Asymptotic properties of covariate-adjusted regression with correlated errors

Damla Şentürk

Department of Statistics, Pennsylvania State University,

University Park, PA 16802, USA

Danh V. Nguyen [1]

Division of Biostatistics, University of California School of Medicine,

Davis, CA 95616, USA

December 28, 2008

Abstract

In covariate-adjusted regression (CAR), the response (Y) and predictors ($X_r, r = 1, \dots, p$) are not observed directly. Estimation is based on n independent observations $\{\tilde{Y}_i, \tilde{X}_{ri}, U_i\}_{i=1}^n$, where $\tilde{Y}_i = \psi(U_i)Y_i$, $\tilde{X}_{ri} = \phi_r(U_i)X_{ri}$ and $\psi(\cdot)$ and $\{\phi_r(\cdot)\}_{r=1}^p$ are unknown functions. In this paper, we discuss the asymptotic properties of this method when the observations are correlated, as in regression models for repeated measurements.

Keywords: General linear model; Longitudinal data; Repeated measures; Multiplicative effects; Varying-coefficient model.

[1] Corresponding author. *Email addresses:* dsenturk@stat.psu.edu (D. Şentürk), ucdnguyen@ucdavis.edu (D.V. Nguyen)

1 Introduction

Şentürk and Müller (2005) proposed the covariate-adjusted regression (CAR) method for the latent multiple linear regression model (*M1*) $Y_i = \gamma_0 + \sum_{r=1}^{p} \gamma_r X_{ri} + e_i$ for n independent observations $\{X_{ri}, Y_i\}_{i=1}^n$, $r = 1, \dots, p$, where γ_r is the slope parameter of the predictor X_r, Y_i is the response and e_i is the error for subject i. Both the response and predictors of interest in (*M1*) are not observed directly but rather they are observed only after multiplicative distortions by an observable covariate U_i: (*M2*) $\tilde{Y}_i = \psi(U_i)Y_i$ and $\tilde{X}_{ri} = \phi_r(U_i)X_{ri}$, where the distorting functions $\psi(\cdot)$ and $\phi_r(\cdot)$ are *unknown/unspecified*. The CAR model (*M1*)-(*M2*) was motivated by the interest to examine the relationship between the response fibrinogen level (Y_i), a protein found in blood plasma, and the predictor serum transferrin level (X_i) in haemodialysis patients (Kaysen et al., 2003, Şentürk and Müller, 2005). However, the corresponding observed protein measurements ($\tilde{Y}_i$, $\tilde{X}_i$) are thought to be multiplicatively distorted by body mass index ($\equiv U_i$), as described by (*M2*). In other words, the objective is estimation of γ_r in the regression relation between Y_i and X_i given in model (*M1*), but the estimation must be based on the available (distorted) data $\{\tilde{X}_{ri}, \tilde{Y}_i, U_i\}_{i=1}^n$. Because the precise effects of U_i are unknown (i.e. the functions $\psi(\cdot)$ and $\phi_r(\cdot)$ are unknown), investigators typically make the simplifying assumption that $\psi(U_i) = \phi_r(U_i) \equiv U_i$, leading to a common normalization via division ($Y_i = \tilde{Y}_i/U_i$ and $X_{ri} = \tilde{X}_{ri}/U_i$) in order to estimate γ_r. The CAR model (*M1*)-(*M2*) accounts for the general unknown effects of U in the estimation of γ_r by allowing $\psi(\cdot)$ and $\phi_r(\cdot)$ to be general (unspecified) smooth functions. Detailed analysis of this data and asymptotic normality of the CAR estimators of γ_r were provided by Şentürk and Müller (2005, 2006).

Uncorrelated errors in (*M1*) is a common assumption in cross-sectional data where observations are taken on different individuals. The covariate-adjusted model (*M1*)-(*M2*) extended to the case of repeated measurements is the general linear model (GLM) with correlated errors within the ith unit of observation (Nguyen and Şentürk, 2008),

$$Y_{it} = \gamma_0 + \sum_{r=1}^{p} \gamma_r X_{rit} + e_{it}, \quad i = 1, \dots, n, \tag{1}$$

where the additional subscript t refers to tth repetition on the ith unit/cluster, $E(e_{it}) = 0$ and $\sigma_{tt'} = cov(e_{it}, e_{it'})$. The observable variables available for the estimation of γ_r are

$$\tilde{Y}_{it} = \psi(U_i)Y_{it}, \quad \text{and} \quad \tilde{X}_{rit} = \phi_r(U_i)X_{rit}, \qquad r = 1, \ldots, p. \tag{2}$$

Without loss of generality, consider (1) in the context of repeated measurements collected at the same time points for all subjects with possibly missing values at some time points. Let $\{s_t\}_{t=1}^{T}$ be the distinct measurement occasions among all measurement times s_{it}, $i = 1, \ldots, n$, $t = 1, \ldots, T_i$. Model (1)-(2) is the (extended) covariate-adjusted GLM. The identifiability condition of vanishing mean distorting effects, $E\{\psi(U_i)\} = 1$ and $E\{\phi_r(U_i)\} = 1$, as in ordinary CAR (Şentürk and Müller, 2005), is needed for estimation.

In this paper we consider the asymptotic properties for the covariate-adjusted GLM with correlated errors model (1)-(2). The paper is organized as follows. The covariate-adjusted estimators are given in Section 2. The asymptotic normality and consistent estimators of the asymptotic variance follow in Section 3 and the main proofs are given in Section 4. Technical conditions and auxiliary lemmas are deferred to an appendix.

