

## Covariate-adjusted linear mixed effects model with an application to longitudinal data

Danh V. Nguyen<sup>a\*†</sup>, Damla Şentürk<sup>b†</sup> and Raymond J. Carroll<sup>c</sup>

<sup>a</sup>Division of Biostatistics, University of California School of Medicine Davis, CA, USA; <sup>b</sup>Department of Statistics, Pennsylvania State University, University Park, PA, USA; <sup>c</sup>Department of Statistics, Texas A&M University College Station, TX, USA

(Received 25 November 2007; final version received 6 May 2008)

Linear mixed effects (LME) models are useful for longitudinal data/repeated measurements. We propose a new class of covariate-adjusted LME models for longitudinal data that nonparametrically adjusts for a normalising covariate. The proposed approach involves fitting a parametric LME model to the data after adjusting for the nonparametric effects of a baseline confounding covariate. In particular, the effect of the observable covariate on the response and predictors of the LME model is modelled nonparametrically via smooth unknown functions. In addition to covariate-adjusted estimation of fixed/population parameters and random effects, an estimation procedure for the variance components is also developed. Numerical properties of the proposed estimators are investigated with simulation studies. The consistency and convergence rates of the proposed estimators are also established. An application to a longitudinal data set on calcium absorption, accounting for baseline distortion from body mass index, illustrates the proposed methodology.

**Keywords:** binning; covariance structure; covariate-adjusted regression (CAR); longitudinal data; mixed model; multiplicative effect; varying coefficient models

*AMS Subject Classification:* 62J05; 62G05

### 1. Introduction

Longitudinal data are common in biomedical and health sciences research, where repeated measurements are obtained over time for each individual. Linear (and nonlinear) mixed effects models are useful for analysing longitudinal data, providing a simple and effective way to incorporate within-subject and between-subject variation and the correlation structure of longitudinal data. Comprehensive overviews of mixed effects models for analysing longitudinal data include [1–3], among others.

When the response and predictors are observed under possibly nonparametric effects of a confounding covariate, a direct application of linear mixed effects (LME) models may lead to biased estimates of the regression relationship of interest. As an example, we consider the relationship

---

\*Corresponding author. Email: ucdnguyen@ucdavis.edu

†The authors contributed equally to the work.

between calcium intake and calcium absorption. Insufficient calcium intake has an important influence on bone health and the risk of osteoporosis [4], especially in women, who comprise ~80% of cases. More recently, low intake has also been associated with the risk of colorectal cancer [5], hypertension [6] and obesity [7]. The resulting health benefits of calcium depend on the body's ability to absorb the ingested calcium, which is partly affected by the body composition, *e.g.*, underweight, overweight, obese. Thus, it is of interest to examine the effect of calcium intake levels on absorption, where the response measurements and the effect of the predictor are both potentially modulated by a common observable confounding factor, body mass index (BMI). BMI is a standard measurement that characterises the body composition taking into account both height and weight ( $\text{BMI} = \text{kg}/\text{m}^2$ ; *e.g.* overweight: 25.0–29.9, obese: 30+). The dual confounding effects of BMI on both the response and the predictor coupled with the uncertainty of the exact form of these effects motivate our proposed approach, which involves an LME model combined with nonparametric modelling of the confounder effects.

More specifically, for the true underlying relationship between repeatedly measured calcium intake ( $X$ ) and absorption ( $Y$ ), we assume/postulate an (unobserved) latent LME model with random intercept and/or slope parameters to accommodate subject-specific variation. The response and predictor modulating factor, namely  $U = \text{BMI}$ , is referred to as the 'distorting' or confounding covariate. The distorting effects of BMI on both the response and predictor are modelled flexibly through the unknown, unspecified *smooth* distorting functions  $\psi(U)$  and  $\phi(U)$  as

$$\tilde{Y} = \psi(U)Y \quad \text{and} \quad \tilde{X} = \phi(U)X, \quad (1)$$

where  $\tilde{Y}$  and  $\tilde{X}$  denote the observed response and predictor variables, respectively. The above multiplicative distortion modelling framework was previously proposed for covariate-adjusted regression (CAR) in the context of linear regression models for cross-sectional data [8]. Note that the formulation in Equation (1) allows for the flexibility of nonparametric modelling of the distortion due to  $U = \text{BMI}$  on absorption and effects of intake, using the general unknown smooth functions  $\psi(\cdot)$  and  $\phi(\cdot)$ , because one does not have *a priori* knowledge of these functional dependencies.

The motivation for the multiplicative distortion form (1) was based on the studies that involve adjustment for body configuration measures, such as  $U = \text{BMI}$ , through division of the main variables of interest by  $U$ , *i.e.*  $\psi(U) = \phi(U) = U$  so that  $Y = \tilde{Y}/U$  and  $X = \tilde{X}/U$ . For many applications, the multiplicative distortion model (1) can be justified, but alternative distortion models, including additive ( $\tilde{Y} = \psi(U) + Y$ ,  $\tilde{X} = \phi(U) + X$ ) and no distortion ( $\tilde{Y} = Y$ ,  $\tilde{X} = X$ ), can be considered. See [8] for a more extended discussion. Although formulated in terms of general multiplicative distortion (1), the CAR estimation method from [8] is actually adaptive to all three distortion cases (multiplicative, additive and no distortion), in that it yields consistent regression coefficient estimates in all these distortion settings. This flexibility to accommodate a variety of distortion models also generalises to the method proposed in this paper. This is further discussed in Section 3.2.

The latent LME model for the underlying response and predictor coupled with their nonparametric distortion models described above is referred to as the covariate-adjusted linear mixed effects (CA-LME) model. The main objective of this paper is to estimate the underlying latent relationship between the response and predictor based on the observed (distorted) longitudinal data  $\{\tilde{Y}, \tilde{X}, U\}$ . This involves estimation of fixed parameters, random effects and variance components of the model. The previous work of Şentürk and Müller [8] involves the estimation of fixed parameters in regression models for cross-sectional data. In the current work, we account for correlation between repeated measurements in the estimation of fixed parameters and propose estimation methods for variance components and subject-specific effects, adjusted for the distorting effects.

We note that the above framework has similarities with measurement error modelling [9] if one views  $\psi(U)$  and  $\phi(U)$  as unobserved errors affecting the response and predictor. The proposed modeling can then be viewed as a multiplicative measurement error model where the error is in both the response and predictor. There is a large literature on additive measurement error modelling, although work on multiplicative measurement error models is limited to multiplicative errors only in the predictors [10,11]. Also, a key difference with the errors in variables setting is that a part of the error, namely  $U$ , is observed in the CA-LME model. This additional information from  $U$  is utilised in our proposed estimation method.

The paper is organised as follows. We describe the main ideas of the CA-LME models with a single predictor in Section 2. Estimation of the fixed effects, subject-specific effects and variance components and the adjustments to eliminate the distortion effects are described in Section 3. In Section 4, we discuss generalisations of the CA-LME model that allow for (1) multiple predictors under general fixed and random effects structures, as in standard LME modelling of directly observed longitudinal data and (2) combination of predictors with and/or without distortion in the model. The asymptotic results for the proposed CA-LME estimators are also given in Section 4. An illustration of the proposed CA-LME methodology to longitudinal data on the effects of calcium intake on absorption, as introduced above, is summarised in Section 5. Numerical properties are investigated in Section 6 and technical details are deferred to an appendix.

## 2. Covariate-adjusted linear mixed effects models

To present the main ideas of CA-LME models, we first consider the case of a single, possibly time-varying, predictor. Suppose that we have repeated measurements from a longitudinal study. Denote the unobserved true response variable of the  $i$ th subject by  $Y_{ij}$  ( $i = 1, \dots, n$ ) at the  $j$ th time point ( $j = 1, \dots, n_i$ ) and the unobserved true predictor variable by  $X_{ij}$ . A simple LME model for the outcome  $Y_{ij}$  is

$$Y_{ij} = (\gamma_0 + \gamma_1 X_{ij}) + (\gamma_{0i} + \gamma_{1i} X_{ij}) + e_{ij}, \quad (2)$$

where  $e_{ij}$  denotes the mean zero random measurement error term for subject  $i$  at occasion  $j$ . For convenience, we may write model (2) using matrix notation as  $\mathbf{Y}_i = \boldsymbol{\chi}_i(\boldsymbol{\gamma} + \boldsymbol{\gamma}_i) + \mathbf{e}_i$ , where the vector of fixed effects or population parameters is  $\boldsymbol{\gamma} = (\gamma_0, \gamma_1)^T$ , the vector of subject-specific random effects is  $\boldsymbol{\gamma}_i = (\gamma_{0i}, \gamma_{1i})^T$ , the vector of response values for subject  $i$  is  $\mathbf{Y}_i = (Y_{i1}, \dots, Y_{in_i})^T$  and the corresponding matrix of predictor values is  $\boldsymbol{\chi}_i$  with the  $j$ th row given by  $\mathbf{X}_{ij}^T = (1, X_{ij})$ . The error term,  $\mathbf{e}_i = (e_{i1}, \dots, e_{in_i})^T$ , is assumed to be  $\mathbf{e}_i \sim (\mathbf{0}, \mathbf{R}_i)$  and the random effects  $\boldsymbol{\gamma}_i \sim (\mathbf{0}, \mathbf{D})$ , where  $\mathbf{R}_i$  and  $\mathbf{D} = (D_{kl})$  ( $k, l = 1, 2$ ) denote the covariance matrices of  $\mathbf{e}_i$  and  $\boldsymbol{\gamma}_i$ , respectively. For simplicity, we may assume that  $\mathbf{R}_i = \sigma^2 \mathbf{I}_{n_i}$ , and define the collection of parameters of interest as  $\boldsymbol{\theta}_i = (\boldsymbol{\gamma}, \boldsymbol{\gamma}_i, \sigma^2, \mathbf{D})$ .

The standard LME model (2) provides a reasonable approximation in many longitudinal data settings, given that the underlying observations  $\{Y_{ij}, X_{ij}\}$  are observable. In this case, standard estimation of the model parameters, namely the fixed effects ( $\boldsymbol{\gamma}$ ), random effects ( $\boldsymbol{\gamma}_i$ ) and variance components ( $\sigma^2$  and  $\mathbf{D}$ ), can be based on maximum likelihood (ML) or restricted maximum likelihood (REML) methods (see *e.g.* [1]). However, these standard estimation approaches are no longer applicable for distortion-prone responses and predictors.

We consider estimation in model (2) under the following longitudinal data distortion framework, where only distorted versions of  $Y_{ij}$  and  $X_{ij}$  are available. More precisely, the estimation must be based on the available distorted response and predictor data,

$$\tilde{Y}_{ij} = \psi(U_i)Y_{ij} \quad \text{and} \quad \tilde{X}_{ij} = \phi(U_i)X_{ij}. \quad (3)$$

The unknown distorting functions  $\{\psi(U_i), \phi(U_i)\}$  are assumed to be smooth functions of the observable confounder  $U$ . The distorted data available for estimation are the collection  $\{\tilde{\mathbf{Y}}_i, \tilde{\boldsymbol{\chi}}_i, U_i\}_{i=1}^n$  for  $n$  subjects, where the  $n_i$ -vector  $\tilde{\mathbf{Y}}_i$  and the matrix of predictors  $\tilde{\boldsymbol{\chi}}_i$  are defined analogously to  $\mathbf{Y}_i$  and  $\boldsymbol{\chi}_i$  above. Also, we let  $T$  denote the number of distinct observation times,  $t_1, \dots, t_T$ .

There are some important consequences of distortion-prone data on the standard ML or REML estimation methods for the LME model (2). As will be illustrated subsequently, the estimates for  $\boldsymbol{\gamma}$  can be severely biased and variance components estimates will also be off-target. Additionally, efficiency considerations, such as incorporating an estimate of the covariance structure between repeated measurements into the standard estimation procedure, do not reduce the apparent bias resulting from the distortion. Because of the data contamination, common methods for estimating the covariance structure in longitudinal data, including parametric and nonparametric approaches (e.g., [12]), cannot capture the true covariance structure of the response,  $\text{Var}(\mathbf{Y}_i) = \mathbf{R}_i + \boldsymbol{\chi}_i \mathbf{D} \boldsymbol{\chi}_i^T$ , due to biases in variance components estimation. We will propose an estimation method to resolve these problems due to the data distortion in Section 3.

Some constraints on the unknown distortion functions are needed for the identifiability of the estimation problem. Similar to the identifiability condition (IC) used for CAR for cross-sectional data, we use the condition that the distortion is mean preserving, *i.e.*, the means of the observed variables  $E(\tilde{Y}_{ij})$  and  $E(\tilde{X}_{rij})$  are the same as that of the underlying variables  $E(Y_{ij})$  and  $E(X_{rij})$ , respectively. This IC is equivalent to the following constraint on the distorting functions [8]:

$$E\{\psi(U_i)\} = 1 \quad \text{and} \quad E\{\phi(U_i)\} = 1. \quad (4)$$

Note that since the mean preserving distortion is an IC, it is not testable. This is analogous to the zero mean errors for additive distortion in the measurement error literature. That is, the IC for multiplicative distortion on  $X$  is  $E(\tilde{X}) = E(X\phi(U)) = E(X)$  since  $E(\phi(U)) = 1$  and for additive distortion, it is  $E(\tilde{X}^*) = E(X + \phi^*(U)) = E(X)$  since  $E(\phi^*(U)) = 0$ .

