Statistica Sinica **33** (2023), 1295-1318
doi:https://doi.org/10.5705/ss.202021.0006

# GLOBALLY ADAPTIVE LONGITUDINAL QUANTILE REGRESSION WITH HIGH DIMENSIONAL COMPOSITIONAL COVARIATES

Huijuan Ma[1], Qi Zheng[2], Zhumin Zhang[3], Huichuan Lai[3] and Limin Peng[4]

[1]*East China Normal University,* [2]*University of Louiswille,*
[3]*University of Wisconsin-Madison and* [4]*Emory University*

*Abstract:* In this work, we propose a longitudinal quantile regression framework that enables a robust characterization of heterogeneous covariate-response associations in the presence of high-dimensional compositional covariates and repeated measurements of both the response and the covariates. We develop a globally adaptive penalization procedure that can consistently identify covariate sparsity patterns across a continuum set of quantile levels. The proposed estimation procedure properly aggregates longitudinal observations over time, and satisfies the sum-zero coefficient constraint needed for a proper interpretation of the effects of compositional covariates. We establish the oracle rate of the uniform convergence and weak convergence of the resulting estimators, and further justify the proposed uniform selector of the tuning parameter in terms of achieving global model selection consistency. We derive an efficient algorithm by incorporating existing R packages to facilitate stable and fast computation. Our extensive simulation studies confirm our theoretical findings. We apply the proposed method to a longitudinal study of cystic fibrosis children, where the associations between the gut microbiome and other diet-related biomarkers are of interest.

*Key words and phrases:* Compositional covariates, globally adaptive penalization, longitudinal data, quantile regression.

## 1. Introduction

Compositional data are frequently encountered in a variety of research fields. Examples include household expenditure compositions in economics, geochemical compositions of rocks in geology, and human microbiome compositions in medical studies. Compositional data consist of proportions bounded between zero and one and sum to one, and are often high dimensional. For instance, human microbiome data are usually captured as percentages (or the relative abundance) of gene sequencing reads (Tyler, Smith and Silverberg (2014)) at a certain

Corresponding author: Limin Peng, Department of Biostatistics and Bioinformatics, Emory University, Atlanta, GA 30322, USA. E-mail: lpeng@emory.edu.

taxonomy level, and the number of operational taxonomy units (e.g., phyla or genus) can range over hundreds, thousands, or even millions. With advancements in technology, an increasing number of studies are collecting such compositional data repeatedly over time. A common question of substantive interest is how these longitudinal compositional measurements are associated with other longitudinal biomarkers or clinical outcomes. This poses a regression problem subject to multiple complications, including a large number of covariates, positiveness and unit-sum constraints on the covariates, and within-subject dependence of the longitudinal observations.

To deal with the high dimensionality of the covariates, a notable line of research has been established in the penalization framework (e.g., Meinshausen and Buhlmann (2006); Zhang and Huang (2008); Kim, Choi and Oh (2008); Lv and Fan (2009); Fan and Lv (2011)). Extensions to longitudinal settings have also been developed (e.g., Wang, Zhou and Qu (2012); Zheng et al. (2018)). When covariates are compositional, given the unit-sum constraint, an increase in one covariate must induce a decrease in another covariate. Applying traditional penalization regression methods without accounting for the compositional nature of the covariates may lead to results that are difficult to interpret. A common strategy for accommodating compositional covariates is to apply a sensible operation to the compositional proportions before incorporating them into a regression model, as in the linear log-contrast model and logistic normal multinomial regression model (Aitchison (1982); Aitchison and Bacon-shone (1984); Aitchison (2003); Xia et al. (2013)). Many studies have focused on covariates that are both compositional and high dimensional. For example, Lin et al. (2014) proposed a Lasso-penalized method for the linear log-contrast regression model that properly accounts for the compositional nature of the covariates. Shi, Zhang and Li (2016) studied an extension of the model of Lin et al. (2014) with a set of linear constraints. Lu, Shi and Li (2019) generalized the model further to a generalized linear log-contrast model, and proposed an $l_1$-penalized likelihood estimation procedure.

However, few works have proposed methods for high-dimensional compositional covariates in a longitudinal setting. Moreover, most existing approaches use a mean-based linear regression, which typically confines covariate effects to be location shifts, and thus can be restrictive for real data. The quantile regression (Koenker and Bassett (1978)), characterized by its flexibility when assessing covariate effects across different quantile levels, has demonstrated promising utility for identifying and depicting dynamic covariate-response associations that often provide useful scientific insights. The modeling strategy of the quantile regression has been incorporated in analyses of longitudinal data under various perspectives (e.g., Koenker (2004); Wang and Fygenson (2009); Ma, Peng and Fu (2019)). In the presence of high-dimensional covariates, many studies (e.g., Li, Liu and Zhu (2007); Zou and Yuan (2008); Wang, Wu and Li (2012); Zheng, Gallagher and Kulasekera (2013); Fan, Fan and Barut (2014)) have examined penalized quantile regression methods. These methods model a single or multiple prespecified quantiles of the response; in other words, are locally concerned. These methods are subject to inherent problems, such as undesirable variability in the variable selection results across neighboring quantile levels, and the potential failure to detect some important variables, owing to an off-target selection of the quantile levels. To address these limitations, Zheng, Peng and He (2015) proposed the perspective of globally concerned quantile regression that enables a simultaneous examination of regression quantiles over a continuum set of quantile levels, and thus reflects the underlying scientific interest in a more robust way. However, although it demonstrates improved stability and "power" of variable selection compared with locally concerned quantile regression approaches, their method is not suitable for handling either longitudinal data or compositional covariates.

In this work, we develop a globally concerned longitudinal quantile regression framework that is tailored to evaluate the effects of high-dimensional longitudinal compositional covariates on longitudinal responses. We consider a longitudinal linear log-contrast quantile regression model, where quantiles of the longitudinal response are linked to the log contrasts of the corresponding compositional covariates. To avoid the shortcomings associated with selecting an irrelevant covariate as the reference in the logcontrasts, we reformulate the model into a symmetric form with a zero-sum constraint of the coefficients, which ensures sensible interpretations of the effects of the compositional covariates. We propose a regularization method, in which a globally adaptive Lasso penalty is imposed on the longitudinal quantile loss function that appropriately aggregates repeated measurements from the same subject. We adapt the *rq.fit.fnc()* function in the existing R package *quantreg* to facilitate the estimation in the presence of the zero-sum constraint of the coefficients.

We conduct theoretical studies for the proposed method in the ultrahigh-dimensional setting, where the number of covariates $p$ may increase exponentially with the sample size $n$ (i.e., $\log p = o(n^b)$, for some $b > 0$), and the number of relevant covariates $s$ also increases with $n$. We attain the uniform convergence rate of the proposed estimator as $O_p(\sqrt{s \log n/n})$, which is the fastest possible rate. Because the longitudinal quantile loss function is not differentiable, to attain this result, we cannot adapt existing works on linear regression-based methods for high-dimensional compositional data, such as that of Lin et al. (2014), which penalizes a smooth least-squares loss function. Instead, we employ theoretical techniques, including chaining theory (Talagrand (2005)), the contraction inequality (Ledoux and Talagrand (1991)), and the empirical process (van der Vaart and Wellner (1996)), as in Zheng, Peng and He (2015). However, these do not address the longitudinal data structure and the compositional constraint for high-dimensional covariates. Therefore, we develop new arguments to account for these special data features. Notably, we properly formulate and establish a crucial Karush–Kuhn–Tucker (KKT) condition tailored to compositional data, which is new in the literature. In addition, we thoroughly justify that penalizing the proposed longitudinal quantile loss function, which adopts the simple working independence assumption, is capable of accommodating longitudinal data with dependent repeated measures.

Our theoretical studies provide useful results not discussed in existing works on high-dimensional compositional covariates based on log-contrast models, such as those of Lin et al. (2014) and Shi, Zhang and Li (2016). For example, our theoretical investigation reveals that the asymptotic behavior of the globally adaptive estimator based on a constrained linear log-contrast quantile regression model is asymptotically equivalent to its unconstrained counterpart, as long as the reference variable for the latter is a truly relevant variable, which is usually not known in advance. In addition, we establish the weak convergence of any linear combination of the proposed estimator to a Gaussian process. We develop a GIC-type uniform tuning parameter selector. We show that the proposed estimation and tuning parameter procedures can correctly identify all globally relevant variables with probability tending to one (i.e., global model selection consistency).

The remainder of this paper proceeds as follows. In Section 2, we introduce a globally concerned framework built on a longitudinal linear log-contrast quantile regression model. Then, we propose a globally adaptive regularization procedure based on a symmetric model representation with a zero-sum coefficient constraint. In Section 3, we present the asymptotic studies for the proposed estimation procedure. In Section 4, we investigate the finite-sample performance of proposed method using simulations. Finally, we demonstrate our methodology by applying it to a longitudinal observational study of children with cystic fibrosis (CF).

## 2. Methodology

### 2.1. Longitudinal linear log-contrast quantile regression model

Consider a longitudinal study with $n$ subjects. Let $Y_i(t)$, $\mathbf{X}_i(t)$, and $\mathbf{W}_i(t)$ denote the longitudinal response, an $r \times 1$ vector of regular covariates with one as the first component, and a $p \times 1$ vector of compositional covariates at time $t$ for subject $i$ $(i = 1, \ldots, n)$, respectively. A component of $\mathbf{X}_i(t)$ may flexibly represent the value of a time-dependent covariate measured at time $t$ or a summary of the covariate history up to time $t$. We consider the setting where $r$ is fixed and $p$ increases with $n$, satisfying $\log p = o(n^b)$, for some $b > 0$. At each time point $t$, the compositional covariates in $\mathbf{W}_i(t)$ are subject to the unit-sum constraint. That is, $\mathbf{W}_i(t)$ belongs to the $(p-1)$-dimensional positive simplex $\mathbb{S}^{p-1} = \{(w_1, \ldots, w_p) : \; w_j > 0, \; j = 1, \ldots, p; \; \sum_{j=1}^{p} w_j = 1\}$. Suppose $Y_i(t)$, $\mathbf{X}_i(t)$, and $\mathbf{W}_i(t)$ are observed at $m_i$ time points, denoted by $\{t_i^{(k)}, k = 1, \ldots, m_i\}$. Define a counting process for the observation time as $N_i(t) = \sum_{k=1}^{m_i} I(t_i^{(k)} \le t)$.

