COVARIATE-ADJUSTED REGRESSION FOR LONGITUDINAL DATA INCORPORATING CORRELATION BETWEEN REPEATED MEASUREMENTS

(Running Title: LONGITUDINAL COVARIATE-ADJUSTED REGRESSION)

DANH V. NGUYEN[1] * AND DAMLA ŞENTÜRK[2]

University of California at Davis and Pennsylvania State University

Summary

We propose an estimation method that incorporates the correlation/covariance structure between repeated measurements in covariate-adjusted regression models for distorted longitudinal data. In this distorted data setting, both the longitudinal response and (possibly time varying) predictors are not directly observable. The unobserved response and predictors are assumed to be distorted/contaminated by unknown functions of a common observable confounder. The proposed estimation methodology adjusts for the distortion effects both in estimation of the covariance structure and the regression parameters using generalized least squares. The finite sample performance of the proposed estimators is studied numerically via simulations. The consistency and convergence rates of the proposed estimators are also established. The proposed method is illustrated with an application to data from a longitudinal study of the cognitive and social development in children.

Key words: binning; clustered data; covariance structure; general linear model; generalized least squares; multiplicative effect; varying-coefficient models.

---

*Author to whom correspondence should be addressed
[1]Division of Biostatistics, MS1C, 1 Shields Ave, University of California, Davis, California 95616, USA
e-mail: ucdnguyen@ucdavis.edu, phone: (530) 754-6510
[2] Department of Statistics, Pennsylvania State University, University Park, Pennsylvania 16802, USA
Acknowledgments. We are grateful to anonymous referees, an associate editor and the editor for constructive suggestions that improved the paper. This publication was made possible by National Institute of Health grants UL1RR024922, RL1AG032119, RL1AG032115 and UL1 RR024146.

1 Introduction

Covariate-adjusted regression (CAR) is an adjustment method for the multiplicative effects of a confounder on the response and the predictor for cross-sectional data (Şentürk & Müller, 2005). It is used when both the response and predictors are not directly observed, but are observed after distortion by unknown functions of an observable confounder. The observed data for the ith subject is modelled as

$$\widetilde{Y}_i = \psi(U_i)Y_i \quad \text{and} \quad \widetilde{X}_i = \phi(U_i)X_i, \quad i = 1, \ldots, n, \tag{1}$$

where $\psi(\cdot)$ and $\phi(\cdot)$ are the unknown smooth distorting functions of the confounder U. Here Y and X denote the unobserved parts of the observed response, $\widetilde{Y}$, and the predictor, $\widetilde{X}$. CAR uncovers the underlying regression relationship between Y and X, adjusted for U, by providing consistent estimates for the parameters in the unobserved model $\mathrm{E}(Y_i|X_i) = \gamma_0 + \gamma_1 X_i$, based on the observed data $\{\widetilde{Y}_i, \widetilde{X}_i, U_i\}_{i=1}^n$. A key advantage of the CAR estimators is that they are consistent under multiplicative distortion (1), as well as additive ($\widetilde{Y}_i = \psi(U_i) + Y_i$ and $\widetilde{X}_i = \phi(U_i) + X_i$) and no distortion ($\widetilde{Y}_i = Y_i$ and $\widetilde{X}_i = X_i$).

In this work we generalize the CAR method of Şentürk & Müller to a longitudinal data setting, where there is a need to adjust for a cross-sectional confounder, such as a baseline confounder. The proposed generalized covariate adjustment procedure involves incorporating the covariance/correlation between repeated measurements on the same individual. This will require an adjusted estimation procedure for the covariance as well as the underlying regression coefficients. We propose a method that incorporates a consistent estimator of the underlying covariance via generalized least squares (GLS). We note that a longitudinal analysis that does not account for the distortions can be biased. Direct estimation of the covariance between repeated measurements based on the observed data will be biased, and an adjustment is needed to eliminate the bias due to the distortion. Thus, standard methods for estimating the covariance, including maximum likelihood (ML) or restricted ML (REML; Searle, Casella & McCulloch, 1992) and nonparametric estimation (e.g., Diggle & Verbyla, 1998), cannot capture the covariance structure. Estimation of the regression parameters and the covariance must both eliminate the distortion. A proposed covariate-adjusted general linear model (CAGLM) estimator, detailed in Section 3, achieves these goals.

The proposed method is illustrated with a longitudinal data set from Curran (1997), who studied the relationship between childhood learning and behavioural development. The main interest is to model the development of early childhood reading skills. Learning is related to, among other factors, the child's antisocial behaviour and the level of parental cognitive and emotional support at home. The relationship between early learning skills and the given explanatory variables is potentially confounded by the mother's age at the start of child rearing, which affects learning, behaviour and the home environment.

The paper is organized as follows. We describe the CAGLM for distorted longitudinal data in Section 2. Estimation, consistency and convergence rates are established in Section 3. In Section 4, the finite sample properties of the estimators are studied via simulations. An application of the proposed methodology to the aforementioned data on childhood reading cognition is illustrated in Section 5. We conclude with a discussion in Section 6. Proofs of the main results are deferred to the Appendix section.

2 Covariate-adjusted general linear models

Let $\mathbf{Y}_i = (Y_{i1}, \ldots, Y_{iT_i})$ be the response measurements at times $(t_{i1}, \ldots, t_{iT_i})$ and $X_{1ij}, \ldots, X_{pij}$ be p explanatory variables for subject i. The $T_i \times (p+1)$ matrix of predictor values for subject i is denoted by $\mathbf{X}_i$, where the jth row is $\mathbf{X}_{ij}^\top = (1, X_{1ij}, \ldots, X_{pij})$. We consider the following underlying linear regression model for the ith subject, $i = 1, \ldots, n$,

$$\mathbf{Y}_i = \mathbf{X}_i \boldsymbol{\gamma} + \mathbf{e}_i, \tag{2}$$

where $\boldsymbol{\gamma} = (\gamma_0, \ldots, \gamma_p)$ and $\mathbf{e}_i = (e_{i1}, \ldots, e_{iT_i})$ are the regression coefficients and random errors with mean zero and covariance matrix $\mathrm{cov}(\mathbf{e}_i) = \boldsymbol{\Sigma}_i$, respectively. Our objectives are estimation of the underlying parameters, $\boldsymbol{\gamma}$, and incorporation of the covariance information in the estimation procedure, based on the distorted response and predictors,

$$\widetilde{Y}_{ij} = \psi(U_i) Y_{ij} \qquad \text{and} \qquad \widetilde{X}_{rij} = \phi_r(U_i) X_{rij}, \qquad \text{for } r = 1, \ldots, p. \tag{3}$$

The distorting functions $\{\psi(\cdot), \phi_r(\cdot)\}_{r=1}^p$ are assumed to be smooth functions of the confounder U. The distorted data available for estimation is the collection $\{\widetilde{\mathbf{Y}}_i, \widetilde{\mathbf{X}}_i, U_i\}_{i=1}^n$ for n subjects. The T_i-vector $\widetilde{\mathbf{Y}}_i$ is the observed vector of measurements for subject i taken at T_i occasions and $\widetilde{\mathbf{X}}_i$ is the matrix of predictor values defined analogously to $\mathbf{X}_i$ above. Also, let T denote the distinct occasions among all t_{ij}, for $1 \leq i \leq n$ and $1 \leq j \leq T_i$.

The model defined by (2)-(3) is a generalization of the CAR model of Şentürk & Müller (2005) for longitudinal data. For the estimation, a similar identifiability condition used for CAR is needed, which is that the distortion is mean preserving, i.e., the means of the observed variables $\mathrm{E}(\widetilde{Y}_{ij})$ and $\mathrm{E}(\widetilde{X}_{rij})$ are the same as those of the underlying variables, $\mathrm{E}(Y_{ij})$ and $\mathrm{E}(X_{rij})$, respectively. We note that this is equivalent to $\mathrm{E}\{\psi(U_i)\} = 1$ and $\mathrm{E}\{\phi_r(U_i)\} = 1$ under distortion (3). See Şentürk & Müller (2005) for details. We refer to the model described above as the CAGLM with correlated errors. Furthermore, as in the case for cross-sectional data, a varying coefficient model (Cleveland, Grosse & Shyu, 1991; Hastie & Tibshirani, 1993) holds between the distorted variables:

$$\widetilde{Y}_{ij} = \beta_0(U_i) + \sum_{r=1}^{p} \beta_r(U_i) \widetilde{X}_{rij} + \epsilon_{ij}, \qquad \text{where } \epsilon_{ij} = \psi(U_i) e_{ij}, \tag{4}$$

$$\beta_0(U_i) = \gamma_0 \psi(U_i), \qquad \text{and} \qquad \beta_r(U_i) = \gamma_r \frac{\psi(U_i)}{\phi_r(U_i)}. \tag{5}$$

Estimation based on model (4) is feasible since $(\widetilde{Y}_{ij}, \widetilde{X}_{rij}, U_i)$ are observable. Our estimation strategy is to first target the varying coefficient functions in (4) and then utilize their relationships to the distorting functions given by (5) to mitigate the distortion effects.

