diagnostic regression analysis*. Oxford, UK: Oxford University Press.

Bamford, S. P., Rojas, A. L., Nichol, R. C., Miller, C. J., Wasserman, L., Genovese, C. R., & Freeman, P. E. (2008). Revealing components of the galaxy population through non-parametric techniques. *Monthly Notices of the Royal Astronomical Society*, *391*(2), 607–616.

Barrientos, A. F., Jara, A., & Quintana, F. A. (2017). Fully nonparametric regression for bounded data using dependent Bernstein polynomials. Technical Report 518.

Bayes, C. L., Bazán, J. L., & De Castro, M. (2017). A quantile parametric mixed regression model for bounded response variables. *Statistics and Its Interface*, *10*(3), 483–493.

Bayes, C. L., Bazán, J. L., & García, C. (2012). A new robust regression model for proportions. *Bayesian Analysis*, *7*(4), 841–866.

Boos, D. D., & Stefanski, L. A. (2013). *Essential statistical inference: Theory and methods*. Springer Texts in Statistics, Vol. *120*. New York, NY: Springer Science & Business Media.

Chen, Y.-C., Genovese, C. R., Tibshirani, R. J., & Wasserman, L. (2016). Nonparametric modal regression. *The Annals of Statistics*, *44*(2), 489–514.

Cheng, R., & Amin, N. (1983). Estimating parameters in continuous univariate distributions with a shifted origin. *Journal of the Royal Statistical Society: Series B (Methodological)*, *45*(3), 394–403.

Cox, D. R., & Hinkley, D. V. (1979). *Theoretical statistics*. Boca Raton, FL: Chapman and Hall/CRC.

deToledo Morrell, L., Stoub, T., Bulgakova, M., Wilson, R., Bennett, D., Leurgans, S., … Turner, D. (2004). MRI-derived entorhinal volume is a good predictor of conversion from MCI to AD. *Neurobiology of Aging*, *25*(9), 1197–1203.

Devanand, D., Pradhaban, G., Liu, X., Khandji, A., De Santi, S., Segal, S., … de Leon, M. J. (2007). Hippocampal and entorhinal atrophy in mild cognitive impairment: Prediction of Alzheimer disease. *Neurology*, *68*(11), 828–836.

Einbeck, J., & Tutz, G. (2006). Modelling beyond regression functions: An application of multimodal regression to speed–flow data. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, *55*(4), 461–475.

Ferrari, S., & Cribari-Neto, F. (2004). Beta regression for modelling rates and proportions. *Journal of Applied Statistics*, *31*(7), 799–815.

Figueroa-Zúñiga, J. I., Arellano-Valle, R. B., & Ferrari, S. L. (2013). Mixed beta regression: A Bayesian perspective. *Computational Statistics & Data Analysis*, *61*, 137–147.

García, C. B. G., Pérez, J. G., & Rambaud, S. C. (2009). The generalized biparabolic distribution. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, *17*(3), 377–396.

Guolo, A., & Varin, C. (2014). Beta regression for time series analysis of bounded data, with application to Canada Google® flu trends. *The Annals of Applied Statistics*, *8*(1), 74–88.

Hahn, E. D. (2008). Mixture densities for project management activity times: A robust approach to PERT. *European Journal of Operational Research*, *188*(2), 450–459.

Hall, P. (1992). On global properties of variable bandwidth density estimators. *The Annals of Statistics*, *20*(2), 762–778.

Hämäläinen, A., Tervo, S., Grau-Olivares, M., Niskanen, E., Pennanen, C., Huuskonen, J., … Soininen, H. (2007). Voxel-based morphometry to detect brain atrophy in progressive mild cognitive impairment. *Neuroimage*, *37*(4), 1122–1131.

Ho, C.-S., Damien, P., & Walker, S. (2017). Bayesian mode regression using mixtures of triangular densities. *Journal of Econometrics*, *197*(2), 273–283.

Hotelling, H. (1931). The generalization of student's ratio. *Annals of Mathematical Statistics*, *2*(3), 360–378.

Jack, C. R., Petersen, R. C., Xu, Y. C., O'Brien, P. C., Smith, G. E., Ivnik, R. J., … Kokmen, E. (1999). Prediction of AD with MRI-based hippocampal volume in mild cognitive impairment. *Neurology*, *52*(7), 1397–1397.

Jauhiainen, A. M., Pihlajamäki, M., Tervo, S., Niskanen, E., Tanila, H., Hänninen, T., … Soininen, H. (2009). Discriminating accuracy of medial temporal lobe volumetry and FMRI in mild cognitive impairment. *Hippocampus*, *19*(2), 166–175.