To obtain a comprehensive and flexible view of how the covariates influence the response, we use quantile regression modeling to formulate the covariate effects on the $\tau$th conditional quantile of $Y(t)$ given $\mathbf{X}(t)$ and $\mathbf{W}(t)$, which is defined as $Q_{Y(t)}\{\tau|\mathbf{X}(t), \mathbf{W}(t)\} = \inf\{y : \Pr\{Y(t) \le y|\mathbf{X}(t), \mathbf{W}(t)\} \ge \tau\}$. However, plugging $\mathbf{W}(t)$ directly into a regression model is problematic, because the components of $\mathbf{W}(t)$ cannot change freely, owning to the unit-sum constraint, making it difficult to interpret the coefficients of $\mathbf{W}(t)$. To deal with the unit-sum constraint, we apply the log-contrast (or log-ratio) transformation of Aitchison and Bacon-shone (1984), which transforms the compositional $\mathbf{W}_i(t)$ from $\mathbb{S}^{p-1}$ to $\mathbf{Z}_i^p(t) \doteq \{\log\{W_{i1}(t)/W_{ip}(t)\}, \ldots, \log\{W_{i,p-1}(t)/W_{ip}(t)\}\}^\top$, where $W_{ij}(t)$ denotes the $j$th component of $\mathbf{W}_i(t)$. The transformation from $\mathbf{W}(t)$ to $\mathbf{Z}_i^p(t)$ is one-to-one and $\mathbf{Z}_i^p(t)$ is freely ranged in $\mathbb{R}^{p-1}$ without any constraint. A log-contrast transformation requires selecting a reference covariate. For $\mathbf{Z}_i^p(t)$, the $p$th component of $\mathbf{W}(t)$, $W_{ip}(t)$, serves as the reference.

We consider the following longitudinal linear log-contrast quantile regression model:

$$Q_{Y_i(t)}\{\tau|\mathbf{X}_i(t), \mathbf{W}_i(t)\} = \mathbf{X}_i(t)^\top \boldsymbol{\alpha}_0(\tau) + \mathbf{Z}_i^p(t)^\top \boldsymbol{\beta}_{0,\setminus p}(\tau) \quad \text{for} \quad \tau \in \Delta, \qquad (2.1)$$

where $\boldsymbol{\alpha}_0(\tau)$ is an $r \times 1$ vector of regression coefficients for $\mathbf{X}_i(t)$, $\boldsymbol{\beta}_{0,\setminus p}(\tau) \doteq \{\beta_{0,1}(\tau), \ldots, \beta_{0,p-1}(\tau)\}^\top$ is a $(p-1) \times 1$ vector of regression coefficients for $\mathbf{Z}_i^p(t)$, and $\Delta \subset (0, 1)$ is a set of quantile levels, prespecified to align with the scientific problem of interest. For example, if we need to identify the variables affecting the center of the response distribution, we can choose $\Delta = [0.4, 0.6]$. If we are interested in the upper tail of the response distribution, we can choose $\Delta = [0.75, 0.9]$. A subtle drawback of model (2.1) is that any variable selection based on the model automatically includes $W_{ip}(t)$ as a relevant covariate, even when $W_{ip}(t)$ is not a relevant variable.

Following the strategy employed in the linear regression setting with compositional covariates (Lin et al. (2014); Shi, Zhang and Li (2016)), we define $\beta_{0,p}(\tau) = -\sum_{j=1}^{p-1} \beta_{0,j}(\tau)$, and re-express model (2.1) as

$$Q_{Y_i(t)}\{\tau|\mathbf{X}_i(t), \mathbf{Z}_i(t)\} = \mathbf{X}_i(t)^\top \boldsymbol{\alpha}_0(\tau) + \mathbf{Z}_i(t)^\top \boldsymbol{\beta}_0(\tau), \quad \text{subject to} \quad \sum_{j=1}^{p} \beta_{0,j}(\tau) = 0, \quad \text{for} \quad \tau \in \Delta. \tag{2.2}$$

Here, $\mathbf{Z}_i(t) = \{\log\{W_{i1}(t)\}, \ldots \log\{W_{ip}(t)\}\}^\top$, and $\boldsymbol{\beta}_0(\tau) = \{\beta_{0,1}(\tau), \ldots, \beta_{0,p-1}(\tau), \beta_{0,p}(\tau)\}^\top$, with $\beta_{0,j}(\tau)$ denoting the $j$th component of $\boldsymbol{\beta}_0(\tau)$. Unlike model (2.1), model (2.2) takes a symmetric form, and does not require choosing the reference covariate. The symmetric form of model (2.2) also enables an estimation that possesses desirable properties such as scale invariance, permutation invariance, and selection invariance (Aitchison (1982); Lin et al. (2014)).

Many longitudinal quantile regression models studied in literature (e.g., Lipsitz et al. (1997); Wang and Fygenson (2009); Sun et al. (2016); Cho, Hong and Kim (2016); Gao and Liu (2020)) bear similar forms to model (2.1) or (2.2), but they do not involve the zero-sum coefficient constraint and were investigated under the locally concerned perspective.

We study a globally concerned framework based on the longitudinal quantile regression model (2.2), where a covariate is considered relevant if it has nonzero effects on the conditional quantiles of $Y(t)$ at some, not necessarily all, quantile levels in $\Delta$. That is, the set of relevant (or active) compositional covariates is defined as

$$S_\Delta = \{j \in \{1, \ldots, p\} : \exists\ \tau \in \Delta,\ |\beta_{0,j}(\tau)| > 0\}.$$

It is clear that $S_\tau \doteq S_{\{\tau\}} \subset S_\Delta$ when $\tau \in \Delta$. The globally concerned perspective warrants a global sparsity assumption, that is, $s \doteq |S_\Delta| = o(n)$, for model identifiability purposes, where $|\cdot|$ denotes the cardinality.

### 2.2. Globally adaptive $L_1$ penalized estimation

The observed longitudinal data can be generally formulated as $\{(Y_i(t)dN_i(t), \mathbf{X}_i(t)dN_i(t), \mathbf{Z}_i(t)dN_i(t)), i = 1, \ldots, n\}$. When $p$ is fixed, model (2.2) without the zero-sum coefficient constraint can be estimated by minimizing the longitudinal quantile loss function,

$$Q(\boldsymbol{\alpha},\boldsymbol{\beta};\tau)=\frac{1}{n}\sum_{i=1}^{n}\int_{0}^{\infty}\rho_{\tau}\{Y_i(t)-\mathbf{X}_i(t)^{\top}\boldsymbol{\alpha}-\mathbf{Z}_i(t)^{\top}\boldsymbol{\beta}\}dN_i(t),$$

where $\rho_\tau(t)=t(\tau-I\{t\le 0\})$ is the $\tau$th quantile loss function. From the definition, $Q(\boldsymbol{\alpha},\boldsymbol{\beta};\tau)$ takes an equal weight summation of the quantile loss function, assessed at all within-subject observation time points. This mimics the idea of constructing a generalized estimating equation (GEE) for longitudinal data under the working independence assumption (Liang and Zeger (1986)). The same strategy is adopted in existing works on longitudinal quantile regressions (e.g., Wang and Fygenson (2009); Sun et al. (2016)). Estimations based on $Q(\boldsymbol{\alpha},\boldsymbol{\beta};\tau)$, like the GEE approach, can properly accommodate longitudinal data with correlated repeated measures.

We propose applying the adaptively weighted $L_1$ regularization to $Q(\boldsymbol{\alpha},\boldsymbol{\beta};\tau)$ to address the high dimensionality of $\mathbf{Z}_i(t)$. This renders a regression coefficient estimator $\hat{\boldsymbol{\gamma}}(\tau)$ as a solution to the following constrained minimization problem:

$$\hat{\boldsymbol{\gamma}}(\tau)\doteq(\hat{\boldsymbol{\alpha}}(\tau)^{\top},\hat{\boldsymbol{\beta}}(\tau)^{\top})^{\top}=\underset{\boldsymbol{\alpha},\boldsymbol{\beta},\sum_{j=1}^{p}\beta_j=0}{\operatorname{argmin}}\left(Q(\boldsymbol{\alpha},\boldsymbol{\beta};\tau)+\lambda\sum_{j=1}^{p}\omega_j(\tau)|\beta_j|\right). \quad (2.3)$$

Aligning with the perspective of globally concerned quantile regression, $\lambda$ is a tuning parameter that is constant over $\tau$ and controls for the global sparsity over $\tau\in\Delta$, namely, $S_\Delta$. Here, $\omega_j(\tau)$ is a nonnegative adaptive weight function that gauges the importance of $Z_{ij}(t)$, the $j$th component of $\mathbf{Z}_i(t)$, for $j=1,\ldots,p$. The adaptive weights may take the following forms: $(w1)$ $\omega_j(\tau)=1/|\check{\boldsymbol{\beta}}_j(\tau)|$; $(w2)$ $\omega_j(\tau)=1/\big(\sup_{\tau\in\Delta}|\check{\boldsymbol{\beta}}_j(\tau)|\big)$; $(w3)$ $\omega_j(\tau)=1/\int_\Delta|\check{\boldsymbol{\beta}}_j(\tau)|d\tau$, where $\check{\boldsymbol{\beta}}(\tau)$ is a uniformly consistent estimator of $\boldsymbol{\beta}_0(\tau)$. As discussed in Zheng, Peng and He (2015), $(w2)$ and $(w3)$ are globally adaptive weights that capture the global impact of a covariate, and may be theoretically and empirically preferable. A uniformly consistent estimator $\check{\boldsymbol{\beta}}(\tau)$ can be obtained by directly adapting the approach of Belloni and Chernozhukov (2011) to high-dimensional longitudinal compositional data (i.e., solving the minimization problem (2.3) with the penalty term and tuning parameter selector presented by Belloni and Chernozhukov (2011)). This can be justified by slightly modifying the proof of Theorem 1 (Section 3), combined with the techniques of Belloni and Chernozhukov (2011).