3 Estimation procedure and consistency

3.1 Two-step estimation procedure

Estimation of $\{\gamma_r\}_{r=0}^p$ proceeds in two steps. Step 1 involves binning the data according to U and then fitting a GLS regression of $\widetilde{Y}$ on $(\widetilde{X}_1, \dots, \widetilde{X}_p)$ within each bin. The coefficients estimated in each bin are the raw estimates of $\{\beta_r(\cdot)\}_{r=0}^p$. Estimates $\{\gamma_r\}_{r=0}^p$ are targeted in step 2 as weighted averages of the raw estimates from step 1. Details follow.

Step 1 of the estimation procedure divides the range of U into m equidistant intervals, $B_1, \dots, B_m$, referred to as bins. Let L_v be the number of subjects falling into bin v, $1 \le v \le m$. To track the observations that fall into a given bin, observations in any bin are marked by a prime. Thus, the data for which $U_i \in B_v$ is given by the collection $\{(U'_{vk}, \widetilde{\mathbf{X}}'_{vk}, \widetilde{\mathbf{Y}}'_{vk}), k = 1, \dots, L_v\} = \{(U_i, \widetilde{\mathbf{X}}_i, \widetilde{\mathbf{Y}}_i), i = 1, \dots, n : U_i \in B_v\}$. The kth data element in bin v (subject k) is $(U'_{vk}, \widetilde{\mathbf{X}}'_{vk}, \widetilde{\mathbf{Y}}'_{vk})$, where $\widetilde{\mathbf{Y}}'_{vk}$, $\widetilde{\mathbf{X}}'_{vk}$, and U'_{vk} are the $T'_{vk} \times 1$ vector of responses, the $T'_{vk} \times (p+1)$ matrix of p predictors, and the confounder, respectively. Similarly, denote the measurement times by $(t'_{vk1}, \dots, t'_{vkT'_{vk}})$.

Raw estimates of $\beta_r(\cdot)$ can be based on weighted least squares (WLS) regression,

$$\widehat{\boldsymbol{\beta}}_v^{wls} = (\widehat{\beta}_{0v}^{wls}, \dots, \widehat{\beta}_{pv}^{wls}) = \left(\sum_{k=1}^{L_v} \widetilde{\mathbf{X}}_{vk}^{\prime\top} \mathbf{W}_{vk}^{-1} \widetilde{\mathbf{X}}'_{vk} \right)^{-1} \sum_{k=1}^{L_v} \widetilde{\mathbf{X}}_{vk}^{\prime\top} \mathbf{W}_{vk}^{-1} \widetilde{\mathbf{Y}}'_{vk} \tag{6}$$

within each bin, where $\mathbf{W}_{vk} = \mathrm{diag}(w_{vk1}, \dots, w_{vkT'_{vk}})$ and w_{vkj} is the number of subjects in bin v that are observed at time t'_{vkj}, $j = 1, \dots, T'_{vk}$. The WLS estimates in (6) do not account for the covariance between repeated measurements. Incorporating the covariance leads to the following GLS estimates,

$$\widetilde{\boldsymbol{\beta}}_v = (\widetilde{\beta}_{0v}, \dots, \widetilde{\beta}_{pv}) = \left(\sum_{k=1}^{L_v} \widetilde{\mathbf{X}}_{vk}^{\prime\top} \boldsymbol{\Sigma}_{vk}^{-1} \widetilde{\mathbf{X}}'_{vk} \right)^{-1} \sum_{k=1}^{L_v} \widetilde{\mathbf{X}}_{vk}^{\prime\top} \boldsymbol{\Sigma}_{vk}^{-1} \widetilde{\mathbf{Y}}'_{vk}, \tag{7}$$

where $\boldsymbol{\Sigma}_{vk}$ denotes the true covariance matrix for subject k in bin v. Since $\boldsymbol{\Sigma}_{vk}$ is unknown, it must be estimated from the distorted data. The details of this estimation will be given in Section 3.2. Assume that $\widehat{\mathbf{V}}_{vk}$ is a consistent estimator for $\boldsymbol{\Sigma}_{vk}$ up to a constant of proportionality. Substituting $\widehat{\mathbf{V}}_{vk}$ into (7), we obtain the GLS estimator

$$\widehat{\boldsymbol{\beta}}_v^{gls} = (\widehat{\beta}_{0v}^{gls}, \ldots, \widehat{\beta}_{pv}^{gls}) = \left(\sum_{k=1}^{L_v} \widetilde{\mathbf{X}}'^{\top}_{vk} \widehat{\mathbf{V}}_{vk}^{-1} \widetilde{\mathbf{X}}'_{vk} \right)^{-1} \sum_{k=1}^{L_v} \widetilde{\mathbf{X}}'^{\top}_{vk} \widehat{\mathbf{V}}_{vk}^{-1} \widetilde{\mathbf{Y}}'_{vk}. \tag{8}$$

In step 2 of the estimation procedure, the covariate-adjusted estimators of $\{\gamma_r\}_{r=0}^p$ are obtained as weighted averages of the raw estimates (6) and (8) from the m bins. This leads to the covariate-adjusted estimators, namely CAR-WLS and CAR-GLS:

$$\widehat{\gamma}_r^{wls} = \frac{1}{\overline{\overline{\widetilde{X}}}_r} \sum_{v=1}^{m} \frac{L_v}{n} \widehat{\beta}_{rv}^{wls} \overline{\overline{\widetilde{X}'}}_{rv}, \qquad \text{and} \qquad \widehat{\gamma}_r^{gls} = \frac{1}{\overline{\overline{\widetilde{X}}}_r} \sum_{v=1}^{m} \frac{L_v}{n} \widehat{\beta}_{rv}^{gls} \overline{\overline{\widetilde{X}'}}_{rv}, \tag{9}$$

where $\overline{\overline{\widetilde{X}}}_r = N^{-1} \sum_{i=1}^n \sum_{j=1}^{T_i} \widetilde{X}_{rij}$ (with $N = \sum_{i=1}^n T_i$) is the overall mean of the predictor variable X_r and $\overline{\overline{\widetilde{X}'}}_{rv} = N_v'^{-1} \sum_{k=1}^{L_v} \sum_{j=1}^{T'_{vk}} \widetilde{X}'_{vrkj}$ (with $N'_v = \sum_{k=1}^{L_v} T'_{vk}$) is the mean of rth predictor in bin v. (Note that $\overline{\overline{\widetilde{X}'}}_{0v} = 1$ and also $\overline{\overline{\widetilde{X}}}_0 = 1$ for the intercept term.) Also note that the quantities in (9) are weighted averages of the form $\widehat{\gamma}_r^* = \sum_{v=1}^m w_v \widehat{\beta}_{rv}^*$ (where the "*" denotes "wls" or "gls") and the weights, $\{w_v\}$, depend on the number of subjects in each bin relative to the total number of subjects (L_v/n). The binning and the weights given in (9) eliminate the impact of the distorting functions. These estimators are motivated by the relations $\mathrm{E}\{\beta_0(U)\} = \gamma_0$ and $\mathrm{E}\{\beta_r(U)\widetilde{X}_r\} = \gamma_r \mathrm{E}(X_r) = \gamma_r \mathrm{E}(\widetilde{X}_r)$.

3.2 Estimation of covariance based on distorted data

Estimation of the covariance of the unobserved errors e_{ij} must be based on the distorted errors $\{\epsilon_{ij} = \psi(U_i)e_{ij}\}$ from the varying coefficient model (4). This involves utilizing the model residuals resulting from all the bins. As before, let $\boldsymbol{\Sigma} = (\sigma_{jj'})$, $1 \le j, j' \le T$, be the $T \times T$ covariance matrix of $\mathbf{e}$. The covariance of $\boldsymbol{\epsilon} = (\epsilon_{i1}, \ldots, \epsilon_{iT})$ from model (4) is

$\text{cov}(\boldsymbol{\epsilon}) = \text{E}\{\psi^2(U)\}\boldsymbol{\Sigma} = \mathbf{V}$. Thus, an estimate of $\mathbf{V}$ using the varying coefficient model residuals can be incorporated into the estimator $\widehat{\gamma}_r^{gls}$, since $\boldsymbol{\Sigma}$ only needs to be targeted up to a constant. More precisely, the estimate of the covariance is incorporated via $\widehat{\boldsymbol{\beta}}_v^{gls} = (\sum_{k=1}^{L_v} \widetilde{\mathbf{X}}_{vk}'^{\top} \widehat{\mathbf{V}}_{vk}^{-1} \widetilde{\mathbf{X}}_{vk}')^{-1} \sum_{k=1}^{L_v} \widetilde{\mathbf{X}}_{vk}'^{\top} \widehat{\mathbf{V}}_{vk}^{-1} \widetilde{\mathbf{Y}}_{vk}'$, where $\widehat{\mathbf{V}}_{vk}$ is the estimator of $\text{cov}(\boldsymbol{\epsilon}_{vk})$.

