HHS Public Access
Author manuscript
Int J Biostat. Author manuscript; available in PMC 2021 November 28.

Published in final edited form as:
Int J Biostat. ; 17(2): 267–282. doi:10.1515/ijb-2020-0010.

Marginal quantile regression for longitudinal data analysis in the presence of time-dependent covariates

I-Chen Chen[1], **Philip M. Westgate**[2]

[1]Division of Field Studies and Engineering, National Institute for Occupational Safety and Health, Cincinnati, OH 45226, USA

[2]Department of Biostatistics, College of Public Health, University of Kentucky, Lexington, KY 40536, USA

Abstract

When observations are correlated, modeling the within-subject correlation structure using quantile regression for longitudinal data can be difficult unless a working independence structure is utilized. Although this approach ensures consistent estimators of the regression coefficients, it may result in less efficient regression parameter estimation when data are highly correlated. Therefore, several marginal quantile regression methods have been proposed to improve parameter estimation. In a longitudinal study some of the covariates may change their values over time, and the topic of time-dependent covariate has not been explored in the marginal quantile literature. As a result, we propose an approach for marginal quantile regression in the presence of time-dependent covariates, which includes a strategy to select a working type of time-dependency. In this manuscript, we demonstrate that our proposed method has the potential to improve power relative to the independence estimating equations approach due to the reduction of mean squared error.

Keywords

Quantile regression models; Modified estimating equations; Empirical covariance matrix; Mean squared error; Time-dependent covariate

1 Introduction

Generalized estimating equations (GEE) [1] are well-known for their use in the marginal analysis of data from longitudinal studies in which measurements contributed from the same subject are correlated over time. As long as a correct mean structure is given, the regression parameters are consistently estimated even when the working correlation structure is misspecified. However, in the presence of certain types of time-dependent covariates, the estimating equations, and thus estimates, can be biased unless an independence working correlation structure is employed [2]. Unfortunately, the resulting regression parameter estimation can be inefficient because not all valid moment conditions are utilized [3, 4]. Therefore, multiple approaches have been proposed to use all valid moments [5, 6, 7]. Most recently, the modified GEE approach proposed by Chen and Westgate [7] has been shown to perform best in terms of improving estimation efficiency.

Methods for the marginal analysis of longitudinal data in the presence of time-dependent covariates have only been developed for the modeling of the mean. An example carried out in this literature focuses on anthropometric screening data from Bouis and Haddad [8], in which the outcome of interest is morbidity index and time-dependent covariates include body mass index (BMI), among others. Unfortunately, modeling the conditional mean of morbidity index may not be ideal because the response distribution is severely right skewed. In this situation, the use of marginal quantile regression represents an appealing alternative approach. In particular, we are interested in how the distribution of the longitudinally measured morbidity index is associated with the time-dependent covariates.

Quantile regression for independent outcomes, introduced by Koenker and Bassett [9], has advantages relative to mean regression in that it does not require parametric assumptions on the error distribution and it is robust to outliers. In addition, quantile regression can provide a thorough description on the entire conditional distribution of a response variable. However, when outcomes are correlated, modeling the within-subject correlation structure can be difficult. Using an independence working correlation structure has been shown to generate consistent estimators of the regression coefficients [10, 11, 12]. This, however, may cause a loss of efficiency, especially when data are highly correlated [13, 14, 15, 16].

Therefore, multiple approaches have recently been proposed for improving regression parameter estimation in marginal quantile regression for longitudinal data [13, 17]. However, the specification of a correlation structure is required for the quasi-score method of Jung [17], and regression parameter estimation from the use of quadratic inference function (QIF) approach of Tang and Leng [13] is not guaranteed to work well even if the correlation structure is correctly specified [18, 19]. Therefore, Fu and Wang [14] suggested a combination of the between- and within- weighted estimating equations under the working exchangeable structure, which was firstly introduced by Stoner and Leroux [20]. Additionally, Fu and Wang [14] extended their approach to allow any type of working correlation structure [19]. As a result, not only does this approach improve estimation performance, but it is robust to different error distributions. In longitudinal studies, some of the covariate values may change over time and cause feed-back effects from the response variable. This topic has not been explored in the literature on marginal quantile regression.

In this manuscript, we first propose an approach for marginal quantile regression in the presence of time-dependent covariates. This proposed method combines the estimating equations approach of Fu *et al*. [19] with the modified GEE approach of Chen and Westgate [7]. In consequence, the proposed approach can achieve notable gains in efficiency when compared with estimating equations under an independence correlation structure. Second, we propose a strategy to select a working type of time-dependency because in practice it may not be the case that the researcher knows the type of time-dependent covariate. In the marginal analysis literature with time-dependency, criteria such as the mean squared error (MSE), taking into account the influences moment conditions have on both the efficiency and bias of regression parameter estimation, can be used to select a working correlation structure [18, 21] or a classification type of time-dependent covariate [22]. We extend the use of the MSE to choose a working classification type such that consistent regression parameter estimation is a result.

This manuscript is organized as follows. Section 2 introduces a marginal quantile regression and types of time-dependent covariates for longitudinal data. In Section 3, we propose the modified estimating equations for quantile regression in the presence of time-dependent covariates. Furthermore, we introduce the approach to selecting a working classification type for time-dependent covariates. In Section 4, we carry out a simulation study to compare the estimation performance and assess the utility of the proposed selection criterion relative to estimating equations with an independence working structure, and Section 5 demonstrates the proposed method in application to the motivating anthropometric screening data [8, 23]. Finally, we give concluding remarks in Section 6.

2 Quantile Regression and Time-Dependent Covariates

2.1 Notation and Quantile Regression

For ease of illustration, suppose we have a longitudinal study in which N independent subjects are repeatedly measured over T distinct time points. However, in general, the number of repeated measurements is allowed to vary across subjects. Let $\boldsymbol{Y}_i = [Y_{i1}, \ldots, Y_{iT}]^T$ denote the observed outcome vector for the ith subject, and assume that the τth conditional quantile of Y_{ij}, $j = 1, \ldots, T$; $i = 1, \ldots, N$ for $\tau \in (0, 1)$ is denoted by $Q_\tau(Y_{ij} \mid \mathbf{x}_{ij}) = \mathbf{x}_{ij}^T \boldsymbol{\beta}^\tau$, where $\mathbf{x}_{ij} = [1, x_{1ij}, \ldots, x_{pij}]^T$ is a vector observed at time point j for subject i, and $\boldsymbol{\beta}^\tau = \left[\beta_0^\tau, \beta_1^\tau, \ldots, \beta_p^\tau\right]^T$ is an unknown vector corresponding to the regression coefficients at the τth quantile. Let $S_{ij}^\tau = \tau - I\left[Y_{ij} \le \mathbf{x}_{ij}^T \boldsymbol{\beta}^\tau\right]$ and $\boldsymbol{S}_i^\tau = \left[S_{i1}^\tau, \ldots, S_{iT}^\tau\right]^T$, where $I(.)$ is an indicator function. The corresponding covariance matrix for $\boldsymbol{S}_i^\tau$ is given by $\boldsymbol{V}_i^\tau = \boldsymbol{A}_i^{1/2} \boldsymbol{R}_i^\tau(\boldsymbol{\alpha}) \boldsymbol{A}_i^{1/2}$, where $\boldsymbol{A}_i = diag[\tau(1 - \tau), \ldots, \tau(1 - \tau)]$ is a diagonal matrix representing the marginal variances, and $\boldsymbol{R}_i^\tau(\boldsymbol{\alpha})$ is a symmetric positive definite correlation matrix with 1 along the diagonal and one or more unknown correlation parameters given by $\boldsymbol{\alpha}$.