2 Covariate-adjusted estimators

We introduce the needed notations and briefly outline the covariate-adjusted GLM estimators of γ_r. For the estimation, a regression of the observed response on the observed predictors, $\tilde{\mathbf{X}}_{it} = (\tilde{X}_{1it}, \ldots, \tilde{X}_{pit})^{\mathrm{T}}$, leads to the observable varying coefficient model,

$$E(\tilde{Y}_{it}|\tilde{\mathbf{X}}_{it}, U_i) = \beta_0(U_i) + \sum_{r=1}^{p} \beta_r(U_i)\tilde{X}_{rit} \tag{3}$$

(Hastie and Tibshirani, 1993), under mutual independence of (e_{it}, U_i, X_{rit}), for $i = 1, \ldots, n$, $t = 1, \ldots, T_i$ and $r = 1, \ldots, p$. The varying coefficient functions in (3) are $\beta_0(u) = \psi(u)\gamma_0$ and $\beta_r(u) = \gamma_r\psi(u)/\phi_r(u)$ and estimation of γ_r is achieved through estimation of the varying coefficient function $\beta_r(\cdot)$ and utilizing these relationships between $\beta_r(\cdot)$ and γ_r (Şentürk and Müller, 2005; Nguyen and Şentürk, 2008). To obtain the raw estimators of $\beta_r(\cdot)$ in (3), a simple binning approach was proposed where the support of U was partitioned into intervals/bins (Nguyen and Şentürk, 2008). Data within each bin is used to fit regressions to obtain raw estimators of $\beta_r(\cdot)$ and the covariate-adjusted estimator of γ_r is based on averaging these raw estimators of $\beta_r(\cdot)$. We now describe in more details this two-step estimation procedure. For convenience of notation the results are presented for the case of complete data ($n_t = n$ for all $t = 1, \ldots, T$).

Step 1: Data binning and raw estimators of $\beta_r(\cdot)$. Let $a \leq U \leq b$, for real numbers $a < b$. The interval $[a, b]$ is divided into m equidistant intervals/bins denoted by $B_1, \ldots, B_m$. The number of subjects in bin B_j is denoted L_j. *Data elements in any given bin are marked by a prime.* For example, the data from the kth subject in B_j observed at time point s_t is $(U'_{jk}, \tilde{X}'_{rjkt}, \tilde{Y}'_{jkt}, X'_{rjkt}, Y'_{jkt}, e'_{jkt})$ for $k = 1, \ldots, L_j$, $r = 1, \ldots, p$ and $t = 1, \ldots, T$. Let $\tilde{\mathbf{X}}'_j$ and $\tilde{\mathbf{Y}}'_j$ be the predictor matrix and response vector for data in bin j, of dimensions $TL_j \times (p+1)$ and $TL_j \times 1$, respectively. After binning the data, a regression of $\tilde{\mathbf{Y}}'_j$ on $\tilde{\mathbf{X}}'_j$ is fitted using data in bin j to obtain the estimators

$$\hat{\boldsymbol{\beta}}_j = (\hat{\beta}_{0j}, \ldots, \hat{\beta}_{pj})^{\mathrm{T}} = (\tilde{\mathbf{X}}_j'^{\mathrm{T}} \tilde{\mathbf{X}}'_j)^{-1} \tilde{\mathbf{X}}_j'^{\mathrm{T}} \tilde{\mathbf{Y}}'_j, \qquad j = 1, \ldots, m. \tag{4}$$

Step 2: Weighted averaging of the raw estimators of $\beta_r(\cdot)$. The covariate-adjusted estimators of γ_r are then obtained as weighted averages of the $\hat{\beta}_j$'s,

$$\hat{\gamma}_= \sum_{j=1}^{m} \frac{L_j}{n} \hat{\beta}_{0j}, \qquad \hat{\gamma}_r = \bar{\bar{\tilde{X}}}_r^{\,-1} \sum_{j=1}^{m} \frac{L_j}{n} \hat{\beta}_{rj} \bar{\tilde{X}}'_{rj}, \quad r = 1, \ldots, p, \tag{5}$$

where $\bar{\bar{\tilde{X}}}_r = (Tn)^{-1} \sum_{i=1}^{n} \sum_{t=1}^{\mathrm{T}} \tilde{X}_{rit}$ and $\bar{\tilde{X}}'_{rj} = (TL_j)^{-1} \sum_{k=1}^{L_j} \sum_{t=1}^{\mathrm{T}} \tilde{X}'_{rjkt}$ are the overall average and the average based on data in bin j for the rth predictor.

Remark: We remark on the estimators for the case of data missing at random in the Appendix section. We note here that the asymptotic results of the next section also hold for a finite amount of data missing completely at random.

3 Asymptotic properties

We state the asymptotic distribution of the covariate-adjusted estimators, $\hat{\gamma}_r$, for a fixed number of time points T. The asymptotic normal distribution of $\hat{\gamma}_r$ and consistent estimators for the asymptotic variance of $\hat{\gamma}_r$ are given in Theorems 1 and 2, respectively.

As is typical in smoothing applications, the number of bins $m = m(n)$ satisfies $m \to \infty$, $n/(m \log n) \to \infty$ and $m/\sqrt{n} \to \infty$ as $n \to \infty$. Further let $\tilde{\boldsymbol{\gamma}}_j$ denote the vector of least squares estimators of $\boldsymbol{\gamma}_j = (\gamma_{0j}, \ldots, \gamma_{pj})^{\mathrm{T}}$ based on the *unobservable* data in bin j,

$$\tilde{\boldsymbol{\gamma}}_j^{\mathrm{T}} = (\tilde{\gamma}_{0j}, \ldots, \tilde{\gamma}_{pj})^{\mathrm{T}} = (\mathbf{X}_j'^{\mathrm{T}} \mathbf{X}_j')^{-1} \mathbf{X}_j'^{\mathrm{T}} \mathbf{Y}_j'^{\mathrm{T}}, \tag{6}$$

where $\mathbf{X}_j'$ and $\mathbf{Y}_j'$ are defined the same way as $\tilde{\mathbf{X}}_j'$ and $\tilde{\mathbf{Y}}_j'$ given in the previous section. (See step 1 of estimation procedure.) For $\hat{\gamma}_r$ and $\tilde{\gamma}_{rj}$ to be well defined, we assume that $\inf_j |\det\{\tilde{\mathbf{P}}_j\}| > \xi$ and $\inf_j |\det\{\mathbf{P}_j\}| > \xi$, where $\tilde{\mathbf{P}}_j \equiv (TL_j)^{-1} \tilde{\mathbf{X}}_j'^{\mathrm{T}} \tilde{\mathbf{X}}_j'$ and $\mathbf{P}_j \equiv (TL_j)^{-1} \mathbf{X}_j'^{\mathrm{T}} \mathbf{X}_j'$, $\xi = \min\{\rho/2, [\inf_j(\phi_1^2(U_j'^*), \ldots, \phi_p^2(U_j'^*))]^p\ \rho/2\}$ with ρ defined in condition 5 (of the Appendix section) and $U_j'^* = L_j^{-1} \sum_{k=1}^{L_j} U_{jk}'$ is the average of the U's in B_j.