We estimate $\mathbf{V}$ as weighted averages of separate covariance estimates from each bin, called *bin-specific covariance estimates*. To present the expression for the estimator of $\sigma_{jj'}$, some additional notation is needed. Because balanced data is not assumed, we define an indicator variable, δ, to track the specific occasions at which the measurements are available for each individual. Let $\delta_{vkj} = 1$ if $\widetilde{Y}_{vkj}'$ is available for subject k (in bin v) at occasion j and $\delta_{vkj} = 0$ otherwise ($k = 1, \ldots, L_v$; $j = 1, \ldots, T$). Denote the T_{vk}'-vector of WLS residuals for subject k in bin v by $\widetilde{\mathbf{r}}_{vk}^* = \widetilde{\mathbf{Y}}_{vk}' - \widetilde{\mathbf{X}}_{vk}' \widehat{\boldsymbol{\beta}}_v^{wls}$ and define the $T \times 1$ vector of augmented residuals $\widetilde{\mathbf{r}}_{vk} = (\widetilde{r}_{vk1}, \ldots, \widetilde{r}_{vkT})$, where $\widetilde{r}_{vkj} = \widetilde{r}_{vkj'}^* I_{\{j'=j\}}$. $I_{\{j'=j\}}$ is the indicator variable that equals 1 if $j' = j$ and is 0 otherwise. We first obtain the bin-specific estimator $\widehat{\sigma}_{vjj'}$ as the sample covariance of the WLS residuals from bin v: $\widehat{\sigma}_{vjj'} = L_{vjj'}^{-1} \sum_{k=1}^{L_v} \widetilde{r}_{vkj} \widetilde{r}_{vkj'} \delta_{vkj} \delta_{vkj'}$, where $L_{vjj'} = \sum_{k=1}^{L_v} \delta_{vkj} \delta_{vkj'}$ is the number of subjects in bin v with measurements at both occasions j and j', and $L_{vj} = L_{vjj} = \sum_{k=1}^{L_v} \delta_{vkj}$ is the number of subjects with measurements at occasion j in bin v. The estimator of $\sigma_{jj'}$ is

$$\widehat{\sigma}_{jj'} = \sum_{v=1}^{m} \frac{L_{vjj'}}{n_{jj'}} \widehat{\sigma}_{vjj'}, \qquad 1 \leq j, j' \leq n, \tag{10}$$

where $n_{jj'}$ is the overall number of subjects observed at occasions j and j', and n_{jj} is the number of subjects observed at occasion j. The quantity $\widehat{\sigma}_{jj'}$ is consistent for $\text{cov}(e_{ij}, e_{ij'})$ up to a constant of proportionality: $\widehat{\sigma}_{jj'} = \text{E}\{\psi^2(U)\}\sigma_{jj'} + O_p(r_n)$ with rate $r_n = \sqrt{(m \log n)/n}$, for the case of a complete design. This is stated formally as Lemma 1 in the Appendix. Thus, $\widehat{\sigma}_{jj'}$ can be incorporated into the CAR-GLS estimator as described in (8) and (9).

3.3 Consistency

The CAR-WLS and CAR-GLS estimators, namely $\widehat{\gamma}_r^{wls}$ and $\widehat{\gamma}_r^{gls}$, are weighted averages of the raw least squares estimators, $\widehat{\boldsymbol{\beta}}_v^{wls}$ and $\widehat{\boldsymbol{\beta}}_v^{gls}$, respectively. The raw estimators for $\{\beta_r(\cdot)\}_{r=0}^p$ must exist in each bin for the proposed estimators of $\{\gamma_r\}_{r=0}^p$ to be well defined. Therefore, it is required that $\widetilde{\mathbb{X}}_v^* = \sum_{k=1}^{L_v} \widetilde{\mathbf{X}}_{vk}'^{\top} \mathbf{W}_{vk}^{-1} \widetilde{\mathbf{X}}_{vk}'$ and $\widetilde{\mathbb{X}}_v = \sum_{k=1}^{L_v} \widetilde{\mathbf{X}}_{vk}'^{\top} \widehat{\mathbf{V}}_{vk}^{-1} \widetilde{\mathbf{X}}_{vk}'$ are nonsingular or, equivalently, that $\det(\widetilde{\mathbb{X}}_v^*)$ and $\det(\widetilde{\mathbb{X}}_v)$ are non-zero. Also, let $\mathbb{X}_v = \sum_{k=1}^{L_v} \mathbf{X}_{vk}'^{\top} \widehat{\mathbf{V}}_{vk}^{-1} \mathbf{X}_{vk}'$ (where $\mathbf{X}_{vk}'$ is the matrix of unobserved predictor values analogous to $\widetilde{\mathbf{X}}_{vk}'$), and define the sequence of events

$$
\begin{aligned}
E_1 &= \left\{\omega \in \Omega : \inf_v |\det(\widetilde{\mathbb{X}}_v^*)| > \zeta \text{ and } \min_v L_v > p\right\}, \\
E_2 &= \left\{\omega \in \Omega : \inf_v |\det(\widetilde{\mathbb{X}}_v)| > \zeta \text{ and } \min_v L_v > p\right\}, \quad \text{and} \\
E_3 &= \left\{\omega \in \Omega : \inf_v |\det(\mathbb{X}_v)| > \zeta \text{ and } \min_v L_v > p\right\},
\end{aligned}
$$

where $(\Omega, \mathcal{F}, \mathcal{P})$ is the underlying probability space. The constant term ζ is given by $\min\{(\inf_v \mathcal{D}_v)^p \rho_1/[2\{\mathrm{E}(\psi^2(U))\}^{p+1}], \rho_1/[2\{\mathrm{E}(\psi^2(U))\}^{p+1}], (\inf_v \mathcal{D}_v)^p \rho_2/2\}$, where $\mathcal{D}_v$ is the set $\{\phi_1^2(U_v'^*), \ldots, \phi_p^2(U_v'^*)\}$, $U_v'^* = L_v^{-1} \sum_{k=1}^{L_v} U_{vk}'$ is the average of the Us in bin v, and the constants ρ_1 and ρ_2 are given in (**C5**). Both $\widehat{\gamma}_r^{wls}$ and $\widehat{\gamma}_r^{gls}$ are well defined on the event $E = E_1 \cap E_2 \cap E_3$ and, in the complete data case, we have the following result.

THEOREM 1. *Under the technical conditions given in the Appendix and on event E,*

$$
\widehat{\gamma}_r^{gls} = \gamma_r + O_p(r_n), \qquad \textit{where } r_n = \sqrt{(m \log n)/n} \textit{ and } r = 0, \ldots, p.
$$

4 Studies of finite sample properties

We examine the performance of the estimators, $\widehat{\boldsymbol{\gamma}}^{wls}$ and $\widehat{\boldsymbol{\gamma}}^{gls}$, in complete and incomplete data settings and in terms of bias, variance and/or mean square error (MSE). We consider the following underlying GLM for the ith individual with T_i repeated measurements,

$\mathbf{Y}_i = \mathbf{X}_i\boldsymbol{\gamma} + \mathbf{e}_i$, where $\mathbf{X}_i$ is the $T_i \times 4$ matrix of predictor values. The jth row of $\mathbf{X}_i$ is given by $(1, X_{1ij}, X_{2ij}, X_{3ij})$ for three predictors. The regression coefficients are $\boldsymbol{\gamma}^\top = (\gamma_0, \ldots, \gamma_3) = (4, -1, 3, -0.2)$, and we take $\mathbf{e}_i \sim N_{T_i}(\mathbf{0}, \boldsymbol{\Sigma}_i)$. The number of repeated measurements on subject i is $1 \leq T_i \leq T = 6$ and the $T \times T$ covariance matrix is

$$\boldsymbol{\Sigma} = \begin{bmatrix} 17.64 & 10.92 & 6.39 & 3.11 & 1.87 & 1.56 \\ & 16.00 & 9.36 & 4.56 & 2.75 & 2.28 \\ & & 12.96 & 6.32 & 3.80 & 3.16 \\ & & & 7.29 & 4.39 & 3.65 \\ & & & & 6.25 & 5.20 \\ & & & & & 10.24 \end{bmatrix}.$$

Note that the covariance between replicate measurements decreases as the time separation increases and the variances are heterogeneous, both common features of longitudinal data.

Also, the time varying predictors, $(X_{1it}, X_{2it}, X_{3it})$, are distributed as $N_3(\boldsymbol{\mu}(j), \boldsymbol{\Delta})$, where the means depend on the measurement times (j) through $\boldsymbol{\mu}(j) = (2, 4j, 5 + j^2)$, for $j \in \{1, \ldots, 6\}$. Their dependence is modelled through the covariance matrix

$$\boldsymbol{\Delta} = \begin{bmatrix} 1.00 & 0.42 & 0.40 \\ & 2.00 & -0.28 \\ & & 4.00 \end{bmatrix}.$$

The distribution of U is $U[2, 8]$ and the distorting functions are $\psi(U) = (15 + U^3)/185$, $\phi_1(U) = (U + 1)^2/39$, $\phi_2(U) = (U + 10)/15$ and $\phi_3(U) = U/5$, where the constants $(185, 39, 15, 5)$ are chosen so that $\{\psi(\cdot), \phi_r(\cdot)\}_{r=1}^3$ satisfy the identifiability constraints.

We simulated 1000 data sets for each n = 50, 70, 100, 150, 250, 350, 500 and 800. For incomplete data, we examined two cases. The first case is non-monotone missing, which aims to model missed observations at follow-up. The second case is monotone missing, which explores missing observations due to dropouts, i.e. if Y_{ij} is missing so are

$Y_{ij+1}, \ldots, Y_{iT}$. The proportion of missing data, generated at random, range from 20% to 60%.