To find the estimate of the regression parameters, $\hat{\boldsymbol{\beta}}^\tau$, we consider the following optimal estimating equations [14, 15, 16, 17]

$$\sum_{i=1}^{N} \boldsymbol{X}_i^T \boldsymbol{\Lambda}_i \boldsymbol{A}_i^{-1/2} \boldsymbol{R}_i^{\tau^{-1}}(\boldsymbol{\alpha}) \boldsymbol{A}_i^{-1/2} \boldsymbol{S}_i^\tau = \mathbf{0}, \tag{1}$$

in which $\Lambda_i = diag[f_{i1}(0), \ldots, f_{iT}(0)]$ with $f_{ij}(0)$ assumed to be a constant can be further eliminated [14]. The score function for the mth component corresponding to $\boldsymbol{\alpha}$, as well as the first partial derivative of the working Gaussian log-likelihood function for $\left(\boldsymbol{S}_1^\tau, \ldots, \boldsymbol{S}_N^\tau\right)$ with respect to the mth component of $\boldsymbol{\alpha}$, can be expressed as [19]

$$\sum_{i=1}^{N} tr\left[\frac{\partial \boldsymbol{R}_i^{\tau^{-1}}(\boldsymbol{\alpha})}{\partial \alpha_m}\left(\boldsymbol{A}_i^{-1/2} \boldsymbol{S}_i^\tau \boldsymbol{S}_i^{\tau^T} \boldsymbol{A}_i^{-1/2} - \boldsymbol{R}_i^\tau\right)\right].$$

The correlation parameter α_m and its corresponding working correlation structure then can be estimated and constructed by optimizing this score function. We note that the asymptotic estimator for $Cov\left(\hat{\beta}^{\tau}\right)$ is hardly obtained due to the involvement of unknown density functions of the errors. As a result, an induced smoothing technique [24, 25] has been commonly used to the marginal quantile regression models [14, 15, 16, 26]

In Equation (1), the $(k+1)$th row corresponds to the estimating equation for β_k^{τ} and is given by

$$\sum_{i=1}^{N}\sum_{s=1}^{T}\sum_{j=1}^{T} x_{kis}v_i^{sj}\left(\tau - I\left[Y_{ij} \le x_{kij}\beta_k^{\tau}\right]\right) = 0,$$

where v_i^{sj}, $i = 1, \ldots, N$ and $s, j = 1, \ldots, T$, is the (s, j)th element of $\boldsymbol{V}_i^{\tau^{-1}}$. If β_k^{τ} corresponds to certain types of time-dependent covariates, as will be specified in the following subsection, then for all s, j we may not have

$$E\left[x_{kis}\left(\tau - I\left[Y_{ij} \le x_{kij}\beta_k^{\tau}\right]\right)\right] = 0. \tag{2}$$

2.2 Types of Time-Dependent Covariates

Four existing types of time-dependent covariates have been introduced in the marginal analysis literature for longitudinal data [5, 27]. In the manner of quantile regression modeling, the kth covariate is classified as a Type I time-dependent covariate if Equation (2) holds for all s, j; $s, j = 1, \ldots, T$, at a given quantile level τ, a Type II if Equation (2) for $s \geqslant j$, a Type III if Equation (2) does not hold for some $s > j$, and a Type IV, which is the opposite of a Type II, if Equation (2) for $s \leqslant j$.

If β_k^{τ} corresponds to a time-dependent covariates which is classified as Type II, III, or IV, then Equation (2) does not hold for some s, j, will result in invalid moments. Pepe and Anderson [2] supported the use of GEE with an independence working correlation structure for marginal mean regression, then the only moment conditions utilized are the ones such that $s = j$ which are always valid regardless of the covariate type. Unfortunately, this safe approach can cause a great efficiency loss if the covariate is not of Type III because additional valid moment conditions are not used [3, 5]. In this situation, approaches allowing the use of all valid moment conditions have been proposed to achieve more efficient parameter estimation [5, 6, 7]. These methods, however, only focus on mean regression and have not been extended to quantile regression when time-dependent covariates exist. Therefore, we propose approaches to improve estimation efficiency and select a working type of time-dependency which is often unknown in practice.

3 Proposed Methods

3.1 Improving Efficiency: Modified Estimating Equations for Quantile Regression

We first propose a modified estimating equations approach for improved efficiency by combining the estimating equations approach of Fu *et al.* [19] with the modified GEE approach of Chen and Westgate [7], which practically takes advantage of GEE's popularity. We replace elements with 0 in the inverse of the correlation matrix and the replacement is executed for each individual biased estimating equation, depending on the covariate type. Specifically, our proposed estimating equations for β_k^τ, $k = 0, 1, \ldots, p$, are given by

$$\sum_{i=1}^{N} \boldsymbol{X}_i^{k+1} \boldsymbol{A}_i^{-1/2} \boldsymbol{R}_i^{\tau *-1}(\boldsymbol{\alpha}) \boldsymbol{A}_i^{-1/2} \boldsymbol{S}_i^{\tau} = \boldsymbol{0}, \tag{3}$$

where $\boldsymbol{X}_i^{k+1}$ is the $(k+1)$th row of $\boldsymbol{X}^T$, and the elements of $\boldsymbol{R}_{ik}^{\tau *-1}(\boldsymbol{\alpha})$, $k = 0, 1, \ldots, p$, are restricted to a certain type of covariate at a given quantile level τ. The modified approach then puts together these estimating equations and estimates regression parameter, correlation parameter, and standard error (SE) in the same nature as with the approach used in marginal quantile regression [19].

We propose to create $\boldsymbol{R}_{ik}^{\tau *-1}$ given in Equation (2) by modifying the inverse of a working correlation structure in general, $\boldsymbol{R}_i^{\tau -1}$, employed in Equation (1) based on the specific type of time-dependent covariate. If parameter k is classified as a Type I time-dependent or time-independent covariate, then the information from all T^2 valid moment conditions is incorporated. Under this circumstance, $\boldsymbol{R}_{ik}^{\tau *-1}$ is equal to $\boldsymbol{R}_i^{\tau -1}$, indicating that the estimating equations from Equations (1) and (2) are identical. When the estimating equation of a parameter corresponds to a Type II time-dependent covariate, $\boldsymbol{R}_{ik}^{\tau *-1}$ is constrained to be a lower triangular matrix such that the $T(T+1)/2$ moment conditions for $s \geqslant j$, $s, j = 1, \ldots, T$, are valid. In other wards, $\boldsymbol{R}_{ik}^{\tau *-1}$ is obtained by making all upper non-diagonal elements equal to 0. With respect to a Type IV time-dependent covariate, a contrast of a Type II, $\boldsymbol{R}_{ik}^{\tau *-1}$ can be obtained by taking $\boldsymbol{R}_i^{\tau -1}$ and making all lower non-diagonal elements equal to 0. Finally, when the parameter corresponds to a Type III time-dependent covariate, $\boldsymbol{R}_{ik}^{\tau *-1}$ is considered to be diagonal matrices in the estimating equation.