### 3. Estimation procedure

#### 3.1. Observable mixed effects varying coefficient model (VCM)

To motivate the proposed estimation method, we first note that the conditional mean of the observable response variable  $\tilde{Y}_{ij}$ , given the subject-specific effects  $\boldsymbol{\gamma}_i$  and data  $\mathcal{D} \equiv \{\tilde{\boldsymbol{\chi}}_i, U_i; i = 1, \dots, n\}$  is,

$$E(\tilde{Y}_{ij} | \mathcal{D}, \gamma_{0i}, \gamma_{1i}) = \beta_0(U_i) + \beta_1(U_i) \tilde{X}_{ij} + b_{0i}(U_i) + b_{1i}(U_i) \tilde{X}_{ij}, \quad (5)$$

where the fixed and subject-specific coefficient functions in (5) are

$$\beta_0(U_i) = \gamma_0 \psi(U_i), \quad \beta_1(U_i) = \gamma_1 \frac{\psi(U_i)}{\phi(U_i)}, \quad (6)$$

and

$$b_{0i}(U_i) = \gamma_{0i} \psi(U_i), \quad b_{1i}(U_i) = \gamma_{1i} \frac{\psi(U_i)}{\phi(U_i)}, \quad (7)$$

respectively. Thus, the above relationship (5) suggests that the regression of the available distorted data  $\{\tilde{Y}_{ij}, \tilde{X}_{ij}, U_i\}$  leads to the following observable varying coefficient model (VCM) with both

fixed and random coefficient functions,

$$\tilde{Y}_{ij} = \beta_0(U_i) + \beta_1(U_i)\tilde{X}_{ij} + b_{0i}(U_i) + b_{1i}(U_i)\tilde{X}_{ij} + \epsilon_{ij}, \quad (8)$$

where  $\epsilon_{ij} \equiv e_{ij}\psi(U_i)$ . Model (8), written more succinctly in matrix notation, is

$$\tilde{\mathbf{Y}}_i = \tilde{\boldsymbol{\chi}}_i\{\boldsymbol{\beta}(U_i) + \mathbf{b}_i(U_i)\} + \boldsymbol{\epsilon}_i, \quad i = 1, \dots, n, \quad (9)$$

with the vectors of functions  $\boldsymbol{\beta}(U_i) = \{\beta_0(U_i), \beta_1(U_i)\}^T$  and  $\mathbf{b}_i(U_i) = \{b_{0i}(U_i), b_{1i}(U_i)\}^T$ , where  $\boldsymbol{\epsilon}_i = (\epsilon_{i1}, \dots, \epsilon_{in_i})^T$ . Also note that given  $U_i = u$ ,  $\mathbf{b}_i(u) = \{b_{0i}(u), b_{1i}(u)\}^T \sim (\mathbf{0}, \tilde{\mathbf{D}}(u))$  and  $\boldsymbol{\epsilon}_i \sim (\mathbf{0}, \tilde{\mathbf{R}}_i(u))$ , where  $\tilde{\mathbf{R}}_i(u) \equiv \text{Var}(\boldsymbol{\epsilon}_i)$  and  $\tilde{\mathbf{D}}(u) \equiv \text{Var}\{\mathbf{b}_i(u)\}$ .

We note that the above observable mixed VCM shares some common features with the class of ‘random varying coefficient’ models [13], where the index of the fixed and subject-specific coefficient functions is the observation time,  $t_{ij}$ , instead of  $U_i$ . The observation that model (8) is a VCM is useful for developing the estimation technique proposed to targets  $\boldsymbol{\gamma}$  and  $\boldsymbol{\gamma}_i$ . More specifically, we will first use a computationally efficient method based on binning to target the varying coefficient functions in Equation (8). Our proposed estimators for  $\boldsymbol{\gamma}$  and  $\boldsymbol{\gamma}_i$  will then be derived using the relations between the varying coefficient functions and the underlying parameters given in Equations (6) and (7).

Note that even though the connection between the CA-LME model in Equation (2) and the VCM in Equation (8) facilitates estimation in the CA-LME model, the proposed CA-LME model is distinct from the VCM in that it is free from, or adjusted for, the effects of  $U$ . The VCM can be viewed as a stratified analysis (with respect to  $U$ ) and such a model aims to address the question, ‘What is the relationship between  $\tilde{Y}$  and  $\tilde{X}$  at varying levels of  $U$ ?’ This is an important question in itself, however this is not the object of inference of interest with respect to the proposed CA-LME model. The object of inference in the CA-LME model is the relationship between  $X$  and  $Y$  (not directly observed). This relationship corresponds to the relation between  $\tilde{Y}$  and  $\tilde{X}$  with the effects of  $U$  removed, *i.e.* free of  $U$ , which is quite different than the target relationship of interest in the VCM. We discuss in further detail the distinction between CA-LME and VCM as well as their limitations, advantages and disadvantages in the discussion section.

Thus, the main contribution of this paper (given in Sections 3.2, 3.3 and 3.4) is the proposal of and the estimation and inference procedures for a latent variable model that explains the regression relation between the variables of primary interest adjusted for  $U$  in the context of repeated (correlated) measurements. Previous work in covariate-adjusted modelling considers a linear regression model for cross-sectional/uncorrelated data, whereas the CA-LME model is designed for correlated data. The proposed estimation procedure is also new where the specific contributions include developing consistent point estimates for fixed effects, estimation of random effects and (within and between subjects) variance components from distorted data.

### 3.2. CA-LME parameter estimation

Estimation of the underlying parameters of interest, namely  $\boldsymbol{\theta}_i = (\boldsymbol{\gamma}, \boldsymbol{\gamma}_i, \sigma^2, \mathbf{D})$ , requires: (a) estimation of the fixed and random varying coefficient functions in the mixed effects VCM (8); (b) estimation and incorporation of the covariance structure among repeated measurements (distinguishing within- and between-subject variation); and (c) adjustment for the distortion effects. The estimation method has three main steps: (1) binning (or ‘stratification’) of the data with respect to the confounding variable  $U$ , (2) fitting an LME model within each bin to obtain bin-specific estimates and (3) aggregating or averaging bin-specific estimates to obtain the CA-LME estimators of  $\boldsymbol{\theta}_i$ . The basic approach to eliminating the distorting effects of  $U$  is to localise the model fitting by using data in each bin only. Details of each step are provided later.

The observable data available for estimation is the collection  $\{\tilde{\mathbf{Y}}_i, \tilde{\boldsymbol{\chi}}_i, U_i\}_{i=1}^n$ . Assuming that the confounding covariate  $U$  is bounded,  $a \leq U \leq b$  where  $a < b$  are real numbers, the initial step of the estimation procedure divides the interval  $[a, b]$  into  $H$  equidistant intervals, denoted by  $B_1, \dots, B_H$  and referred to as bins. Let  $L_v$  be the number of subjects falling into bin  $v$ , for  $v = 1, \dots, H$ . To track the data corresponding to subjects falling into a given bin, observations in any given bin are marked by a prime. More specifically, the data for which  $U_i \in B_v$ , for  $i = 1, \dots, n$ , are given by the collection  $\{(U'_{vk}, \tilde{\boldsymbol{\chi}}'_{vk}, \tilde{\mathbf{Y}}'_{vk}), k = 1, \dots, L_v, \}$ . Let  $n'_{vk}$  denote the number of repeated measurements for subject  $k$  in bin  $v$ , then the data corresponding to the  $k$ th subject falling in bin  $v$  is  $(U'_{vk}, \tilde{\boldsymbol{\chi}}'_{vk}, \tilde{\mathbf{Y}}'_{vk})$ , where  $\tilde{\mathbf{Y}}'_{vk} = (\tilde{Y}'_{vk1}, \dots, \tilde{Y}'_{vkn'_{vk}})^T$ ,  $\tilde{\boldsymbol{\chi}}'_{vk}$  is the  $n'_{vk} \times 2$  matrix of predictor values and  $U'_{vk}$  is the confounder. For example, if subjects 2 and 3 fall into the bin  $v = 7$ , then  $k = 1, 2, L_7 = 2$  and  $n'_{v1} = n_2$  with  $\tilde{\mathbf{Y}}'_{71} = \tilde{\mathbf{Y}}_2, \tilde{\boldsymbol{\chi}}'_{71} = \tilde{\boldsymbol{\chi}}_2$ , and  $U'_{71} = U_2$ . Similarly,  $n'_{v2} = n_3, \tilde{\mathbf{Y}}'_{72} = \tilde{\mathbf{Y}}_3, \tilde{\boldsymbol{\chi}}'_{72} = \tilde{\boldsymbol{\chi}}_3$  and  $U'_{72} = U_3$ .

After binning the data, we approximate the mixed effects VCM (9), specifically  $\tilde{\mathbf{Y}}_i = \tilde{\boldsymbol{\chi}}_i\{\boldsymbol{\beta}(U_i) + \mathbf{b}_i(U_i)\} + \boldsymbol{\epsilon}_i$ , for  $i = 1, \dots, n$ , local to bin  $v$  ( $v = 1, \dots, H$ ). This is achieved by fitting an LME model using the data within each bin. More precisely, for each bin  $v = 1, \dots, H$ , we fit the following LME model,

$$\tilde{\mathbf{Y}}'_{vk} = \tilde{\boldsymbol{\chi}}'_{vk}\{\boldsymbol{\beta}_v + \mathbf{b}_{vk}\} + \boldsymbol{\epsilon}'_{vk}, \quad k = 1, \dots, L_v, \tag{10}$$

where  $\boldsymbol{\beta}_v = (\beta_{0v}, \beta_{1v})^T, \mathbf{b}_{vk} = (b_{0vk}, b_{1vk})^T$  and  $\boldsymbol{\epsilon}'_{vk} = (\epsilon_{vk1}, \dots, \epsilon_{vkn'_{vk}})^T$  denote *bin-specific* vectors of fixed effects, subject-specific effects and errors for subject  $k$  in bin  $v$ . Also, denote the bin-specific covariance matrices by  $\tilde{\mathbf{D}}_v \equiv \text{Var}(\mathbf{b}_{vk})$  and  $\tilde{\mathbf{R}}_{vk} \equiv \text{Var}(\boldsymbol{\epsilon}'_{vk})$ .

Model (10) can be interpreted as an approximation to model (9) ‘local’ to bin  $v$ . Note that because model (10) is based on data local/specific to bin  $v$ , the resulting fixed effects and variance components estimates in model (10), denoted by  $\hat{\boldsymbol{\beta}}_v, \hat{\tilde{\mathbf{D}}}_v$  and  $\hat{\tilde{\mathbf{R}}}_{vk}$  target  $\boldsymbol{\beta}(\cdot), \tilde{\mathbf{D}}(\cdot)$  and  $\tilde{\mathbf{R}}_k(\cdot)$  evaluated in the specific neighbourhood associated with bin  $B_v$ . It follows from the standard LME model theory that the estimators of  $\boldsymbol{\beta}_v$  and  $\mathbf{b}_{vk}$  are

$$\hat{\boldsymbol{\beta}}_v = \mathbf{M}_v^{-1} \left\{ \sum_{k=1}^{L_v} \tilde{\boldsymbol{\chi}}_{vk}{}^T \tilde{\boldsymbol{\Omega}}_{vk} \tilde{\mathbf{Y}}'_{vk} \right\}, \tag{11}$$

and

$$\hat{\mathbf{b}}_{vk} = \tilde{\mathbf{D}}_v \tilde{\boldsymbol{\chi}}_{vk}{}^T \tilde{\boldsymbol{\Omega}}_{vk} (\tilde{\mathbf{Y}}'_{vk} - \tilde{\boldsymbol{\chi}}'_{vk} \hat{\boldsymbol{\beta}}_v), \tag{12}$$

where  $\tilde{\boldsymbol{\Omega}}_{vk} = \tilde{\mathbf{V}}_{vk}^{-1}, \tilde{\mathbf{V}}_{vk} = \tilde{\mathbf{R}}_{vk} + \tilde{\boldsymbol{\chi}}_{vk} \tilde{\mathbf{D}}_v \tilde{\boldsymbol{\chi}}_{vk}{}^T$  and  $\mathbf{M}_v = \sum_{k=1}^{L_v} \tilde{\boldsymbol{\chi}}_{vk}{}^T \tilde{\boldsymbol{\Omega}}_{vk} \tilde{\boldsymbol{\chi}}_{vk}$ . When  $\tilde{\mathbf{D}}_v$  and  $\tilde{\mathbf{R}}_{vk}$  are unknown, they can be estimated using the REML (or ML) method [3]. The REML estimators of  $\tilde{\mathbf{D}}_v$  and  $\tilde{\mathbf{R}}_{vk}$ , denoted by  $\hat{\tilde{\mathbf{D}}}_v$  and  $\hat{\tilde{\mathbf{R}}}_{vk}$ , are substituted into (11) and (12).

The CA-LMEs estimators of the fixed effects,  $\gamma_0$  and  $\gamma_1$ , involve averaging of the bin-specific estimated fixed varying coefficient functions, namely  $\{\hat{\boldsymbol{\beta}}_v\}_{v=1}^H$ , to eliminate the distortion across  $U$ . The CA-LME estimators are given by

$$\hat{\gamma}_r = \frac{1}{\bar{X}_r} \frac{1}{T} \sum_{j=1}^T \frac{1}{m_j} \sum_{v=1}^H \sum_{k \in I_{vj}} \hat{\boldsymbol{\beta}}_{rv} \tilde{X}'_{rvkj}, \quad r = 0, 1, \tag{13}$$

where  $\bar{X}_r = T^{-1} \sum_{j=1}^T m_j^{-1} \sum_{i \in I_j} \tilde{X}_{rij}$  is the overall mean of the predictor variable  $X_r$  (with  $X_{0ij} \equiv 1, X_{1ij} = X_{ij}$ ). Here  $m_j$  denotes the number of subjects observed at time  $t_j$ . Also,  $I_j$  denotes the set of indices of subjects that are observed at time  $t_j$ . Similarly,  $I_{vj}$  is the set of indices of subjects in bin  $B_v$  observed at time  $t_j$ . Note that the distortion effects of  $U$  cancel

in the CA-LME estimators given in Equation (13) and  $\hat{\gamma}_r$  targets the fixed effect parameter  $\gamma_r$ , since  $E\{\beta_r(U)\tilde{X}_{rj}\} = \gamma_r E\{\psi(U)X_{rj}\} = \gamma_r E(X_{rj}) = \gamma_r E(\tilde{X}_{rj})$ , for all  $j = 1, \dots, T$  and  $r = 0, 1$ . The CA-LME estimators,  $\hat{\gamma}_r$ , can be interpreted as method of moments estimators, because  $\tilde{X}_r$  targets  $E(\tilde{X}_r)$  and  $\hat{\gamma}_r \tilde{X}_r$  targets  $E\{\beta_r(U)\tilde{X}_r\}$ .