The simulation study suggests that the parameter estimates are fairly robust to the number of bins, m. Supplemental Table A (http://dnguyen.ucdavis.edu/.html/sup_carglm.html) illustrates the minimal effect of m on the estimation. This insensitivity to m was also observed in the case of CAR (Şentürk & Müller, 2006). We also note that for high amounts of missing and small sample size, the covariance matrix estimate may not be positive definite. We adopted a "ridge regression" modification by adding a small constant, $\lambda = 0.2$, to the diagonal elements of $\widehat{\mathbf{V}}$ when the minimum eigenvalue was not positive.

FIGURE 1 ABOUT HERE

Figure 1 displays the estimates of bias, variance, and MSE of the estimators, $\widehat{\gamma}_r^{wls}$ and $\widehat{\gamma}_r^{gls}$. Displayed are results for estimating γ_1 with complete data and with 30% missing. We report on this case in more details, since the other cases are similar. The results suggest that CAR-GLS outperforms the CAR-WLS for small to moderate n ($n \leq 250$). The favorable relative performance of CAR-GLS to CAR-WLS also holds as the missing rate increases, as displayed in Figure 2 for MSE at $n = 70$ to 500. As expected, both methods deteriorate (MSE increases) as the amount of missing data increases. Also, although CAR-GLS is superior, the difference in MSE between CAR-GLS and CAR-WLS narrows as the amount of missing data increases. This is illustrated in Figure 3 by the differences in absolute bias, $|\text{bias}(\widehat{\gamma}_r^{wls})| - |\text{bias}(\widehat{\gamma}_r^{gls})|$, for 0% to 50% (non-monotone) missing. The largest relative gain for CAR-GLS is for complete data and the gain reduces as the missing data rate increases.

FIGURE 2 ABOUT HERE

Additional simulations were conducted for various covariance structures and levels of distortion (low, medium, high) at the suggestion of a reviewer. Details of these simulations are available at http://dnguyen.ucdavis.edu/.html/sup_carglm.html. The estimation results described above hold for different covariance structures (exponential, heterogeneous compound symmetry, hybrid structure defined as exponential plus heterogeneous compound symmetry and variance components only). Results from the simulations involving different levels of distortion are also similar. As expected, the nominal levels of bias, variance and MSE are lowest for the low distortion setting.

FIGURE 3 ABOUT HERE

5 Data example

The underlying relationship between the development of early learning and behavioural conduct in children has various practical implications, such as the initiation of early intervention strategies to improve academic performance. Many studies are devoted to better understanding the impact of early childhood antisocial behaviour on learning ability (Curran, 1997; Patterson, 1986; Reid, 1993). This relationship is complex and is related to the school and home environments. Although early childhood antisocial behaviour may potentially impact learning, it is intertwined with the parental provision of emotional and cognitive support at home. These relationships are potentially confounded by the parent's readiness to provide emotional and cognitive support, reflected in the age of the parent at the start of child rearing. The data set was obtained from the National Longitudinal Study of Youth (NLSY), consisting of 405 children within the first two years of entry into primary school and followed over an eight-year period. Measurements were taken at two-year intervals between 1986 and 1992. Detailed descriptions of the data set, study design, inclusion criteria and data availability can be found in Curran (1997). We analyze a subset of 202 girls, where measurements on all 202 children were made at time 1, 179 at time 2, 131 at time 3 and 129 at time 4. Complete data were available for 105 children.

We illustrate the proposed method by considering the regression model between reading recognition skill ($\widetilde{Y}$) on antisocial behaviour score ($\widetilde{X}_1$), level of emotional support ($\widetilde{X}_2$) and level of cognitive stimulation ($\widetilde{X}_3$), adjusted for the mother's age ($\mathrm{MA} = U$) at the first visit. The time varying predictor $\widetilde{X}_1$ is measured at the four occasions along with the response $\widetilde{Y}$. The remaining two predictors are cross-sectional (evaluated on the first visit). A regression of $\widetilde{Y}$ on $(\widetilde{X}_1, \widetilde{X}_2, \widetilde{X}_3)$ leads to the varying coefficient model $Y_{ij} = \beta_0(\mathrm{MA}_i) + \beta_1(\mathrm{MA}_i)\widetilde{X}_{1ij} + \beta_2(\mathrm{MA}_i)\widetilde{X}_{2i} + \beta_3(\mathrm{MA}_i)\widetilde{X}_{3i} + \epsilon_{ij}$. Table 1 gives the estimates of the underlying regression parameters $\{\gamma_r\}_{r=0}^4$ from an unadjusted (standard) analysis using REML estimation, which incorporates the estimate of the unstructured covariance matrix. Also provided in Table 1 are the estimates from the adjusted analysis using CAR-GLS ($\widehat{\gamma}_r^{gls}$), as described in Section 3.2. The coefficient estimate for antisocial behaviour is negative, suggesting that an increased antisocial behaviour is associated with a reduction in learning, although this is not significant in both the unadjusted and adjusted analysis. Emotional support at home is also not significant, although the point estimate for the unadjusted analysis is higher. However, cognitive stimulation, without adjusting for the mother's age, is significantly related to the child's reading skill. This is in contrast to the CAR-GLS analysis, which suggests that the effect of cognitive stimulation on reading skill is explained by the mother's readiness to provide cognitive support; thus, it becomes insignificant after adjusting for the mother's age. We note that the analysis using CAR-WLS and the standard analysis (GLM-REML) method with the mother's age included as an additional predictor in the regression both lead to the same conclusion as the CAR-GLS method for this data (results not provided). Including U in the regression modelling as an additional predictor has been shown to adjust for only additive linear effects of U. Hence, it is a more restrictive adjustment method than CAR-GLS. See Şentürk & Müller (2005) for a comparative analysis.

TABLE 1 ABOUT HERE

We note that although the main goal of the proposed method is to uncover the partial relationship between $\widetilde{Y}$ and $\widetilde{X}_r$ adjusted for U, it can also be used to graphically examine the effects of U. The distortion on Y, i.e. the form of $\psi(U)$, is given by the y-intercept of the varying coefficient model, $\beta_0(U) = \gamma_0\psi(U)$. Distortion functions of the predictors, $\{\phi_r(U)\}_{r=1}^p$, can be targeted similarly using the estimates of the varying coefficient functions with $\{\widehat{\gamma}_r\widehat{\beta}_0(U)\}/\{\widehat{\beta}_r(U)\widehat{\gamma}_0\}$, since $\beta_r(U) = \gamma_r\psi(U)/\phi_r(U)$. Note that the distortion function "estimates" will be identifiable if γ_0 and γ_r are different from zero. For the data example, since γ_1, γ_2 and γ_3 are not significant, we cannot estimate $\phi_r(U)$ explicitly. However, the slight curve in the estimate of $\beta_0(U)$, plotted in Figure 4, suggests that, as the mother's age increases, the response is amplified.

FIGURE 4 ABOUT HERE

The confidence intervals for the CAR-GLS method described above are bootstrap percentile confidence intervals, given by the $100(\alpha/2)$th and the $100(1-\alpha/2)$th percentiles of B bootstrap estimates. We generated $B = 1000$ bootstrap samples from the original data to obtain the interval estimates for CAR-GLS. The estimated nonparametric densities of the standardized 1000 bootstrap estimates of $\{\gamma_r\}_{r=0}^3$ are close to the standard normal density. We also examined the coverage levels of bootstrap intervals used above via simulation. Under the simulation model, we generated 1000 bootstrap samples for each $n = 50$, 70, 100, 150 and 250. The estimated coverage values approach the targeted levels of 80%, 90%, and 95%. For example, in the complete data case, the average observed coverage levels are within about 1.34% of the targeted levels. For data with 20%-30% missing, the average coverage levels are about 1.4% to 1.85% of the targeted levels.

6 Discussion

An advantage of GLM for continuous longitudinal data is the ease of interpretation. However, for distorted data, adjustment for the confounding on the response and predictors is needed for proper interpretation. We have provided a covariate-adjusted estimation procedure for longitudinal data that incorporates the covariance structure to improve the performance of the estimators. The proposed estimation method has several appealing flexibilities. For instance, the estimators were derived without distributional assumptions (e.g. normality), and the proposed estimation procedure is flexible in handling different forms of distorted longitudinal data. Also, the implementation is not difficult, since routinely available least squares fitting tools can be applied to the data within each bin. The simulation studies indicate that the estimation algorithm is effective under missing data.

We note that the binning algorithm used corresponds to a local constant approximation of the varying coefficient functions. One difference with binning compared to a local constant approximation via kernel functions is that, by considering non-overlapping windows, substantially less computation is involved. Nevertheless, it still incorporates the full information in the data. We would expect the binning algorithm to perform similarly to a local constant approximation via kernel functions, where improvements in both methods can be expected in terms of bias and variance (for small to moderate sample sizes) if a local linear approximation is used.

Note that CAR-GLS also has features reminiscent of path analysis or structural equation modelling (SEM), which are popular methods in analyzing social science data. SEM is based on multiple regression relations where latent variables are also utilized similar to CAR-GLS. While there are similarities, CAR-GLS has different flexibilities for which

SEM was not designed. For example, CAR-GLS models the effects of U nonparametrically as being multiplicative or additive, whereas in SEM the effects are modelled as additive and linear. However, SEM allows for a set of regression relations, a feature for which CAR-GLS is not designed. Generalization of the proposed method from a single underlying regression model to a set of regression relations, as in SEM, is an interesting open problem.