3.2 Selection of Working Classification Type for Time-Dependency

Use of the approach just proposed requires data analysts know the covariate's type of time-dependency, although this is likely unknown in practice. We now propose an approach to select a working type of time-dependency with the goal of producing the least variable regression parameter estimate possible. We note that although more than one type of time-dependent covariate can be chosen at any given quantile level τ, for simplicity of notation we assume there is only one time-dependent covariate of unknown type.

To choose a working type for this covariate, we first consider an estimated MSE given by

$$\widehat{MSE}\left(\hat{\beta}_c^\tau\right) = \widehat{Cov}\left(\hat{\beta}_c^\tau\right) + \left(\hat{\beta}_c^\tau - \hat{\beta}_{III}^\tau\right)\left(\hat{\beta}_c^\tau - \hat{\beta}_{III}^\tau\right)^T, \tag{4}$$

where $\hat{\beta}_c^\tau$ is the vector of regression parameter estimates in which the time-dependent covariate is assumed to be Type c, $c = I, II, III$, or IV, and $\widehat{Cov}\left(\hat{\beta}_c^\tau\right)$ denotes an empirically estimated covariance matrix of $\hat{\beta}_c^\tau$. We note that $\widehat{Cov}\left(\hat{\beta}_z\right)$ can be obtained by using the induced smoothing method [24]. In Equation (3), we replace the unknown β^τ with $\hat{\beta}_{III}^\tau$ because $\hat{\beta}_{III}^\tau \xrightarrow{p} \beta^\tau$, thus providing a consistent bias estimate, which is $\left(\beta_c^\tau - \beta^\tau\right)$. Here, the estimate of bias is followed by the defined β_c^τ such that $\hat{\beta}_c^\tau \xrightarrow{p} \beta_c^\tau$. As $N \to \infty$, $\widehat{Cov}\left(\hat{\beta}_c^\tau\right) \to 0$ and $\widehat{MSE}\left(\hat{\beta}_c^\tau\right) \to \left(\beta_c^\tau - \beta^\tau\right)\left(\beta_c^\tau - \beta^\tau\right)^T$. Therefore, if a given working covariate type yields bias, then asymptotically this type will not be chosen when using the selection approach. Specifically, if the truth is of Type I, then any working type produces consistent regression parameter estimation and can be chosen through this approach. If the true type is II (IV), then this approach method will choose either II (IV) or III. Moreover, asymptotically our method will choose Type III if this is the true type.

In order to utilize this estimated MSE to select a working classification type, we propose choosing the type that occurs with the smallest value for the trace of an empirical covariance matrix, $tr\left[\widehat{MSE}\left(\hat{\beta}_c^\tau\right)\right]$. We note that this criterion has been proven to perform well for the selection of a working covariate type [22]. In addition, the true variance of a corresponding regression parameter estimate relies upon the complex probabilities of each type being chosen, and therefore $\widehat{Cov}\left(\hat{\beta}_c^\tau\right)$ can result in a biased estimate of the variance. In consequence, cluster bootstrapped SEs should be adopted for statistical inference [22, 28, 29]. Note that the empirical coverage probabilities of 95% confidence intervals using bootstrapped SEs resulted in near-nominal coverage, although the results are not shown in the simulation study.

4 Simulation Study

4.1 Study Description

We now compare the performances of our proposed selection approach for covariate type of time-dependency to the use of an independence working correlation structure, which treats unknown types of time-dependency as Type III, in the marginal quantile analysis. The selection approach is demonstrated with the modified estimating equations method using a first-order autoregressive (AR-1) working correlation structure, as AR-1 may be preferred over other structures such as exchangeable in a longitudinal study [30].

Three scenarios are carried out in the simulation study, corresponding to true Type I, II, and III time-dependent covariates, with results presented in Tables 1–3, respectively. Each scenario has the same marginal model given by $Y_{ij} = \beta_0 + \beta_1 x_{ij} + \epsilon_{ij}$, $i = 1, \ldots, N$; $j = 1, \ldots, T$. A fourth scenario extends the scenario with a true Type II covariate by adding two additional types of covariates. Specifically, the model includes a time-independent binary indicator, a known type I time-dependent covariate corresponding to time itself, and a Type II covariate (Table 4). The data generation depending on the covariate type are described in the following paragraph. The number of subjects (N) used in any given setting is 100 and each subject contributes 5 repeated measurements (T). Each setting is conducted through 1,000 simulations using R version 3.6.2 [31]. Furthermore, models are based on previous marginal mean regression literature for time-dependent covariates [5, 7, 30, 32] and marginal quantile regression literature [14, 16, 19]. Although marginal quantile models including multiple types of time-dependent covariates were also studied, results were similar and are not presented.

When the time-dependent covariate is either Type I, II, or III, data are generated from $Y_{ij} = \tilde{\beta}_0 + \tilde{\beta}_1 x_{1ij} + \tilde{\beta}_2 x_{1i,\,j-1} + \gamma_i + \epsilon_{ij}$ and $x_{1ij} = \kappa x_{1i,j-1} + \theta\gamma_i + \delta_{ij}$, $i = 1, \ldots, 100$; $j = 1, \ldots, 5$, where $\tilde{\boldsymbol{\beta}} = [0,\ 1,\ 1]^T$, and random effects, γ_i and δ_{ij}, are mutually independent and normally distributed with mean 0 and variance 1 [5, 30]. Note that $Var(\gamma_i) = \sigma_\gamma^2$ and $Var(\delta_{ij}) = \sigma_\delta^2$. The covariate is of Type I if $\tilde{\beta}_2 = \theta = 0$, while the covariate is of Type II if $\theta = 0$. Additionally, x_{i0} follows a normal distribution with mean 0 and variance $\left(\theta^2\sigma_\gamma^2 + \sigma_\delta^2\right)/\left(1 - \kappa^2\right)$ because the time process for x_{ij} is stationary. Here let $\kappa = 0.5$ and $\theta = 1.5$. The marginal mean given by $E\left[Y_{ij} \mid x_{1ij}\right] = \tilde{\beta}_0 + \left\{\tilde{\beta}_1 + \kappa\tilde{\beta}_2 + \left[\left(\theta^2\sigma_\gamma^2\right)(1 + \kappa)/\theta\left(\theta^2\sigma_\gamma^2 + \sigma_\delta^2\right)\right]\right\}x_{1ij}$ gives true values of $\tilde{\beta}_0 = 0$ for the marginal intercept, and $\tilde{\beta}_1 = 1$, $\tilde{\beta}_1 + \kappa\tilde{\beta}_2 = 1.5$, and $\tilde{\beta}_1 + \kappa\tilde{\beta}_2 + \left[\left(\theta^2\sigma_\gamma^2\right)(1 + \kappa)/\theta\left(\theta^2\sigma_\gamma^2 + \sigma_\delta^2\right)\right] = 2.19$ for the marginal parameters corresponding to the Type I, II, and III covariates, correspondingly. In scenario 4, $x_{2ij} = j$ and $x_{3ij} \sim Bernoulli(0.6)$ added to the above model [7]. The true values of $\tilde{\beta}_3 = 0.5$ and $\tilde{\beta}_4 = 1.5$ are corresponded to these time-independent and Type I time-dependent covariates, respectively.