We note here that the proposed estimators in Equation (13) are consistent under multiplicative distortion (Theorem 1). Note also that an alternative way to handle multiplicative distortion in practice would be to transform it into additive distortion by taking logarithms where the underlying latent model would then be holding between the logarithm of the underlying variables. However, if the distortion is, in fact, additive taking logarithms would not deal with the problem. Even though the proposed estimators are formulated for multiplicative distortion, they are also consistent under additive distortion ( $\tilde{Y} = \psi(U) + Y$  and  $\tilde{X}_r = \phi_r(U) + X_r$ ) as well as the trivial case of no distortion as is shown next. The fact that the proposed estimators are consistent under different types of distortion and no distortion is an advantage of the proposed method over taking logarithms, since it is difficult (and more typically not possible) to justify one type of distortion over another *a priori*.

The consistency of the estimators given in Equation (13) for the additive distortion and no distortion cases follows from the fact that both of these cases also lead to a mixed VCM between the observed variables, as described for the multiplicative distortion in Section 3.1 above. More precisely, in the mixed VCMs derived from these two cases,  $\beta_1(U) = \gamma_1$  (in both models),  $\beta_0(U) = \gamma_0$  for the no distortion case and  $\beta_0(U) = \gamma_0 - \gamma_1\phi(U) + \psi(U)$  for the additive distortion case. Hence, it follows that  $E\{\beta_r(U)\tilde{X}_{rj}\} = \gamma_r E(\tilde{X}_{rj})$  holds for all cases, because  $E\{\phi(U)\} = E\{\psi(U)\} = 0$  under additive effects guaranteeing no distortion on average, *i.e.*  $E(\tilde{X}) = E(X)$  and  $E(\tilde{Y}) = E(Y)$ . Using this information Şentürk and Müller [8] proposed a diagnostic tool to check whether the underlying distortion is additive by checking whether the slope function of the fitted VCM is equal to a constant. Similarly, for the case of no-distortion, one can check whether the slope and the  $y$ -intercept functions are both constant functions. This diagnostic tool provides information on the type of distortion based on the data and is implemented in Section 5 via the proposed bootstrap hypothesis test of Şentürk and Müller. This is helpful in identifying cases where a simpler estimation approach can be adopted when the model reduces to additive or no distortion.

Next, we consider the estimators (predictors) of the subject-specific effects for the CA-LME model, namely  $\gamma_{ri}$  for the  $i$ th individual. Recall from (7) that we have

$$\gamma_{ri} = b_{ri}(U_i) \frac{\phi_r(U_i)}{\psi(U_i)} = b_{ri}(U_i) \frac{\gamma_r}{\beta_r(U_i)}, \quad r = 0, 1,$$

where the last equality above follows directly from Equation (6). Assume without loss of generality that the  $i$ th subject is the  $k$ th individual in bin  $v$ . Then the estimator of the random effect coefficient function,  $b_{ri}(\cdot)$ , is targeted by the subject-specific estimator  $\hat{b}_{rvk}$  from bin  $v$ , given by Equation (12). Targeting  $\beta_r(U_i)$  and  $\gamma_r$  by  $\hat{\beta}_{rv}$  and  $\hat{\gamma}_r$ , we arrive at the following plug-in estimator of  $\gamma_{ri}$  for individual  $k$  in bin  $v$

$$\hat{\gamma}_{rvk} = \hat{b}_{rvk} \frac{\hat{\gamma}_r}{\hat{\beta}_{rv}}, \quad r = 0, 1. \quad (14)$$

### 3.3. Covariate-adjusted estimators of variance components

Similar to the estimation of the fixed and random effects, adjustments are also needed for the estimation of the variance components. The effects of the data distortion on the variance components estimation can be seen from the mixed effects VCM (8). Thus, given  $U_i = u$ ,

the within-subject covariance matrix is  $\tilde{\mathbf{R}}_i(u) \equiv \text{Var}\{\epsilon_i\} = \psi^2(u)\sigma^2\mathbf{I}_{n_i}$  and the between-subject covariance matrix is

$$\tilde{\mathbf{D}}(u) \equiv \text{Var}\{\mathbf{b}_i(u)\} = \begin{bmatrix} \psi^2(u)D_{11} & \left\{ \frac{\psi^2(u)}{\phi(u)} \right\} D_{12} \\ \left\{ \frac{\psi^2(u)}{\phi(u)} \right\} D_{12} & \left\{ \frac{\psi^2(u)}{\phi^2(u)} \right\} D_{22} \end{bmatrix}.$$

The above calculation shows the direct relationship between the variance components of the true unobserved LME model (*i.e.*  $\sigma^2$  and  $\mathbf{D}$ ) and the corresponding ones in the mixed effects VCM (8), at  $U_i = u$ . Thus, the bin-specific (REML) estimators of the variance components, namely  $\hat{\hat{\mathbf{D}}}_v$  and  $\hat{\hat{\mathbf{R}}}_{v\mathbf{k}}$ , target  $\tilde{\mathbf{D}}(\cdot)$  and  $\tilde{\mathbf{R}}_i(\cdot)$  evaluated in the specific neighbourhood of  $U$ , respectively.

Consider the average of the unadjusted bin-specific between-subject variance components estimators, *i.e.*  $\tilde{\mathbf{D}} = (\tilde{D}_{ll'}) = H^{-1} \sum_{v=1}^H \hat{\hat{\mathbf{D}}}_v$ , for  $l, l' = 1, 2$ , which target

$$\begin{pmatrix} \lambda_{11}D_{11} & \lambda_{12}D_{12} \\ \lambda_{12}D_{12} & \lambda_{22}D_{22} \end{pmatrix},$$

where  $\lambda_{11} = E\{\psi^2(U)\}$ ,  $\lambda_{12} = E\{\psi^2(U)/\phi(U)\}$  and  $\lambda_{22} = E\{\psi^2(U)/\phi^2(U)\}$ . To obtain the underlying between-subject variance components,  $D_{ll'}$ , we estimate the coefficients  $\lambda_{ll'}$  and make the required adjustments. The estimated adjustment coefficients are

$$\hat{\lambda}_{11} = \hat{\gamma}_0^{-2} \sum_{v=1}^H \frac{L_v}{n} \hat{\beta}_{0v}^2, \quad \hat{\lambda}_{22} = \hat{\gamma}_1^{-2} \sum_{v=1}^H \frac{L_v}{n} \hat{\beta}_{1v}^2 \quad \text{and} \quad \hat{\lambda}_{12} = \hat{\gamma}_1^{-2} \mu_{\tilde{X}_r}^{-1} \sum_{v=1}^H \frac{L_v}{n} \hat{\beta}_{1v}^2 \mu_{\tilde{X}_r, v},$$

where  $\mu_{\tilde{X}_r} = n^{-1} \sum_{i=1}^n n_i^{-1} \sum_{j=1}^{n_i} \tilde{X}_{rij}$  and  $\mu_{\tilde{X}_r, v} = L_v^{-1} \sum_{k=1}^{L_v} n'_{vk} \sum_{j=1}^{n'_{vk}} \tilde{X}'_{rvkj}$ . Hence, the covariate-adjusted between-subject variance and covariance estimators are obtained as  $\hat{D}_{ll'} = \hat{\lambda}_{ll'}^{-1} \tilde{D}_{ll'}$ , for  $l, l' = 1, 2$ . Similarly, for the within-subject variance, averaging the bin-specific within-subject variances, namely  $\bar{\sigma}^2 = H^{-1} \sum_{v=1}^H \hat{\sigma}_v^2$ , will target  $\lambda_{11}\sigma^2$ , where  $\hat{\sigma}_v^2$  is the bin-specific within-subject variance estimate obtained from bin  $v$ . Therefore, the adjusted within-subject variance estimator that targets  $\sigma^2$  is  $\hat{\sigma}^2 = \hat{\lambda}_{11}^{-1} \bar{\sigma}^2$ .

Finally, we note that it is not difficult to see that the estimators of the required adjustment factors, namely  $\{\hat{\lambda}_{11}, \hat{\lambda}_{22}, \hat{\lambda}_{12}\}$  given above, are sample averages/moments corresponding to the expectations  $\{\lambda_{11}, \lambda_{22}, \lambda_{12}\}$ , which follows from Equation (6). That is, we have from Equation (6) that  $\lambda_{11} = \gamma_0^{-2} E\{\beta_0^2(U)\}$ ,  $\lambda_{22} = \gamma_1^{-2} E\{\beta_1^2(U)\}$  and  $\lambda_{12} = \gamma_1^{-2} E\{\beta_1^2(U)\tilde{X}_1\}/E(\tilde{X}_1)$ . We note that showing the consistency of  $\hat{\lambda}_{ll'}$  for  $\lambda_{ll'}$  involves similar arguments as in the proof of Theorem 2 in [14] and, therefore, is omitted in this work.

#### 4. Model generalisation and consistency

For simplicity of exposition, we have considered an LME model with a random intercept and slope and one predictor variable as the underlying model (2). Generalisation to a model containing (a) multiple predictors, (b) a mixture of distorted and undistorted predictors, and/or (c) more complex random effects and covariance structures would broaden the applicability of the proposed method. We now consider a more general CA-LME model that incorporates extensions (a), (b) and (c). Denote  $p$  distorted predictors by  $\tilde{X}_{rij}$ , for  $r = 1, \dots, p$ , where  $\tilde{X}_{rij} = \phi_r(U_i)X_{rij}$ . Note that the distortion on each predictor is modelled flexibly by allowing a *different* distorting function,  $\phi_r(\cdot)$ ,

corresponding to the  $r$ th predictor. Furthermore, to accommodate undistorted predictors in the regression model, such as age or time, let  $W_{rij}, r = 1, \dots, q$ , be a set of  $q$  undistorted predictors. For instance, descriptive and graphical analysis may suggest a model with baseline age and a quadratic trend in time,  $t_{ij}^2$ . In such a case, one may choose to include the terms  $W_{1ij} = t_{ij}^2$  and  $W_{2ij} = \text{age}_i$  in the model as undistorted predictors.

Without loss of generality, we may assume that random effects are associated with the first  $p_1$  distorted predictors ( $p_1 \leq p$ ) and the first  $q_1$  ( $q_1 \leq q$ ) undistorted predictors. Then the underlying LMEs model for individual  $i$  at time  $t_j$  is

$$Y_{ij} = \sum_{r=0}^p \gamma_r X_{rij} + \sum_{r=1}^q \delta_r W_{rij} + \sum_{r=0}^{p_1} \gamma_{ri} X_{rij} + \sum_{r=1}^{q_1} \delta_{ri} W_{rij} + e_{ij},$$

or equivalently

$$\mathbf{Y}_i = \boldsymbol{\chi}_i \boldsymbol{\gamma} + \mathbf{Z}_i \boldsymbol{\gamma}_i + \mathbf{e}_i, \tag{15}$$

where  $\boldsymbol{\chi}_i$  is an  $n_i \times (p + q + 1)$  predictor matrix for the fixed effects with the  $j$ th row given by  $(1, X_{1ij}, \dots, X_{pij}, W_{1ij}, \dots, W_{qij})$ ,  $\mathbf{Z}_i$  is an  $n_i \times (p_1 + q_1 + 1)$  predictor matrix corresponding to the random effects with the  $j$ th row given by  $(1, X_{1ij}, \dots, X_{p_1ij}, W_{1ij}, \dots, W_{q_1ij})$ ,  $\boldsymbol{\gamma} = (\gamma_0, \gamma_1, \dots, \gamma_p, \delta_1, \dots, \delta_q)^T$  and  $\boldsymbol{\gamma}_i = (\gamma_{0i}, \gamma_{1i}, \dots, \gamma_{p_1i}, \delta_{1i}, \dots, \delta_{q_1i})^T$ . The corresponding observable mixed effects VCM, generalising Equation (8), is

$$\tilde{Y}_{ij} = \sum_{r=0}^p \beta_r(U_i) \tilde{X}_{rij} + \sum_{r=1}^q \eta_r(U_i) W_{rij} + \sum_{r=0}^{p_1} b_{ri}(U_i) \tilde{X}_{rij} + \sum_{r=1}^{q_1} g_{ri}(U_i) W_{rij} + \epsilon_{ij}, \tag{16}$$

where  $\tilde{X}_{0ij} \equiv 1$  and the fixed and random varying coefficient functions are

$$\begin{aligned} \beta_r(U_i) &= \gamma_r \frac{\psi(U_i)}{\phi_r(U_i)}, \quad 0 \leq r \leq p, & \eta_r(U_i) &= \delta_r \psi(U_i), \quad 1 \leq r \leq q, \\ b_{ri}(U_i) &= \gamma_{ri} \frac{\psi(U_i)}{\phi_r(U_i)}, \quad 0 \leq r \leq p_1 & \text{and} & \quad g_{ri}(U_i) = \delta_{ri} \psi(U_i), \quad 1 \leq r \leq q_1, \end{aligned}$$

with  $\phi_0(U_i) \equiv 1$ . The LME model local to bin  $v$ , approximating model (16), is

$$\tilde{\mathbf{Y}}'_{vk} = \tilde{\boldsymbol{\chi}}'_{vk} \boldsymbol{\beta}_v + \tilde{\mathbf{Z}}'_{vk} \mathbf{b}_{vk} + \boldsymbol{\epsilon}'_{vk}, \quad k = 1, \dots, L_v, \tag{17}$$

where  $\boldsymbol{\beta}_v = (\beta_{0v}, \beta_{1v}, \dots, \beta_{pv}, \eta_{1v}, \dots, \eta_{qv})^T$  and  $\mathbf{b}_{vk} = (b_{0vk}, b_{1vk}, \dots, b_{p_1vk}, g_{1vk}, \dots, g_{q_1vk})^T$  are coefficient vectors,  $\boldsymbol{\epsilon}'_{vk} = (\epsilon_{vk1}, \dots, \epsilon_{vkn'_{vk}})^T$  is the vector of bin-specific errors, and  $\tilde{\boldsymbol{\chi}}'_{vk}$  and  $\tilde{\mathbf{Z}}'_{vk}$  are the fixed and random predictor matrix for individual  $k$  in bin  $v$  analogous to the predictor matrices  $\boldsymbol{\chi}_i$  and  $\mathbf{Z}_i$  defined above. The coefficients  $\beta$ 's and  $\eta$ 's correspond to the distorted and undistorted fixed effects, and the  $b$ 's and  $g$ 's correspond to distorted and undistorted random effects.