Finally, we note that the estimated coefficients in the data example of Section 5 could change if other covariates, such as the mother's education level, were added to the model. This issue, which can be viewed as related to model mis-specification, is important to the regression modelling process generally and is, therefore, applicable to covariate-adjusted modelling as well. For the data analyzed, other important predictor variables, such as the mother's level of education, were not available. However, conceptually, such relevant variables should be investigated in covariate-adjusted modelling in the same way as in traditional regression modelling.

Appendix

Proofs

The following definitions are used in the proofs. 1. The $(p+1) \times 1$ vector of ones is denoted by $\mathbf{1}$ and the $(p+1) \times (p+1)$ matrix with all entries equal to one is denoted by $\mathbf{J}$. 2. $U_v^M = a + (2v-1)\{(b-a)/(2m)\}$ is the midpoint of bin B_v. 3. $\mathbf{\Delta}_{v1} = (\psi(U_v'^*), \psi(U_v'^*)\phi_1(U_v'^*), \ldots, \psi(U_v'^*)\phi_p(U_v'^*))$. 4. $\mathbf{\Delta}_{v2} = (\psi(U_v'^*), \psi(U_v'^*)/\phi_1(U_v'^*), \ldots, \psi(U_v'^*)/\phi_p(U_v'^*))^\top$. 5. $\mathbf{A} \boxdot \mathbf{B}$ denotes the Hadamard product of two matrices, whose (i,j)th element is equal to the product of the (i,j)th elements of the matrices $\mathbf{A}$ and $\mathbf{B}$.

The following technical conditions are assumed.

(C1) The covariate U is bounded, $-\infty < a \leq U \leq b < \infty$, for real numbers $a < b$. The density $f(u)$ of U satisfies $\inf_{a \leq u \leq b} f(u) > c_1 > 0$, $\sup_{a \leq u \leq b} f(u) < c_2 < \infty$ for real numbers c_1 and c_2, and is uniformly Lipschitz continuous.

(C2) The variables (e_{ij}, U_i, X_{rij}) are mutually independent for $r = 1, \dots, p$, $i = 1, \dots, n$, and $j = 1, \dots, T$.

(C3) For some bound $B \in \mathbb{R}$, $\sup_{1\le i\le n, 1\le r\le p, 1\le j\le T} |X_{rij}| \le B$ and $\mathrm{E}(X_r) \ne 0$.

(C4) The contamination functions, $\psi(\cdot)$ and $\{\phi_r(\cdot)\}_{r=1}^p$, are twice continuously differentiable, satisfying $\mathrm{E}\{\psi(U)\} = 1$, $\mathrm{E}\{\phi_r(U)\} = 1$, $\phi_r(\cdot) > 0$, and $\mathrm{E}\{\psi^2(U)\} > 0$.

(C5) As $n \to \infty$, $(nT^2)^{-1}\sum_{i=1}^n \mathbf{X}_i^\top \mathbf{\Sigma}^{-1}\mathbf{X}_i \to \mathbb{C}_1$ and $(nT^2)^{-1}\sum_{i=1}^n \mathbf{X}_i^\top \mathbf{X}_i \to \mathbb{C}_2$ in probability, where $\rho_1 = |\det(\mathbb{C}_1)| > 0$ and $\rho_2 = |\det(\mathbb{C}_2)| > 0$. Also, the covariance matrix $\mathbf{\Sigma} = \mathrm{cov}(\mathbf{e})$ is positive definite.

(C6) The functions $h_1(u) = \int x g_1(x, u)dx$ and $h_2(u) = \int x g_2(x, u)dx$ are uniformly Lipschitz, where $g_1(\cdot,\cdot)$ and $g_2(\cdot,\cdot)$ are the joint density functions of (X_{rj}, U) and $(X_{rj}e_j, U)$ for $1 \le j \le T$ and $1 \le r \le p$, respectively.

(C7) $\mathrm{E}(e_j^\lambda) < \infty$ for $\lambda > 4$ and $1 \le j \le T$.

The above conditions are similar to those in Şentürk & Müller (2006), and more detailed explanations are given there. We note that (**C6**) and (**C7**) are needed for the uniform consistency of certain functional forms over all the bins, leading to results in Lemma 3.

For the asymptotic results below, we assume a balanced design so that $\widehat{\mathbf{V}}_{vk} = \widehat{\mathbf{V}}$. The following boundedness considerations are used in the following proofs. As $\widetilde{X}_{rij}$ is bounded, since $X_{rij} = O(1)$, U has compact support, $\phi_r(\cdot)$ is continuous for $1 \le r \le p$, and $\tilde{X}'_{rvkj}$ is also bounded. Defining $\lambda_{0vk} = \psi(U'_{vk}) - \psi(U'^*_v)$, $\lambda_{rvk} = \phi_r(U'_{vk}) - \phi_r(U'^*_v)$ and $\lambda'_{rvk} = \psi(U'_{vk})/\phi_r(U'_{vk}) - \psi(U'^*_v)/\phi_r(U'^*_v)$ for $1 \le k \le L_v$ and $1 \le r \le p$, we obtain the following boundedness results for $1 \le r \le p$, using Taylor expansions: $\sup_{v,k} |U'_{vk} - U'^*_v| \le (b-a)/m$; $\sup_{v,k} |\lambda_{0vk}| = O(m^{-1})$; $\sup_{v,k} |\lambda_{rvk}| = O(m^{-1})$; $\sup_{v,k} |\lambda'_{rvk}| = O(m^{-1})$. Also, the following Lemmas will be used in the proof of Theorem 1.

LEMMA 1. *Under conditions (C1)-(C6) for complete data on event E_1,*

$$\widehat{\sigma}_{jj'} = E\{\psi^2(U)\}\sigma_{jj'} + O_p(r_n), \quad \textit{for} \;\; 1 \le j, j' \le n, \quad \textit{where} \;\; r_n = \sqrt{(m \log n)/n}.$$

Proof of Lemma 1. It has been shown in Şentürk & Nguyen (2006) that

$$\widehat{\beta}_{0v}^{wls} = \psi(U_v'^*)\gamma_0 + O_p(r_n) \quad \text{and} \qquad \widehat{\beta}_{rv}^{wls} = \frac{\psi(U_v'^*)}{\phi_r(U_v'^*)}\gamma_r + O_p(r_n), \;\; \text{for} \;\; r = 1, \ldots, p.$$

Consequently, substituting the above expansions into the residual $\widetilde{r}_{vkj}$ gives:

$$\begin{aligned} \widetilde{r}_{vkj} &= \widetilde{Y}_{vkj}' - \widehat{\beta}_{0v}^{wls} - \sum_{r=1}^{p} \widehat{\beta}_{rv}^{wls} \widetilde{X}_{rvkj}' = \psi(U_v'^*)\left\{ Y_{vkj}' - \gamma_0 - \sum_{r=1}^{p} \gamma_r X_{rvkj}' \right\} + O_p(r_n) \\ &= \psi(U_v'^*)e_{vkj}' + O_p(r_n), \end{aligned}$$

where Y_{vkj}', X_{rvkj}', and e_{vkj}' are the response, predictor and error variables for subject k at time j in bin v. Thus, the the covariance estimator given in (10) is

$$\begin{aligned} \widehat{\sigma}_{jj'} &= n^{-1} \sum_{v=1}^{m} \sum_{k=1}^{L_v} \psi^2(U_v'^*)e_{vkj}'e_{vkj'}' + O_p(r_n) = n^{-1} \sum_{i=1}^{n} \psi^2(U_i)e_{ij}e_{ij'} + O_p(r_n) \\ &= \mathrm{E}\{\psi^2(U)\}\sigma_{jj'} + O_p(r_n), \end{aligned}$$

which follows from the Law of Large Numbers.

LEMMA 2. *Under conditions (C1-C5) for complete data on event E,* $\sup_v |\widetilde{\mathbb{X}}_v^{-1} - \mathbf{\Phi}_v \boxdot \mathbb{X}_v^{-1}| = O_p(m^{-1})\mathbf{J}$, *where*

$$\mathbf{\Phi}_v = \begin{bmatrix} 1 & 1/\phi_1(U_v'^*) & \cdots & 1/\phi_p(U_v'^*) \\ 1/\phi_1(U_v'^*) & 1/\phi_1^2(U_v'^*) & \cdots & 1/\{\phi_1(U_v'^*)\phi_p(U_v'^*)\} \\ \vdots & \vdots & \ddots & \vdots \\ 1/\{\phi_1(U_v'^*)\phi_p(U_v'^*)\} & 1/\{\phi_2(U_v'^*)\phi_p(U_v'^*)\} & \cdots & 1/\phi_p^2(U_v'^*) \end{bmatrix}.$$

Proof of Lemma 2. The proof follows similarly to Lemma 3 of Şentürk & Müller (2006), noting that the elements of $\mathbf{V}^{-1}$ are $O_p(1)$ given that $\mathbf{V}^{-1}$ exists.