For the following theorems, we define the following additional notations:

1. $\lambda_\psi \equiv E\{\psi^2(U)\}$, $\lambda_{\phi_r} \equiv E\{\phi_r^2(U)\}$, $\lambda_{\psi\phi_r} \equiv E\{\psi(U)\phi_r(U)\}$, and $\sigma_\psi^2 \equiv \mathrm{var}\{\psi(U)\}$.
2. $\zeta_r \equiv T^{-2} \sum_{t=1}^{T} \sum_{t'=1}^{T} E(X_{rt} X_{rt'})$.
3. $\mathcal{X}$ is a $p+1 \times p+1$ matrix where the $(r+1, q+1)$ entry is $\sum_t E(X_{rt} X_{qt})$, for $r, q = 0, \ldots, p$.
4. $\mathcal{X}_{tt'} \equiv E(\mathbf{X}_t^{\mathrm{T}} \mathbf{X}_{t'})$, where $\mathbf{X}_t^{\mathrm{T}} = (1, X_{1t}, \ldots, X_{pt})$.
5. $\mathcal{M} \equiv \mathcal{X}^{-1} (T^{-2} \sum_{t=1}^{\mathrm{T}} \sum_{t'=1}^{\mathrm{T}} \sigma_{tt'} \mathcal{X}_{tt'}) \mathcal{X}^{-1^{\mathrm{T}}}$.
6. $\tilde{\mathbf{P}}_{jt} \equiv L_j^{-1} \tilde{\mathbf{X}}_{jt}'^{\mathrm{T}} \tilde{\mathbf{X}}_{jt}'$, where $\tilde{\mathbf{X}}_{jt}'$ is the $L_j \times (p+1)$ predictor matrix containing the observations in bin j only from time point s_t.
7. Also, define the scalar $\tilde{P}_{rjk} \equiv T^{-2} \sum_t \sum_{t'} \tilde{X}_{rjkt}' \tilde{X}_{rjkt'}'$.

Denote convergence in distribution and probability by $\xrightarrow{\mathcal{D}}$ and $\xrightarrow{p}$. The (i, j)th element of a matrix $\mathbf{A}$ is denoted $[\mathbf{A}]_{ij}$.

Theorem 1. *Under conditions (1-8) given in the Appendix, as $n \to \infty$,*

$$\sqrt{n}(\hat{\gamma}_r - \gamma_r) \xrightarrow{\mathcal{D}} \mathbb{N}(0, \sigma_r^2), \qquad 0 \le r \le p,$$

where $\sigma_0^2 = \gamma_0^2\sigma_\psi^2 + \lambda_\psi[\mathcal{M}]_{11}$, $\sigma_r^2 = (EX_r)^{-2}\{(\gamma_r^2\lambda_\psi - 2\gamma_r^2\lambda_{\psi\phi_r} + \gamma_r^2\lambda_{\phi_r})\zeta_r + (EX_r)^2\lambda_\psi[\mathcal{M}]_{r+1,r+1}\}$, *and* $EX_r = T^{-1}\sum_t EX_{rt}$.

We remark here that the asymptotic normality of the ordinary CAR estimators in the i.i.d. case (Theorem 1 of Şentürk and Müller (2006)) can be obtained as a direct corollary to Theorem 1 above with $T = 1$. Also, note that the asymptotic variance, namely σ_r^2, depends on the covariance between repeated measurements $\sigma_{tt'}$. Thus, large sample inference, such as confidence interval construction, will require estimation of this covariance (or correlation) between repeated measurements. Theorem 2 below provides consistent estimators for σ_r^2, which includes an estimator for the covariance term $\sigma_{tt'}$.

Theorem 2. *Under the conditions (1-8) given in the Appendix, as* $n \to \infty$,

$$\hat{\sigma}_r^2 \xrightarrow{p} \sigma_r^2, \qquad 0 \le r \le p,$$

where $\hat{\sigma}_{tt'} = n^{-1}\sum_j\sum_k(\tilde{Y}'_{jkt} - \hat{\beta}_{0j} - \sum_{r=1}^p \hat{\beta}_{rj}\tilde{X}'_{rjkt})(\tilde{Y}'_{jkt'} - \hat{\beta}_{0j} - \sum_{r=1}^p \hat{\beta}_{rj}\tilde{X}'_{rjkt'})$,

$$\hat{\sigma}_0^2 = \left(\sum_{j=1}^m L_j n^{-1}\hat{\beta}_{0j}^2 - \hat{\gamma}_0^2\right) + \sum_{j=1}^m L_j n^{-1}\left[\tilde{\mathbf{P}}_j^{-1}T^{-2}\sum_t\sum_{t'}\hat{\sigma}_{tt'}\tilde{\mathbf{P}}_{jt}\tilde{\mathbf{P}}_j^{-1}\right]_{11},$$

$$\hat{\sigma}_r^2 = \bar{\bar{X}}_r^{-2}\Bigg\{n^{-1}\sum_{j=1}^m \hat{\beta}_{rj}^2\sum_{k=1}^{L_j}\tilde{P}_{rjk} - 2\hat{\gamma}_r n^{-1}\sum_{j=1}^m \hat{\beta}_{rj}\sum_{k=1}^{L_j}\tilde{P}_{rjk} + \hat{\gamma}_r^2 n^{-1}\sum_{j=1}^m\sum_{k=1}^{L_j}\tilde{P}_{rjk}$$
$$+ \sum_{j=1}^m L_j n^{-1}\bar{\bar{X}}_{rj}'^2\left[\tilde{\mathbf{P}}_j^{-1}T^{-2}\sum_t\sum_{t'}\hat{\sigma}_{tt'}\tilde{\mathbf{P}}_{jt}\tilde{\mathbf{P}}_j^{-1}\right]_{r+1,r+1}\Bigg\}.$$

The variance estimators are motivated by the identifiability conditions described in the Introduction section, the form of $\beta_r(\cdot)$, Lemma 1 and 2(a.) in the appendix. The terms in $\{\hat{\sigma}_r^2\}_{r=0}^p$ are sample versions of the corresponding population quantities given in the asymptotic variance of Theorem 1.