Furthermore, let $\epsilon_{ij} = q + e_{ij}$ and the use of q is to guarantee $p(\epsilon_{ij} \leqslant 0) = \tau \in \{0.1, 0.25, 0.5, 0.75, 0.9\}$, the quantile level. Four cases are accounted for $\boldsymbol{e}_i = [e_{i1}, \ldots, e_{i5}]^T$: cases (1)-(3) assume that $\boldsymbol{e}_i$ follows multivariate normal distribution, multivariate Student's t-distribution with three degrees of freedom, and multivariate log-normal distribution, correspondingly, generating data using a true AR-1 structure with a correlation parameter 0.7; in order to create correlated heteroscedastic errors, cases (4) assumes $e_{ij} = 0.25(1+|x_{ij}|)\zeta_{ij}$, where $\boldsymbol{\zeta}_i = [\zeta_{i1}, \ldots, \zeta_{i5}]^T$ follows multivariate normal distribution with the same combinations as cases (1)-(3).

In order to examine differences in estimation performances, in Tables 1–4 we present empirical biases corresponding to either the reference approach with an independence working structure or our proposed approach with an exchangeable or AR-1 structure, empirical MSEs of estimates for β_1, and ratios of empirical MSEs, which we refer to as relative efficiencies (REs). For any given RE, the numerator is the MSE resulting from the use of reference approach, and the denominator is the MSE resulting from the use of our approach. The empirical powers for both approaches are also provided. Furthermore, we present the number of times a working covariate type is selected out of the 1,000 simulations. Note that we do not use a Type IV time-dependent covariate in our simulation study as results are comparable to those corresponding to a true Type II covaraite due to the resemblance in definitions. In the Supplementary Materials, we provide empirical coverage probabilities (CPs) of 95% confidence intervals.

4.2 Results

Results corresponding to either a true Type I, II, or III time-dependent covariate (Tables 1, 2, and 3, respectively) show that the proposed selection approach used with the modified estimating equations method is more efficient than the approach incorporating an independence working correlation structure, i.e. use of working Type III, in the presence of within-subject correlation (cases 1–4). The REs ranged from 1.16 to 1.36, 1.03 to 1.13, and 1.03 to 1.11, correspondingly, over scenarios 1–3. When correlated heteroscedastic errors were accounted for (case 4), the results, in terms of REs and selection frequencies, were similar to those with errors following correlated parametric distributions (cases 1–3). In Table 4, RE results corresponding to a true Type II covariate in the multiple regression model were also similar to results observed in Table 2.

Results show, particularly in Tables 1, 2, and 4 with respect to Types I and II time-dependencies, that our proposed approach can notably improve power via the reduction in MSEs. In Table 1 for Type I time-dependency, empirical power increases range from approximately 0.04 to 0.20, with the majority of increases being greater than 0.10. In Tables 2 and 4 for Type II time-dependency, overall gains in power were not as high, but were still very notable and ranged from approximately 0.02 to 0.11. Finally, we note that in Supplementary Material we demonstrate that both methods are similar in terms of the validity of inference, and hence power gains with our proposed approach are due to reductions in MSEs.

In general, the proposed approach demonstrated improvements for any given quantile level. Improvements are due to the fact that the proposed method chose either Type I or Type II, relative to the inefficient choice of Type III, the majority of the time when Type I or II was the truth. Finally, bias in estimates was often negligible and similar for the two methods.

Additionally, our proposed approach demonstrated consistent results, relative to the independence estimating equations approach, in terms of REs and selection frequencies for any given quantile level. The RE results corresponding to cases 1–4 and five quantile levels under the three scenario settings also demonstrated that, given a higher within-subject correlation, the proposed selection method, in general, resulted in the greater efficiency and chose most often the desired type of covariate. The results with respect to REs and selection frequencies were comparable regardless of the given correlation structure. Our selection approach had efficiency gains when the true Type I or II was under consideration. When Type I was the truth, simulated biases were negligible as expected because all moments are valid in such settings. When the truth was Type II, Type I was the only specification type that can result in bias. In such settings, Type I was never selected with our proposed method because the bias was negligible. Furthermore, our approach ensured that when Type III was the truth, Type III was correctly chosen as the working type in the majority of the simulations and hence the resulting bias was negligible and comparable to the bias resulting from the independence estimating equations.

5 Application

We adopt the anthropometric screening data from the children study in the Philippines [8, 23] to examine the association between anthropometric factors and morbidity index over time. The data were severely skewed to the right and were originally obtained from 448 households from 1984 to 1985 [8]. Then, a subset of data containing 370 children (⩽ 14 years) was used as the final data [5, 23], in which each child had measurements at three time points with four months between each subsequent measurement. Children with incomplete measurements were excluded, and only one child per household was selected for eliminating statistical correlation resulted from household clustering [23].

We use the marginal model suggested in the existing literature [5, 6, 7, 32], but employ marginal quantile regression at three quantile levels, τ = 0.25, 0.50, and 0.75, given by

$$Y_{ij} = \beta_0 + \beta_1 BMI_{ij} + \beta_2 \text{ Age }_{ij} + \beta_3 Female_i + \beta_4 SR2_{ij} + \beta_5 SR3_{ij} + \epsilon_{ij}; \quad j = 1, 2, 3,$$

where Y_{ij}, as presented below, is the ith child's morbidity index during the jth four-month interval, and the morbidity index was conducted through the logistic transformation [5, 23].

$$Y_{ij} = \log\left(\frac{\text{days child was sick in last 2 weeks prior to time } j + 0.5}{14.5 - \text{ days child was sick in last 2 weeks prior to time } j}\right).$$

Three covariates, including age in months and two indicators for survey rounds 2 and 3, are categorized as the known Type I time-dependent covariates, whereas BMI's classification type of time-dependency is unknown and is the main focus of this analysis.

As in the simulation study, we analyze this data using the independence estimating equations method and our modified method with an AR-1 correlation structure, and select a working type for BMI through the use of our selection approach under three given quantiles. Table 5 gives the estimates of regression parameters and corresponding cluster bootstrapped SEs using 2,000 cluster bootstrap samples, as well as the working type for BMI selected by our method. We note that although it is common practice to utilize BMI z-scores, we do not do so in order to stay consistent with the time-dependent covariate literature which uses this data [5, 6, 7, 27, 32].

The proposed approach assigns a working Type III classification for BMI at the first quartile (25th quantile) and median (50th quantile), whereas a working Type I classification is chosen at the third quartile (75th quantile) based on the smallest criterion value. Although there is an apparent discrepancy in the type chosen, we note that the goal is not specifically to choose the true type; rather, for any given quantile, the goal is to select the type that results in the lowest MSE. Essentially, the discrepancy is due to different types being estimated to yield a smaller MSE at different quantiles. Note that at the 25th and 50th quantile levels both approaches produce similar results in terms of regression parameter and

SE estimates for BMI due to the choice of Type III. Furthermore, our proposed approach produces smaller SE estimates than the reference approach at the 75th quantile, thus revealing our proposed method's potential for efficiency improvement. For the other time-dependent covariates of known type, smaller SE estimates are obtained using the proposed method. The use of a marginal quantile analysis provides a more complete description with respect to BMI by investigating different quantiles for the right-skewed morbidity index distribution, rather than the marginal mean analysis which gives support for the use of a working Type I specification [22].