LEMMA 3. *Under conditions (C1)-(C7) specified above and with complete data, on event E:*

(a) $\sup_v |(L_v T^2)^{-1}\mathbb{X}_v^{-1} - E\{\psi^2(U)\}\mathcal{S}^{-1}| = O_p(r_n)\mathbf{J}$,
where $\mathcal{S}$ is a $(p+1)\times(p+1)$ matrix with the (r,r')th entry $T^{-2}\sum_{j=1}^{T}\sum_{j'=1}^{T}\mathbf{\Sigma}_{jj'}^{-1}E(X_{rj}X_{r'j'})$, $0 \le r, r' \le p$, and $1 \le j, j', \le T$, and

(b) $\sup_v |(L_v T^2)^{-1}\sum_{k=1}^{L_v}\mathbf{X}_{vk}'^{\top}\widehat{\mathbf{V}}^{-1}\mathbf{e}_{vk}'| = O_p(r_n)\mathbf{1}$.

Proof of Lemma 3. Let $m_{rr'jj'}(U_v^M) = \mathrm{E}(X_{rj}X_{r'j'})$ and $\widehat{m}_{rr'jj'}(U_v^M) = L_v^{-1}\sum_{k=1}^{L_v} X'_{rvkj}X'_{r'vkj'}$. The (r,r')th entry of the $(p+1)\times(p+1)$ matrix $(L_v T^2)^{-1}\mathbb{X}_v$ is $T^{-2}\sum_{j=1}^{T}\sum_{j'=1}^{T}\widehat{\mathbf{V}}_{jj'}^{-1}$ $\widehat{m}_{rr'jj'}(U_v^M)$, for $0 \le r, r' \le p$. Given $\widehat{\mathbf{V}}^{-1}$ exists, it follows from Lemma 1 that $\widehat{\mathbf{V}}_{jj'}^{-1} = \mathbf{\Sigma}_{jj'}^{-1}[\mathrm{E}\{\psi^2(U)\}]^{-1} + O_p(r_n)$. Thus, $T^{-2}\sum_{j=1}^{T}\sum_{j'=1}^{T}\widehat{\mathbf{V}}_{jj'}^{-1}\widehat{m}_{rr'jj'}(U_v^M) = T^{-2}\sum_{j=1}^{T}\sum_{j'=1}^{T}$ $[\mathbf{\Sigma}_{jj'}^{-1}\widehat{m}_{rr'jj'}(U_v^M)/\mathrm{E}\{\psi^2(U)\}] + O_p(r_n)$. Note that $\widehat{m}_{rr'jj'}(U_v^M)$ is a Nadaraya-Watson kernel estimator (Nadaraya, 1964; Watson, 1964) with a uniform kernel $K(\cdot) = (1/2)\mathbf{1}_{[-1,1]}$ and $h = (b-a)/m$. From the uniform consistency of Nadaraya-Watson estimators with kernels of compact support, we have that $\sup_{a\le u\le b}|\hat{m}_{rr'jj'}(u) - m_{rr'jj'}(u)| = O_p(\sqrt{r_n})$. This implies

$$\sup_j |\hat{m}_{rr'jj'}(U_{nj}^M) - m_{rr'jj'}(U_{nj}^M)| = O_p(\sqrt{r_n}), \tag{11}$$

and

$$\sup_v \left| T^{-2}[1/\mathrm{E}\{\psi^2(U)\}]\left\{\sum_{j=1}^{T}\sum_{j'=1}^{T}\mathbf{\Sigma}_{jj'}^{-1}\widehat{m}_{rr'jj'}(U_v^M) - \sum_{j=1}^{T}\sum_{j'=1}^{T}\mathbf{\Sigma}_{jj'}^{-1}m_{rr'jj'}(U_v^M)\right\}\right| = O_p(r_n).$$

Thus,

$$\sup_v \left| T^{-2}\sum_{j=1}^{T}\sum_{j'=1}^{T}\widehat{\mathbf{V}}_{jj'}^{-1}\widehat{m}_{rr'jj'}(U_v^M) - T^{-2}\sum_{j=1}^{T}\sum_{j'=1}^{T}\mathbf{\Sigma}_{jj'}^{-1}(1/\mathrm{E}\{\psi^2(U)\})\mathrm{E}(X_{rj}X_{r'j'})\right| = O_p(r_n)$$

and Lemma 3(a) follows. For part (b) of the Lemma, consider the rth entry of the $(p+1)\times$

1 vector $(L_vT^2)^{-1}\sum_{k=1}^{L_v}\mathbf{X}_{vk}'^{\top}\widehat{\mathbf{V}}^{-1}\mathbf{e}_{vk}'$, namely $T^{-2}\sum_{j=1}^{T}\sum_{j'=1}^{T}\widehat{\mathbf{V}}_{jj'}^{-1}L_v^{-1}\sum_{k=1}^{L_v}X_{rvkj}'e_{r'vkj'}'$. With the same reasoning used for part (a) with the uniform consistency of the Nadaraya-Watson kernel estimators, it follows that $\sup_v|T^{-2}\sum_{j=1}^{T}\sum_{j'=1}^{T}\widehat{\mathbf{V}}_{jj'}^{-1}L_v^{-1}\sum_{k=1}^{L_v}X_{rvkj}'e_{r'vkj'}'-T^{-2}\sum_{j=1}^{T}\sum_{j'=1}^{T}\mathbf{\Sigma}_{jj'}^{-1}\mathrm{E}(X_{rj}e_{j'}|U)|=O_p(r_n)$, where $\mathrm{E}(X_{rj}e_j|U)=0$, and part (b) follows.

Proof of Theorem 1. By Lemma 1, Lemma 3 and the boundedness properties, we have

$$\sup_v\left|\frac{1}{L_vT^2}\left(\sum_{k=1}^{L_v}\widetilde{\mathbf{X}}_{vk}'^{\top}\widehat{\mathbf{V}}^{-1}\widetilde{\mathbf{Y}}_{vk}'-\mathbf{\Delta}_{v1}\boxdot\sum_{k=1}^{L_v}\mathbf{X}_{vk}'^{\top}\widehat{\mathbf{V}}^{-1}\mathbf{Y}_{vk}'\right)\right|=O_p(r_n)\mathbf{1},\tag{12}$$

where

$$\begin{aligned}\sum_{k=1}^{L_v}\widetilde{\mathbf{X}}_{vk}'^{\top}\widehat{\mathbf{V}}^{-1}\widetilde{\mathbf{Y}}_{vk}' &= \left(\sum_j\sum_{j'}\widehat{\mathbf{V}}_{jj'}^{-1}\sum_k\tilde{Y}_{vkj}',\ldots,\sum_j\sum_{j'}\widehat{\mathbf{V}}_{jj'}^{-1}\sum_k\tilde{X}_{pvkj'}'\tilde{Y}_{vkj}'\right)\\ \sum_{k=1}^{L_v}\mathbf{X}_{vk}'^{\top}\widehat{\mathbf{V}}^{-1}\mathbf{Y}_{vk}' &= \left(\sum_j\sum_{j'}\widehat{\mathbf{V}}_{jj'}^{-1}\sum_k Y_{vkj}',\ldots,\sum_j\sum_{j'}\widehat{\mathbf{V}}_{jj'}^{-1}\sum_k X_{pvkj'}'Y_{vkj}'\right).\end{aligned}$$

Let $\widetilde{\boldsymbol{\gamma}}_v=(\sum_{k=1}^{L_v}\mathbf{X}_{vk}'^{\top}\mathbf{V}^{-1}\mathbf{X}_{vk}')^{-1}\sum_{k=1}^{L_v}\mathbf{X}_{vk}'^{\top}\mathbf{V}^{-1}\mathbf{Y}_{vk}'$. Under event E, Lemma 2 together with (12) imply that $\sup_v\left|\widehat{\boldsymbol{\beta}}_v^{gls}-\mathbf{\Delta}_{v2}\boxdot\widetilde{\boldsymbol{\gamma}}_v\right|=O_p(r_n)\mathbf{1}$. Note that $\widetilde{\boldsymbol{\gamma}}_v=\gamma+(\sum_{k=1}^{L_v}\mathbf{X}_{vk}'^{\top}\mathbf{V}^{-1}\mathbf{X}_{vk}')^{-1}\sum_{k=1}^{L_v}\mathbf{X}_{vk}'^{\top}\mathbf{V}^{-1}\mathbf{e}_{vk}'$, where $\mathbf{e}_{vk}'=(e_{vk1}',\ldots,e_{vkT}')$ is the error vector for individual k in bin v. Using Lemma 2, it holds that $\sup_v\left|\widehat{\boldsymbol{\beta}}_v^{gls}-\mathbf{\Delta}_{v2}\boxdot\boldsymbol{\gamma}\right|=O_p(r_n)\mathbf{1}$. Thus, we have that the intercept estimator is

$$\widehat{\gamma}_0^{gls}=\sum_{v=1}^{m}\frac{L_v}{n}\widehat{\beta}_{0v}=\sum_{v=1}^{m}\frac{L_v}{n}\psi(U_v'^*)\gamma_0+O_p(r_n)=\frac{\gamma_0}{n}\sum_{i=1}^{n}\psi(U_i)+O_p(r_n)=\gamma_0+O_p(r_n).$$