4 Proofs

We provide the proofs of Theorems 1 and 2. The following notations are used in the proofs and are given here for reference: (1) $\mathbf{J}$ is a $(p+1)\times(p+1)$ matrix of ones. (2) $\mathbf{1}$ denotes a $(p+1)\times 1$ vector of ones. (3) $\mathbf{\Delta}_j^{(1)} = \{\psi(U_j'^*), \psi(U_j'^*)\phi_1(U_j'^*), \ldots, \psi(U_j'^*)\phi_p(U_j'^*)\}^{\mathrm{T}}$. (4) $\mathbf{\Delta}_j^{(2)} = \{\psi(U_j'^*), \psi(U_j'^*)/\phi_1(U_j'^*), \ldots, \psi(U_j'^*)/\phi_p(U_j'^*)\}^{\mathrm{T}}$. (5) R_{jrkt} denotes the $(r, T(k-1)+t)$th element of the matrix $\mathbf{P}_j^{-1}\mathbf{X}_j'^{\mathrm{T}}$ and similarly $R_{jrk(i)t}$ denotes the $(r, T(k-1)+t)$th element of the matrix $\mathbf{P}_{j(i)}^{-1}\mathbf{X}_{j(i)}'^{\mathrm{T}}$. These quantities are further elaborated on in the proof of Theorem 1. In addition, we define the Hadamard product for two matrices of the same dimension, denoted by $\mathbf{A}\boxdot\mathbf{B}$, with the (i,j)th element equal to the product of the (i,j)th elements of $\mathbf{A}$ and $\mathbf{B}$.

Proof of Theorem 1. Since $\tilde{X}_{jkt}'$ is bounded, $\sup_{1\le j\le m}|\tilde{\mathbf{P}}_j| = O(1)\mathbf{J}$ and $\sup_{1\le j\le m}|\tilde{\mathbf{P}}_j^{-1}| = O(1)\mathbf{J}$. By Lemma 2 $(b.)$, it holds that $\sup_j |\{(TL_j)^{-1}\tilde{\mathbf{X}}_j'^{\mathrm{T}}\tilde{\mathbf{Y}}_j'\} - [\boldsymbol{\Delta}_j^{(1)}\boxdot\{(TL_j)^{-1}\mathbf{X}_j'^{\mathrm{T}}\mathbf{Y}_j'\}]| = O_p(m^{-1})\mathbf{1}$. This result combined with Lemma 1 implies

$$\sup_j \left|\hat{\boldsymbol{\beta}}_j - \boldsymbol{\Delta}_j^{(2)} \boxdot \tilde{\boldsymbol{\gamma}}_j\right| = O_p(m^{-1})\mathbf{1}. \tag{7}$$

where $\tilde{\boldsymbol{\gamma}}_j$ is given in (6). Hence we have

$$\sqrt{n}(\hat{\gamma}_0 - \gamma_0) = \sum_j^m \frac{L_j}{\sqrt{n}}\psi(U_j'^*)\left\{\gamma_0 + [(TL_j\mathbf{P}_j)^{-1}\mathbf{X}_j'^{\mathrm{T}}\mathbf{e}_j']_1\right\} - \sqrt{n}\gamma_0 + O_p\left(\frac{\sqrt{n}}{m}\right), \tag{8}$$

where $\mathbf{e}_j' = (e_{j11}', \ldots, e_{j1T}', \ldots, e_{jL_j1}, \ldots, e_{jL_jT})^{\mathrm{T}}$. By the boundedness properties, Lemma 2 $(a.)$ and $(b.)$, (8) can be further simplified to

$$\sqrt{n}(\hat{\gamma}_0 - \gamma_0) = \sum_{j=1}^m\sum_{k=1}^{L_j}\frac{1}{T\sqrt{n}}\sum_{t=1}^{\mathrm{T}}\left[\gamma_0\psi(U_{jk}') + \psi(U_{jk}')e_{jkt}'R_{j1kt}\right] - \sqrt{n}\gamma_0 + O_p\left(\frac{\sqrt{n}}{m}\right) \tag{9}$$

where R_{jrkt} denotes the $(r, T(k-1)+t)$th element of the matrix $\mathbf{P}_j^{-1}\mathbf{X}_j'^{\mathrm{T}}$.

In (9) the sum is over all bins and over all points within the bins. Therefore, it is equal to the sum over all data points, summed up in a random order. Thus, let $\mathbf{X}_{j(i)}'$ and $L_{j(i)}$ be the matrix $\mathbf{X}_j'$ and the number of points in the jth bin such that $\{U_i \in B_j; i = 1, \ldots, n\}$, respectively. Further, let $\mathbf{P}_{j(i)} \equiv (TL_{j(i)})^{-1}\mathbf{X}_{j(i)}'^{\mathrm{T}}\mathbf{X}_{j(i)}$ and $R_{jrk(i)t}$ be the $(r, T(k-1)+t)$th element of the matrix $\mathbf{P}_{j(i)}^{-1}\mathbf{X}_{j(i)}'^{\mathrm{T}}$, where $U_i = U_{jk}'$ is the kth element in the ordered sample $(U_{j1}', \ldots, U_{jL_j}') \in B_j$. Using these notations and due to the equivalence in the summation over all subjects, we may express (9) as

$$\sqrt{n}(\hat{\gamma}_0 - \gamma_0) = \sum_{\substack{i=1\\ j,k}}^n \frac{1}{T\sqrt{n}}\sum_{t=1}^{\mathrm{T}}\left[\gamma_0\psi(U_i) + \psi(U_i)e_{it}R_{j1k(i)t} - \gamma_0\right] + O_p\left(\frac{\sqrt{n}}{m}\right).$$