To solve the constrained minimization problem in (2.3), we first write the objective function as a classical quantile loss function. Let $\mathbf{e}_j$ be a $p$-dimensional vector with the $j$th component equal to one and all others equal to zero, for $j = 1, \ldots, p$. In addition, for any integer $m \geq 2$, denote the $m$-vector of ones and zeros by $\mathbf{1}_m$ and $\mathbf{0}_m$, respectively. Because $\rho_\tau(u) + \rho_\tau(-u) = |u|$,

$$\lambda \sum_{j=1}^{p} \omega_j(\tau)|\beta_j| = \sum_{j=1}^{p} \{\rho_\tau(Y_j^* - \mathbf{X}_j^{*\top}\boldsymbol{\alpha} - \mathbf{Z}_j^{*\top}\boldsymbol{\beta}) + \rho_\tau(Y_{p+j}^* - \mathbf{X}_{p+j}^{*}{}^\top\boldsymbol{\alpha} - \mathbf{Z}_{p+j}^{*}{}^\top\boldsymbol{\beta})\},$$

where $(Y_j^*, \mathbf{X}_j^*, \mathbf{Z}_j^*) = (0, \mathbf{0}_r, \lambda\omega_j(\tau)\mathbf{e}_j)$ and $(Y_{p+j}^*, \mathbf{X}_{p+j}^*, \mathbf{Z}_{p+j}^*) = (0, \mathbf{0}_r, -\lambda\omega_j(\tau)\mathbf{e}_j)$. Letting $\boldsymbol{\gamma} = (\boldsymbol{\alpha}^\top, \boldsymbol{\beta}^\top)^\top$, we then formulate the equality constraint $\sum_{j=1}^p \beta_j = 0$ as two inequality constraints, $\sum_{j=1}^p \beta_j \geq 0$ (or expressed as $(\mathbf{0}_r^\top, \mathbf{1}_p^\top)^\top\boldsymbol{\gamma} \geq 0$ in matrix form) and $-\sum_{j=1}^p \beta_j \geq 0$ (or expressed as $(\mathbf{0}_r^\top, -\mathbf{1}_p^\top)^\top\boldsymbol{\gamma} \geq 0$ in matrix form). Then, the quantile regression problem in (2.3) with the linear inequality constraints can be solved using the existing function *rq.fit.fnc()* in the R package *quantreg*, and the augmented data set $\{Y_i(t_i^{(k)}), \mathbf{X}_i(t_i^{(k)}), \mathbf{Z}_i(t_i^{(k)}),\ k = 1, \ldots, m_i; i = 1, \ldots, n\}$, coupled with $\{(Y_j^*, \mathbf{X}_j^*, \mathbf{Z}_j^*), (Y_{p+j}^*, \mathbf{X}_{p+j}^*, \mathbf{Z}_{p+j}^*),\ j = 1, \ldots, p\}$.

The set of relevant compositional covariates, $S_\Delta$, is estimated by

$$\hat{S}_\Delta \doteq \{j \in \{1, \ldots, p\} : \exists\ \tau \in \Delta, |\hat{\beta}_j(\tau)| > 0\}.$$

### 2.3. Tuning parameter selection

Tuning parameter selection plays an important role in variable selection. In the globally concerned setting, a critical idea is to set $\lambda$ as a common tuning parameter across all $\tau \in \Delta$ as a means to control the overall model complexity and avoid overall fitting. We adapt the generalized information criterion (GIC) (Nishii (1984); Fan and Tang (2013)) to the setting of globally concerned longitudinal quantile regression with compositional covariates.

Specifically, we propose the following uniform selector of the tuning parameter by minimizing

$$\mathrm{GIC}(\lambda) = \int_\Delta \log \hat{\sigma}_\lambda(\tau) d\tau + (|\hat{S}_\lambda| - 1)\phi_n,$$

where $\hat{S}_\lambda = \{j \in \{1, \ldots, p\} : \sup_{\tau\in\Delta} |\hat{\beta}_{j,\lambda}(\tau)| \neq 0\}$,

$$\hat{\sigma}_\lambda(\tau) = \frac{1}{n}\sum_{i=1}^{n} \int_0^\infty \rho_\tau\{Y_i(t) - \mathbf{X}_i(t)^\top\hat{\boldsymbol{\alpha}}_\lambda(\tau) - \mathbf{Z}_i(t)^\top\hat{\boldsymbol{\beta}}_\lambda(\tau)\}dN_i(t),$$

and $\phi_n$ is a sequence converging to zero with $n$. Here, $\hat{\beta}_{j,\lambda}(\tau)$, $\hat{\boldsymbol{\alpha}}_\lambda(\tau)$, and $\hat{\boldsymbol{\beta}}_\lambda(\tau)$ represent the proposed estimates for $\beta_j(\tau)$, $\boldsymbol{\alpha}(\tau)$, and $\boldsymbol{\beta}(\tau)$, respectively, with the tuning parameter fixed at $\lambda$. A popular choice of $\phi_n$ is $n^{-1}\log(p)\log(\log(n))$.

Note that the model size pertaining to the compositional covariates is $|\hat{S}_\lambda| - 1$, owning to the zero-sum constraint.

As shown in Theorem 3, with a properly chosen $\phi_n$ and a reasonable upper bound imposed on the model size, the proposed tuning parameter $\hat{\lambda}$, which is the minimizer of GIC($\lambda$) with respect to $\lambda$, can consistently identify the true model $S_\Delta$. In other words, with probability tending to one, $\hat{S}_{\hat{\lambda}} = S_\Delta$.

### 2.4. Grid-based approximation

With finite sample sizes, minimizing (2.3) for all $\tau \in \Delta$ yields estimates that are exactly piecewise constant functions of $\tau$. However, although the exact breakpoints of these piecewise constant functions can be identified by adapting the procedure of Koenker and d'Orey (1987) and Portnoy (1991), the computation expense can be overwhelming in the ultrahigh-dimensional cases. Therefore, we approximate $\hat{\boldsymbol{\alpha}}(\cdot)$ and $\hat{\boldsymbol{\beta}}(\cdot)$ using piecewise constant functions that jump only at the grid points of a prespecified sufficiently fine $\tau$-grid in $\Delta$ to alleviate the computation burden. Let $\mathcal{S}_n$ denote the $\tau$-grid in $\Delta$, for $\tau_0 < \tau_1 < \ldots < \tau_{M(n)}$, and define the size of $\mathcal{S}_n$ as $\|\mathcal{S}_n\| = \max\{\tau_k - \tau_{k-1} : k = 1, \ldots, M(n)\}$. The grid-based approximations are given by $\hat{\boldsymbol{\alpha}}^{\mathcal{S}_n}(\cdot) = \sum_{k=1}^{M(n)} \hat{\boldsymbol{\alpha}}(\tau_k) I(\tau_{k-1} < \tau \leq \tau_k)$, and $\hat{\boldsymbol{\beta}}^{\mathcal{S}_n}(\cdot) = \sum_{k=1}^{M(n)} \hat{\boldsymbol{\beta}}(\tau_k) I(\tau_{k-1} < \tau \leq \tau_k)$. With a certain smoothness assumption for $\boldsymbol{\alpha}_0(\cdot)$ and $\boldsymbol{\beta}_0(\cdot)$, we can show that $(\hat{\boldsymbol{\alpha}}^{\mathcal{S}_n}(\cdot)^\top, \hat{\boldsymbol{\beta}}^{\mathcal{S}_n}(\cdot)^\top)^\top$ and $(\hat{\boldsymbol{\alpha}}(\cdot)^\top, \hat{\boldsymbol{\beta}}(\cdot)^\top)^\top$ have the same convergence rate and asymptotic distribution if $\|\mathcal{S}_n\|$ converges to zero at the rate $o((ns)^{-1/2})$.

## 3. Theoretical Results

Without loss of generality, we assume that $r$, the number of usual covariates, is finite. Let $S_\Delta = \{1, \ldots, s\}$ and use $S_\Delta^c = \{s+1, \ldots, p\}$ to denote the collection of all irrelevant compositional variables. We allow the number of compositional covariates $p_n \doteq p$ and the true model size $s_n \doteq s$ to increase with the sample size $n$. For ease of presentation, we often omit the subscript $n$ when it is clear from the context.

Let $\mathbf{V}_i(t) = (\mathbf{X}_i(t)^\top, \mathbf{Z}_i(t)^\top)^\top$ and $\boldsymbol{\gamma}(\tau) = (\boldsymbol{\alpha}(\tau)^\top, \boldsymbol{\beta}(\tau)^\top)^\top$, satisfying $\sum_{j=1}^{p} \beta_j(\tau) = 0$. Thus, $\boldsymbol{\gamma}_0(\tau) = (\boldsymbol{\alpha}_0(\tau)^\top, \boldsymbol{\beta}_0(\tau)^\top)^\top$. We decompose $\mathbf{Z}_i(t)$ into $(\mathbf{Z}_{ia}(t)^\top, \mathbf{Z}_{ib}(t)^\top)^\top$ and $\mathbf{V}_i(t)$ into $(\mathbf{V}_{ia}(t)^\top, \mathbf{V}_{ib}(t)^\top)^\top$, where $\mathbf{Z}_{ia}(t) = (Z_{i,1}(t), \ldots, Z_{i,s}(t))^\top$, $\mathbf{V}_{ia}(t) = (\mathbf{X}_i(t)^\top, \mathbf{Z}_{ia}(t)^\top)^\top$, and $\mathbf{V}_{ib}(t) = \mathbf{Z}_{ib}(t) = (Z_{i,s+1}(t), \ldots, Z_{i,p}(t))^\top$. Similarly, $\boldsymbol{\beta}(\tau) = (\boldsymbol{\beta}_a(\tau)^\top, \boldsymbol{\beta}_b(\tau)^\top)^\top$ and $\boldsymbol{\gamma}(\tau) = (\boldsymbol{\gamma}_a(\tau)^\top, \boldsymbol{\gamma}_b(\tau)^\top)^\top$, where $\boldsymbol{\beta}_a(\tau) = (\beta_1(\tau), \ldots, \beta_s(\tau))^\top$, $\boldsymbol{\gamma}_a(\tau) = (\boldsymbol{\alpha}(\tau)^\top, \boldsymbol{\beta}_a(\tau)^\top)^\top$, and $\boldsymbol{\gamma}_b(\tau) = \boldsymbol{\beta}_b(\tau) = (\beta_{s+1}(\tau), \ldots, \beta_p(\tau))^\top$. The regularity conditions (C1)–(C5) are stated in Section S1 of the Supplementary

Material.