6 Concluding Remarks

Covariate values in a longitudinal study may change over time. Marginal mean regression analyses for longitudinal data have been widely introduced when covariates are time-variant. However, for some real-world data the use of mean regression models may be sensitive to skewness and outliers in the data. In such cases, the use of marginal quantile analysis for modeling the conditional quantiles of the response variable is recommended. Therefore, we proposed a new approach for marginal quantile regression in order to improve regression parameter estimation and hence power. Our proposed method was shown to be superior in regard to power in the presence of Type I or II time-dependency.

Although for simplicity we only considered independence and AR-1 working correlation structures in the manuscript, other structures with less parsimonious forms are available as well, including exchangeable and Toeplitz correlation matrices. We note that with our modified approach, the working structure is technically not an actual correlation structure because some non-zero elements of $\boldsymbol{R}_i^{\tau^{-1}}$ corresponding to invalid moment conditions are replaced with zeros. In this situation, $\boldsymbol{R}_{ik}^{\tau*^{-1}}$ will not be the inverse of a true correlation matrix when β_k^τ corresponds to a Type II or IV.

Our simulation study and application example were analyzed via marginal quantile regression models with balanced repeated measurements. Nonetheless, the proposed estimation approach and selection approach in this manuscript are applicable to subjects with varying repeated measurements. Future study can be extended to improve efficiency of estimation performance of composite marginal quantile regression [26], which has been proposed when multiple quantiles share common characteristics, in the presence of time-varying covariates. Furthermore, approaches using a general stationary autocorrelation structure [16] and a selection technique, via the use of a Gaussian pseudolikelihood in substitution for a parametric likelihood [19], to decide the most adequate working correlation structure have been suggested to prevent the specification of any specific working correlation structures. Simultaneously selecting a working correlation structure and deciding a covariate type of time-dependency can be further developed.

Coverage probabilities were sub-nominal in some settings (see Supplementary Material). Further work is needed with the existing and proposed methods to ensure proper CPs in all scenarios, although CP is not the inferential gain of focus in this manuscript.

The corresponding R code and functions for implementing the discussed approaches in this manuscript are given in Supporting Information and can be acquired by contacting the author at okv0@cdc.gov.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

References

[1]. Liang KY and Zeger SL. Longitudinal data analysis using generalized linear models. Biometrika 1986; 73: 13–22.

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

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

[4]. Wang YG and Carey V. Working correlation structure misspecification, estimation and covariate design: implications for generalised estimating equations performance. Biometrika 2003; 90: 29–41.

[5]. Lai TL and Small D. Marginal regression analysis of longitudinal data with time-dependent covariates: a generalized method-of-moments approach. Journal of the Royal Statistical Society: Series B 2007; 69: 79–99.

[6]. Zhou Y, Lefante J, Rice J et al. Using modified approaches on marginal regression analysis of longitudinal data with time-dependent covariates. Statistics in Medicine 2014; 33: 3354–3364. [PubMed: 24723212]

[7]. Chen IC and Westgate PM. Improved methods for the marginal analysis of longitudinal data in the presence of time-dependent covariates. Statistics in Medicine 2017; 36: 2533–2546. [PubMed: 28436045]

[8]. Bouis HE and Haddad LJ. Effects of agricultural commercialization on land tenure, household resource allocation, and nutrition in the philippines. Research Report 79, International Food Policy Research Institute, Washington DC, 1990.

[9]. Koenker R and Bassett G. Regression quantiles. Econometrica 1978; 46: 33–50.

[10]. Chen L, Wei LJ and Parzen MI. Quantile regression for correlated observations, in: Proceedings of the Second Seattle Symposium in Biostatistics: Analysis of Correlated Data. New York: Springer, 2003.

[11]. Yin G and Cai J. Quantile regression models with multivariate failure time data. Biometrics 2005; 61: 151–161. [PubMed: 15737088]

[12]. Wang HJ and Zhu Z. Empirical likelihood for quantile regression model with longitudinal data. Journal of Statistical Planning and Inference 2011; 141: 1603–1615.

[13]. Tang CY and Leng C. Empirical likelihood and quantile regression in longitudinal data analysis. Biomerika 2011; 98: 1001–1006.

[14]. Fu L and Wang YG. Quantile regression for longitudinal data with a working correlation model. Computational Statistics and data Analysis 2012; 56: 2526–2538.

[15]. Leng C and Zhang W. Smoothing combined estimating equations in quantile regression for longitudinal data. Statistics and Computing 2014; 24: 123–136.

[16]. Lu X and Fan Z. Weighted quantile regression for longitudinal data. Computational Statistics 2015; 30: 569–592.

[17]. Jung SH. Quasi-likelihood for median regression models. Journal of American Statistical Association 1996; 91: 251–257.

[18]. Westgate PM. Criterion for the simultaneous selection of a working correlation structure and either generalized estimating equations or the quadratic inference function approach. Biometrical Journal 2014; 56: 461–476. [PubMed: 24431030]

[19]. Fu L, Wang YG and Zhu M. A gaussian pseudolikelihood approach for quantile regression with repeated measurements. Computational Statistics and data Analysis 2015; 84: 41–53.

[20]. Stoner JA and Leroux BG. Analysis of clustered data: a combined estimating equations approach. Biometrika 2002; 89: 567–578.

[21]. Westgate PM. Improving the correlation structure selection approach for generalized estimating equations and balanced longitudinal data. Statistics in Medicine 2014; 33: 2222–2237. [PubMed: 24504841]

[22]. Chen IC and Westgate PM. A novel approach to selecting classification types for time-dependent covariates in the marginal analysis of longitudinal data. Statistical Methods in Medical Research 2019; 28: 3176–3186. [PubMed: 30203725]

[23]. Bhargava A Modelling the health of filipino children. Journal of the Royal Statistical Society: Series A 1994; 157: 417–432.

[24]. Brown BM and Wang YG. Standard errors and covariance matrices for smoothed rank estimators. Biometrika 2005; 92: 149–158.

[25]. Pang L, Lu W and Wang HJ. Variance estimation in censored quantile regression via induced smoothing. Computational Statistics and Data Analysis 2012; 56: 785–796. [PubMed: 22547899]

[26]. Yang CC, Chen YH and Chang HY. Composite marginal quantile regression analysis for longitudinal adolescent body mass index data. Statistics in Medicine 2017; 36: 3380–3397. [PubMed: 28574584]

[27]. Lalonde TL, Wilson JR and Yin J. Gmm logistic regression models for longitudinal data with time-dependent covariates and extended classifications. Statistics in Medicine 2014; 33: 4756–4769. [PubMed: 25130989]

[28]. Moulton LH and Zeger SL. Analyzing repeated measures on generalized linear models via the bootstrap. Biometrics 1989; 45: 381–394.

