Stochastic Functional Estimates in Longitudinal Models with Interval-Censored Anchoring Events

Chenghao Chu[1], Ying Zhang[2], Wanzhu Tu[1]

[1] Department of Biostatistics, Indianan University, Indianapolis, USA

[2] Department of Biostatistics, University of Nebraska Medical Center, Omaha, USA

Abstract

Timelines of longitudinal studies are often anchored by specific events. In the absence of fully observed the anchoring event times, the study timeline becomes undefined, and the traditional longitudinal analysis loses its temporal reference. In this paper, we considered an analytical situation where the anchoring events are interval-censored. We demonstrated that by expressing the regression parameter estimators as stochastic functionals of a plug-in estimate of the unknown anchoring event time distribution, the standard longitudinal models could be extended to accommodate the situation of less well-defined timelines. We showed that for a broad class of longitudinal models, the functional parameter estimates are consistent and asymptotically normally distributed with a $\sqrt{n}$ convergence rate under mild regularity conditions. Applying the developed theory to linear mixed-effects models, we further proposed a hybrid computational procedure that combines the strengths of the Fisher's scoring method and the expectation-expectation (EM) algorithm, for model parameter estimation. We conducted a simulation study to validate the

This is the author's manuscript of the article published in final edited form as:

Chu, C., Zhang, Y., & Tu, W. (2020). Stochastic functional estimates in longitudinal models with interval-censored anchoring events. Scandinavian Journal of Statistics, 47(3), 638–661. https://doi.org/10.1111/sjos.12419

asymptotic properties and to assess the finite sample performance of the proposed method. A real data analysis was used to illustrate the proposed method. The method fills in a gap in the existing longitudinal analysis methodology for data with less well defined timelines.

Keywords: Empirical process, Interval censoring, Longitudinal data, Nonparametrics, Pseudo-likelihood.

1 Introduction

An implicit assumption for longitudinal studies is that they have well-defined timelines. In clinical research, study timelines usually start at pre-specified events. For example, timelines of randomized trials start at the initiation of treatment therapies. In observational studies of cancer, time of oncogenesis, when cancer cells first emerge in organ tissues, marks the beginning of a clinical disease. In longitudinal studies, events such as therapeutic initiation and oncogenesis define the starting points of study timelines. Herein, we refer to these events as the anchoring events. Although the anchoring events themselves are not the outcomes of interest, without the anchoring event times, all outcome measures lose their temporal reference. For example, one would not be able to assess the rate of tumor growth without knowing the time of oncogenesis.

In many studies, however, the times at which the anchoring events occur are not directly observed and are subject to interval-censoring. For example, in studies of pubertal growth, researchers are interested in assessing the rates of weight change immediately before and after the pubertal growth spurt (PGS), the time point at which the bodily height increase reaches its maximum velocity (Tanner and Whitehouse, 1976). The exact time of PGS in an individual, however, is unavailable – it is only known to occur within a time interval. This leads us to an interval censoring situation: For a given subject, we have a random observation interval $(L, R]$, which covers the anchoring-event time $T \in (L, R]$.

In the pubertal growth study, had we known the exact PGS times (T) in the study participants, we would be able to estimate the rates of weight change around the PGS, using a piecewise longitudinal regression model. But in the absence of T, the standard models are no longer applicable. To remedy the situation, researchers proposed a number of *ad hoc* methods to impute T, including mid-point and model-based imputation methods (Shankar et al., 2005; Tu et al., 2009). However, the uncertainty associated with the imputation process not only inflates the standard errors in estimating the model parameters, but may also lead to biased estimation. Alternatively, Robinson et al. (2010) and van den Hout et al. (2013) developed joint modeling approaches by making parametric assumptions for T. However misspecified parametric distributions often lead to biased estimation results.

In this paper, we propose a new semiparametric approach to estimate the longitudinal model parameters in the situation with interval-censored anchoring-event times without making a parametric assumption for T. Writing the observed data as $\boldsymbol{D} = (\mathbf{Y}, \mathbf{W}, L, R)$, where $\mathbf{Y}$ and $\mathbf{W}$ are respectively the vectors of longitudinally observed outcome and covariates, we consider the following model

$$\boldsymbol{Y}\big|(\boldsymbol{W}, L, R, T) \sim \phi(\boldsymbol{Y}\big|\boldsymbol{W}, L, R, T; \boldsymbol{\theta}),$$

where $\phi(\cdot)$ is a density function with parameters $\boldsymbol{\theta}$ when the anchoring-event time T is known. The focus of the current research is to estimate $\boldsymbol{\theta}$ when T is known inside the interval $(L, R]$ but subject to an unknown cumulation distribution function (CDF) F.