In Theorem 1, we show that the proposed estimator is uniformly consistent over $\Delta$ with the convergence rate $O_p(\sqrt{(r+s)\log n/n})$, which is the fastest possible and is as good as that of the globally adaptive estimator of Zheng, Peng and He (2015). For a single $\tau$ or a finite number of $\tau$, we establish a faster convergence rate, $O_p(\sqrt{(r+s)/n})$, as stated in Corollary 1.

**Theorem 1.** *Suppose conditions* (C1)–(C5) *(stated in the Supplementary Material) hold. Furthermore, we assume that* $n/((r+s)^3 \log^2 \max\{n, r+p\}) \to \infty$ *and*

$$\sup_{j>r+s, \boldsymbol{\delta} \in R_{r+s-1}} \frac{E[\int_0^\infty \{V_{ij}(t)\mathbf{V}_i(t)^\top \boldsymbol{\delta}\}^2 dN_i(t)]}{\|\boldsymbol{\delta}\|^2} = o\left(\frac{\log \max\{n, r+p\}}{(r+s)\log n}\right).$$

*If* $r+s = o(n^{1/3})$, $\sup_{j\in S_\Delta, \tau\in\Delta} \lambda w_j(\tau) = O_p(\sqrt{n\log n})$, $\lambda/(\sqrt{r+s}\log\max\{n, r+p\}) \to \infty$, *and* $(\inf_{j>r+s,\tau\in\Delta} w_j(\tau))^{-1}\sqrt{n}/\sqrt{(r+s)\log\max\{n,r+p\}} = O_p(1)$, *then the proposed estimator satisfies*

$$\sup_{\tau\in\Delta} \|\hat{\boldsymbol{\gamma}}(\tau) - \boldsymbol{\gamma}_0(\tau)\| = O_p\left(\sqrt{\frac{(r+s)\log n}{n}}\right).$$

**Corollary 1.** *Suppose the conditions in Theorem* 1 *hold. Then, the proposed estimator satisfies*

$$\|\hat{\boldsymbol{\gamma}}(\tau_0) - \boldsymbol{\gamma}_0(\tau_0)\| = O_p\left(\sqrt{\frac{r+s}{n}}\right).$$

In Theorem 2, we establish the weak convergence of the proposed estimator.

**Theorem 2.** *Suppose the conditions in Theorem* 1 *hold. If* $(r+s)^3 \log^4 n = o(n)$, *for any given* $\boldsymbol{\xi} \in R_{r+s-1}$ *and* $\|\boldsymbol{\xi}\| = 1$, *we have the following results:*

*(a) If* $\sqrt{n/\{(r+s)\log n\}} \inf_{1\le j\le s, \tau\in\Delta} |\beta_{0j}(\tau)| \to \infty$, *then*

$$n^{1/2}\boldsymbol{\xi}^\top \left[\mathbf{H}_\tau \{\hat{\boldsymbol{\gamma}}(\tau) - \boldsymbol{\gamma}_0(\tau)\} + \frac{\lambda}{n}\boldsymbol{\varpi}(\tau)\right]$$

*converges weakly to a mean-zero Gaussian process with covariance*

$$\begin{aligned}\boldsymbol{\Sigma}(\tau,\tau') = {} & E\{h_{n,\boldsymbol{\xi},\tau}(\mathbf{V}(t), Y) h_{n,\boldsymbol{\xi},\tau'}(\mathbf{V}(t), Y)\} \\ & - E\{h_{n,\boldsymbol{\xi},\tau}(\mathbf{V}(t), Y)\} E\{h_{n,\boldsymbol{\xi},\tau'}(\mathbf{V}(t), Y)\},\end{aligned}$$

*where* $h_{n,\boldsymbol{\xi},\tau}(\mathbf{V}(t), Y) = \int_0^\infty \boldsymbol{\xi}^\top \mathbf{V}(t)\psi_\tau\{Y(t) - \mathbf{V}(t)^\top \boldsymbol{\gamma}_0(\tau)\} dN(t)$, $\psi_\tau(u) = \tau - I(u<0)$,

$$\mathbf{H}_\tau = \begin{pmatrix} E[\int_0^\infty f_{t,\tau}\{0|\mathbf{V}_i(t)\}\mathbf{V}_{ia}(t)\mathbf{V}_{ia}(t)^\top dN_i(t)] & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix},$$

$\boldsymbol{\varpi}(\tau) = \left(\mathbf{0}_r^\top, (\boldsymbol{\omega}(\tau) \circ sign(\boldsymbol{\beta}_0(\tau)))^\top, \mathbf{0}_{p-s}^\top\right)^\top$, $\circ$ *denotes the Hadamard product, and* $\boldsymbol{\omega}(\tau) = (\omega_1(\tau), \ldots, \omega_p(\tau))^\top$*;*

*(b) If* $\sup_{\tau\in\Delta} n^{-1/2}\{\sum_{j\in S_\tau} \lambda w_j^2(\tau)\}^{1/2} = o_p(1)$*, then* $n^{1/2}\boldsymbol{\xi}^\top \mathbf{H}_\tau\{\hat{\gamma}(\tau) - \boldsymbol{\gamma}_0(\tau)\}$ *converges weakly to a mean-zero Gaussian process with covariance* $\boldsymbol{\Sigma}(\tau, \tau')$*.*

To establish the asymptotic properties of the GIC tuning parameter selector, we assume the following condition (C5+), which is an enhanced version of (C5) presented in the Supplementary Material:

(C5+) (a)

$$\begin{aligned} 0 < \Lambda_{\min} &:= \inf_{\boldsymbol{\delta}\in R_\ell, \ell\le r+\kappa, \boldsymbol{\delta}\neq\mathbf{0}} \frac{\boldsymbol{\delta}^\top E[\int_0^\infty \mathbf{V}_i(t)\mathbf{V}_i(t)^\top dN_i(t)]\boldsymbol{\delta}}{\|\boldsymbol{\delta}\|^2} \\ &\le \sup_{\boldsymbol{\delta}\in R_\ell, \ell\le r+\kappa, \boldsymbol{\delta}\neq\mathbf{0}} \frac{\boldsymbol{\delta}^\top E[\int_0^\infty \mathbf{V}_i(t)\mathbf{V}_i(t)^\top dN_i(t)]\boldsymbol{\delta}}{\|\boldsymbol{\delta}\|^2} := \Lambda_{\max} < \infty. \end{aligned}$$

(b)

$$q' := \sup_{\boldsymbol{\delta}\in R_\ell, \ell\le r+\kappa, \boldsymbol{\delta}\neq\mathbf{0}} \frac{E[\int_0^\infty |\mathbf{V}_i(t)^\top\boldsymbol{\delta}|^2 dN_i(t)]^{3/2}}{E[\int_0^\infty |\mathbf{V}_i(t)^\top\boldsymbol{\delta}|^3 dN_i(t)]} > 0,$$

where $R_\ell = \{\boldsymbol{\delta} = (\boldsymbol{\delta}_{\mathbf{x}}^\top, \boldsymbol{\delta}_{\mathbf{z}}^\top)^\top : \boldsymbol{\delta}_{\mathbf{x}} \in \mathbb{R}^r, \sum_{j=1}^p \delta_{zj} = 0, \|\boldsymbol{\delta}_{\mathbf{z}}\|_0 \le \ell - r\}$, with $\|\cdot\|_0$ denoting the $L_0$ norm.

In addition, we set a model size upper bound, denoted by $\kappa$, with $s < \kappa < p$,

$$\xi_n = \min\left\{\min_{1\le j\le r}\int_\Delta |\alpha_{0j}(\tau)|d\tau, \min_{1\le j\le s}\int_\Delta |\beta_{0j}(\tau)|d\tau\right\},$$

which measures the minimal overall effect of the usual and compositional relevant variables on the conditional distribution. Theorem 3 and Corollary 2 present the consistency of the tuning parameter selection based on the GIC.

**Theorem 3.** *Suppose the conditions in Theorem* 1 *and* (C5+) *hold. Furthermore,* $\log(r+p)/n = o(\phi_n)$*,* $\phi_n = o(\xi_n^{5/2})$*, and* $\kappa n^{-1}\log\max\{n, r+p\} = o(\xi_n^3)$*. Then,*

$$P\left(\inf_{S\neq S_\Delta, |S|\le\kappa} GIC(S) > GIC(S_\Delta)\right) \to 1.$$

**Corollary 2.** *Under the same conditions as in Theorem* 3, *if*

$$\Big\{\inf_{j>r+s,\tau\in\Delta} w_j(\tau)\Big\}^{-1}\frac{\sqrt{n}}{\sqrt{(r+s)\log\max\{n,r+p\}}}=O_p(1)$$

*and* $\sup_{\tau\in\Delta,j\in S_\tau} w_j(\tau)=O_p(\sqrt{n}/(\sqrt{r+s}\log\max\{n,r+p\}))$, *then* $P\big(\hat{S}_{\hat{\lambda}}=S_\Delta\big)\to 1$.