[29]. Sherman M and le Cessie S. A comparison between bootstrap methods and generalized estimating equations for correlated outcomes in generalized linear models. Communications in Statistics-Simulation and Computation 1997; 26: 901–925.

[30]. Diggle PJ, Heagerty PJ, Liang KY et al. The Analysis of Longitudinal Data. 2nd ed. New York: Oxford University Press, 2002.

[31]. R Core Team. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria, 2019. URL https://www.R-project.org/.

[32]. Leung DHY, Small DS, Qin J et al. Shrinkage empirical likelihood estimator in longitudinal analysis with time-dependent covariates–application to modeling the health of filipino children. Biometrics 2013; 69: 624–632. [PubMed: 23845158]

Table 1:

Results for Cases 1–4 in which one Type I time-dependent covariate is incorporated.

		τ=0.10		τ=0.25		τ=0.50		τ=0.75		τ=0.90	
		EX	***AR*-1**	***EX***	***AR*-1**	***EX***	***AR*-1**	***EX***	***AR*-1**	***EX***	***AR*-1**
	$Bias_I$	*−.0026*	*−.0026*	*.0008*	*.0008*	*.0029*	*.0029*	*.0003*	*.0003*	*.0043*	*.0043*
	$Bias_P$	*−.0012*	*−.0006*	*.0026*	*.0017*	*.0017*	*.0020*	*.0005*	*.0012*	*.0012*	*.0025*
	MSE_P	*.0092*	*.0091*	*.0069*	*.0071*	*.0064*	*.0065*	*.0072*	*.0071*	*.0092*	*.0091*
	RE	*1.245*	*1.259*	*1.289*	*1.258*	*1.257*	*1.249*	*1.254*	*1.273*	*1.266*	*1.276*
Case (1)	$Power_I$	*0.635*	*0.635*	*0.714*	*0.714*	*0.734*	*0.734*	*0.691*	*0.691*	*0.652*	*0.652*
	$Power_P$	*0.785*	*0.785*	*0.861*	*0.849*	*0.872*	*0.877*	*0.840*	*0.847*	*0.791*	*0.818*
	Type I	590	571	586	561	552	554	534	544	577	624
	Type II	262	305	264	314	280	296	296	323	275	255
	Type III	148	124	150	125	168	150	170	133	148	121
	$Bias_I$	*−.0089*	*−.0089*	*.0067*	*.0067*	*.0015*	*.0015*	*−.0003*	*−.0003*	*.0025*	*.0025*
	$Bias_P$	*−.0060*	*−.0062*	*.0055*	*.0048*	*.0010*	*.0017*	*.0004*	*.0007*	*.0010*	*.0023*
	MSE_P	*.0146*	*.0145*	*.0095*	*.0093*	*.0083*	*.0081*	*.0097*	*.0097*	*.0149*	*.0148*
	RE	*1.237*	*1.250*	*1.212*	*1.241*	*1.181*	*1.222*	*1.243*	*1.247*	*1.223*	*1.232*
Case (2)	$Power_I$	*0.454*	*0.454*	*0.649*	*0.649*	*0.659*	*0.659*	*0.608*	*0.608*	*0.482*	*0.482*
	$Power_P$	*0.627*	*0.652*	*0.779*	*0.781*	*0.811*	*0.832*	*0.764*	*0.782*	*0.649*	*0.661*
	Type I	544	583	568	569	547	551	533	546	577	581
	Type II	256	263	263	309	284	330	286	325	246	271
	Type III	200	154	169	122	169	119	181	129	177	148
	$Bias_I$	*.0041*	*.0041*	*−.0004*	*−.0004*	*.0056*	*.0056*	*.0023*	*.0023*	*.0115*	*.0115*
	$Bias_P$	*.0026*	*.0029*	*.0001*	*−.0002*	*.0032*	*.0040*	*.0016*	*.0030*	*.0093*	*.0126*
	MSE_P	*.0073*	*.0070*	*.0069*	*.0067*	*.0077*	*.0076*	*.0121*	*.0118*	*.0444*	*.0446*
	RE	*1.311*	*1.363*	*1.232*	*1.274*	*1.195*	*1.212*	*1.199*	*1.228*	*1.172*	*1.166*
Case (3)	$Power_I$	*0.700*	*0.700*	*0.735*	*0.735*	*0.727*	*0.727*	*0.581*	*0.581*	*0.329*	*0.329*
	$Power_P$	*0.870*	*0.872*	*0.866*	*0.861*	*0.859*	*0.868*	*0.714*	*0.730*	*0.368*	*0.386*
	Type I	600	588	502	512	543	547	530	550	470	517
	Type II	248	296	310	344	279	316	305	308	258	229
	Type III	152	116	188	144	178	137	165	142	272	254
	$Bias_I$	*.0014*	*.0014*	*−.0015*	*−.0015*	*.0004*	*.0004*	*.0015*	*.0015*	*−.0025*	*−.0025*
	$Bias_P$	*.0015*	*.0005*	*.0007*	*−.0002*	*.0003*	*−.0015*	*.0012*	*.0012*	*−.0014*	*−.0038*
	MSE_P	*.0078*	*.0074*	*.0061*	*.0061*	*.0050*	*.0048*	*.0058*	*.0058*	*.0073*	*.0072*
	RE	*1.219*	*1.281*	*1.253*	*1.246*	*1.236*	*1.291*	*1.262*	*1.253*	*1.276*	*1.294*
Case (4)	$Power_I$	*0.716*	*0.716*	*0.765*	*0.765*	*0.812*	*0.812*	*0.776*	*0.776*	*0.690*	*0.690*
	$Power_P$	*0.857*	*0.874*	*0.897*	*0.890*	*0.912*	*0.914*	*0.893*	*0.896*	*0.860*	*0.855*
	Type I	485	486	493	483	521	525	484	462	491	471
	Type II	328	399	322	380	319	352	318	388	315	386
	Type III	187	115	185	137	160	123	198	150	194	143

τ - quantile level; EX - exchangeable; AR-1 - first-order autoregressive;

$Bias_I$ and $Bias_P$ - empirical biases of the method with an independence structure and the proposed method;

MSE_P - empirical mean squared error (MSE) of the proposed approach;

RE - relative efficiency or ratio of the empirical MSE from the estimation method with an independence structure to the MSE from the proposed method;

$Power_I$ and $Power_P$ - empirical powers of the reference approach and the proposed approach, respectively;

Types I-III - the number of times out of 1,000 simulations that the given covariate type was selected.

Table 2:

Results for Cases 1–4 in which one Type II time-dependent covariate is incorporated.