The solution to Equation (17) is similar to the single-predictor case given by Equations (11) and (12) in Section 3.2. More specifically, the estimator of  $\boldsymbol{\beta}_v$  is as given by Equation (11), except that the matrix  $\tilde{\boldsymbol{\chi}}'_{vk}$  is replaced by the one defined above for the more general fixed and random effects structure (after Equation (15)). The estimator for the subject-specific effects given by Equation (12) is modified accordingly as  $\hat{\mathbf{b}}_{vk} = \tilde{\mathbf{D}}_v \tilde{\mathbf{Z}}'_{vkT} \tilde{\Omega}_{vk} (\tilde{\mathbf{Y}}'_{vk} - \tilde{\boldsymbol{\chi}}'_{vk} \hat{\boldsymbol{\beta}}_v)$ , where  $\tilde{\Omega}_{vk} = \tilde{\mathbf{V}}_{vk}^{-1}$  and  $\tilde{\mathbf{V}}_{vk} = \tilde{\mathbf{R}}_{vk} + \tilde{\mathbf{Z}}'_{vk} \tilde{\mathbf{D}}_v \tilde{\mathbf{Z}}'_{vkT}$ . As before, we replace the unknown  $\tilde{\mathbf{D}}_v$  and  $\tilde{\mathbf{R}}_{vk}$  with their REML estimators,  $\hat{\mathbf{D}}_v$  and  $\hat{\mathbf{R}}_{vk}$ , based on data within bin  $v$ . Note that the implementation requires no new ideas from the simple case because the fitting can be achieved by common software packages (e.g. R, SPLUS, SAS) as in the simple case. One only needs to specify the additional columns of data corresponding to the predictors in  $\tilde{\boldsymbol{\chi}}'_{vk}$  and  $\tilde{\mathbf{Z}}'_{vk}$ .

The CA-LME estimators of the fixed and subject-specific effects are computed similar to Equations (13) and (14), respectively, as

$$\hat{\gamma}_r = \frac{1}{T \bar{X}_r} \sum_{j=1}^T \frac{1}{m_j} \sum_{v=1}^H \sum_{k \in I_{vj}} \hat{\beta}_{rv} \tilde{X}'_{rvkj}, \quad 1 \leq r \leq p, \quad \tilde{\delta}_r = \frac{1}{T} \sum_{j=1}^T \frac{1}{m_j} \sum_{v=1}^H \sum_{k \in I_{vj}} \tilde{\eta}_{rv}, \quad 1 \leq r \leq q,$$

$$\hat{\gamma}_{rvk} = \tilde{b}_{rvk} \frac{\hat{\gamma}_r}{\hat{\beta}_{rv}}, \quad 0 \leq r \leq p_1 \quad \text{and} \quad \tilde{\eta}_{rvk} = \tilde{g}_{rvk} \frac{\tilde{\delta}_r}{\tilde{\eta}_{rv}}, \quad 1 \leq r \leq q_1.$$

The covariate-adjusted estimation of the variance components proceeds similarly as described in Section 3.3 and details for the more general case are in Appendix 1.

We state the consistency result of the proposed CA-LME model estimators,  $\hat{\gamma}_0, \dots, \hat{\gamma}_p, \tilde{\delta}_1, \dots, \tilde{\delta}_q$ , when the total number of subjects  $n$  and the number of subjects observed at time  $j$ ,  $m_j$  ( $j = 1, \dots, T$ ) tend to infinity for a fixed number of total time points  $T$ . The proof is deferred to Appendix 2. Consistency is established for the data missing completely at random (unbalanced data). For  $m_0 = \inf_j m_j$ , assume that  $m_0 \rightarrow \infty$ . As is typical for smoothing, the total number of bins  $H$  satisfies  $H \rightarrow \infty$  and  $m_0 / \{H \log(m_0)\} \rightarrow \infty$  as  $m_0 \rightarrow \infty$ . The CA-LME model estimators are averages of the bin-specific varying coefficient function estimators:  $\hat{\beta}_v = (\hat{\beta}_{0v}, \hat{\beta}_{1v}, \dots, \hat{\beta}_{pv}, \tilde{\eta}_{1v}, \dots, \tilde{\eta}_{qv})^T = \mathbf{M}_v^{-1} \sum_{k=1}^{L_v} \tilde{\mathbf{X}}_{vk}^T \tilde{\Omega}_{vk} \tilde{\mathbf{Y}}'_{vk}$ ,  $v = 1, \dots, H$ . These estimators for the varying coefficient functions, and therefore  $\tilde{\Omega}_{vk} = \tilde{\mathbf{V}}_{vk}^{-1}$  must exist for each bin in order for the proposed estimators to be well defined. Therefore, it is required that  $\mathbf{M}_v$  and  $\tilde{\mathbf{V}}_{vk}$  be nonsingular.

**THEOREM 1** *Under the technical conditions given in Appendix 2,*

$$\hat{\gamma}_r = \gamma_r + O_p(m_0^{-1/2}) + O(H^{-1}), \quad 0 \leq r \leq p$$

$$\hat{\delta}_r = \delta_r + O_p(m_0^{-1/2}) + O(H^{-1}), \quad 1 \leq r \leq q.$$

## 5. CA-LME model of calcium absorption and intake

The relationship between the effects of calcium intake and absorption is important for a variety of human health issues and disease prevention. For example, adequate calcium intake is important in preventing osteoporosis, a condition characterised by porous and fragile bones associated with increased risks of bone fractures especially in women. More recent studies, although preliminary, have linked increased calcium intake to lowered blood pressure, decreased risk of colon adenomas, and reduced weight gain over time. Although the outcome of complex diseases are typically multifactorial in nature, one relationship of interest is between calcium intake and absorption, which we explore in this section.

In particular, we are interested in examining the relationship between the effect of calcium intake levels on calcium absorption, where the response measurements and the effect of the predictor are both potentially modulated by BMI. Because different body compositions (*e.g.* underweight, overweight, obese) may partly influence individuals' ability to absorb the ingested calcium, BMI is a suitable marker for body composition that accounts for both height and weight ( $\text{BMI} = \text{kg}/\text{m}^2$ ). Also, because absorption continually declines with age our underlying LME model for calcium absorption for individual  $i$  at measurement occasion  $j$  is

$$ca.abs_{ij} = \gamma_0 + \gamma_{0i} + (\gamma_1 + \gamma_{1i})\text{intake}_{ij} + \delta_1 \text{age}_{ij} + e_{ij},$$

where the parameters of interest include  $\boldsymbol{\gamma} = (\gamma_0, \gamma_1)^T$ ,  $\boldsymbol{\gamma}_i = (\gamma_{0i}, \gamma_{1i})^T$ ,  $\delta_1$ ,  $\sigma^2 = \text{Var}(e_{ij})$  and  $\mathbf{D} = \text{Cov}(\boldsymbol{\gamma}_i)$  as previously described in Section 2. Estimation is based on the observed data,

denoted by  $\{\widetilde{ca.abs}_{ij}, \widetilde{intake}_{ij}, \widetilde{age}_{ij}\}$ , and BMI for each subject ( $bmi_i$ ). Note that age is incorporated as an undistorted fixed effect. The data are from a longitudinal study on  $n = 188$  female subjects [15], of age between 35 and 45 at the beginning of the study and repeated measurements were taken every 5-year intervals for a total of four occasions ( $1 \leq n_i \leq 4, j = 1, \dots, n_i$ ). The number of subjects with 1, 2, 3 and 4 repeated measurements were 31, 41, 50 and 66, respectively.

Figure 1 displays the varying coefficient estimates utilised in arriving at the parameter estimates of the CA-LME model. The observed calcium intake and age both have a negative relationship with the observed calcium absorption, while the effect of the observed intake seems to decline with increased BMI. The  $y$ -intercept and slope varying coefficient functions for observed calcium intake are both found to be different than a constant function using the bootstrap test of Şentürk and Müller [8] ( $p$ -values: 0.0239, 0.0153, respectively). This implies that the distortion does not reduce to the additive or the no distortion case from the goodness-of-fit discussion outlined in Section 3.2.

Table 1(A) summarises the parameter estimates for the CA-LME model,  $\hat{\gamma}_0, \hat{\gamma}_1$  and  $\tilde{\delta}_1$ . As a baseline comparison, Table 1(A) also displays the estimates of the underlying regression parameters using a standard LME (REML estimation) with the same fixed and random effects structure as the CA-LME model. The only difference is that the effects of BMI are ignored and estimates are obtained from an unadjusted LME model. We first note that given calcium intake, there is a significant natural decline in absorption with age that is consistent between the CA-LME analysis ( $\tilde{\delta}_1 = -0.0035$ ) and the unadjusted-LME analysis (age slope estimate also  $-0.0035$ ), where the ratios of estimates to standard errors (SEs) are approximately  $-4.807$  and  $-5.151$ , respectively. The significant, but small, absolute magnitude in the estimated

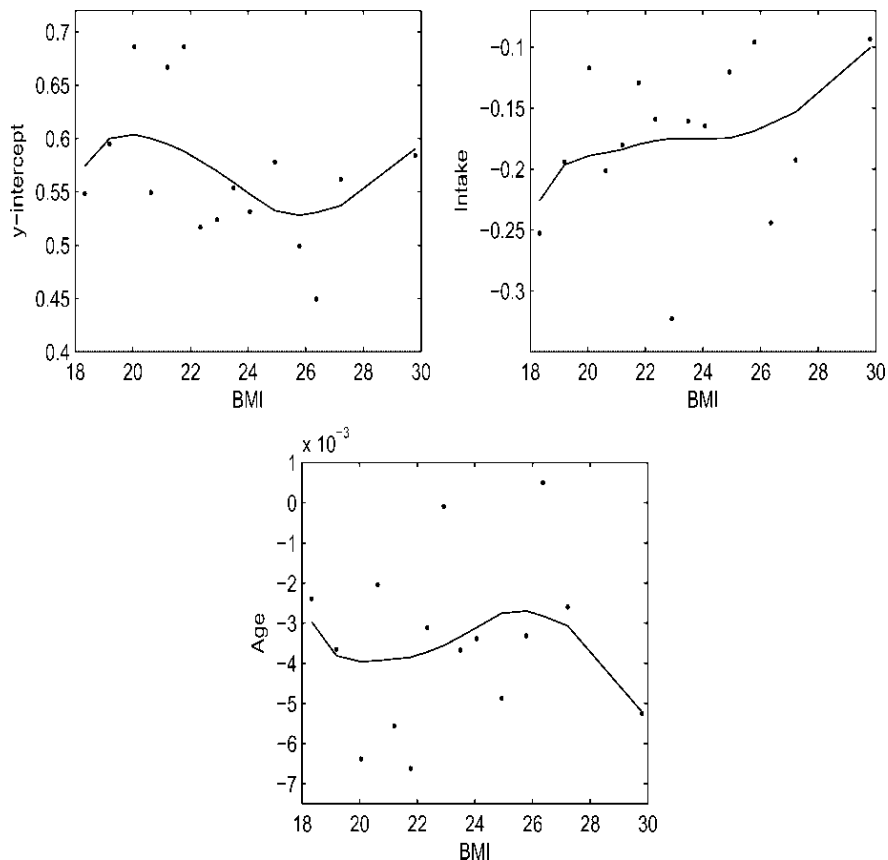


Figure 1. Plot of  $\hat{\beta}_0(\cdot), \hat{\beta}_1(\cdot)$  and  $\hat{\eta}(\cdot)$  for the calcium absorption data.

Table 1. Estimation of fixed effects and variance components for the unobserved latent model of calcium absorption as a function of calcium intake and age:  $E(ca.abs_{ij}) = \gamma_0 + \gamma_{0i} + (\gamma_1 + \gamma_{1i})intake_{ij} + \delta_1 age_{ij}$ , based on observed data  $\{\widehat{ca.abs}_{ij}, \widehat{intake}_{ij}, age_{ij}\}_{j=1}^{n_i}$  for  $n = 188$  women and  $1 \leq n_i \leq 4$ .