For any $1\le l\le s$, we use $\mathbf{Z}_i^l(t)$ to denote the log-ratio transformed $\mathbf{W}_i(t)$ when the reference is the $l$th component; that is, $\mathbf{Z}_i^l(t)$ is the vector $\mathbf{Z}_i(t)-Z_{i,l}(t)\mathbf{1}_p$, with the $l$th component removed. We also define $\mathbf{V}_i^l(t)=(\mathbf{X}_i(t)^\top,\mathbf{Z}_i^l(t)^\top)^\top$. Let $\boldsymbol{\gamma}_{\backslash l}(\tau)=(\boldsymbol{\alpha}(\tau)^\top,\boldsymbol{\beta}_{\backslash l}(\tau)^\top)^\top$, where $\boldsymbol{\beta}_{\backslash l}(\tau)=(\beta_1(\tau),\ldots,\beta_{l-1}(\tau),\beta_{l+1}(\tau),\ldots,\beta_p(\tau))^\top$. Let $\hat{\gamma}_{\backslash l}(\tau)$ be the solution of the following unconstrained minimization problem:

$$\frac{1}{n}\sum_{i=1}^n\int_0^\infty\rho_\tau\{Y_i(t)-\mathbf{V}_i^l(t)^\top\gamma_{\backslash l}\}dN_i(t)+\lambda\sum_{j=1,j\ne l}^p\omega_j(\tau)|\beta_j|, \tag{3.1}$$

where $\boldsymbol{\gamma}_{\backslash l}=(\alpha_1,\ldots,\alpha_r,\beta_1,\ldots,\beta_{l-1},\beta_{l+1},\ldots,\beta_p)^\top$. Then, the globally adaptive unconstrained estimator $\hat{\boldsymbol{\gamma}}_l^u(\tau)$ with the $l$th component as the reference is $(\hat{\gamma}_{1,\backslash l}(\tau),\ldots,\hat{\gamma}_{r,\backslash l}(\tau),\hat{\gamma}_{r+1,\backslash l}(\tau),\ldots,\hat{\gamma}_{r+l-1,\backslash l}(\tau),-\sum_{k=1,k\ne l}^p\hat{\gamma}_{r+k,\backslash l}(\tau),\hat{\gamma}_{r+l+1,\backslash l}(\tau),\ldots,\hat{\gamma}_{r+p,\backslash l}(\tau))^\top$. We state the asymptotic properties of $\hat{\boldsymbol{\gamma}}_l^u(\tau)$ in the following theorem:

**Theorem 4.** *Under the same conditions as in Theorem* 2, *if* $(r+s)^3\log^4 n=o(n)$ *and* $\sup_{\tau\in\Delta,j\in S_\tau} n^{-1/2}\lambda w_j(\tau)=o_p(1)$, *then, for any given* $\boldsymbol{\xi}\in R_{r+s-1}$, $\|\boldsymbol{\xi}\|=1$, *and* $1\le l\le s$, *we have*

- *(a)* $n^{1/2}\boldsymbol{\xi}^\top\mathbf{H}_\tau\{\hat{\boldsymbol{\gamma}}_l^u(\tau)-\boldsymbol{\gamma}_0(\tau)\}$ *converges weakly to a mean-zero Gaussian process with covariance* $\boldsymbol{\Sigma}(\tau,\tau')$ *and* $P(\sup_{\tau\in\Delta}\|\hat{\boldsymbol{\gamma}}_{b,l}^u(\tau)\|_\infty=0)\to 1$;
- *(b)* $n^{1/2}\boldsymbol{\xi}^\top\{\hat{\boldsymbol{\gamma}}_l^u(\tau)-\boldsymbol{\gamma}_0(\tau)\}$ *and* $n^{1/2}\boldsymbol{\xi}^\top\{\hat{\boldsymbol{\gamma}}(\tau)-\boldsymbol{\gamma}_0(\tau)\}$ *are asymptotically equivalent.*

Theorem 4 indicates that the proposed constrained estimator is asymptotically equivalent to an unconstrained estimator that uses a relevant variable as the reference. However, the latter approach requires preliminary knowledge about the truly relevant variables, which may not be available in practice.

By our theorems, the technical constraints for $s$ include $(r+s)^3\log^2\max\{n,r+p\}=o(n)$ and $(r+s)^3\log^4 n=o(n)$. When $p=O(n^a)$ $(a>0)$, we can allow $s$ to be close to, but smaller than $o(n^{1/3})$, which is the fastest model size growth rate derived in Welsh (1989) and He and Shao (2000) for an unpenalized quantile regression estimator to achieve asymptotic normality. Proofs of the theorems are provided in the Supplementary Material (Section S4).

## 4. Simulation Studies

In this section, we carry out simulation studies to evaluate the finite-sample performance of the proposed method. We consider the sample size $n = 100$ and generate $Y(t)$ based on the assumed quantile regression model with $r = 4$ and $p = 400$. Specifically, we generate the longitudinal observation times $t_i^{(k)}$, for $k = 1, \ldots, m_i$, from a standard Poisson process, where $m_i$ is the integer part of $2+U_i$ with $U_i \sim Uniform(0, 2)$. With $r = 4$, we generate $X_{i1}$ from $Uniform(0, 1)$ and $X_{i2}$ from $Bernoulli(0.5)$. For each observed time point $t = t_i^{(k)}$, we first generate a $p$-dimensional vector $\tilde{\mathbf{Z}}_i(t) = (\tilde{Z}_{i1}(t), \ldots, \tilde{Z}_{ip}(t))^\top$ from a multivariate normal distribution $N_p(\mathbf{0}, \mathbf{\Sigma})$, where $\mathbf{\Sigma} = (\rho^{|i-j|})$, with $\rho = 0.5$. Next, we set $\check{Z}_{ij}(t) = \Phi(\tilde{Z}_{ij}(t))$, for $j \neq 7$ and $\check{Z}_{i7}(t) = -\Phi(\tilde{Z}_{i7}(t))$, and then standardize $\check{Z}_{ij}(t)$ so that its second moment is equal to one, where $\Phi(\cdot)$ is the standard normal distribution function and $j = 1, \ldots, p$. The standardized $\check{Z}_{ij}(t)$ $(j = 1, \ldots, p)$ form the covariate vector $\mathbf{Z}_i(t)$.

To generate the longitudinal responses, we consider the following four setups:

Setup (I): Data are generated from a longitudinal linear model with independent homogeneous errors,

$$Y_i(t) = -X_{i1} + X_{i2} - t + \mathbf{Z}_i(t)^\top \mathbf{b} + \epsilon_i(t),$$

where $\mathbf{b} = (1, 0.8, 0.9, 1, 2, -1.5, -4.2, 0, \ldots, 0)^\top$, $\epsilon_i(t) \sim N(0, 1)$ for any $t > 0$, and $\epsilon_i(t)$ and $\epsilon_i(t')$ are independent for $t > 0$, $t' > 0$, and $t \neq t'$.

Setup (II): Data are generated from a longitudinal linear model with dependent homogeneous errors,

$$Y_i(t) = -X_{i1} + X_{i2} - t + \mathbf{Z}_i(t)^\top \mathbf{b} + a_i + \epsilon_i(t),$$

where $\mathbf{b} = (1, 0.8, 0.9, 1, 2, -1.5, -4.2, 0, \ldots, 0)^\top$, $a_i \sim N(0, 1/2)$, $\epsilon_i(t) \sim N(0, 1/2)$ for $t > 0$, $\epsilon_i(t)$ and $\epsilon_i(t')$ are independent for $t > 0$, $t' > 0$, and $t \neq t'$, and $a_i$ and $\epsilon_i(t)$ are independent for $t > 0$.

Setup (III): Data are generated from a longitudinal linear model with independent heterogeneous errors,

$$Y_i(t) = -X_{i1} + X_{i2} - t + \mathbf{Z}_i(t)^\top \mathbf{b}_1 + (X_{i1} + \mathbf{Z}_i(t)^\top \mathbf{b}_2)\epsilon_i(t),$$

where $\mathbf{b}_1 = \mathbf{b} = (1, 0.9, 0.75, 0.5, 0.8, 1, -4.95, 0, \ldots, 0)^\top$, $\mathbf{b}_2 = (0, 0.25, 0, 1, 0, 0, -1.25, 0, \ldots, 0)^\top$, $\epsilon_i(t) \sim N(0, 1)$ for any $t > 0$, and $\epsilon_i(t)$ and $\epsilon_i(t')$ are independent for $t > 0$, $t' > 0$, and $t \neq t'$.

Setup (IV): Data are generated from a longitudinal linear model with dependent heterogeneous errors,

$$Y_i(t) = -X_{i1} + X_{i2} - t + \mathbf{Z}_i(t)^\top \mathbf{b}_1 + (X_{i1} + \mathbf{Z}_i(t)^\top \mathbf{b}_2)(a_i + \epsilon_i(t)),$$

where $\mathbf{b}_1 = \mathbf{b} = (1, 0.8, 0.9, 1, 2, -1.5, -4.2, 0, \ldots, 0)^\top$, $\mathbf{b}_2 = (0, 0.2, 0, 0.1, 0, 0, -0.3, 0, \ldots, 0)^\top$, $a_i \sim N(0, 1/2)$ and $\epsilon_i(t) \sim N(0, 1/2)$ for $t > 0$, $\epsilon_i(t)$ and $\epsilon_i(t')$ are independent for $t > 0$, $t' > 0$, and $t \neq t'$, and $a_i$ and $\epsilon_i(t)$ are independent for $t > 0$.

Under Setups (I) and (II), we can show that

$$Q_{Y_i(t)}\{\tau|\mathbf{X}_i(t), \mathbf{Z}_i(t)\} = Q_e(\tau) - X_{i1} + X_{i2} - t + \mathbf{Z}_i(t)^\top \mathbf{b},$$

where $Q_e(\tau)$ is the $\tau$th quantile of the standard normal distribution. Under Setups (III) and (IV), we can show that

$$Q_{Y_i(t)}\{\tau|\mathbf{X}_i(t), \mathbf{Z}_i(t)\} = \{-1 + Q_e(\tau)\}X_{i1} + X_{i2} - t + \mathbf{Z}_i(t)^\top \{\mathbf{b}_1 + \mathbf{b}_2 Q_e(\tau)\}.$$

In all setups, the true regression coefficients for $\mathbf{Z}_i(t)$ satisfy the zero-sum constraint at each $\tau$.

We evaluate the finite-sample performance of the proposed globally adaptive Lasso estimators with weights ($w2$) and ($w3$), denoted by $\text{AW}_2$ and $\text{AW}_3$, respectively. We set $\Delta = [0.1, 0.9]$ and the $\tau$-grid $\mathcal{S}_n$ as $\{0.1 < 0.125 < \cdots < 0.9\}$. We select the tuning parameter $\lambda$ using a GIC criterion with $\phi_n = \log(\log n) \log p/n$, except for that in the initial estimator. The candidate values for $\lambda$ include $N/4$ equally spaced grid points between $N/150$ and $N/15$, where $N = \sum_{i=1}^n m_i$ is the total number of longitudinal observations. We adapt the method of Belloni and Chernozhukov (2011) over $\Delta$ to get the estimator $\check{\boldsymbol{\beta}}(\tau)$ for calculating the adaptive weight functions.