Jørgensen, B. (1987). Exponential dispersion models. *Journal of the Royal Statistical Society: Series B (Methodological)*, *49*(2), 127–145.

Kemp, G. C., & Silva, J. S. (2012). Regression towards the mode. *Journal of Econometrics*, *170*(1), 92–101.

Killiany, R., Hyman, B., Gomez-Isla, T., Moss, M., Kikinis, R., Jolesz, F., … Albert, M. (2002). MRI measures of entorhinal cortex vs hippocampus in preclinical AD. *Neurology*, *58*(8), 1188–1196.

Kueper, J. K., Speechley, M., & Montero-Odasso, M. (2018). The Alzheimers disease assessment scale–cognitive subscale (ADAS-Cog): Modifications and responsiveness in pre-dementia populations. A narrative review. *Journal of Alzheimer's Disease*, *63*(2), 423–444 .

Lee, M.-J. (1989). Mode regression. *Journal of Econometrics*, *542*(3), 337–349.

Lee, M.-J. (1993). Quadratic mode regression. *Journal of Econometrics*, *57*(1–3), 1–19.

Liu, J., Zhang, R., Zhao, W., & Lv, Y. (2013). A robust and efficient estimation method for single index models. *Journal of Multivariate Analysis*, *122*, 226–238.

McCullagh, P., (2019). *Generalized linear models*. London, UK: Routledge.

Migliorati, S., Di Brisco, A. M., & Ongaro, A. (2018). A new regression model for bounded responses. *Bayesian Analysis*, *13*(3), 845–872.

Rosen, W. G., Mohs, R. C., & Davis, K. L. (1984). A new rating scale for Alzheimer's disease. *The American Journal of Psychiatry*, *141*(11), 1356–1364.

Schmidt, M., (1996). *Rey auditory verbal learning test: A handbook*. Los Angeles, CA: Western Psychological Services.

Smithson, M., & Verkuilen, J. (2006). A better lemon squeezer? Maximum-likelihood regression with beta-distributed dependent variables. *Psychological methods*, *11*(1), 54.

Tombaugh, T. N., & McIntyre, N. J. (1992). The mini-mental state examination: A comprehensive review. *Journal of the American Geriatrics Society*, *40*(9), 922–935.

Verkuilen, J., & Smithson, M. (2012). Mixed and mixture regression models for continuous bounded responses using the beta distribution. *Journal of Educational and Behavioral Statistics*, *37*(1), 82–113.

Wang, X., Chen, H., Cai, W., Shen, D., & Huang, H. (2017). Regularized modal regression with applications in cognitive impairment prediction. In I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, & R. Garnett (Eds.), *Advances in neural information processing systems 30* (pp. 1448–1458). Red Hook, NY: Curran Associates.

Whitwell, J. L., Shiung, M. M., Przybelski, S., Weigand, S. D., Knopman, D. S., Boeve, B. F., … Jack, C. (2008). MRI patterns of atrophy associated with progression to AD in amnestic mild cognitive impairment. *Neurology*, *70*(7), 512–520.

Yang, H., & Yang, J. (2014). A robust and efficient estimation and variable selection method for partially linear single-index models. *Journal of Multivariate Analysis*, *129*, 227–242.

Yao, W., & Li, L. (2014). A new regression model: Modal linear regression. *Scandinavian Journal of Statistics*, *41*(3), 656–671.

Yu, K., & Aristodemou, K. (2012). Bayesian mode regression. *arXiv preprint arXiv:1208.0579*.

Zhang, R., Zhao, W., & Liu, J. (2013). Robust estimation and variable selection for semiparametric partially linear varying coefficient model based on modal regression. *Journal of Nonparametric Statistics*, *25*(2), 523–544.

Zhao, W., Zhang, R., Liu, J., & Lv, Y. (2014). Robust and efficient variable selection for semiparametric partially linear varying coefficient model based on modal regression. *Annals of the Institute of Statistical Mathematics*, *66*(1), 165–191.

Zhou, H., & Huang, X. (2019). Bandwidth selection for nonparametric modal regression. *Communications in Statistics—Simulation and Computation*, *48*, 968–984.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of the article.

How to cite this article: Zhou H, Huang X, for the Alzheimer's Disease Neuroimaging Initiative. Parametric mode regression for bounded responses. *Biometrical Journal*. 2020;1–19. https://doi.org/10.1002/bimj.202000039