Parameter	CA-LME		Unadjusted-LME	
(A) Fixed effects estimates and (SE)				
	Estimate	S.E.	Estimate	S.E.
Intercept ( $\gamma_0$ )	0.5698	0.0059	0.5575	0.0262
Intake ( $\gamma_1$ )	-0.1705	0.0041	-0.1489	0.0114
Age ( $\delta_1$ )	-0.0035	0.0001	-0.0035	0.0005
(B) Variance components estimates				
$d_{11} = \text{Var}(\gamma_{0i})$	0.0067		0.0042	
$d_{22} = \text{Var}(\gamma_{1i})$	0.0084		0.0012	
$d_{12} = \text{Cov}(\gamma_{0i}, \gamma_{1i})$	-0.0062		-0.0020	
$\sigma^2$	0.0042		0.0046	

SEs corresponding to CA-LME model estimates are obtained using 400 bootstrap samples.

coefficient for age reflects the age range of the study subjects (between 35 and 64 across all measurement occasions), unlike in young adulthood where there is typically a rapid decline in absorption after reaching peak bone mass and density for skeletal development (*e.g.*, see [16]). Our main parameter of interest is  $\gamma_1$  for calcium intake and the unadjusted-LME model yields a significant estimate of  $-0.1489$  (SE, 0.0114). Similarly, the CA-LME model which accounts for the possible distortion effects of BMI, yields the estimate of  $\hat{\gamma}_1 = -0.1705$ . Although both estimates of the effect of calcium intake on calcium absorption is clearly significant and in the same direction, the estimate from the CA-LME model is  $\sim 1.89$  SEs lower, relative to the unadjusted estimate. Thus, the CA-LME model analysis suggests a stronger inverse (negative) relationship between calcium intake and absorption after adjusting for body configuration, BMI. We note that this inverse relationship between calcium intake and absorption is consistent with previous studies (*e.g.*, see [17]). Since bin specific estimators given in Equation (11) are heteroskedastic, we use the wild bootstrap method with 400 bootstrap replications to obtain the SEs for the CA-LME estimates (Table 1). Details of the bootstrap implementation are provided in Appendix 3. For this data, we note that the precision did not improve with increased replication beyond  $\sim 100$ .

The within- and between-subject variance parameter estimates are given in Table 1(B). The estimate of within-subject variance is smaller than the variation in the calcium intake slope parameter among individuals (as well as the intercept) for the CA-LME model. Both CA-LME and the unadjusted-LME models suggest a strong correlation between subject-specific intake slopes and intercepts ( $\text{Corr}(\widehat{\gamma_{0i}}, \widehat{\gamma_{1i}}) = -0.83$  and  $-0.92$ , respectively). The within-subject variance estimates are similar for both models ( $\hat{\sigma}^2 \approx 0.004$ ), although variation in the calcium intake slope is appreciably lower when not adjusting for BMI. Also, the number of bins used in the analysis of the calcium data was 15 for a sample size of  $n = 188$ . The coefficients estimates are similar for different number of bins used and this is further explored in the simulation study given in the next section.

## 6. Numerical properties

To evaluate the performance of the proposed method, we consider a model with a mixture of fixed and random effects and distorted and undistorted predictors. The underlying LME model,

analogous to the model used for calcium absorption is

$$Y_{ij} = \gamma_0 + \gamma_{0i} + (\gamma_1 + \gamma_{1i})X_{ij} + \delta_1 W_{ij} + e_{ij}, \tag{18}$$

where  $(\gamma_0, \gamma_1, \delta_1) = (1.5, 2.0, 0.75)$  and the subject-specific effects are obtained as bivariate normal random variates:  $\boldsymbol{\gamma}_i \sim \text{Normal}(\mathbf{0}, \mathbf{D})$  with  $D_{11} = \text{Var}(\gamma_{0i}) = 0.5625$ ,  $D_{22} = \text{Var}(\gamma_{1i}) = 1.0$  and  $D_{12} = \text{Cov}(\gamma_{0i}, \gamma_{1i}) = 0.375$ ; thus,  $\text{Corr}(\gamma_{0i}, \gamma_{1i}) = 0.5$ . The undistorted predictor,  $W_{ij}$ , is taken to be the time points  $t_{ij}$ . The within-subject covariance matrix is  $\mathbf{R}_i = 0.5^2 \mathbf{I}_{n_i}$ . The predictor values are obtained as  $X_{ij} \sim \text{Normal}[1.5(t_{ij} + 1)^2, 1]$ , where the sequence of time points are  $t_{ij} = j/(T + 1)$ , for  $j = 1, \dots, T = 6$ .

To mimic unbalanced data, a common feature in longitudinal data, we randomly removed a proportion,  $\pi$ , of observations from the complete data. We implemented the simulation study for different configurations of sample size and missing rate,  $(n, \pi)$ , with  $n$  ranging from 100 to 800 and  $\pi$  from 0.1 to 0.4. Each simulation configuration was replicated 1000 times. The distorted data are  $\tilde{Y}_{ij} = \psi(U_i)Y_{ij}$  and  $\tilde{X}_{ij} = \phi(U_i)X_{ij}$  with  $\psi(u) = u(u/4 + 3)/v_1$  and  $\phi(u) = (3u - 1)^2/v_2$ . The constants  $v_1$  and  $v_2$  are normalising constants so that the distorting functions satisfy the IC (4). The confounding variable  $U_i$  was obtained as  $U_i \sim \text{Uniform}[2,6]$ .

The main CA-LME estimation results are summarised in Table 2 for the fixed effects ( $\boldsymbol{\gamma}$ ) and Table 3 for the variance components. Given are the estimates based on the CA-LME analysis, the unadjusted analysis which is the standard LME model fitted to the observed data, and the benchmark (optimal) analysis which is the latent LME model (18) fitted to the underlying (unobserved) data  $(Y_{ij}, X_{ij}, W_{ij})$ . The last approach is clearly the optimal one, but these estimates would not be available in practice since  $(Y_{ij}, X_{ij})$  are not directly observable. However, for simulation studies they can be used as the benchmark to compare the proposed CA-LME model analysis and the unadjusted analysis. Generally, as expected, unadjusted LME estimates are biased, whereas the CA-LME estimates target the true parameters. More specifically, the fixed effects estimates in Table 2 suggest that the CA-LME estimates are close to the benchmark LME estimates (LME-benchmark, Table 2) and the bias decreases as sample size increases. Furthermore, the standard unadjusted LME estimates can be severely biased for all sample size configurations.

Similarly, the unadjusted LME estimates of the true variance components ( $\sigma^2$  and  $\mathbf{D}$ ) are generally off-target and the absolute biases are many folds larger relative to the CA-LME and benchmark estimates (Table 3). Also, similar to the fixed effects parameter estimates, the variance components estimates for CA-LME track closely the benchmark estimates. In evaluating the

Table 2. Simulation study: estimation of fixed effects for unbalanced data with 20% missing.

Fixed effects ( $\gamma_0 = 1.5, \gamma_1 = 2, \delta_1 = 0.75$ )										
$n$	CA-LME			Benchmark-LME			Unadjusted-LME			
	True	1.5	2	0.75	1.5	2	0.75	1.5	2	0.75
(A) Mean										
100		1.4956	2.0044	0.7489	1.4948	1.9983	0.7504	1.5337	2.2316	0.8790
200		1.4997	1.9988	0.7523	1.4989	1.9996	0.7508	1.5395	2.2298	0.8830
400		1.4982	2.0003	0.7480	1.4991	2.0006	0.7501	1.5401	2.2334	0.8802
800		1.4997	1.9985	0.7478	1.4996	1.9984	0.7491	1.5395	2.2304	0.8802
(B) Standard deviation										
100		0.1156	0.1244	0.1726	0.0972	0.1081	0.1519	0.1165	0.1517	0.1716
200		0.0861	0.0825	0.1217	0.0716	0.0739	0.1105	0.0870	0.1010	0.1209
400		0.0595	0.0575	0.0845	0.0503	0.0490	0.0770	0.0602	0.0712	0.0854
800		0.0434	0.0424	0.0611	0.0363	0.0363	0.0550	0.0438	0.0518	0.0617

Numbers given are (A) averages and (B) standard deviation across 1000 simulations.

Table 3. Simulation study: estimation of variance components for unbalanced data with 20% missing.

Variance components: $\sigma^2 = 0.25$ , $D_{11} = 0.563$ , $D_{22} = 1$ , $D_{12} = 0.375$												
$n/10$	CA-LME				Benchmark-LME				Unadjusted-LME			
	$\sigma^2$	$D_{11}$	$D_{22}$	$D_{12}$	$\sigma^2$	$D_{11}$	$D_{22}$	$D_{12}$	$\sigma^2$	$D_{11}$	$D_{22}$	$D_{12}$
(A) Mean												
10	0.236	0.589	1.008	0.359	0.250	0.569	1.010	0.382	0.289	0.920	0.182	1.735
20	0.237	0.584	0.986	0.356	0.250	0.562	1.006	0.376	0.289	0.910	0.175	1.730
40	0.240	0.566	0.981	0.360	0.250	0.563	1.002	0.375	0.289	0.907	0.175	1.729
80	0.242	0.561	0.985	0.367	0.250	0.563	0.998	0.376	0.289	0.912	0.175	1.719
(B) Standard deviation												
10	0.033	0.190	0.226	0.112	0.021	0.144	0.145	0.103	0.035	0.226	0.145	0.350
20	0.023	0.125	0.157	0.078	0.015	0.097	0.103	0.071	0.025	0.159	0.096	0.252
40	0.014	0.086	0.108	0.057	0.010	0.069	0.073	0.052	0.018	0.114	0.070	0.181
80	0.010	0.064	0.073	0.039	0.007	0.050	0.049	0.035	0.012	0.085	0.049	0.124

Numbers given are (A) averages and (B) standard deviation across 1000 simulations.

performance of the proposed estimators, we also examined the variance and mean square error (MSE) in estimating the fixed effects and variance components. From the fixed effects results displayed in Figure 2, we make the following observations regarding this specific simulation study: (1) the variability in the proposed CA-LME estimators is much lower in the estimation of  $\gamma_1$  ( $X$ -slope) and is similar for  $\gamma_0$  (intercept) and  $\eta_1$  ( $W$ -slope) compared to the unadjusted

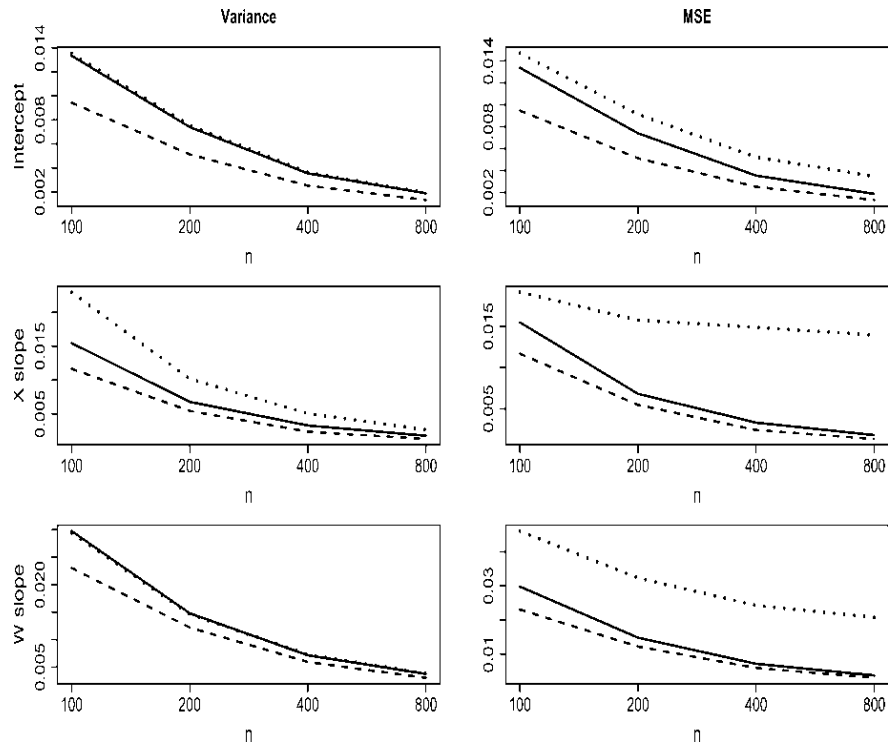


Figure 2. Estimated variance and MSE for estimation of fixed effects coefficients, specifically for the intercept ( $\gamma_0$ ),  $X$  slope ( $\gamma_1$ ) and  $W$  slope ( $\delta_1$ ), corresponding to the benchmark (dashed line), CA-LME (solid line) and unadjusted LME (dotted line) estimates. Due to the large MSE for the unadjusted LME estimates of the  $X$  slope, we plotted  $1/4 \times$  MSE (of the unadjusted estimates) so that it would be on a similar order as the CA-LME and benchmark MSEs (column 2, row 2).

LME analysis; (2) lower variance combined with the small bias (Table 2) result in substantially reduced MSE for the CA-LME estimators, where the estimated MSE tracks the optimal MSE closely as  $n$  increases. These results also hold for the covariate-adjusted variance components (results not shown).

The general pattern of results described above for 20% missing data is similar to the cases with 10%, 30% and 40% missing data. The implementation of the binning algorithm requires specification of the number of bins  $H$ . There are some practical restrictions on the choice of the number of bins,  $H$ . From the practical point of view,  $H$  should be chosen such that there are enough points to fit the corresponding mixed effects models in each bin. From a theoretical point of view,  $H \rightarrow \infty$  and  $m_0/(H \log(m_0)) \rightarrow \infty$ . This means that the number of bins and the number of points within each bin both increase with sample size. Thus, we performed a sensitivity analysis to analyse the effects of the number of bins on the proposed estimators. For the current longitudinal data setting, our study suggests that given the minimal requirement that there are enough data to fit the LME model within each bin, the CA-LME estimates are fairly robust to the bin choice  $H$ . This was also reported in the case of linear regression models for cross-sectional data [14]. For the simulation results reported, the average number of bins were 10, 19, 29 and 45 corresponding to sample size  $n = 100, 200, 400$  and  $800$ , respectively. For each of these sample sizes, Monte Carlo simulation suggests that the corresponding range of bins where the estimates are robust are (8, 14), (10, 21), (20, 33) and (30, 54), respectively. For example, Table 4 illustrates the robustness of the proposed estimators for  $n = 200$  for the number of bins ranging from 10 to 21. Given are the bias and MSE for each parameter of the model.