We compare $\text{AW}_2$ and $\text{AW}_3$ with the locally concerned adaptive Lasso estimator at a single predetermined quantile level $\tau = 0.2, 0.5$, or $0.8$, denoted by $\text{SS}(\tau)$, as well as with the pointwise approach, which simply combines the estimates from $\text{SS}(\tau)$ over $\tau \in \Delta$, and is denoted by PS. We also consider four other benchmark estimation procedures, namely, ALasso (i), ALasso (ii), ALasso (iii), and ALasso (iv). The ALasso (i) estimators are the unconstrained estimators obtained by minimizing (3.1), with the reference, the $l$-th component, chosen randomly. The ALasso (ii) estimators are the globally adaptive estimators derived from model (2.2) without considering the zero-sum constraint. That is, the ALasso (ii) estimators, $(\hat{\boldsymbol{\alpha}}(\tau)^{(\mathrm{ii})}, \hat{\boldsymbol{\beta}}(\tau)^{(\mathrm{ii})})$, are obtained as

$$\underset{\boldsymbol{\alpha},\boldsymbol{\beta}}{\operatorname{argmin}} \left\{ \frac{1}{n} \sum_{i=1}^{n} \int_{0}^{\infty} \rho_\tau \{Y_i(t) - \mathbf{X}_i(t)^\top \boldsymbol{\alpha} - \mathbf{Z}_i(t)^\top \boldsymbol{\beta}\} dN_i(t) + \lambda \sum_{j=1}^{p} \omega_j(\tau) |\beta_j| \right\}.$$

The ALasso (iii) estimator is obtained by fitting the log-contrast model based on the relevant variables selected by the ALasso (ii) approach. The ALasso (iv) estimator is obtained by solving the minimization problem (2.3) without including the zero-sum constraint, using the selected relevant variables to fit a log-contrast model, and then selecting the tuning parameter using the GIC criterion and determining the final estimator.

We assess the variable selection performance of the different methods described above in terms of the mean number of correctly identified relevant variables (NCN), mean number of incorrectly selected variables (NIN), percentage of under-fitted models (PUF), percentage of correctly fitted models (PCF), and percentage of over-fitted models (POF). To evaluate the global estimation accuracy over $\tau \in \Delta$, we consider three average estimation errors, $\mathrm{AEE}_{\ell_1}$, $\mathrm{AEE}_{\ell_2}$, and $\mathrm{AEE}_{\ell_\infty}$, where

$$\mathrm{AEE}_{\ell_q} \doteq \frac{1}{|\Delta|} \int_{\Delta} \|\hat{\boldsymbol{\beta}}(\tau) - \boldsymbol{\beta}^*(\tau)\|_q \, d\tau.$$

For $\mathrm{SS}(\tau)$, we calculate the average estimation errors by extrapolating the coefficient estimate as the constant value of the whole coefficient function over $\tau \in \Delta$. To assess how well the estimated coefficients satisfy the zero-sum constraint, we adopt the criterion SUM, which is defined as $\mathrm{SUM} = \sum_{j=1}^{p} \beta_j(\tau_*)$, where $\beta_j(\cdot)$ denotes the estimated coefficient function and $\tau_* = \mathrm{argmax}_{\tau\in\Delta} |\sum_{j=1}^{p} \beta_j(\tau)|$. Better performance is indicated by NCN closer to seven, the true number of relevant covariates, PCF closer to 100%, NIN, PUF, and POF closer to zero, smaller $\mathrm{AEE}_{\ell_1}$, $\mathrm{AEE}_{\ell_2}$, and $\mathrm{AEE}_{\ell_\infty}$, and SUM closer to or equal to zero.

The simulation results for setups (I)–(IV) are presented in Table S1, Table S2, Table S3, and Table 1, respectively, where Tables S1–S3 are provided in the Supplementary Material. Simulation results are summarized based on 300 replicates. As seen from these tables, the proposed estimators with the globally adaptive weights, $\mathrm{AW}_2$ and $\mathrm{AW}_3$, perform well in all setups, where the error terms can be homogeneous or heterogeneous, and can be independent or dependent across different time points. In all setups, the PCFs based on these estimators are around or above 85%, and the zero-sum constraint is always met by the estimated coefficient functions. As shown by additional simulations reported in the Supplementary Material (see Tables S4–S5), the PCFs can increase further as the variance of the longitudinal error decreases. In setups (I) and (II), where the effects of $\mathbf{Z}(t)$ are constant over $\tau$, the estimation accuracy is comparable between the proposed globally adaptive estimators and the local estimators, SS($\tau$), for $\tau = 0.2, 0.5, 0.8$. However, the variable selection based on SS($\tau$) is more likely to miss relevant variables, as reflected by the higher PUFs, particularly when $\tau = 0.2$ or 0.8. In setups (III) and (IV), where the effects of $\mathbf{Z}(t)$ are not constant over $\tau$, SS($\tau$) performs much worse in terms of variable selection than do $\text{AW}_2$ and $\text{AW}_3$. This may lead to a deterioration in the average estimation errors for SS($\tau$) observed in setups (III) and (IV). In all setups, the pointwise method produces average estimation errors similar to those of $\text{AW}_2$ and $\text{AW}_3$. However, the pointwise method tends to overfit, with a POF equal to 31.7% in setup (I), 26.3% in setup (II), and 23% in setups (III) and (IV).

The results for the globally adaptive estimators under ALasso (i) show a common overfitting problem associated with adopting the unconstrained log-contrast model. This is because the ALasso (i) procedure automatically includes the reference compositional covariate, which may not be a truly relevant covariate. The results under ALasso (ii) suggest that the underlying zero-sum constraint of the coefficients is not satisfied if it is not carefully accounted for in the estimation procedure. In such a situation, interpreting the resulting coefficient estimates as the effects of compositional covariates is problematic. The ALasso (iii) approach renders satisfactory rates of correct fitting, but yields larger estimation errors compared with those of the proposed method. The ALasso (iv) method tends to overfit, with the percentages of overfitting above 25%. The enlarged estimation errors and the overfitting behavior reflect the disadvantage of handling the zero-sum constraint separately from the model estimation and variable selection. In summary, the simulation results show the importance of the proposed globally adaptive estimators, as well as their satisfactory empirical performance.

## 5. A Real-Data Example

We applied the proposed method to a longitudinal data set from the *Feeding Infants Right... from the STart* (FIRST) study. The FIRST study is an ongoing perspective observational study that has enrolled and followed up on children with CF from the neonatal period. In this study, various diet-related biomarkers were collected repeatedly at prespecified CF care visits. For example, fecal specimens were collected at approximately 2, 4, 6, 8, and 12 months of age for

Table 1. Simulation results under Setup (IV) with dependent heterogeneous errors.

| | $AEE_{\ell_1}$ | $AEE_{\ell_2}$ | $AEE_{\ell_\infty}$ | NCN | NIN | PUF (%) | PCF (%) | POF (%) | SUM |
|---|---|---|---|---|---|---|---|---|---|
| | | | | Proposed | | | | | |
| $AW_2$ | 2.261 | 1.024 | 0.694 | 6.923 | 0.040 | 7.7 | 88.3 | 4.0 | 0.000 |
| $AW_3$ | 2.297 | 1.041 | 0.707 | 6.877 | 0.017 | 12.0 | 86.7 | 1.3 | 0.000 |
| SS(0.2) | 2.828 | 1.275 | 0.841 | 6.110 | 0.007 | 57.3 | 42.0 | 0.7 | 0.000 |
| SS(0.5) | 1.908 | 0.863 | 0.577 | 6.670 | 0.017 | 29.0 | 69.3 | 1.7 | 0.000 |
| SS(0.8) | 2.623 | 1.206 | 0.829 | 6.273 | 0.023 | 43.3 | 54.7 | 2.0 | 0.000 |
| PS | 2.277 | 1.034 | 0.702 | 6.970 | 0.257 | 3.0 | 74.0 | 23.0 | 0.000 |
| | | | | ALasso (i) | | | | | |
| $AW_2$ | 2.456 | 1.067 | 0.706 | 6.917 | 1.030 | 8.0 | 0.7 | 91.3 | 0.000 |
| $AW_3$ | 2.491 | 1.084 | 0.718 | 6.860 | 1.010 | 13.3 | 0.7 | 86.0 | 0.000 |
| | | | | ALasso (ii) | | | | | |
| $AW_2$ | 2.326 | 1.064 | 0.725 | 6.913 | 0.033 | 8.7 | 88.3 | 3.0 | 1.916 |
| $AW_3$ | 2.352 | 1.076 | 0.734 | 6.857 | 0.017 | 14.0 | 84.3 | 1.7 | -2.315 |
| | | | | ALasso (iii) | | | | | |
| $AW_2$ | 2.549 | 1.146 | 0.770 | 6.913 | 0.033 | 8.7 | 88.3 | 3.0 | 0.000 |
| $AW_3$ | 2.668 | 1.199 | 0.811 | 6.857 | 0.017 | 14.0 | 84.3 | 1.7 | 0.000 |
| | | | | ALasso (iv) | | | | | |
| $AW_2$ | 2.294 | 1.030 | 0.685 | 6.963 | 0.380 | 3.7 | 63.7 | 32.7 | 0.000 |
| $AW_3$ | 2.305 | 1.035 | 0.690 | 6.957 | 0.347 | 4.3 | 65.0 | 30.7 | 0.000 |

each child. Gut microbiome composition data were then extracted from the fecal specimens using 16S rRNA gene pyrosequencing, and comprise the relative abundance of 364 unique genera subject to the unit-sum constraint. The level of calprotectin, a biomarker for inflammation in the gastrointestinal (GI) tract, was also tracked over time, and recorded in units of micrograms per gram of stool. In our analysis of the FIRST data set, the specific question of interest is how the gut microbiome composition is associated with the calprotectin level over time. Identifying the subcompositional bacterial taxa linked to the variations in calprotectin can provide useful insights into the early CF disease mechnisam.