Since the likelihood for the observed data $\boldsymbol{D}$ can be expressed as a functional of F of the anchoring-event time T and the model parameters $\boldsymbol{\theta}$, we propose a two-stage estimation method: In the first step, we estimate F using the nonparametric maximum likelihood method (NPMLE) for interval-censored data (Groeneboom and Wellner, 1992); in the second step, we plug in the NPMLE of the CDF, $\hat{F}_n$, into the likelihood of the observed data to obtain model parameter estimates $\hat{\boldsymbol{\theta}}_n$. This two-stage estimation method is essentially a semiparametric maximum pseudolikelihood estimation method (SPMPLE), and the estimates of the model parameters can be viewed as stochastic functionals of $\hat{F}_n$, i.e., $\hat{\boldsymbol{\theta}}_n = \mathbb{Q}_n(\hat{F}_n)$. In general, the asymptotic normality is not automatically assured for this type of functional estimators. In fact, previous research showed that for *nonparamet-*

ric models with *interval-censored* data, functional estimates tended to have convergence rates that are slower than $n^{1/2}$ (Zhang et al., 2016), but for linear mixed-effects models with *right-censored* data, functional estimates could still achieve an asymptotic normality (Kong et al., 2017).

In this paper, we examined the asymptotic behavior of stochastic functional estimates $\hat{\boldsymbol{\theta}}_n = \mathbb{Q}_n(\hat{F}_n)$. We showed that for a broad class of longitudinal models, asymptotic normality could indeed be achieved under fairly mild regularity conditions. To the best of our knowledge, this is the first systematic study of stochastic function estimates in semiparametric longitudinal models when the anchoring-event times are interval censored. In pursuing the research, we have put forward a theoretical foundation upon which the traditional longitudinal models can be extended to situations where the study timelines are less well defined.

The rest of paper is organized as follows. In Section 2, we propose the functional estimates of the model parameters from semiparametric pseudolikelihood analysis for longitudinal data with interval-censored anchoring-event times and study their asymptotic properties. In Section 3, we present an adaptive numerical algorithm for parameter estimation in the context of the linear mixed-effects models. Simulation studies and real data analysis are presented in Sections 4 and 5, respectively. We conclude the paper in Section 6 with a discussion of the proposed method. Technical details, including a more generally applicable asymptotic theorem for the semiparametric Z-estimators that can be used to prove the theorem given in Section 2, are presented in the Appendix and online supplementary material.

2 A two-stage semiparametric pseudolikelihood method

2.1 The model and parameter estimation

We consider a generic setting where the anchoring events that define the study timeline in a longitudinal study are interval-censored. From each subject, we observe a response vector $\boldsymbol{Y}$, a covariate vector $\boldsymbol{W}$ from longitudinally repeated measurements, and a censoring interval $(L, R]$ that contains the unobserved anchoring-event time T, i.e., $L < T \leq R$.

Given $\boldsymbol{W}, L, R$ and the unobserved T, the conditional density of $\boldsymbol{Y}$ can be modeled as

$$\boldsymbol{Y} \big| (\boldsymbol{W}, L, R, T) \sim \phi(\boldsymbol{Y} \big| \boldsymbol{W}, L, R, T; \boldsymbol{\theta}), \tag{1}$$

for a known density function ϕ with parameter $\boldsymbol{\theta}$, whose true value $\boldsymbol{\theta}_0$ is of the interest. Here, the conditional density function ϕ can be any continuous or discrete distributions. For the main theoretical result to hold, we only require ϕ to satisfy a set of regularity conditions (see Section 2.2). In most of the modeling situations, ϕ is assumed to be a member of the exponential family of distributions, though the proposed modeling structure does allow other distributions. When ϕ is the density function of a normal distribution, the model can be written in the familiar form of linear mixed-effects models, which we shall examine as a special case with greater details in Section 3.