		τ=0.10		τ=0.25		τ=0.50		τ=0.75		τ=0.90	
		EX	***AR*-1**	***EX***	***AR*-1**	***EX***	***AR*-1**	***EX***	***AR*-1**	***EX***	***AR*-1**
Case (1)	*Bias* $_I$	−.0044	−.0044	−.0000	−.0000	.0001	.0001	−.0012	−.0012	.0047	.0047
	Bias $_P$	−.0077	−.0042	−.0006	−.0003	−.0004	.0005	−.0024	−.0009	.0020	.0053
	MSE $_P$	.0131	.0124	.0104	.0102	.0092	.0092	.0110	.0109	.0128	.0122
	RE	1.054	1.109	1.082	1.103	1.125	1.114	1.101	1.110	1.067	1.115
	Power $_I$	0.890	0.890	0.944	0.944	0.955	0.955	0.935	0.935	0.891	0.891
	Power $_P$	0.956	0.945	0.979	0.977	0.983	0.979	0.979	0.978	0.960	0.950
	Type I	40	0	14	0	4	0	12	0	30	0
	Type II	769	792	730	781	707	739	731	760	747	810
	Type III	191	208	256	219	289	261	257	240	223	190
Case (2)	*Bias* $_I$	−.0083	−.0083	.0015	.0015	.0006	.0006	−.0014	−.0014	.0021	.0021
	Bias $_P$	−.0116	−.0057	.0005	.0030	−.0002	.0013	−.0041	−.0010	−.0025	.0021
	MSE $_P$	.0185	.0172	.0138	.0130	.0117	.0116	.0143	.0138	.0182	.0169
	RE	1.033	1.112	1.049	1.107	1.081	1.092	1.048	1.085	1.054	1.136
	Power $_I$	0.741	0.741	0.882	0.882	0.912	0.912	0.885	0.885	0.777	0.777
	Power $_P$	0.859	0.840	0.944	0.940	0.965	0.971	0.947	0.938	0.865	0.847
	Type I	82	3	34	0	18	0	24	0	65	2
	Type II	686	762	724	787	685	732	707	758	693	740
	Type III	232	235	242	213	297	268	269	242	242	258
Case (3)	*Bias* $_I$	.0041	.0041	−.0015	−.0015	.0039	.0039	.0011	.0011	.0042	.0042
	Bias $_P$	.0006	.0039	−.0025	−.0008	.0017	.0025	−.0006	.0032	−.0057	.0031
	MSE $_P$	.0126	.0120	.0107	.0106	.0113	.0112	.0146	.0138	.0375	.0368
	RE	1.077	1.128	1.096	1.111	1.074	1.086	1.056	1.113	1.058	1.079
	Power $_I$	0.933	0.933	0.952	0.952	0.944	0.944	0.832	0.832	0.466	0.466
	Power $_P$	0.978	0.978	0.979	0.981	0.969	0.974	0.898	0.899	0.511	0.519
	Type I	35	0	14	0	10	0	31	0	165	22
	Type II	798	845	683	758	685	732	715	759	546	662
	Type III	167	155	303	242	305	268	254	241	289	316
Case (4)	*Bias* $_I$	.0008	.0008	−.0031	−.0031	−.0020	−.0020	.0029	.0029	−.0023	−.0023
	Bias $_P$	−.0008	.0011	−.0035	−.0033	−.0022	−.0010	.0022	.0025	−.0047	−.0021
	MSE $_P$	.0115	.0114	.0091	.0090	.0073	.0073	.0086	.0085	.0115	.0109
	RE	1.112	1.124	1.070	1.084	1.109	1.104	1.096	1.106	1.075	1.127
	Power $_I$	0.937	0.937	0.966	0.966	0.974	0.974	0.967	0.967	0.948	0.948
	Power $_P$	0.973	0.975	0.985	0.990	0.989	0.988	0.985	0.992	0.983	0.979
	Type I	16	0	9	0	0	0	8	0	18	0
	Type II	802	847	722	786	699	706	758	817	754	815
	Type III	182	153	269	214	301	294	234	183	228	185

τ - quantile level; EX - exchangeable; AR-1 - first-order autoregressive;

$Bias_I$ and $Bias_P$ - empirical biases of the method with an independence structure and the proposed method;

MSE_P - empirical mean squared error (MSE) of the proposed approach;

RE - relative efficiency or ratio of the empirical MSE from the estimation method with an independence structure to the MSE from the proposed method;

$Power_I$ and $Power_P$ - empirical powers of the reference approach and the proposed approach, respectively;

Types I-III - the number of times out of 1,000 simulations that the given covariate type was selected.

Table 3:

Results for Cases 1–4 in which one Type III time-dependent covariate is included.

		τ=0.10		τ=0.25		τ=0.50		τ=0.75		τ=0.90	
		EX	***AR*-1**	***EX***	***AR*-1**	***EX***	***AR*-1**	***EX***	***AR*-1**	***EX***	***AR*-1**
Case (1)	$Bias_I$	.0098	.0098	.0140	.0140	.0158	.0158	.0163	.0163	.0115	.0115
	$Bias_P$	.0088	.0083	.0134	.0126	.0154	.0145	.0158	.0150	.0101	.0097
	MSE_P	.0019	.0019	.0014	.0014	.0012	.0012	.0016	.0016	.0018	.0018
	RE	1.042	1.049	1.032	1.040	1.047	1.058	1.029	1.032	1.045	1.051
	$Power_I$	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	$Power_P$	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	Type I	2	49	3	22	2	23	0	19	9	45
	Type II	141	297	57	244	55	253	68	254	137	301
	Type III	857	654	940	734	943	724	932	727	854	654
Case (2)	$Bias_I$	.0079	.0079	.0152	.0152	.0173	.0173	.0129	.0129	.0080	.0080
	$Bias_P$	.0060	.0056	.0147	.0138	.0167	.0156	.0122	.0113	.0058	.0054
	MSE_P	.0027	.0027	.0019	.0019	.0016	.0016	.0018	.0018	.0028	.0028
	RE	1.048	1.046	1.037	1.045	1.049	1.050	1.042	1.049	1.056	1.051
	$Power_I$	0.999	0.999	1.000	1.000	1.000	1.000	1.000	1.000	0.998	0.998
	$Power_P$	0.999	0.999	1.000	1.000	1.000	1.000	1.000	1.000	0.999	0.999
	Type I	3	32	1	7	0	5	0	4	9	28
	Type II	223	380	83	269	42	203	81	268	219	353
	Type III	774	588	916	724	958	792	919	728	772	619
Case (3)	$Bias_I$	.0200	.0200	.0195	.0195	.0187	.0187	.0076	.0076	.0089	.0089
	$Bias_P$	.0193	.0186	.0187	.0182	.0184	.0174	.0060	.0051	.0041	.0038
	MSE_P	.0017	.0017	.0012	.0012	.0015	.0015	.0024	.0024	.0063	.0063
	RE	1.035	1.045	1.052	1.064	1.045	1.051	1.035	1.024	1.109	1.116
	$Power_I$	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.914	0.914
	$Power_P$	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.941	0.938
	Type I	3	36	14	104	1	16	1	1	13	21
	Type II	30	190	43	175	32	221	163	350	383	459
	Type III	967	774	943	721	967	763	836	649	604	520
Case (4)	$Bias_I$	.0128	.0128	.0164	.0164	.0169	.0169	.0148	.0148	.0123	.0123
	$Bias_P$	.0111	.0106	.0155	.0148	.0159	.0155	.0133	.0129	.0106	.0099
	MSE_P	.0029	.0029	.0021	.0021	.0018	.0018	.0020	.0020	.0026	.0026
	RE	1.035	1.032	1.039	1.051	1.053	1.058	1.054	1.065	1.048	1.047
	$Power_I$	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	$Power_P$	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	Type I	2	18	5	39	6	86	4	40	0	25
	Type II	124	329	82	291	112	298	105	277	126	313
	Type III	874	653	913	670	882	616	891	683	874	662

τ - quantile level; EX - exchangeable; AR-1 - first-order autoregressive;

$Bias_I$ and $Bias_P$ - empirical biases of the method with an independence structure and the proposed method;

MSE_P - empirical mean squared error (MSE) of the proposed approach;

RE - relative efficiency or ratio of the empirical MSE from the estimation method with an independence structure to the MSE from the proposed method;

$Power_I$ and $Power_P$ - empirical powers of the reference approach and the proposed approach, respectively;

Types I-III - the number of times out of 1,000 simulations that the given covariate type was selected.