The final data set includes 135 subjects and a total of 328 longitudinal records, after excluding seven children with a low birth weight. Table S6 in the Supplementary Material presents the basic summary statistics by gender, number of longitudinal records, and calprotectin levels. The results show that 56% of the subjects are boys, and about 50% of the subjects have three or four longitudinal records. Furthermore, the calprotectin levels present a skewed distribution, with the median (= 64.5) considerably smaller than the mean (= 111.2). In this case, adopting longitudinal quantile regression modeling can deliver a more comprehensive and robust view about how the gut microbiome composition influences calprotectin levels.

In our analysis, we implement the proposed globally adaptive methods with the adaptive weights ($w2$) and ($w3$) and $\Delta = (0.2, 0.8]$ (denoted by $\text{AW}_2$ and $\text{AW}_3$, respectively), the locally concerned adapive-Lasso method $\text{SS}(\tau)$ with $\tau = 0.2, 0.3, \ldots, 0.8$, and the pointwise method (denoted by PS), which is a union set for $\text{SS}(\tau)$, with $\tau = 0.2, 0.225, \ldots, 0.8$. We include gender as a regular covariate. The compositional covariates are the relative abundance of the 364 genera measured from the gut microbiome samples. We exclude six genera that have a relative abundance below the detection limit in all samples. In addition, we replace all non-detectable relative abundance with an extremely small constant $10^{-20}$, which is much smaller than the minimum nonzero relative abundance captured in our data set, $4.418 \times 10^{-6}$. For the tuning parameter selection, the candidate values of $\lambda$ include $N/4$ equally spaced grid points between $N/150$ and $N/15$, where $N = 328$. To avoid selecting boundary $\lambda$, $\phi_n$ in the GIC is chosen as $\log(\log n)\log p/(20n)$ for the globally concerned and locally concerned quantile regressions. Estimates below $10^{-4}$ are shrunk to zero.

To evaluate each method, we compute the prediction errors as follows. We first randomly split the 135 subjects into a training set of size 120 and a testing set of size 15. We apply the method to the training data set and obtain the estimator of $(\boldsymbol{\alpha}_0(\tau)^\top, \boldsymbol{\beta}_0(\tau)^\top)^\top$, denoted by $(\hat{\boldsymbol{\alpha}}^{\text{train}}(\tau)^\top, \hat{\boldsymbol{\beta}}^{\text{train}}(\tau)^\top)^\top$. Then, we calculate the prediction error in the testing set as

$$\text{PE}(\Delta) = \frac{\sum_{i \in \mathcal{T}} \int_\Delta \int_0^\infty \rho_\tau\{Y_i(t) - \mathbf{X}_i(t)^\top \hat{\boldsymbol{\alpha}}^{\text{train}}(\tau) - \mathbf{Z}_i(t)^\top \hat{\boldsymbol{\beta}}^{\text{train}}(\tau)\} dN_i(t) d\tau}{\sum_{i=1}^n 1\{i \in \mathcal{T}\}},$$

where $\mathcal{T}$ denotes the test set. For $\text{SS}(\tau)$, we calculate $\text{PE}(\Delta)$ by treating the coefficient estimate as a constant-valued function over $\tau \in \Delta$.

Table 2 lists the genus sets selected using the different methods. The average prediction errors (PEs) and the corresponding standard deviations (within parentheses) are also presented. The PE calculations are based on 200 random splits of the training and test sets. Table 2 shows that the selected genus sets vary considerably across the locally concerned methods, $\text{SS}(\tau)$, with different choices of $\tau$. These observations suggest that some genera may have varying effects on different quantiles of calprotectin level, and, may also, in part, reflect the variable selection instability associated with $\text{SS}(\tau)$ (Zheng, Peng and He (2015)). For example, the genus "g115" may only affect median calprotectin, but not the lower or upper quantiles. In contrast, the proposed globally concerned methods give

Table 2. Analysis of the FIRST data set.

| $\tau$ | Method | Selected Genus Sets | PE |
|---|---|---|---|
| $[0.2, 0.8]$ | $AW_2$ | g50 g93 g115 g137 g147 g152 g162 g178 g184 g197 g204 g210 g213 g219 g297 g319 g370 | 0.5279 (0.0837) |
| | $AW_3$ | g50 g93 g115 g147 g152 g162 g178 g197 g204 g210 g213 g219 g297 g319 g370 | 0.5271 (0.0837) |
| | PS | g14 g32 g50 g64 g93 g115 g119 g137 g147 g152 g153 g162 g178 g183 g184 g188 g193 g197 g199 g204 g210 g213 g219 g297 g319 g370 | 0.5278 (0.0826) |
| 0.2 | SS | g147 g153 g213 | 0.7893 (0.1420) |
| 0.3 | SS | None | 0.6833 (0.1206) |
| 0.4 | SS | g50 g93 g119 g147 g162 g183 g197 g199 g204 g213 g297 | 0.6135 (0.1038) |
| 0.5 | SS | g14 g115 g137 g147 g193 g197 g204 g213 g219 g297 g319 | 0.5855 (0.0976) |
| 0.6 | SS | None | 0.5960 (0.0981) |
| 0.7 | SS | g147 g152 g178 g184 g197 g204 g213 g297 g319 g370 | 0.6621 (0.1025) |
| 0.8 | SS | g32 g147 g152 g162 g178 g197 g204 g213 | 0.7926 (0.1235) |

robust and parsimonious selections of genus sets. For example, the selected genus sets are almost identical between $AW_2$ and $AW_3$. The selected genera are mostly also selected by one of the $SS(\tau)$. Naively pooling the results from the $SS(\tau)$, as shown by the PS method, leads to selecting an excessive number of genera (i.e., 26 genera). Some genera selected by $SS(\tau)$, but not by $AW_2$ or $AW_3$, are possibly "false positives" as suggested by the apparent overfitting behavior of the PS method demonstrated in the simulation studies. Moreover, the proposed method $AW_3$ yields the smallest prediction error. The prediction error of $AW_2$ is close to the second smallest value. The locally concerned $SS(\tau)$ methods produce larger prediction errors, because they neglect important genera that do not show effects at the $\tau$th quantile, but are relevant to other quantiles. In summary, the proposed globally adaptive methods strike the best balance between parsimonious variable selections and accurate predictions, while retaining sensible interpretations by satisfying the zero-sum constraint of the coefficients.

## 6. Conclusion

In this work, we develop a globally concerned longitudinal quantile regression framework that accommodates high-dimensional compositional covariates. The proposed method achieves the oracle convergence rate and the global model selection consistency, while enjoying interpretative advantages.

The longitudinal quantile regression model presented here assumes that no covariate effects change over time. To accommodate temporal covariate effects, model (2.1) or (2.2) can be extended with the regression coefficients formulated as bivariate functions of $\tau$ and $t$. Intuitively, this can be achieved by combining the proposed method with the strategy of Park and He (2017). That is, the longitudinal loss function can be modified by incorporating spline approximations of the regression coefficient functions, with the penalty term adjusted accordingly. Nevertheless, this approach may be computationally prohibitive because of the additional high-dimensional layer induced by the spline approximations. Specifically, suppose there are $L$ spline basis functions and $L = O(n^{1/5})$. Based on the proposed estimation for model (2.1), the computational complexity is about $O(n^2 \cdot p \cdot M(n))$, based on the result of Klee and Minty (1972) for the simplex algorithm. When considering the spline-based estimation for the extended model with time-varying coefficients, we expect that the computational intensity will be roughly equivalent to that of fitting a quantile regression model for a data set with sample size $nM(n)$ and covariate dimension $pL$, which is about $O(n^2 M(n)^2 pL)$. Given $M(n) = O(n)$, as suggested by Zheng, Peng and He (2015), tackling the more flexible model with time-varying coefficients would require $O(n^{6/5})$ times the computational effort needed for the proposed model (2.1), which can be computationally prohibitive for high-dimensional applications. How to address such an obstacle merits future research.

After applying the proposed method to a real data set, assessing the adequacy of model (2.1) with the prespecified quantile index set $\Delta$ and the selected relevant variables may be of practical interest. To this end, we can adapt the model-checking strategy of Peng and Huang (2008), and consider the stochastic process

$$K_n(\tau) = n^{-1/2} \sum_{i=1}^{n} \int_0^\infty W(\mathbf{V}_i(t))\psi_\tau\{Y_i(t) - \mathbf{X}_i(t)^\top \hat{\boldsymbol{\alpha}}(\tau) - \mathbf{Z}_i(t)^\top \hat{\boldsymbol{\beta}}(\tau)\} dN_i(t)$$

as an analogue of the martingale-based diagnostic process employed by Peng and Huang (2008), where $\psi_\tau(u) = \tau - I(u < 0)$. Here, $W(\cdot)$ is a known bounded function and $\mathbf{V}_i(t) = (\mathbf{X}_i(t)^\top, \mathbf{Z}_i(t)^\top)^\top$. A lack-of-fit test statistic can be constructed based on $\sup_{\tau\in\Delta} |K_n(\tau)|$. Following the lines of Peng and Huang (2008), the corresponding p-value can be obtained by using a properly designed resampling scheme to approximate the distribution of $K_n(\cdot)$ under model assumption (2.1).