Next, to assess the subject-specific estimators  $\{\hat{\gamma}_{0i}, \hat{\gamma}_{1i}\}$ , we define the following average residual sum of squares error (ARSSE) quantities of prediction error,

$$ARSSE_1 = \frac{1}{n} \sum_{i=1}^n \frac{1}{n_i} \sum_{j=1}^{n_i} \left( Y_{ij} - \hat{Y}_{ij}^{(1)} \right)^2 \quad \text{and} \quad ARSSE_2 = \frac{1}{n} \sum_{i=1}^n \frac{1}{n_i} \sum_{j=1}^{n_i} \left( Y_{ij} - \hat{Y}_{ij}^{(2)} \right)^2,$$

where  $\hat{Y}_{ij}^{(1)} = \hat{\gamma}_0 + \hat{\gamma}_1 X_{ij} + \tilde{\delta}_1 W_{ij}$  is the population fitted values and  $\hat{Y}_{ij}^{(2)} = (\hat{\gamma}_0 + \hat{\gamma}_{0i}) + (\hat{\gamma}_1 + \hat{\gamma}_{1i}) X_{ij} + \tilde{\delta}_1 W_{ij}$  is the fitted values with the addition of subject-specific random effects. We examine the proportionate reduction in error (PRE) due to the addition of the random effects, namely  $PRE = (ARSSE_1 - ARSSE_2)/ARSSE_1$ . For example, with  $n = 200$  subjects, the minimum, mean and maximum PRE among the 1000 simulation replications are 0.56, 0.87 and 0.96 with a standard deviation of 0.04. Thus, the addition of the CA-LME predicted subject-specific effects reduce, on average, about 87% reduction in prediction error. Note that this is a substantial reduction, especially given that we have demonstrated that the bias in the population estimates  $(\hat{\gamma}_0, \hat{\gamma}_1, \tilde{\delta}_1)$  is very small (Table 2). The distribution of the PREs, for each sample size (20% missing), is summarised in Figure 3. The results show appreciable reduction in error for all sample sizes and, as expected, this improves with sample size.

Table 4. Effects of the number of bins  $H$  on estimation ( $\gamma_0 = 1.5, \gamma_1 = 2.0, \delta = 0.75$ ).

# bin H	Estimates				MSE	
10	1.4967	1.9988	0.7497	0.0060	0.0083	0.0110
13	1.4938	1.9879	0.7514	0.0053	0.0073	0.0125
16	1.4992	1.9975	0.7565	0.0075	0.0070	0.0120
19	1.5007	1.9880	0.7481	0.0066	0.0066	0.0137
21	1.5052	2.0048	0.7681	0.0069	0.0059	0.0127

Results given are averages over simulation runs.

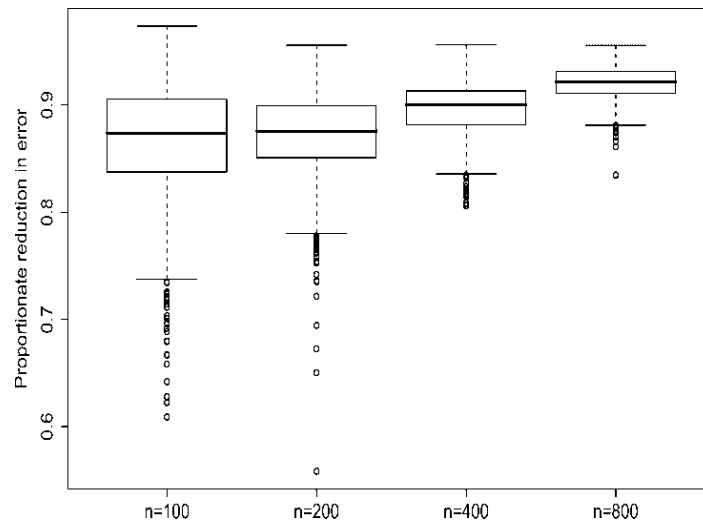


Figure 3. PRE due to the inclusion of estimated subject-specific effects. Displayed are the results for 1000 Monte Carlo data sets (20% missing data) for each sample size  $n$ .

Finally, we note that proposed method can be implemented in standard statistical software with LME model routines, including SAS, R or Splus since they involve fitting LME models after binning by  $U$ . The computational cost is modest as it involves mainly fitting the  $H$  LME models.

## 7. Discussion

As described in Section 3.1, the VCM can be applied to study the relationship between the *observed* variables  $\tilde{Y}$  and  $\tilde{X}$  at varying levels of  $U$ , although this is not the objective of inference for CAR (or the CA-LME) model. Thus, the main distinction (difference) between the two approaches of analysing the data with VCMs and with CAR is that in the latter approach,  $U$  is viewed as a ‘nuisance’ parameter. Hence in the CAR analysis, the result is a model that is free from, or adjusted for, the effects of  $U$  (by removing its effect). The object of inference in the CAR/CA-LME model is the relationship between  $X$  and  $Y$  (not directly observed). Thus, if one were interested in inference for the relationship between  $\tilde{Y}$  and  $\tilde{X}$  and how this relationship/effect is modified by  $U$ , the VCM is the preferred/suitable modelling approach and CAR would not be suitable for this specific aim. Also, we emphasise that in the VCM analysis,  $U$  is an important part of the modelling where one of the main aims is to recover (understand) the effects of  $U$  via the varying coefficient functions, *i.e.* the  $\beta_r(\cdot)$ ’s.

Although VCMs and CAR models both serve useful, but different, purposes (as described above) the applicability of CAR can seem limited, when viewed from a latent variable modelling perspective. From a latent variable modelling perspective, it is true that not being able to check the linearity of the underlying model is a limitation that it shares with other latent variable models. However, the applicability and advantage of CAR is in (biomedical) applications where the adjustments (for  $U$ ) is due to general measurement errors induced by  $U$  and, therefore,  $U$  is nuisance (so that the relationship of interest is between  $X$  and  $Y$ ). The most common example of this in the literature is the adjustment for  $U$  via division (which originally motivated the CAR method). That is, the motivation for the treatment of  $U$  as a nuisance variable and, hence, for our covariate-adjusted approach comes from studies that involve adjustment via division by body configuration measures (such as BMI or body surface area). An example is the 2002 study of Kaysen et al. [18] where albumin turnover and protein catabolic rate were among the variables

that were adjusted for body surface area via division. After the adjustment, researchers proceed to analyse the linear regression relation between the adjusted variables. Here the assumption is that the effects of  $U$  on the variables are removed by the division. Thus, in such cases, CAR has an advantage over the simple adjustment via division, in that the distortion effects of  $U$  are in fact unknown (*a priori*) and CAR provides a way to model this uncertainty.

Also, we point out here that for the particular simulation set-up considered in Section 6, it was observed that the variance of CA-LME estimators is lower than or approximately equal to those of the unadjusted estimators. Even though we know that the unadjusted estimators do not target the underlying regression parameters (bias does not decrease with increasing  $n$ ), since the variance formulas are not derived for CA-LME estimators, one cannot claim that their variance will always be larger compared to the variance of the CA-LME estimators generally. Also, unlike other typical studies where the bias and variance trade-off of two estimators can be studied via a tuning parameter, like the bandwidth, the current setting does not lend itself to such a simple analysis. With respect to this issue, there are (at least) two key considerations that explains the complicated bias-variance relationship. First, the effects of the distortion functions on the variance of both estimators are different, which makes it difficult to predict which variance will be larger in a given set-up (*i.e.* generally). Second, since the unadjusted estimators do not target the underlying regression parameters, they are estimating a different quantity. Hence, the estimators can potentially have very different variances. In our study, we have two estimators that have completely different forms due to the distortion and they target two different quantities.

Finally, while estimation of the varying coefficient functions (*i.e.* the  $\beta(\cdot)$ 's) are a natural by-product of the proposed estimation procedure, estimation of  $\phi(\cdot)$  and  $\psi(\cdot)$  is not. Although this can be viewed as an advantage of the covariate adjusted estimation approach, *i.e.* that without having to know the form of distortions, or without targeting these distortion functions, the underlying relationship can be targeted directly; these functions can provide information graphically. Formulating an alternative estimation algorithm to also target the distortion functions is a challenge for further research.

## Acknowledgements

We are grateful to two reviewers and an Associate Editor for many detailed suggestions which substantially improved the paper. Support for this work includes the National Institute of Health (NIH) grants UL1RR024922, RL1AG032119 and RL1AG032115, National Institute of Child Health and Human Development grant HD036071, NIEHS grant P01-ES011269-06 and grant UL1 RR024146 from the National Center for Research Resources, a component of NIH.

## References

- [1] P.J. Diggle, P. Heagerty, K.-Y. Liang, and S.L. Zeger, *Analysis of Longitudinal Data*, 2nd ed., Oxford University Press, Oxford, 2002.
- [2] G.M. Fitzmaurice, N.M. Laird, and J.H. Ware, *Applied Longitudinal Analysis*, John Wiley & Sons Inc., New Jersey, 2004.
- [3] M. Davidian and D. Giltinan, *Nonlinear Models for Repeated Measurement Data*, Chapman and Hall, New York, 1995.
- [4] B. Dawson-Hughes, *The role of calcium in the treatment of osteoporosis*, in *Osteoporosis*, R. Marcus, D. Feldman, and J. Kelsey, eds., Academic Press, San Diego, 1996, pp. 1159–1168.
- [5] K. Wu, W.C. Willett, C.S. Fuchs, G.A. Colditz, and E.L. Giovannucci, *Calcium intake and risk of colon cancer in women and men*, *J. Natl. Cancer Inst.* 94 (2002), pp. 437–446.
- [6] P.S. Allender, J.A. Cutler, D. Follmann, F.P. Cappuccio, J. Pryer, and P. Elliott, *Dietary calcium and blood pressure: a meta-analysis of randomized clinical trials*, *Ann. Intern. Med.* 124 (1996), pp. 825–831.
- [7] M.B. Zemel, *Regulation of adiposity and obesity risk by dietary calcium: mechanisms and implications*, *J. Amer. Coll. Nut.* 21 (2002), pp. 146S–151S.
- [8] D. Şentürk and H.G. Müller, *Covariate adjusted regression*, *Biometrika* 92 (2005), pp. 59–74.
- [9] R.J. Carroll, D. Ruppert, L.A. Stefanski, and C.M. Crainiceanu, *Measurement Error in Nonlinear Models: a Modern Perspective*, 2nd ed., Chapman and Hall CRC Press, Boca Raton, 2006.

- [10] J.T. Hwang, *Multiplicative error-in-variables models with applications to recent data released by the US Department of Energy*, J. Amer. Statist. Assoc. 81 (1986), pp. 680–688.
- [11] S. Iturria, R.J. Carroll, and D. Firth, *Multiplicative measurement error estimation: estimating equations*, J. Roy. Statist. Soc. B 61 (1999), pp. 547–561.
- [12] P.J. Diggle and A. Verbyla, *Nonparametric estimation of covariance structure in longitudinal data*, Biometrics 54 (1998), pp. 401–415.
- [13] H. Wu and H. Liang, *Backfitting random varying-coefficient models with time-dependent smoothing covariates*, Scand. J. Stat. 31 (2004), pp. 3–19.
- [14] D. Şentürk and H.G. Müller, *Inference for covariate adjusted regression via varying coefficient models*, Ann. Statist. 34 (2006), pp. 654–679.
- [15] C.S. Davis, *Statistical Methods for the Analysis of Repeated Measurements*, Springer-Verlag, New York, 2002.
- [16] V. Matkovic, *Calcium metabolism and calcium requirements during skeletal modeling and consolidation of bone mass*, Amer. J. Clin. Nut. 54 (1991), pp. 245S–605S.
- [17] R.P. Heaney, R.R. Recker, M.R. Stegman, and A.J. Moy, *Calcium absorption in women: relationships to calcium intake, estrogen status, and age*, J. Bone Miner. Res. 4 (1989), pp. 469–475.
- [18] G.A. Kaysen, J.A. Dubin, H.G. Müller, W.E. Mitch, L.M. Rosales, N.W. Levin, and the Hemo Study Group, *Relationship among inflammation nutrition and physiologic mechanisms establishing albumin levels in hemodialysis patients*, Kidney Int. 61 (2002), pp. 2240–2249.