Table 4:

Results for Cases 1–4 in which one time-independent, one Type I, and one Type II time-dependent covariates are incorporated. Only results corresponding to the Type II covariate are shown.

		τ=0.10		τ=0.25		τ=0.50		τ=0.75		τ=0.90	
		EX	***AR*-1**	***EX***	***AR*-1**	***EX***	***AR*-1**	***EX***	***AR*-1**	***EX***	***AR*-1**
Case (1)	$Bias_I$	−.0073	−.0073	−.0023	−.0023	−.0093	−.0093	−.0030	−.0030	−.0106	−.0106
	$Bias_P$	−.0091	−.0060	−.0051	−.0029	−.0101	−.0100	−.0048	−.0028	−.0149	−.0102
	MSE_P	.0150	.0153	.0101	.0098	.0089	.0088	.0110	.0107	.0154	.0148
	RE	1.159	1.139	1.116	1.150	1.101	1.106	1.080	1.119	1.110	1.156
	$Power_I$	0.849	0.849	0.941	0.941	0.946	0.946	0.919	0.919	0.846	0.846
	$Power_P$	0.927	0.917	0.987	0.988	0.976	0.979	0.972	0.972	0.914	0.922
	Type I	41	0	15	0	7	0	18	0	59	1
	Type II	588	628	677	741	691	771	657	734	577	650
	Type III	371	372	308	259	302	229	325	266	364	349
Case (2)	$Bias_I$	−.0106	−.0106	−.0082	−.0082	−.0006	−.0006	−.0036	−.0036	−.0044	−.0044
	$Bias_P$	−.0175	−.0102	−.0105	−.0080	−.0015	−.0001	−.0055	−.0027	−.0087	−.0037
	MSE_P	.0255	.0248	.0144	.0143	.0109	.0110	.0132	.0127	.0230	.0228
	RE	1.109	1.140	1.112	1.123	1.144	1.143	1.094	1.142	1.122	1.134
	$Power_I$	0.682	0.682	0.843	0.843	0.919	0.919	0.875	0.875	0.689	0.689
	$Power_P$	0.791	0.768	0.946	0.937	0.965	0.968	0.943	0.941	0.782	0.776
	Type I	105	5	23	0	14	0	28	0	94	6
	Type II	529	643	652	730	696	749	676	735	560	645
	Type III	366	352	325	270	290	251	296	265	346	349
Case (3)	$Bias_I$	−.0113	−.0113	−.0025	−.0025	−.0038	−.0038	−.0008	−.0008	−.0095	−.0095
	$Bias_P$	−.0134	−.0094	−.0035	−.0020	−.0060	−.0039	−.0060	−.0003	−.0143	−.0092
	MSE_P	.0136	.0133	.0103	.0101	.0105	.0103	.0175	.0168	.0521	.0546
	RE	1.105	1.131	1.107	1.134	1.091	1.114	1.074	1.117	1.108	1.058
	$Power_I$	0.909	0.909	0.950	0.950	0.945	0.945	0.763	0.763	0.461	0.461
	$Power_P$	0.969	0.975	0.983	0.991	0.970	0.978	0.872	0.844	0.521	0.519
	Type I	50	0	7	0	15	0	54	1	224	66
	Type II	608	663	674	720	700	768	624	712	422	559
	Type III	342	337	319	280	285	232	322	287	354	375
Case (4)	$Bias_I$	−.0109	−.0109	−.0064	−.0064	−.0035	−.0035	−.0057	−.0057	−.0088	−.0088
	$Bias_P$	−.0147	−.0091	−.0081	−.0051	−.0038	−.0028	−.0058	−.0044	−.0139	−.0086
	MSE_P	.0145	.0139	.0114	.0109	.0097	.0096	.0112	.0107	.0150	.0136
	RE	1.100	1.148	1.057	1.100	1.098	1.113	1.081	1.131	1.015	1.119
	$Power_I$	0.878	0.878	0.936	0.936	0.966	0.966	0.942	0.942	0.896	0.896
	$Power_P$	0.938	0.942	0.962	0.967	0.990	0.993	0.971	0.982	0.950	0.947
	Type I	91	0	23	0	10	0	17	0	82	1
	Type II	556	653	662	727	680	735	657	716	541	636

	τ=0.10		τ=0.25		τ=0.50		τ=0.75		τ=0.90	
	EX	***AR*-1**	***EX***	***AR*-1**	***EX***	***AR*-1**	***EX***	***AR*-1**	***EX***	***AR*-1**
Type III	353	347	315	273	310	265	326	284	377	363

τ - quantile level; EX - exchangeable; AR-1 - first-order autoregressive;

$Bias_I$ and $Bias_P$ - empirical biases of the method with an independence structure and the proposed method;

MSE_P - empirical mean squared error (MSE) of the proposed approach;

RE - relative efficiency or ratio of the empirical MSE from the estimation method with an independence structure to the MSE from the proposed method;

$Power_I$ and $Power_P$ - empirical powers of the reference approach and the proposed approach, respectively;

Types I-III - the number of times out of 1,000 simulations that the given covariate type was selected.

Table 5:

Parameter estimates, empirical and cluster bootstrapped standard error estimates (in parentheses), and working types of covariate for BMI resulting from analyses of the anthropometric dataset.

	Independence			Proposed*		
Variable	**$\tau = 0.25$**	**$\tau = 0.50$**	**$\tau = 0.75$**	**$\tau = 0.25$**	**$\tau = 0.50$**	**$\tau = 0.75$**
BMI	−0.20 (0.002)	−0.18 (0.003)	−0.05 (0.024)	−0.20 (0.002)	−0.18 (0.003)	−0.05 (0.019)
Age	−0.01 (0.001)	−0.01 (0.001)	−0.03 (0.005)	−0.01 (0.001)	−0.01 (0.001)	−0.03 (0.005)
Gender	−0.02 (0.021)	−0.02 (0.030)	0.41 (0.267)	−0.02 (0.026)	−0.01 (0.036)	0.42 (0.221)
SR 2	−0.08 (0.025)	−0.08 (0.036)	−0.69 (0.328)	−0.07 (0.021)	−0.06 (0.033)	−0.66 (0.248)
SR 3	0.002 (0.027)	0.05 (0.035)	0.42 (0.374)	0.01 (0.024)	0.06 (0.034)	0.47 (0.281)
Type				III	III	I

τ - quantile level; SR - survey round; Type - working covariate type for BMI.

*Note that the standard error estimates are obtained using the cluster bootstrapped method.