Following the idea of the weighted GEE (Liang and Zeger (1986)) and the quasi-likelihood approach for a median regression (Jung (1996)), we can incorporate within-subject correlations of repeated measures to further improve the estimation efficiency of the proposed method. Specifically, consider a weighted penalized estimating equation,

$$n^{-1/2}\sum_{i=1}^{n}\mathbf{V}_i^{\top}\mathbf{Q}_i(\tau;\boldsymbol{\alpha},\boldsymbol{\beta})^{-1}\boldsymbol{S}_i(\tau;\boldsymbol{\alpha},\boldsymbol{\beta})+\lambda\sum_{j=1}^{p}\omega_j(\tau)\mathrm{sign}(\beta_j)=0,$$

subject to constraint $\sum_{j=1}^{p}\beta_j=0$, where $\mathbf{V}_i=(\mathbf{V}_i(t_i^{(1)})\ldots,\mathbf{V}_i(t_i^{(m_i)}))^{\top}$, $\boldsymbol{S}_i(\tau;\boldsymbol{\alpha},\boldsymbol{\beta})$ $=(S_{i1}(\tau;\boldsymbol{\alpha},\boldsymbol{\beta}),\ldots,S_{i,m_i}(\tau;\boldsymbol{\alpha},\boldsymbol{\beta}))^{\top}$ with $S_{ik}(\tau;\boldsymbol{\alpha},\boldsymbol{\beta})=I(Y_i(t_i^{(k)})-\mathbf{X}_i(t_i^{(k)})^{\top}\boldsymbol{\alpha}-\mathbf{Z}_i(t_i^{(k)})^{\top}\boldsymbol{\beta}\le 0)-\tau$, and $\mathbf{Q}_i(\tau;\boldsymbol{\alpha},\boldsymbol{\beta})$ is a working covariance matrix that approximates the covariance of $\mathbf{S}_i(\tau;\boldsymbol{\alpha},\boldsymbol{\beta})$. When $\mathbf{Q}_i(\tau;\boldsymbol{\alpha},\boldsymbol{\beta})$ is an identity matrix $\mathbf{I}_{m_i}$, solving this estimating equation is equivalent to minimizing (2.3), which adopts the working independence assumption. However, note that the weighted estimating equation loses the nice monotonicity property possessed by the unweighted version. In addition, the covariance of $\mathbf{Q}_i(\tau;\boldsymbol{\alpha},\boldsymbol{\beta})$ is often unknown in practice, and its empirical estimate may not be stable when the sample size is not large, as in the FIRST data set. One possible way to alleviate the computational issue is to adopt an iterative algorithm in which we first solve the weighted estimating equation, with the parameters $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ in the weight function $\mathbf{Q}_i(\tau;\boldsymbol{\alpha},\boldsymbol{\beta})^{-1}$ fixed, and then update the weight function using the resulting parameter estimates. In this case, the estimating equation in each iteration is still monotone. Applying this strategy may improve the estimation efficiency, while still being computationally viable. Investigating such a weighted method is left to future research.

## Supplementary Material

Detailed proofs of the lemmas and theorems and additional simulation studies are provided in the online Supplementary Material.

## Acknowledgments

The authors are grateful to the associate editor and two referees for their helpful comments and suggestions. This work was partially supported by the National Institutes of Health grants (R01HL113548, R01DK209692, R03AG067611, and R21AG070659), National Natural Science Foundation of China (11901200, 71931004), and Shanghai Pujiang program (19PJ1403400).

## References

Aitchison, J. (1982). The statistical analysis of compositional data. *Journal of the Royal Statistical Society. Series B (Methodological)* **44**, 139–177.

Aitchison, J. (2003). *The Statistical Analysis of Compositional Data.* Blackburn Press, Caldwell, New Jersey.

Aitchison, J. and Bacon-shone, J. (1984). Log contrast models for experiments with mixtures. *Biometrika* **71**, 323–330.

Belloni, A. and Chernozhukov, V. (2011). $\ell_1$-penalized quantile regression in high-dimensional sparse models. *The Annals of Statistics* **39**, 82–130.

Cho, H., Hong, H. G. and Kim, M.-O. (2016). Efficient quantile marginal regression for longitudinal data with dropouts. *Biostatistics* **17**, 561–575.

Fan, J., Fan, Y. and Barut, E. (2014). Adaptive robust variable selection. *The Annals of Statistics* **42**, 324–351.

Fan, J. and Lv, J. (2011). Non-concave penalized likelihood with NP-dimensionality. *IEEE Transactions on Information Theory* **57**, 5467–5484.

Fan, Y. and Tang, C. Y. (2013). Tuning parameter selection in high dimensional penalized likelihood. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **75**, 531–552.

Gao, X. and Liu, Q. (2020). Sparsity identification in ultra-high dimensional quantile regression models with longitudinal data. *Communications in Statistics-Theory and Methods* **49**, 4712–4736.

He, X. and Shao, Q.-M. (2000). On parameters of increasing dimensions. *Journal of Multivariate Analysis* **73**, 120–135.

Jung, S.-H. (1996). Quasi-likelihood for median regression models. *Journal of the American Statistical Association* **91**, 251–257.

Kim, Y., Choi, H. and Oh, H.-S. (2008). Smoothly clipped absolute deviation on high dimensions. *Journal of the American Statistical Association* **103**, 1665–1673.

Klee, V. and Minty, G. (1972). *How Good Is the Simplex Algorithm? In Inequalities.* (Edited by 0. Shisha). Academic Press, New York.

Koenker, R. (2004). Quantile regression for longitudinal data. *Journal of Multivariate Analysis* **91**, 74–89.

Koenker, R. and Bassett, G. (1978). Regression quantiles. *Econometrica* **46**, 33–50.

Koenker, R. and d'Orey, V. (1987). Algorithm as 229: Computing regression quantiles. *Journal of the Royal Statistical Society, Series C (Applied Statistics)* **36**, 383–393.

Ledoux, M. and Talagrand, M. (1991). *Probability in Banach Space: Isoperimetry and Processes. Ergebnisse der Mathematik und ihrer Grenzgebiete.* Springer, Berlin.

Li, Y., Liu, Y. and Zhu, J. (2007). Quantile regression in reproducing kernel Hilbert spaces. *Journal of American Statistical Association* **102**, 255–268.

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

Lin, W., Shi, P., Feng, R. and Li, H. (2014). Variable selection in regression with compositional covariates. *Biometrika* **101**, 785–797.

Lipsitz, S. R., Fitzmaurice, G. M., Molenberghs, G. and Zhao, L. P. (1997). Quantile regression methods for longitudinal data with drop-outs: Application to CD4 cell counts of patients infected with the human immunodeficiency virus. *Journal of the Royal Statistical Society: Series C (Applied Statistics)* **46**, 463–476.

Lu, J., Shi, P. and Li, H. (2019). Generalized linear models with linear constraints for microbiome compositional data. *Biometrics* **75**, 235–244.

Lv, J. and Fan, Y. (2009). A unified approach to model selection and sparse recovery using regularized least squares. *The Annals of Statistics* **37**, 3498–3528.

Ma, H., Peng, L. and Fu, H. (2019). Quantile regression modeling of latent trajectory features with longitudinal data. *Journal of Applied Statistics* **46**, 2884–2904.

Meinshausen, N. and Buhlmann, P. (2006). High-dimensional graphs and variable selection with the Lasso. *The Annals of Statistics* **34**, 1436–1462.

Nishii, R. (1984). Asymptotic properties of criteria for selection of variables in multiple regression. *The Annals of Statistics* **12**, 758–765.

Park, S. and He, X. (2017). Hypothesis testing for regional quantiles. *Journal of Statistical Planning and Inference* **191**, 13–24.

Peng, L. and Huang, Y. (2008). Survival analysis with quantile regression models. *Journal of the American Statistical Association* **103**, 637–649.

Portnoy, S. (1991). Asymptotic behavior of the number of regression quantile breakpoints. *SIAM Journal on Scientific and Statistical Computing* **12**, 867–883.

Shi, P., Zhang, A. and Li, H. (2016). Regression analysis for microbiome compositional data. *The Annals of Applied Statistics* **10**, 1019–1040.

Sun, X., Peng, L., Manatunga, A. and Marcus, M. (2016). Quantile regression analysis of censored longitudinal data with irregular outcome-dependent follow-up. *Biometrics* **72**, 64–73.

Talagrand, M. (2005). *The Generic Chaining*. Springer, Berlin.

Tyler, A. D., Smith, M. I. and Silverberg, M. S. (2014). Analyzing the human microbiome: A "how to" guide for physicians. *The American Journal of Gastroenterology* **109**, 983–993.

van der Vaart, A. and Wellner, J. (1996). *Weak Convergence and Empirical Processes: With Applications to Statistics*. Springer, New York.

Wang, H. and Fygenson, M. (2009). Inference for censored quantile regression models in longitudinal studies. *The Annals of Statistics* **37**, 756–781.

Wang, L., Wu, Y. and Li, R. (2012). Quantile regression for analyzing heterogeneity in ultrahigh dimension. *Journal of the American Statistical Association* **101**, 1418–1429.

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

Welsh, A. (1989). On $M$-processes and $M$-estimation. *The Annals of Statistics* **17**, 337–361.

Xia, F., Chen, J., Fung, W. and Li, H. (2013). A logistic normal multinomial regression model for microbiome compositional data analysis. *Biometrics* **69**, 1053–1063.

Zhang, C.-H. and Huang, J. (2008). The sparsity and bias of the Lasso selection in high-dimensional linear regression. *The Annals of Statistics* **36**, 1567–1594.

Zheng, Q., Gallagher, C. and Kulasekera, K. B. (2013). Adaptive penalized quantile regression for high dimensional data. *Journal of Statistical Planning and Inference* **143**, 1029–1038.

Zheng, Q., Peng, L. and He, X. (2015). Globally adaptive quantile regression with ultra-high dimensional data. *The Annals of Statistics* **43**, 2225–2258.

Zheng, X., Fu, B., Zhang, J. and Qin, G. (2018). Variable selection for longitudinal data with high-dimensional covariates and dropouts. *Journal of Statistical Computation and Simulation* **88**, 712–725.

Zou, H. and Yuan, M. (2008). Composite quantile regression and the oracle model selection theory. *The Annals of Statistics* **36**, 1108–1126.

Huijuan Ma

KLATASDS-MOE, Academy of Statistics and Interdisciplinary Sciences, East China Normal University, Shanghai 200062, China.

E-mail: hjma@fem.ecnu.edu.cn

Qi Zheng

Department of Bioinformatics and Biostatistics, University of Louiswille, Louisville, KY 40202, USA.

E-mail: qi.zheng@louisville.edu

Zhumin Zhang

Department of Nutritional Sciences, University of Wisconsin-Madison, Madison, WI 53706, USA.

E-mail: zhangz@surgery.wisc.edu

Huichuan Lai

Department of Nutritional Sciences, University of Wisconsin-Madison, Madison, WI 53706, USA.

E-mail: hlai@wisc.edu

Limin Peng

Department of Biostatistics and Bioinformatics, Emory University, Atlanta, GA 30322, USA.

E-mail: lpeng@emory.edu

(Received January 2021; accepted November 2021)