### Appendix 1: Covariate-adjusted variance components

Denote the vector of distorted and undistorted random effects for the underlying LME model (15) by  $\boldsymbol{\gamma}_i = (\boldsymbol{\gamma}_i^{(1)\top}, \boldsymbol{\gamma}_i^{(2)\top})^\top$ , where  $\boldsymbol{\gamma}_i^{(1)} = (\gamma_{0i}, \dots, \gamma_{p_1i})^\top$  and  $\boldsymbol{\gamma}_i^{(2)} = (\delta_{1i}, \dots, \delta_{q_1i})^\top$ . We express the covariance matrix of these random effects,  $\mathbf{D} \equiv \text{Var}(\boldsymbol{\gamma}_i)$ , as

$$\mathbf{D} = (D_{ll'}) = \begin{pmatrix} \mathbf{D}^{(11)} & \mathbf{D}^{(12)} \\ \mathbf{D}^{(21)} & \mathbf{D}^{(22)} \end{pmatrix}, \quad 1 \leq l, l' \leq p_1 + q_1 + 1,$$

where  $\mathbf{D}^{(11)} = \text{Var}(\boldsymbol{\gamma}_i^{(1)})$ ,  $\mathbf{D}^{(22)} = \text{Var}(\boldsymbol{\gamma}_i^{(2)})$ ,  $\mathbf{D}^{(12)} = \text{Cov}(\boldsymbol{\gamma}_i^{(1)}, \boldsymbol{\gamma}_i^{(2)})$ . The observable VCM is Equation (16) where the random varying coefficients at  $U_i = u$  is denoted by  $\mathbf{b}_i(u) = \{\mathbf{b}_i^{(1)\top}(u), \mathbf{b}_i^{(2)\top}(u)\}^\top$  with  $\mathbf{b}_i^{(1)}(u) = \{b_{0i}(u), \dots, b_{p_1i}(u)\}^\top$  and  $\mathbf{b}_i^{(2)}(u) = \{g_{1i}(u), \dots, g_{q_1i}(u)\}^\top$ . Direct calculation yields,

$$\tilde{\mathbf{D}}(u) = \text{Var}\{\mathbf{b}_i(u)\} = \begin{pmatrix} \tilde{\mathbf{D}}^{(11)}(u) & \tilde{\mathbf{D}}^{(12)}(u) \\ \tilde{\mathbf{D}}^{(21)}(u) & \tilde{\mathbf{D}}^{(22)}(u) \end{pmatrix},$$

where

$$\begin{aligned} \tilde{\mathbf{D}}^{(11)}(u) &= \text{Var}\{\mathbf{b}_i^{(1)}(u)\} \\ &= \begin{pmatrix} \psi^2(u)D_{11} & \frac{\psi^2(u)}{\phi_1(u)}D_{12} & \cdots & \frac{\psi^2(u)}{\phi_{p_1}(u)}D_{1,p_1+1} \\ \frac{\psi^2(u)}{\phi_1(u)}D_{12} & \frac{\psi^2(u)}{\phi_1^2(u)}D_{22} & \cdots & \frac{\psi^2(u)}{\phi_1(u)\phi_{p_1}(u)}D_{2,p_1+1} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\psi^2(u)}{\phi_{p_1}(u)}D_{p_1+1,1} & \frac{\psi^2(u)}{\phi_1(u)\phi_{p_1}(u)}D_{p_1+1,2} & \cdots & \frac{\psi^2(u)}{\phi_{p_1}^2(u)}D_{p_1+1,p_1+1} \end{pmatrix}, \\ \tilde{\mathbf{D}}^{(12)}(u) &= \text{Cov}\{\mathbf{b}_i^{(1)}(u), \mathbf{b}_i^{(2)}(u)\} = \begin{pmatrix} \psi^2(u)D_{1,p_1+2} & \cdots & \psi^2(u)D_{1,p_1+q_1+1} \\ \frac{\psi^2(u)}{\phi_1(u)}D_{2,p_1+2} & \cdots & \frac{\psi^2(u)}{\phi_1(u)}D_{2,p_1+q_1+1} \\ \vdots & \ddots & \vdots \\ \frac{\psi^2(u)}{\phi_{p_1}(u)}D_{2,p_1+2} & \cdots & \frac{\psi^2(u)}{\phi_{p_1}(u)}D_{2,p_1+q_1+1} \end{pmatrix}, \end{aligned}$$

and  $\tilde{\mathbf{D}}^{(22)}(u) = \text{Var}\{\mathbf{b}_i^{(2)}(u)\} = \psi^2(u)\mathbf{D}^{(22)}$ . Therefore, the average of the variance components estimates from each bin, namely  $\tilde{\mathbf{D}} = (\tilde{D}_{ll'}) = H^{-1} \sum_{v=1}^H \hat{\mathbf{D}}_v$ , ( $1 \leq l, l' \leq p_1 + q_1 + 1$ ), can be adjusted to obtain the covariate-adjusted variance components estimators. From the above expression for  $\tilde{\mathbf{D}}(u)$ , the required adjustment coefficients involve

$$\lambda_0 \equiv E\{\psi^2(U)\}, \quad \lambda_{rr} \equiv E\left\{\frac{\psi^2(U)}{\phi_r^2(U)}\right\}, \quad 1 \leq r \leq p_1,$$

$$\lambda_r \equiv E\left\{\frac{\psi^2(U)}{\phi_r(U)}\right\}, \quad 1 \leq r \leq p_1 \quad \text{and} \quad \lambda_{rs} \equiv E\left\{\frac{\psi^2(U)}{\phi_r(U)\phi_s(U)}\right\}, \quad 1 \leq r, s \leq p_1, r \neq s.$$

Similar to Section 3.3, it can be shown that  $\lambda_0 = \gamma_0^{-2}E\{\beta_0^2(U)\}$ ,  $\lambda_{rr} = \gamma_r^{-2}E\{\beta_r^2(U)\}$ ,  $\lambda_r = \gamma_r^{-2}E\{\beta_r^2(U)\tilde{X}_r\}/E(\tilde{X}_r)$  and  $\lambda_{rs} = (\gamma_r\gamma_s)^{-1}E\{\beta_r(U)\beta_s(U)\}$ . The estimators for the adjustment coefficients are:

$$\hat{\lambda}_0 = \hat{\gamma}_0^{-2} \sum_{v=1}^H \frac{L_v}{n} \hat{\beta}_{0v}^2, \quad \hat{\lambda}_r = \hat{\gamma}_r^{-2} \mu_{\tilde{X}_r}^{-1} \sum_{v=1}^H \frac{L_v}{n} \hat{\beta}_{rv}^2 \mu_{\tilde{X}_r v}, \quad 1 \leq r \leq p_1,$$

$$\hat{\lambda}_{rr} = \hat{\gamma}_r^{-2} \sum_{v=1}^H \frac{L_v}{n} \hat{\beta}_{rv}^2, \quad 1 \leq r \leq p_1 \quad \text{and} \quad \hat{\lambda}_{rs} = (\hat{\gamma}_r \hat{\gamma}_s)^{-1} \sum_{v=1}^H \frac{L_v}{n} \hat{\beta}_{rv} \hat{\beta}_{sv}, \quad 1 \leq r, s \leq p_1, r \neq s.$$

They are sample moments corresponding to theoretical moments  $\{\lambda_0, \lambda_{rr}, \lambda_r, \lambda_{rs}\}$ .

## Appendix 2: Proof

*Technical conditions.* The following assumptions are made.

- (C1) The covariate  $U$  is bounded below and above:  $-\infty < a \leq U \leq b < \infty$ , for real numbers  $a < b$ . The density  $f(u)$  of  $U$  satisfies  $\inf_{a \leq u \leq b} f(u) > 0$ ,  $\sup_{a \leq u \leq b} f(u) < \infty$  and is uniformly Lipschitz; that is there exists a real number  $M$  such that  $\sup_{a \leq u \leq b} |f(u+c) - f(u)| \leq M|c|$  for any real number  $c$ .
- (C2) The needed dependence/independence structure is that  $U$  is independent of  $e_j$ ,  $X_{rj}$  is independent of  $e_j$  and  $U$ ; and  $W_{sj}$  is independent of  $e_j$  for  $r = 1, \dots, p$ ,  $s = 1, \dots, q$ ,  $j = 1, \dots, T$ .
- (C3) For the predictors,  $\sup |X_{rij}| \leq B_1$  and  $\sup |W_{rij}| \leq B_2$  for some bounds  $B_1, B_2 \in \mathbb{R}^+$  and where sup is taken over  $1 \leq i \leq n$ ,  $1 \leq r \leq p$ ,  $1 \leq j \leq n_i$ . In addition,  $T^{-1} \sum_{j=1}^T E(X_{rj}) \neq 0$  for  $1 \leq r \leq p$ .
- (C4) The functions  $\psi(\cdot)$  and  $\phi_r(\cdot)$ ,  $1 \leq r \leq p$ , are twice continuously differentiable, satisfying  $E\{\psi(U)\} = 1$ ,  $E\{\phi_r(U)\} = 1$  and  $|\phi_r(\cdot)| > c$ , for some constant  $c > 0$ .
- (C5) Define  $\mathbf{\Omega}_i^*$  to be the normalisation of  $\mathbf{\Omega}_i$ , where the  $j_1$ th row and  $j_2$ th column element of  $\mathbf{\Omega}_i^*$  is given as  $(\mathbf{\Omega}_i^*)_{j_1 j_2} = (\mathbf{\Omega}_i)_{j_1 j_2} / m_{j_1 j_2}$ , and  $m_{j_1 j_2}$  denotes the number of subjects observed at both times  $t_{j_1}$  and  $t_{j_2}$ . As  $\inf_{j_1 j_2} m_{j_1 j_2} \rightarrow \infty$ ,  $\sum_{i=1}^n \mathbf{\chi}_i^T \mathbf{\Omega}_i^* \mathbf{\chi}_i \rightarrow \mathcal{C}$  in probability where the limiting  $p + q + 1 \times p + q + 1$  matrix  $\mathcal{C}$  is nonsingular.

*Proof of Theorem 1* We first introduce some boundedness considerations. Since  $X_{rij}$  and  $W_{rij}$  are assumed to be bounded (C3) and  $U$  has compact support (C1),  $\tilde{X}_{rij}$  are also bounded. The matrices  $\tilde{\mathbf{V}}_{vk}$  and  $\mathbf{M}_v = \sum_{k=1}^{L_v} \tilde{\mathbf{\chi}}_{vk}^T \tilde{\mathbf{\Omega}}_{vk} \tilde{\mathbf{\chi}}_{vk}'$  have been assumed to be invertible previously; more precisely, assume  $\sup_{1 \leq v \leq H} \mathbf{M}_v^{-1} = O(1) \mathbf{1}_{(p+q+1) \times (p+q+1)}$ , where  $\mathbf{1}_{a \times b}$  is an  $a \times b$  dimensional matrix of ones. We define  $\Delta_{vk} = \psi(U'_{vk}) - \psi(U_v^*)$  and  $\Delta'_{rvk} = \psi(U'_{vk})/\phi_r(U'_{vk}) - \psi(U_v^*)/\phi_r(U_v^*)$  for  $1 \leq k \leq L_v$ ,  $1 \leq r \leq p$ ,  $1 \leq v \leq H$ , where  $U_v^* = L_v^{-1} \sum_{k=1}^{L_v} U'_{vk}$ , is the

average of the  $U$ 's in  $B_v$ . The following boundedness results for  $1 \leq r \leq p$ ,  $1 \leq k \leq L_v$  can be obtained using Taylor expansions:  $\sup_{v,k} |U'_{vk} - U_v'^*| \leq (b - a)/H$ ;  $\sup_{v,k} |\Delta_{vk}| = O(H^{-1})$ ;  $\sup_{v,k} |\Delta'_{rvk}| = O(H^{-1})$ .

The coefficient function estimators  $\hat{\beta}_v$  can be expressed as

$$\hat{\beta}_v = \mathbf{M}_v^{-1} \left\{ \sum_{k=1}^{L_v} \tilde{\mathbf{X}}_{vk}^T \tilde{\Omega}_{vk} \tilde{\mathbf{Y}}'_{vk} \right\} = \mathbf{M}_v^{-1} \left\{ \sum_{k=1}^{L_v} \sum_{j_1=1}^{n'_{vk}} \sum_{j_2=1}^{n'_{vk}} (\tilde{\Omega}_{vk})_{j_2 j_1} \right.$$

$$\times \left[ \begin{array}{c} \sum_{r=0}^p \beta_r(U'_{vk}) \tilde{\mathbf{X}}'_{rvkj_1} + \sum_{r=0}^{p_1} b_{rvk}(U'_{vk}) \tilde{\mathbf{X}}'_{rvkj_1} \\ \vdots \\ \tilde{\mathbf{X}}'_{pvkj_2} \left[ \left\{ \sum_{r=0}^p \beta_r(U'_{vk}) + \sum_{r=0}^{p_1} b_{rvk}(U'_{vk}) \right\} \tilde{\mathbf{X}}'_{rvkj_1} \right] \\ \vdots \\ \mathbf{W}'_{qvkj_2} \left[ \left\{ \sum_{r=0}^p \beta_r(U'_{vk}) + \sum_{r=0}^{p_1} b_{rvk}(U'_{vk}) \right\} \tilde{\mathbf{X}}'_{rvkj_1} \right] \end{array} \right] + \sum_{k=1}^{L_v} \sum_{j_1=1}^{n'_{vk}} \sum_{j_2=1}^{n'_{vk}} (\tilde{\Omega}_{vk})_{j_2 j_1}$$

$$\times \left[ \begin{array}{c} \sum_{r=1}^q \eta_r(U'_{vk}) \mathbf{W}'_{rvkj_1} + \sum_{r=1}^{q_1} g_{rvk}(U'_{vk}) \mathbf{W}'_{rvkj_1} + \epsilon'_{vkj_1} \\ \vdots \\ \tilde{\mathbf{X}}'_{pvkj_2} \left[ \left\{ \sum_{r=1}^q \eta_r(U'_{vk}) + \sum_{r=1}^{q_1} g_{rvk}(U'_{vk}) \right\} \mathbf{W}'_{rvkj_1} \right] + \epsilon'_{vkj_1} \\ \vdots \\ \mathbf{W}'_{qvkj_2} \left[ \left\{ \sum_{r=1}^q \eta_r(U'_{vk}) + \sum_{r=1}^{q_1} g_{rvk}(U'_{vk}) \right\} \mathbf{W}'_{rvkj_1} \right] + \epsilon'_{vkj_1} \end{array} \right] \Big\}.$$

Next we expand  $\beta_r(U'_{vk})$  and  $\eta_r(U'_{vk})$  in the above formulation around  $U_v'^*$ , the average of the  $U$ 's in bin  $B_v$ . The remainder terms from this expansion, when averaged to form  $\hat{\gamma}_r$ , will be shown to be either  $o_p(m_0^{-1/2})$  or  $o_p(H^{-1})$ , while the leading term will be shown to be  $\gamma_r + O_p(m_0^{-1/2}) + O(H^{-1})$ , and hence Theorem 1 follows. Details are given below.