Hence $\sqrt{n}(\hat{\gamma}_0 - \gamma_0) \equiv \sum_{i=1}^n Z_{0i} + O_p(\sqrt{n/m})$, i.e. $\sqrt{n}(\hat{\gamma}_0 - \gamma_0)$ is asymptotically equivalent to $S_{0n} \equiv \sum_{i=1}^n Z_{0i}$. The asymptotic normality of $\sqrt{n}(\hat{\gamma}_0 - \gamma_0)$ follows by noting that

$\{S_{0\ell} = \sum_{i=1}^{\ell} Z_{0i}, F_{0\ell}, 1 \le \ell \le n\}$ is a mean zero martingale, where $F_{0\ell}$ is the the σ-field generated by $\{e_{\ell t}, U_\ell, L_{j(\ell)}, \mathbf{X}'_{j(\ell)}\}$, and $S_{0n} \xrightarrow{\mathcal{D}} \mathbb{N}(0, \sigma_0^2)$ under the conditions of Lemma 3.

We establish the asymptotic normality of $\sqrt{n}(\hat{\gamma}_r - \gamma_r)$, $r \ge 1$ next. Write $\hat{\gamma}_r$ as $\hat{\gamma}_r = \widehat{\theta}/\overline{\theta}$, where $\widehat{\theta} \equiv \sum_{j=1}^m L_j n^{-1} \hat{\beta}_{rj} \bar{\tilde{X}}'_{rj}$ and $\overline{\theta} \equiv \sum_{j=1}^m L_j n^{-1} \bar{\tilde{X}}'_{rj}$. We show that the pair $(\widehat{\theta}, \overline{\theta})$ is asymptotically bivariate normal using the Cramer-Wold device so that the asymptotic normality of $\sqrt{n}(\hat{\gamma}_r - \gamma_r)$ will follow by the Delta method. For

$$\sqrt{n}(V_{1n}, V_{2n})^{\mathrm{T}} \equiv \sqrt{n}\left(\widehat{\theta} - \gamma_r EX_r,\ \overline{\theta} - EX_r\right)^{\mathrm{T}} \xrightarrow{\mathcal{D}} \mathbb{N}_2(\{0,0\}^{\mathrm{T}}, \mathbf{\Sigma}_r).$$

to hold, it is sufficient to show that $\sqrt{n}(aV_{1n} + bV_{2n})$, for real coefficients a and b, is asymptotically normal. Using similar calculations as for (9), it can be shown that $\sqrt{n}(aV_{1n} + bV_{2n})$ is asymptotically equivalent to a sum of martingale differences:

$$\begin{aligned}\sqrt{n}(aV_{1n} + bV_{2n}) &= \sum_{\substack{i=1\\ j,k}}^{n} \frac{1}{T\sqrt{n}} \sum_{t=1}^{\mathrm{T}} [a\gamma_r \psi(U_i) X_{rit} + a\bar{X}'_{rj(i)} \psi(U_i) e_{it} R_{jrk(i)t} - a\gamma_r EX_r \\ &+ b\phi_r(U_i) X_{rit} - bEX_r] + O_p\left(\frac{\sqrt{n}}{m}\right) \equiv \sum_{i=1}^{n} Z_{ri} + O_p\left(\frac{\sqrt{n}}{m}\right),\end{aligned}$$

where $\bar{X}'_{rj(i)} = (TL_j)^{-1} \sum_{k=1}^{L_j} \sum_{t=1}^{\mathrm{T}} \tilde{X}'_{rj(i)kt}$. Thus $\sqrt{n}(aV_{1n} + bV_{2n})$ is asymptotically equivalent to $S_{rn} = \sum_{i=1}^n Z_{ri}$. It can be verified that $\{S_{r\ell} = \sum_{i=1}^{\ell} Z_{ri}, F_{r\ell}, 1 \le \ell \le n\}$ is a mean zero martingale for $n \ge 1$, as in the case for S_{0l} above. It follows from Lemma 4 that $S_{rn} \xrightarrow{\mathcal{D}} \mathbb{N}(0, (a,b)\mathbf{\Sigma}_r(a,b)^{\mathrm{T}})$.

Proof of Theorem 2. Consider $\hat{\sigma}_0^2 = \hat{Q}_1 - \hat{\gamma}_0^2 + \hat{Q}_2$ and $\hat{\sigma}_r^2 = (\hat{Q}_3 - 2\hat{Q}_4 + \hat{Q}_5 + \hat{Q}_6)/\bar{\tilde{X}}_r^2$, where each of the terms $\hat{Q}_1, \ldots, \hat{Q}_6$ are defined and analyzed below. It holds by using Lemma 2 (a.) and (b.) that $\sup_j |\tilde{\boldsymbol{\gamma}}_j - \boldsymbol{\gamma}| = o_p(1)\mathbf{1}$, where $\boldsymbol{\gamma} = (\gamma_0, \gamma_1, \ldots, \gamma_p)^{\mathrm{T}}$. This together with the uniforn consistency of $\hat{\boldsymbol{\beta}}_j$ given in (7) implies

$$\sup_j \left|\hat{\boldsymbol{\beta}}_j - \mathbf{\Delta}_j^{(2)} \boxdot \boldsymbol{\gamma}\right| = o_p(1)\mathbf{1}. \tag{10}$$