Similarly, the CAR-GLS estimator for γ_r is

$$\widehat{\gamma}_r^{gls}=\frac{1}{\overline{\overline{\widetilde{X}}}_r}\sum_{v=1}^{m}\frac{L_v}{n}\widehat{\beta}_{rv}^{gls}\overline{\widetilde{X}'}_{rv}=\frac{\gamma_r}{\overline{\overline{\widetilde{X}}}_r}\sum_{v=1}^{m}\frac{L_v}{n}\frac{\psi(U_v'^*)}{\phi_r(U_v'^*)}\phi_r(U_v'^*)\frac{1}{T}\sum_{j=1}^{T}\frac{1}{L_v}\sum_{k=1}^{L_v}X_{rvkj}'+O_p(r_n)$$

Note that, since $\overline{\overline{\widetilde{X}}}_r=T^{-1}\sum_{j=1}^{T}\mathrm{E}(X_{rj})+O_p(n^{-1/2})$ and $\sup_v|T^{-1}\sum_{j=1}^{T}L_v^{-1}\sum_{k=1}^{L_v}X_{rvkj}'-$

$T^{-1}\sum_{j=1}^{T}\mathrm{E}(X_{rj})| = O_p(r_n)$, by (11) in the proof of Lemma 3 below, we have

$$\widehat{\gamma}_r^{gls} = \gamma_r \sum_{v=1}^{m} \frac{L_v}{n}\psi(U_v'^*) + O_p(r_n) = \frac{\gamma_r}{n}\sum_{i=1}^{n}\psi(U_i) + O_p(r_n) = \gamma_r + O_p(r_n).$$

References

Cleveland, W.S., Grosse, E. & Shyu, W.M. (1991). Local regression models. In *Statistical Models in S*, eds J.M. Chambers & T.J. Hastie, pp. 309–376. Pacific Grove, CA: Wadsworth & Brooks.

Curran, P.J. (1997). Comparing three modern approaches to longitudinal data analysis: An examination of a single developmental sample. Supporting documentation for the 1997 meeting of the Society for Research on Child Development, Washington, D.C. http://www.unc.edu/ curran/srcd-docs/srcdmeth.pdf

Diggle, P.J. & Verbyla, A.P. (1998). Nonparametric estimation of covariance structure in longitudinal data. *Biometrics* **54**, 401–415.

Hastie, T. & Tibshirani, R. (1993). Varying coefficient models. *J. Roy. Statist. Soc. Ser. B* **55**, 757–96.

Nadaraya, E.A. (1964). On estimating regression. *Theory Probab. Appl.* **9**, 141-142.

Patterson, G.R. (1986). Performance models for antisocial boys. *American Psychologist* **41**, 432–444.

Reid, J.B. (1993). Prevention of conduct disorder before and after school entry: Relating interventions to developmental findings. *Development and Psychopathology* **5**, 243–262.

Searle, S.R., Casella, G. & McCulloch, C.E. (1992). *Variance Components.* New York: John Wiley.

Şentürk, D. & Müller, H.G. (2005). Covariate adjusted regression. *Biometrika* **92**, 59–74.

Şentürk, D. & Müller, H.G. (2006). Inference for covariate adjusted regression via varying coefficient models. *Ann. Statist.* **34**, 654–679.

Şentürk, D. & Nguyen, D.V. (2006). Covariate-adjusted general linear models for longitudinal data. Technical report, Department of Statistics, Pennsylvania State University, University Park, PA, USA.

Watson, G.S. (1964). Smooth regression analysis. *Sankhyā Ser. A* **26** 101–116.

FIGURE LEGENDS

1. **Bias, variance and MSE.** Plots of absolute value of the estimated bias, the variance, and the mean square error (MSE) for the CAR-WLS and CAR-GLS estimators, as a function of sample size $n/10$. Results are displayed for (top) complete/balanced data, (middle) non-monotone missing/incomplete data with 30% missing, and (bottom) monotone missing/incomplete data with 30% missing.

2. **Effect of missing rate on MSE.** Plots of MSEs for CAR-WLS and CAR-GLS estimation as the percentage of missing data increases from 20% to 60%. Displayed are MSE results for estimating γ_1 at sample sizes $n = 70$, 150, 250, and 500.

3. **Effect of missing rate on bias difference.** Plots of the difference in absolute bias between CAR-WLS and CAR-GLS estimation of γ_3, $|\text{Bias}(\widehat{\gamma}_3^{wls})| - |\text{Bias}(\widehat{\gamma}_3^{gls})|$, versus sample size n for varying amount of missing data: 0% (complete data), 20%, 30%, 40%, and 50% missing.

4. **Estimate of $\beta_0(\cdot)$.** Plot of the estimated varying coefficient function $\widehat{\beta}_0(age)$ in the CAR-GLS model for the data analysis.

Table 1: Estimated relationship between reading cognition skills and the covariates antisocial behaviour, cognitive stimulation and emotional support. (A) Underlying regression parameter estimates from REML estimation for a general unstructured covariance matrix. (B) CAR-GLS estimation, adjusted for the mother's age at the first visit (U = MA) in the estimation of both the regression parameters and the covariance structure between repeated measurements. The corresponding 95% confidence interval estimates for the CAR-GLS are based on 1000 bootstrap samples.

	(A) GLM-REML: $\widehat{\gamma}_{\text{unadj.}}$		(B) CAR-GLS: $\widehat{\gamma}^{gls}$	
Coefficient	Point estimate	95% Interval estimate	Point estimate	95% Interval estimate
Intercept	2.4883	(1.9421, 3.0345)	3.6806	(2.8707, 4.4046)
Antisoc.	−0.0260	(−0.0668, 0.0148)	−0.0348	(−0.0760, 0.0156)
Cogn.	0.0604	(0.0118, 0.1091)	0.0563	(−0.0287, 0.1401)
Emot.	0.0284	(−0.0239, 0.0808)	0.0176	(−0.0515, 0.0979)