Making use of the remainder terms  $\Delta_{vk}$  and  $\Delta'_{rvk}$  defined above, we have

$$\hat{\beta}_v = \mathbf{M}_v^{-1} \mathbf{M}_v \left\{ \gamma_0 \psi(U_v'^*), \gamma_1 \frac{\psi(U_v'^*)}{\phi_1(U_v'^*)}, \dots, \gamma_p \frac{\psi(U_v'^*)}{\phi_p(U_v'^*)}, \delta_1 \psi(U_v'^*), \dots, \delta_q \psi(U_v'^*) \right\}^T$$

$$+ \mathbf{M}_v^{-1} \mathcal{S}_{XZ} \left\{ \gamma_{0vk} \psi(U'_{vk}), \gamma_{1vk} \frac{\psi(U'_{vk})}{\phi_1(U'_{vk})}, \dots, \gamma_{p_1vk} \frac{\psi(U'_{vk})}{\phi_{p_1}(U'_{vk})}, \delta_{1vk} \psi(U'_{vk}), \dots, \delta_{q_1vk} \psi(U'_{vk}) \right\}^T$$

$$\begin{aligned}
 & + \mathbf{M}_v^{-1} \sum_{k=1}^{L_v} \sum_{j_1=1}^{n'_{vk}} \sum_{j_2=1}^{n'_{vk}} (\tilde{\Omega}_{vk})_{j_2 j_1} \begin{bmatrix} \gamma_0 \Delta_{vk} + \sum_{r=1}^p \gamma_r \Delta'_{rvk} \tilde{X}'_{rvkj_1} + \sum_{r=1}^q \delta_r \Delta_{vk} W'_{rvkj_1} \\ \vdots \\ \tilde{X}'_{pvkj_2} (\gamma_0 \Delta_{vk} + \sum_{r=1}^p \gamma_r \Delta'_{rvk} \tilde{X}'_{rvkj_1} + \sum_{r=1}^q \delta_r \Delta_{vk} W'_{rvkj_1}) \\ \vdots \\ W'_{qvkj_2} (\gamma_0 \Delta_{vk} + \sum_{r=1}^p \gamma_r \Delta'_{rvk} \tilde{X}'_{rvkj_1} + \sum_{r=1}^q \delta_r \Delta_{vk} W'_{rvkj_1}) \end{bmatrix} \\
 & + \mathbf{M}_v^{-1} \sum_{k=1}^{L_v} \sum_{j_1=1}^{n'_{vk}} \sum_{j_2=1}^{n'_{vk}} (\tilde{\Omega}_{vk})_{j_2 j_1} \begin{bmatrix} \psi(U'_{vk}) e'_{vkj_1} \\ \vdots \\ \tilde{X}'_{pvkj_2} \psi(U'_{vk}) e'_{vkj_1} \\ \vdots \\ W'_{qvkj_2} \psi(U'_{vk}) e'_{vkj_1} \end{bmatrix},
 \end{aligned}$$

where  $\mathcal{S}_{XZ} \equiv \sum_{k=1}^{L_v} \tilde{\mathbf{X}}_{vk}^T \tilde{\Omega}_{vk} \tilde{\mathbf{Z}}'_{vk}$ . Thus, the CA-LME estimators  $\hat{\gamma}_r$  becomes

$$\begin{aligned}
 \hat{\gamma}_r &= \frac{\gamma_r}{T \bar{\tilde{X}}_r} \sum_{j=1}^T \frac{1}{m_j} \sum_{v=1}^H \sum_{k \in I_{vt}} \frac{\psi(U_v^{*})}{\phi_r(U_v^{*})} \tilde{X}'_{rvkj} + \frac{1}{T \bar{\tilde{X}}_r} \sum_{j=1}^T \frac{1}{m_j} \sum_{v=1}^H \{\mathbf{M}_v^{-1}\}_{r+1,1} \\
 & \times \sum_{k=1}^{L_v} \sum_{j_1=1}^{n'_{vk}} \sum_{j_2=1}^{n'_{vk}} (\tilde{\Omega})_{j_2 j_1} \tilde{X}'_{rvkj} \tilde{\mathbf{Z}}'_{vkj_1} \mathbf{b}_{vk} + \dots + \frac{1}{T \bar{\tilde{X}}_r} \sum_{j=1}^T \frac{1}{m_j} \sum_{v=1}^H \{\mathbf{M}_v^{-1}\}_{r+1,p} \\
 & \times \sum_{k=1}^{L_v} \sum_{j_1=1}^{n'_{vk}} \sum_{j_2=1}^{n'_{vk}} (\tilde{\Omega})_{j_2 j_1} \tilde{X}'_{pvkj_2} \tilde{X}'_{rvkj} \tilde{\mathbf{Z}}'_{vkj_1} \mathbf{b}_{vk} + \dots + \frac{1}{T \bar{\tilde{X}}_r} \sum_{j=1}^T \frac{1}{m_j} \sum_{v=1}^H \{\mathbf{M}_v^{-1}\}_{r+1,q} \\
 & \times \sum_{k=1}^{L_v} \sum_{j_1=1}^{n'_{vk}} \sum_{j_2=1}^{n'_{vk}} (\tilde{\Omega})_{j_2 j_1} W'_{qvkj_2} \tilde{X}'_{rvkj} \tilde{\mathbf{Z}}'_{vkj_1} \mathbf{b}_{vk} + \frac{1}{T \bar{\tilde{X}}_r} \sum_{j=1}^T \frac{1}{m_j} \sum_{v=1}^H \{\mathbf{M}_v^{-1}\}_{r+1,1} \\
 & \times \sum_{k=1}^{L_v} \sum_{j_1=1}^{n'_{vk}} \sum_{j_2=1}^{n'_{vk}} (\tilde{\Omega})_{j_2 j_1} \tilde{X}'_{rvkj} A_{rvkj_1} + \dots + \frac{1}{T \bar{\tilde{X}}_r} \sum_{j=1}^T \frac{1}{m_j} \sum_{v=1}^H \{\mathbf{M}_v^{-1}\}_{r+1,p} \\
 & \times \sum_{k=1}^{L_v} \sum_{j_1=1}^{n'_{vk}} \sum_{j_2=1}^{n'_{vk}} (\tilde{\Omega})_{j_2 j_1} \tilde{X}'_{pvkj_2} \tilde{X}'_{rvkj} A_{rvkj_1} + \dots + \frac{1}{T \bar{\tilde{X}}_r} \sum_{j=1}^T \frac{1}{m_j} \sum_{v=1}^H \{\mathbf{M}_v^{-1}\}_{r+1,q} \\
 & \times \sum_{k=1}^{L_v} \sum_{j_1=1}^{n'_{vk}} \sum_{j_2=1}^{n'_{vk}} (\tilde{\Omega})_{j_2 j_1} W'_{qvkj_2} \tilde{X}'_{rvkj} A_{rvkj_1} + \frac{1}{T \bar{\tilde{X}}_r} \sum_{j=1}^T \frac{1}{m_j} \sum_{v=1}^H \{\mathbf{M}_v^{-1}\}_{r+1,1} \\
 & \times \sum_{k=1}^{L_v} \sum_{j_1=1}^{n'_{vk}} \sum_{j_2=1}^{n'_{vk}} (\tilde{\Omega})_{j_2 j_1} \tilde{X}'_{rvkj} \psi(U'_{vk}) e_{vkj_1} + \dots + \frac{1}{T \bar{\tilde{X}}_r} \sum_{j=1}^T \frac{1}{m_j} \sum_{v=1}^H \{\mathbf{M}_v^{-1}\}_{r+1,p}
 \end{aligned}$$

$$\begin{aligned} & \times \sum_{k=1}^{L_v} \sum_{j_1=1}^{n'_{vk}} \sum_{j_2=1}^{n'_{vk}} (\tilde{\Omega})_{j_2 j_1} \tilde{X}'_{pvkj_2} \tilde{X}'_{rvkj_1} \psi(U'_{vk}) e_{vkj_1} + \dots + \frac{1}{T \tilde{X}_r} \sum_{j=1}^T \frac{1}{m_j} \sum_{v=1}^H \{\mathbf{M}_v^{-1}\}_{r+1,q} \\ & \times \sum_{k=1}^{L_v} \sum_{j_1=1}^{n'_{vk}} \sum_{j_2=1}^{n'_{vk}} (\tilde{\Omega})_{j_2 j_1} W'_{qvkj_2} \tilde{X}'_{rvkj_1} \psi(U'_{vk}) e_{vkj_1} \\ & \equiv P_1 + (P_2 + \dots + P_{p+q+3}) + (P_{p+q+4}, \dots, P_{2p+2q+5}) + (P_{2p+2q+6} + \dots + P_{3p+3q+7}), \end{aligned}$$

where  $A_{rvkj_1} = \gamma_0 \Delta_{vk} + \sum_{r=1}^p \gamma_r \Delta'_{rvk} \tilde{X}'_{rvkj_1} + \sum_{r=1}^q \delta_r \Delta_{vk} W'_{rvkj_1}$  and  $\tilde{\mathbf{Z}}'_{vkj_1} = (1, \tilde{X}'_{1vkj_1}, \dots, \tilde{X}'_{p_1 vkj_1}, W'_{1vkj_1}, \dots, W'_{q_1 vkj_1})^T$ . Let us analyse each term separately. Term  $P_1$  becomes

$$\begin{aligned} P_1 &= \frac{\gamma_r}{T \tilde{X}_r} \sum_{j=1}^T \frac{1}{m_j} \sum_{v=1}^H \sum_{k \in I_{v_i}} \frac{\psi(U'_{vk})}{\phi_r(U'_{vk})} \tilde{X}'_{rvkj} + O(H^{-1}) \\ &= \frac{\gamma_r}{T \tilde{X}_r} \sum_{j=1}^T \frac{1}{m_j} \sum_{i \in I_j} \frac{\psi(U_i)}{\phi_r(U_i)} \tilde{X}'_{rij} + O(H^{-1}) \\ &= \gamma_r + O_p(m_0^{-1/2}) + O(H^{-1}). \end{aligned}$$

Note that  $E(P_2|U, \tilde{\chi}, L_v, \chi) = 0$  and since  $\{L_v^{-1} \mathbf{M}_v\}^{-1}$  is bounded uniformly in  $v$ ,

$$\begin{aligned} \text{var}(P_2|U, \tilde{\chi}, L_v, \chi) &= \frac{1}{T^2 \tilde{X}_r^2} \sum_{j=1}^T \frac{1}{m_j^2} \sum_{v=1}^H (L_v^{-1} \mathbf{M}_v^{-1})_{r+1,1} \\ & \times \frac{1}{L_v^2} \sum_{k=1}^{L_v} \sum_{j_1=1}^{n'_{vk}} \sum_{j_2=1}^{n'_{vk}} (\tilde{\Omega}_{vk})_{j_2 j_1}^2 \tilde{X}'_{rvkj_2} \sum_{s_1=1}^{p_1+q_1+1} \sum_{s_2=1}^{p_1+q_1+1} (\tilde{\mathbf{D}}_v)_{s_1 s_2} (\tilde{\mathbf{Z}}'_{vkj_1})_{s_1} (\tilde{\mathbf{Z}}'_{vkj_1})_{s_2} \\ & \approx \sum_{j=1}^T \frac{1}{m_j^2} \sum_{v=1}^H \frac{1}{L_v} \sum_{k=1}^{L_v} \tilde{X}'_{rvkj}{}^2 = o_p(m_0^{-1}). \end{aligned}$$

Thus,  $P_2 = o_p(m_0^{-1/2})$  and so is  $P_3, \dots, P_{p+q+3}$ . With similar expansions and using the fact that  $\Delta_{vk}, \Delta'_{rvk}$  are  $O(H^{-1})$ ,  $P_{p+q+4}, \dots, P_{2p+2q+5} \approx \sum_{j=1}^T m_j^{-1} \sum_{v=1}^H L_v^{-1} \sum_{k=1}^{L_v} O(H^{-1}) = o_p(H^{-1})$ . Similar considerations of conditional mean and variance can be used to show that  $P_{2p+2q+6}, \dots, P_{3p+3q+7} \approx o_p(m_0^{-1/2})$ . Thus, it follows that  $\hat{\gamma}_r = \gamma_r + O_p(m_0^{-1/2}) + O(H^{-1})$  for  $0 \leq r \leq p$ . Showing that  $\hat{\delta}_r = \delta_r + O_p(m_0^{-1/2}) + O(H^{-1})$  follows closely the derivation above and therefore is omitted here.

### Appendix 3: Bootstrap estimate of SEs

To obtain SE estimates for the data analysis of Section 5, the wild bootstrap is used as it is more suitable for heteroskedastic cases that commonly arise in nonparametric regression. For the CAR model, the point estimators for the fixed effects are obtained by averaging heteroskedastic estimators coming from each bin given by Equation (11). Bin specific estimators are indeed heteroskedastic since their variance depends on  $U$ . The wild bootstrap algorithm implemented is as follows.

- (1) Within each bin  $v$ , bin specific estimates are obtained based on the original data and the resulting residual vectors are  $\hat{\epsilon}'_k$  for subject  $k$  in bin  $v$  ( $k = 1, \dots, L_v$ ).

- (2) We multiply each residual belonging to a specific subject and repetition (*i.e.* each component of  $\hat{\epsilon}'_k$ ) by a random variable sampled from the two-point distribution attaching masses  $(\sqrt{5} + 1)/2\sqrt{5}$  and  $(\sqrt{5} - 1)/2\sqrt{5}$  to the points  $-(\sqrt{5} - 1)/2$  and  $(\sqrt{5} + 1)/2$  to obtain the wild bootstrap residual vectors  $\hat{\epsilon}'_{k^*}$ ,  $k = 1, \dots, L_v$ . (These wild bootstrap residuals approximate the variance and skewness of the residuals for each subject.)
- (3) The new responses in bin  $v$ ,  $\tilde{Y}_{vk^*}$ , are obtained from the wild bootstrap residuals; and bootstrap bin-specific estimates (11) are obtained based on the bootstrap data.

We repeat the above procedure to obtain  $B = 400$  bootstrap samples to estimate the SE. (SE estimates in this case stabilise after  $B = 100$  bootstrap samples.) ■