For the consistency of $\hat{\sigma}_0^2$, we have that

$$\hat{Q}_1 \equiv \sum_{j=1}^{m} \frac{L_j}{n} \hat{\beta}_{0j}^2 = \gamma_0^2 \sum_{j=1}^{m} \frac{L_j}{n} \{\psi^2(U_j'^*) + o_p(1)\} = \frac{\gamma_0^2}{n} \sum_{i=1}^{n} \psi^2(U_i) + o_p(1) = \gamma_0^2 \lambda_\psi + o_p(1),$$

and $\hat{Q}_2 \equiv \sum_{j=1}^m L_j n^{-1}[\tilde{\mathbf{P}}_j^{-1} T^{-2} \sum_t \sum_{t'} \hat{\sigma}_{tt'} \tilde{\mathbf{P}}_{jt} \tilde{\mathbf{P}}_j^{-1}]_{11} \xrightarrow{p} \lambda_\psi [\mathcal{M}]_{11}$. The convergence of $\hat{Q}_1$ follows from (10), boundedness considerations and the law of large numbers. The convergence of $\hat{Q}_2$ follows from Lemma 1, Lemma 2 (a.) and (C5). Thus, $\hat{\sigma}_0^2 \xrightarrow{p} \sigma_0^2$ follows since $\hat{\gamma}_0 \xrightarrow{p} \gamma_0$. For the consistency of $\hat{\sigma}_r^2$, $r \geq 1$, note that

$$\hat{Q}_3 \equiv n^{-1} \sum_{j=1}^m \hat{\beta}_{rj}^2 \sum_{k=1}^{L_j} \tilde{P}_{rjk} = \gamma_r^2 \lambda_\psi \zeta_r + o_p(1),$$

$$\hat{Q}_4 \equiv \hat{\gamma}_r n^{-1} \sum_{j=1}^m \hat{\beta}_{rj} \sum_{k=1}^{L_j} \tilde{P}_{rjk} = \gamma_r \lambda_{\psi\phi} \zeta_r + o_p(1),$$

$$\hat{Q}_5 \equiv \hat{\gamma}_r^2 n^{-1} \sum_{j=1}^m \sum_{k=1}^{L_j} \tilde{P}_{rjk} = \gamma_r^2 \lambda_\phi \zeta_r + o_p(1),$$

$$\hat{Q}_6 \equiv \sum_{j=1}^m \frac{L_j}{n} \bar{\tilde{X}}_{rj}^{\prime 2} \left[\tilde{\mathbf{P}}_j^{-1} \frac{1}{T^2} \sum_t \sum_{t'} \hat{\sigma}_{tt'} \tilde{\mathbf{P}}_{jt} \tilde{\mathbf{P}}_j^{-1} \right]_{r+1,r+1} \xrightarrow{p} \{EX_r\}^2 \lambda_\psi [\mathcal{M}]_{r+1,r+1},$$

and $\hat{\sigma}_{tt'} = n^{-1} \sum_i \psi^2(U_i) e_{it} e_{it'} + o_p(1) = \lambda_\psi \sigma_{tt'} + o_p(1)$, where $\zeta_r = T^{-2} \sum_t \sum_{t'} E(X_{rt} X_{rt'})$. The results follow since $\hat{\gamma}_r \xrightarrow{p} \gamma_r$ and $\bar{\tilde{X}}_r^2 \xrightarrow{p} (EX_r)^2$.

Appendix: Auxiliary results

Technical conditions. The following assumptions are made. (C1) The covariate U is bounded : $-\infty < a \leq U \leq b < \infty$. The density $f(u)$ of U satisfies $\inf_{a \leq u \leq b} f(u) > 0$, $\sup_{a \leq u \leq b} f(u) < \infty$, and is uniformly Lipschitz. (C2) The variables (e_t, U, X_{rt}) are mutually independent for $r = 1, \ldots, p$, $t = 1, \ldots, T$. (C3) For the predictors, $\sup_{i,r,t} |X_{rit}| \leq B$ for some bound $B \in \mathbb{R}$. (C4) The functions $\psi(\cdot)$ and $\phi_r(\cdot)$, $1 \leq r \leq p$, are twice continuously differentiable, satisfying $E\{\psi(U)\} = 1$, $E\{\phi_r(U)\} = 1$, $\phi_r(\cdot) > 0$, $1 \leq r \leq p$. (C5) Let $\mathbf{X}_{it}' = (1, X_{1it}', \ldots, X_{pit}')^{\mathrm{T}}$ contain p components of the ith subject observed at time s_t, and $\mathbf{X} = (\mathbf{X}_{11}', \ldots, \mathbf{X}_{1T}', \ldots, \mathbf{X}_{n1}', \ldots, \mathbf{X}_{nT}')^{\mathrm{T}}$ of size $Tn \times (p+1)$. As $n \to \infty$, $(Tn)^{-1} \mathbf{X}^{\mathrm{T}} \mathbf{X} \xrightarrow{p} \mathcal{X}$, where the limiting matrix $\mathcal{X}$ is nonsingular; i.e. $\rho \equiv |\det(\mathcal{X})| > 0$. (C6) The function $h(u) = \int x g(x, u) dx$ is uniformly Lipschitz, where $g(\cdot, \cdot)$ is the joint density function of $\tilde{X}_{rt}$ and U for $r = 1, \ldots, p$ and $t = 1, \ldots, T$. (C7) The function $h(u) = \int x g(x, u) dx$ is uniformly Lipschitz, where $g(\cdot, \cdot)$ is the joint density function of $\tilde{X}_{rt} e_t$ and U for $r = 1, \ldots, p$ and $t = 1, \ldots, T$. (C8) The error term satisfies $E|e_t^\lambda| < \infty$ for $\lambda > 4$ and $t = 1, \ldots, T$.

Lemma 1. *Under C1-C6, it holds that* $\sup_j |\tilde{\mathbf{P}}_j^{-1} - [\mathbf{\Phi}_j \boxdot \mathbf{P}_j^{-1}]| = O(m^{-1})\mathbf{J}$, *where the* (k,l) *element of* $\mathbf{\Phi}_j$ *is* $[\mathbf{\Phi}_j]_{kl} = \{\phi_{k-1}(U_j'^{*})\phi_{l-1}(U_j'^{*})\}^{-1}$, *for* $1 \le k,l \le p+1$ *and* $\phi_0(U_j'^{*}) \equiv 1$.

Proof. The proof follows closely to Lemma 3 of Şentürk and Müller (2006) and is omitted.