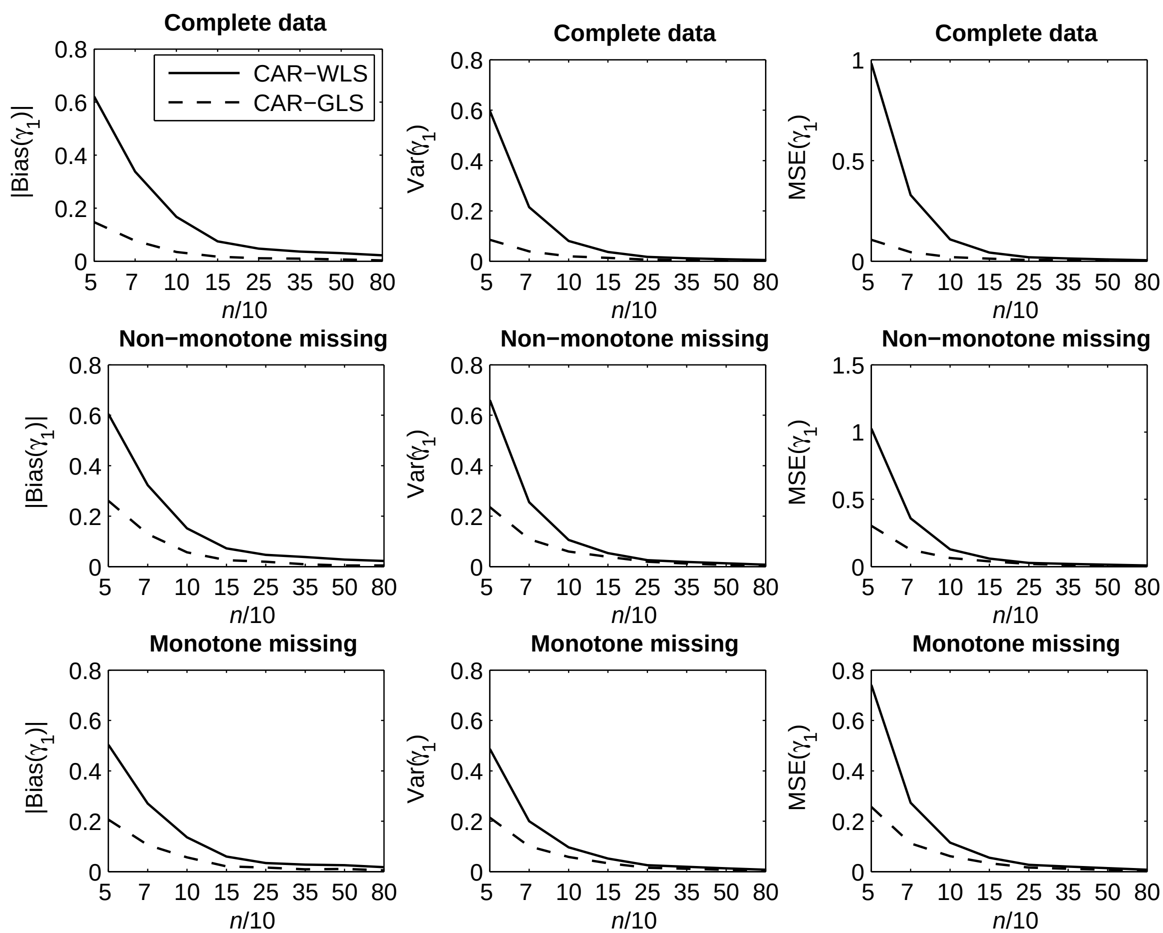

Figure 1: **Bias, variance and MSE.**

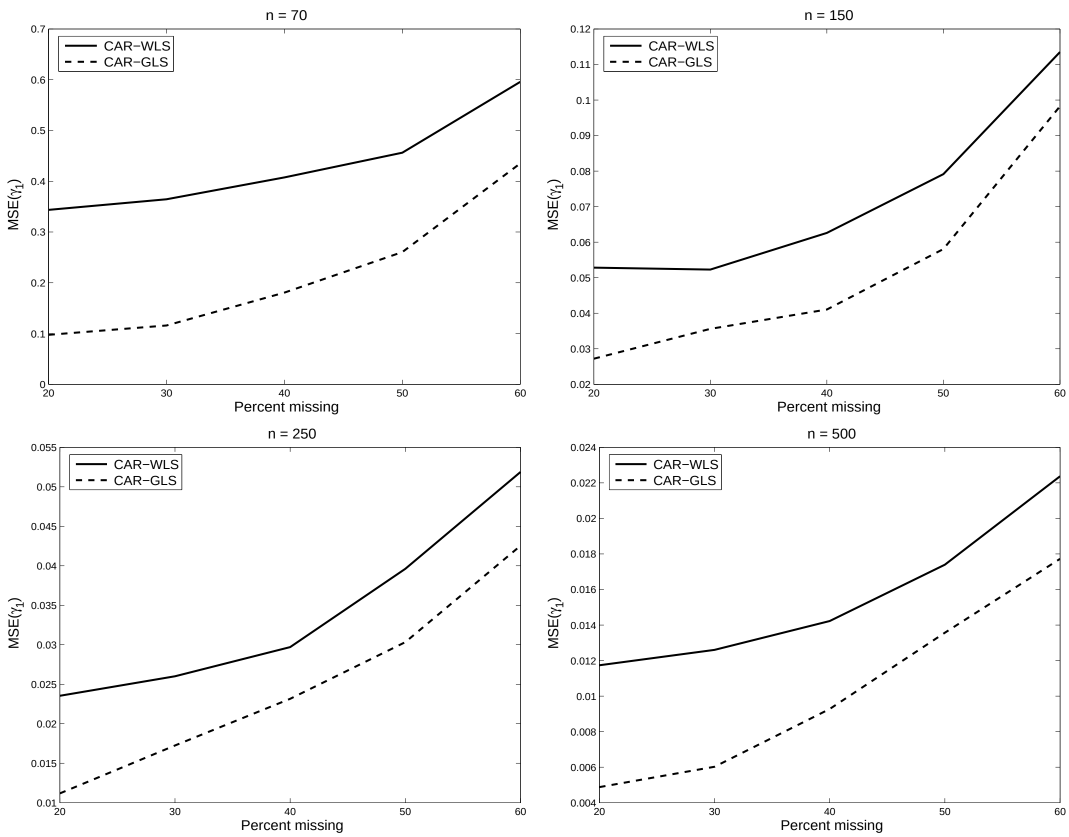

Figure 2: **Missing rate on MSE.**

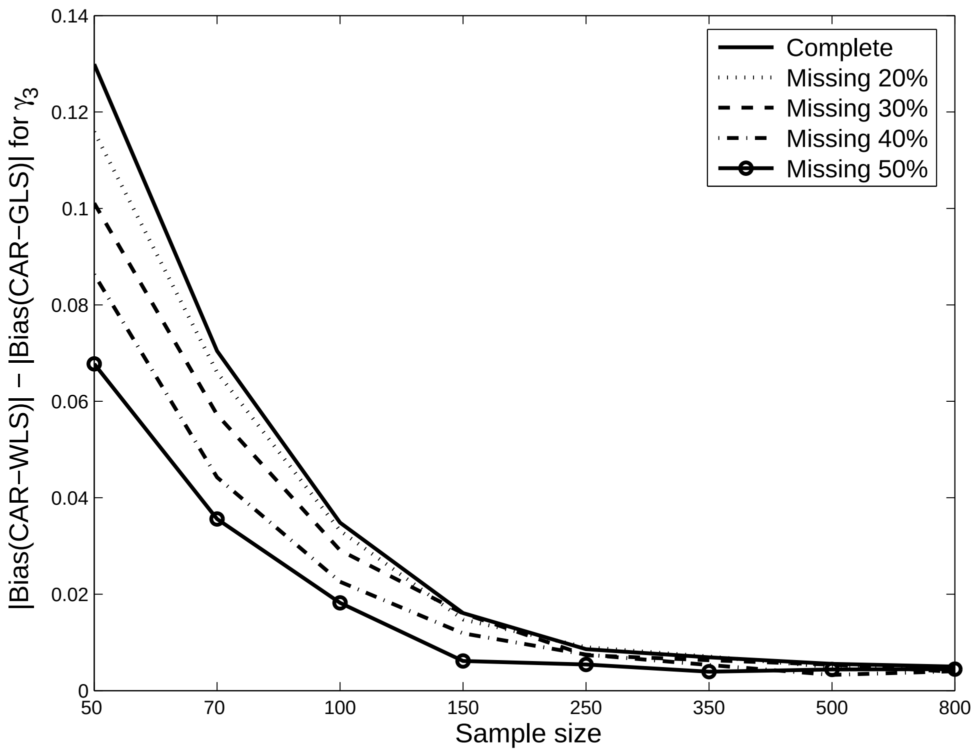

Figure 3: **Missing rate on bias difference.**

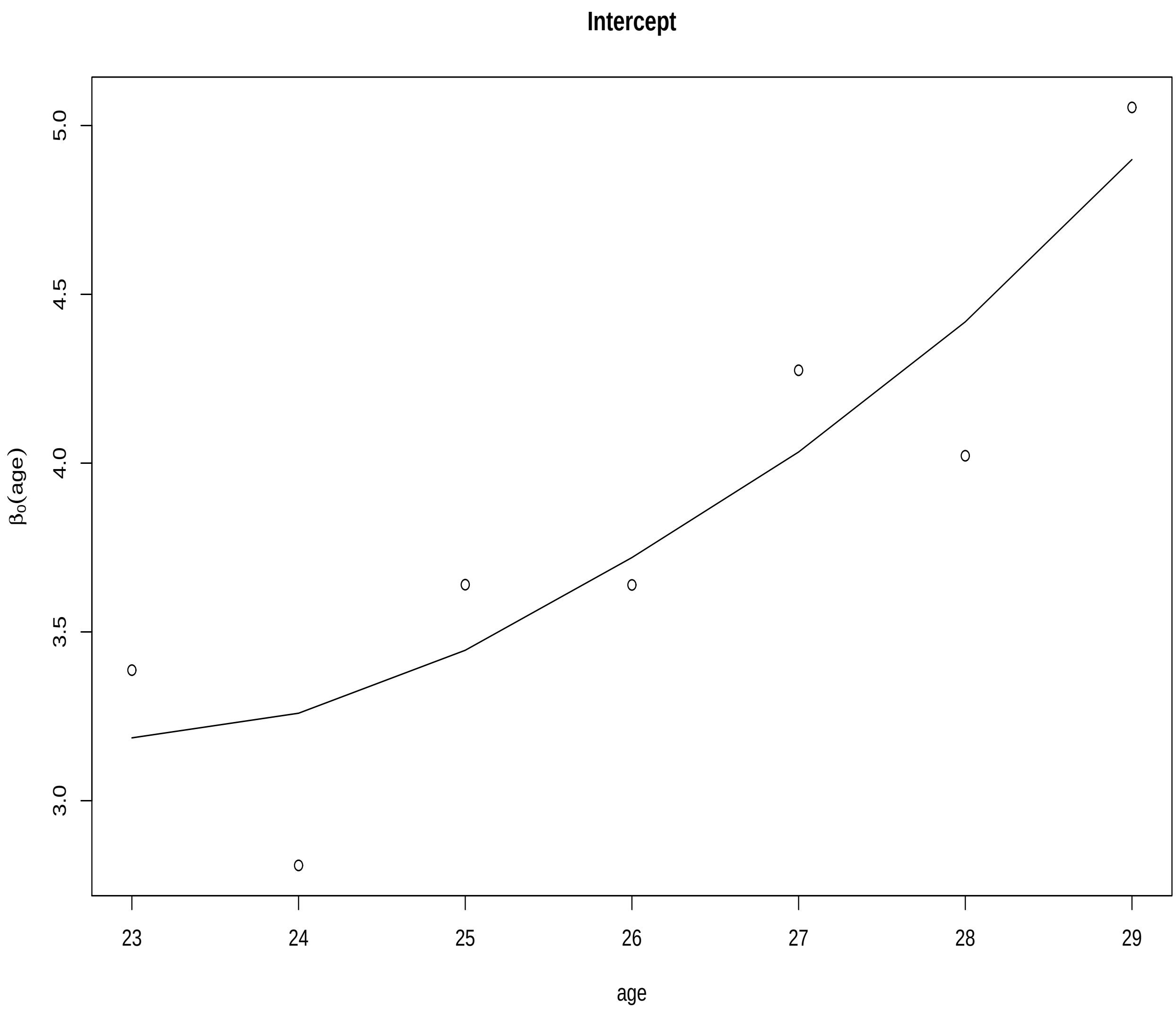

Figure 4: **Estimate of $\beta_0(\cdot)$.**