Lemma 2. *Under conditions (1-8), for a sequence* r_n *such that* $r_n = O_p\{(m\log n)/n\}$, *we have: (a.)* $\sup_j |\mathbf{P}_j^{-1} - \mathcal{X}^{-1}| = O_p(\sqrt{r_n})\mathbf{J}$ *and (b.)* $\sup_j |(TL_j)^{-1}\mathbf{X}_j'^{T}\mathbf{e}_j'\ | = O_p(\sqrt{r_n})\mathbf{1}$.

Proof. Note that we can express the matrix $\mathbf{P}_j$ in terms of the sample moments within bins $X_{rj}'^{(\ell)} \equiv (TL_j)^{-1}\sum_{k=1}^{L_j}\sum_{t=1}^{\mathrm{T}} X_{rjkt}'^{\ell}$, and $(X_{rj}'X_{r'j}')^{(\ell)} \equiv (TL_j)^{-1}\sum_{k=1}^{L_j}\sum_{t=1}^{\mathrm{T}}(X_{rjkt}'X_{r'jkt}')^{\ell}$ for $1 \le r, r' \le p$. Then the proof proceeds as in Lemma 4 of Şentürk and Müller (2006).

Lemma 3. *Under conditions (1-8), the martingale differences* $Z_{0\ell}$ *satisfy the conditions: (a.)* $\sum_{\ell=1}^{n} E\{Z_{0\ell}^2 I(|Z_{0\ell}| > \epsilon)\} \to 0$ *for all* $\epsilon > 0$ *and (b.)* $\Delta_0^2 = \sum_{\ell=1}^{n} Z_{0\ell}^2 \xrightarrow{p} \sigma_0^2$ *for* $\sigma_0^2 > 0$.

Proof. Let $w_{0t} = 1/\sqrt{n}$ and $v_{0t} \equiv T^{-1}\sum_t \alpha_{1ntt} + \alpha_{2ntt}e_{tt}$, where $\alpha_{1ntt} = \gamma_0\psi(U_t) - \gamma_0$ and $\alpha_{2ntt} = \psi(U_t)R_{j1k(t)t}$. Define $Z_{0t} = w_{0t}v_{0t}$. Note that $E(v_{0t}) = 0$ and that $\sup|\alpha_{1ntt}| < c_1$ and $\sup|\alpha_{2ntt}| < c_2$ for some $c_1, c_2 > 0$, where the sup is taken over $1 \le t \le n, 1 \le t \le T$. Similar calculations as in the proof of Lemma 1 in Şentürk and Müller (2006) gives: (A) $\sum_{t=1}^{n} E\{Z_{0t}^2 I(|Z_{0t}| > \epsilon)\} \le n^{-1}\sum_{t=1}^{n}\{E(v_{0t}^4)\}^{1/2}\{P(v_{0t}^2 > n\epsilon^2)\}^{1/2}$, for $\epsilon > 0$, and (B) $P(v_{0t}^2 > n\epsilon^2) \le P(T^{-2}\sum_t\sum_{t'} c_1^2 + c_2^2 e_{tt}e_{tt'} + 2c_1c_2|e_{tt'}| > n\epsilon^2)$. Lemma 3 (a.) follows since $E(v_{0t}^4)$ is bounded uniformly in n and t because e_{tt} has finite fourth moment by condition C8 and since $P(T^{-2}\sum_t\sum_{t'} c_1^2 + c_2^2 e_{tt}e_{tt'} + 2c_1c_2|e_{tt'}| > n\epsilon^2) \to 0$ uniformly in n and t.

Next, consider Δ_0^2 in Lemma 3(b.). It is $\Delta_0^2 = T_1 + \cdots + T_4$, where $T_1 = \gamma_0^2\{n^{-1}\sum_\ell \psi^2(U_\ell)\} + \gamma_0^2 - 2\gamma_0^2\{n^{-1}\sum_\ell \psi(U_\ell)\}$, $T_2 = 2\gamma_0 T^{-1}\sum_{t=1}^{\mathrm{T}} n^{-1}\ \sum_\ell \psi^2(U_\ell)e_{\ell t}R_{j1k(\ell)t}$, $T_3 = -2\gamma_0 T^{-1}\sum_{t=1}^{\mathrm{T}} n^{-1}\sum_\ell \psi(U_\ell)e_{\ell t}R_{j1k(\ell)t}$, $T_4 = T^{-2}\sum_t\sum_{t'} n^{-1}\sum_\ell \psi^2(U_\ell)e_{\ell t}e_{\ell t'}R_{j1k(\ell)t} \times R_{j1k(\ell)t'}$. From the law of large numbers, $T_1 \xrightarrow{p} \gamma_0^2\sigma_\psi^2$. Both the second and third terms are $O_p(n^{-1/2})$ since $E(T_2|U, \mathbf{X}, L_j) = 0$ and $\mathrm{var}(T_2|U, \mathbf{X}, L_j) = O(n^{-1})$. For T_4, it can be shown using Lemma 2 (a.) that $T_4 = \lambda_\psi[\mathcal{M}]_{11} + o_p(1)$. Thus, $\Delta_0^2 \xrightarrow{p} \gamma_0^2\sigma_\psi^2 + \lambda_\psi[\mathcal{M}]_{11}$.

Lemma 4. *Under conditions (1-8), the martingale differences* $Z_{r\ell}$ *satisfy the following conditions: (a.)* $\sum_{\ell=1}^{n} E\{Z_{r\ell}^2 I(|Z_{r\ell}| > \epsilon)\} \to 0$ *for all* $\epsilon > 0$ *and (b.)* $\Delta_r^2 = \sum_{\ell=1}^{n} Z_{r\ell}^2 \xrightarrow{p} (a,b)\mathbf{\Sigma}_r(a,b)^T$ *for* $(a,b)\mathbf{\Sigma}_r(a,b)^T > 0$.

Proof. Lemma 4(a.) follows similarly to Lemma 3(a.). The term Δ_r^2 simplifies to $\Delta_r^2 = T_1 + \cdots + T_7 + o_p(1)$, where $T_1 = -(a^2\gamma_r^2 + 2ab\gamma_r + b^2)(EX_r)^2$, $T_2 = [a^2\gamma_r^2\lambda_\psi + 2ab\gamma_r\lambda_{\psi\phi_r}b^2\lambda_{\phi_r}]\zeta_r$, $T_3 = 2a^2\gamma_r T^{-2}\sum_t\sum_{t'} n^{-1}\sum_\ell \psi^2(U_\ell)e_{\ell t}\bar{X}'_{rj(\ell)}X'_{r\ell t'}R_{jrk(\ell)t}$, $T_4 = -2a^2\gamma_r EX_r T^{-1}$ $\sum_t n^{-1}\sum_\ell \psi(U_\ell)e_{\ell t}\bar{X}'_{rj(\ell)}R_{jrk(\ell)t}$, $T_5 = 2abT^{-2}\sum_t\sum_{t'} n^{-1}\sum_\ell \psi(U_\ell)\phi_r(U_\ell)\, e_{\ell t}\bar{X}'_{rj(\ell)}X'_{r\ell t'}R_{jrk(\ell)t}$, $T_6 = -2abEX_rT^{-1}\sum_t n^{-1}\sum_\ell \psi(U_\ell)e_{\ell t}\bar{X}'_{rj(\ell)}R_{jrk(\ell)t}$, and $T_7 = a^2T^{-2}\sum_t\sum_{t'} n^{-1}\sum_\ell \psi^2(U_\ell)$ $e_{\ell t}e_{\ell t'}\bar{X}'^2_{rj(\ell)}R_{jrk(\ell)t} \times R_{jrk(\ell)t'} + o_p(1)$. It can be shown, similar to the proof of Lemma 3, that $T_3, \ldots, T_6$ are $O_p(n^{-1/2})$. The remaining term can be shown using Lemma 2(a.) to be $T_7 = a^2\lambda_\psi(EX_r)^2[\mathcal{M}]_{r+1,r+1} + o_p(1)$. Thus, we have $\Delta_r^2 \xrightarrow{p} (a,b)\boldsymbol{\Sigma}_r(a,b)^{\mathrm{T}}$, where elements of the matrix $\boldsymbol{\Sigma}_r$ are: $[\boldsymbol{\Sigma}_r]_{11} = \gamma_r^2\lambda_\psi\zeta_r - \gamma_r^2(EX_r)^2 + \lambda_\psi(EX_r)^2[\mathcal{M}]_{r+1,r+1}$, $[\boldsymbol{\Sigma}_r]_{12} = \gamma_r\lambda_{\phi_r\psi}\zeta_r - \gamma_r(EX_r)^2$ and $[\boldsymbol{\Sigma}_r]_{22} = \lambda_{\phi_r}\zeta_r - (EX_r)^2$.

The case of data missing at random. As we remarked at the end of Section 2, the asymptotic results also hold for data missing completely at random. The notation becomes more complicated with unequal number of repetitions for subjects, and the main changes to the estimators are as follow. The bin-specific regression coefficients in (4) become $\hat{\boldsymbol{\beta}}_j = (\tilde{\mathbf{X}}_j'^{\mathrm{T}}\mathbf{W}_j\tilde{\mathbf{X}}_j')^{-1}\tilde{\mathbf{X}}_j'^{\mathrm{T}}\mathbf{W}_j\tilde{\mathbf{Y}}_j'$, where $\tilde{\mathbf{X}}_j'$ becomes a matrix of size $\sum_{k=1}^{L_j} T_{jk} \times (p+1)$, $\tilde{\mathbf{Y}}_j'$ becomes a vector of size $\sum_{k=1}^{L_j} T_{jk} \times 1$ and $\mathbf{W}_j$ is a $\sum_{k=1}^{L_j} T_{jk} \times \sum_{k=1}^{L_j} T_{jk}$ diagonal weight matrix with elements $\{1/(TL_{j1}), \ldots, 1/(TL_{jT_{j1}}), \ldots, 1/(TL_{j1}),\ \ldots, 1/(TL_{jT_{jL_j}})\}$. Here T_{jk} denotes the number of measurements for the kth subject in bin j and L_{jt} denotes the number of subjects in bin j that are observed at time s_t. The covariate-adjusted estimators are then given as $\hat{\gamma}_r = (\sum_{j=1}^m \hat{\beta}_{rj}, \hat{\mu}_{\tilde{X}'_{rj}})/\hat{\mu}_{\tilde{X}_r}$, for $r = 0, \ldots, p$, where $\hat{\mu}_{\tilde{X}_r} = T^{-1}\sum_{t=1}^{\mathrm{T}} n_t^{-1}\sum_{i\in I_t}\tilde{X}_{rit}$, $\hat{\mu}_{\tilde{X}'_{rj}} = T^{-1}\sum_{t=1}^{\mathrm{T}} n_t^{-1}\sum_{k\in I_{jt}}\tilde{X}'_{rjkt}$ and $\tilde{X}_{0it} = \tilde{X}'_{0jkt} = 1$. Here n_t denotes the number of subjects observed at time s_t. The summation index I_t denotes the set of subject indices who are observed at time s_t. The index set I_{jt} denotes the set of subjects who are observed at time s_t in bin j.

Acknowledgment

We are grateful to two anonymous referees for their reviews of the manuscript. Support for this work includes the National Institute of Health grants UL1RR024922, RL1AG032119 and RL1AG032115, National Institute of Child Health and Human Development grant HD036071 and grant UL1 RR024146 from the National Center for Research Resources.

References

Hastie, T., Tibshirani, R., 1993. Varying coefficient models. J. R. Statist. Soc. B. 55, 757–96.

Kaysen, G. A., Dubin, J. A., Müller, H. G., Mitch, W. E., Rosales, L. M., Levin, N. W. & the Hemo Study Group, 2002. Relationship among inflammation nutrition and physiologic mechanisms establishing albumin levels in hemodialysis patients. Kidney Int. 61, 2240–2249.

Nguyen, D.V. and Şentürk, D., 2008. Covariate-adjusted regression for longitudinal data incorporating correlation between repeated measurements. Australian and New Zealand Journal of Statistics, in-press.

Şentürk, D., Müller, H. G., 2005. Covariate-adjusted regression. Biometrika 92, 75–89.

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