In traditional longitudinal models, the parameter $\boldsymbol{\theta}$ is defined only when $\boldsymbol{W}, L, R$, and T are fully observed; when T is not observed, the true value $\boldsymbol{\theta}_0$ cannot be estimated from Model (1). To estimate $\boldsymbol{\theta}_0$ in the absence of T, we focus on the density function of $\boldsymbol{Y}$ given the observed data $\boldsymbol{W}$, L, R, and $L < T \leq R$, i.e.,

$$\boldsymbol{Y} \big| (\boldsymbol{W}, L, R, L < T \leq R) \sim \int \phi(\boldsymbol{Y} \big| \boldsymbol{W}, L, R, t; \boldsymbol{\theta}) \; dF_{T|(\boldsymbol{W}, L, R, L < T \leq R; \boldsymbol{\theta})}(t),$$

where $F_{T|(\boldsymbol{W},L,R,L<T\leq R;\boldsymbol{\theta})}$ is the conditional CDF of T given the observed covariates $\boldsymbol{W}$, L, R, $L < T \leq R$, and parameter $\boldsymbol{\theta}$.

We assume that T is conditionally independent of $\boldsymbol{W}$, given L and R; $(L, R]$ is an independent censoring interval and its distribution is not informative to $\boldsymbol{\theta}$. These assumptions lead to $F_{T|(\boldsymbol{W},L,R,L<T\leq R;\boldsymbol{\theta})}(t) = 1(L < t \leq R)\big(F_0(t) - F_0(L)\big)\Big/\big(F_0(R) - F_0(L)\big)$, where F_0 denotes the true but often unknown CDF of the anchoring-event time T. Then we have

$$\begin{aligned} \boldsymbol{Y}\big|(\boldsymbol{W}, L, R, L<T\leq R) \;\; &\sim \;\; \int_L^R \frac{\phi(\boldsymbol{Y}\big|\boldsymbol{W}, L, R, t; \boldsymbol{\theta})dF_0(t)}{F_0(R) - F_0(L)} \\ &\propto \;\; \int_L^R \phi(\boldsymbol{Y}\big|\boldsymbol{W}, L, R, t; \boldsymbol{\theta})dF_0(t). \end{aligned} \tag{2}$$

Since F_0 is also unknown, we estimate it with a semiparametric pseudolikelihood approach. Given a random sample $\{\boldsymbol{D}_i = (\boldsymbol{Y}_i, \boldsymbol{W}_i, L_i, R_i) : i = 1, \cdots, n\}$, i.e., in the absence of known anchoring event time T_i, $i = 1, \cdots, n$, we first estimate F_0 through the NPMLE method with the interval-censored data $\{(L_i, R_i] : i = 1, \cdots, n\}$. That is, we estimate F_0 by the step function $\hat{F}_n$ that has jumps only at the points $\{L_1, R_1, L_2, R_2, \cdots\}$ and maximizes the nonparametric likelihood function

$$\mathcal{L}(F) = \prod_{i=1}^{n} \big(F(R_i) - F(L_i)\big),$$

over the class $\mathcal{F}$ of one-dimensional CDFs. According to Groeneboom and Wellner (1992), the NPMLE $\hat{F}_n$ converges to F_0 in the supremum norm with a convergence rate of $n^{1/3}$. Zhang and Jamshidian (2004) provided an overview of the efficient and robust algorithms for calculating $\hat{F}_n$.

With the estimated CDF $\hat{F}_n$, we then estimate $\boldsymbol{\theta}_0$ by $\hat{\boldsymbol{\theta}}_n$, the maximizer of the pseudolikelihood

$$\mathcal{L}_n(\boldsymbol{\theta}, \hat{F}_n) = \sum_{i=1}^{n} \log \left(\int_{L_i}^{R_i} \phi(\boldsymbol{Y}_i \big| \boldsymbol{W}_i, L_i, R_i, t; \boldsymbol{\theta}) d\hat{F}_n(t) \right). \tag{3}$$

The algorithmic efficiency for computing the parameter estimates $\hat{\boldsymbol{\theta}}_n$ from Model (3) depends on the specific distribution ϕ. In Section 3, we provide a hybrid algorithm in the case of Gaussian linear mixed-effects models.

2.2 The asymptotic properties of $\hat{\boldsymbol{\theta}}_n$

The estimator $\hat{\boldsymbol{\theta}}_n$ is a stochastic functional of the NPMLE $\hat{F}_n$. Consider the following stochastic functional $\mathbb{Q}_n$, which maps a CDF $F \in \mathcal{F}$ to

$$\begin{aligned} \mathbb{Q}_n(F) &= \arg\max_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \left(\mathcal{L}_n(\boldsymbol{\theta}, F)\right) \\ &= \arg\max_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \left\{ \sum_{i=1}^{n} \log \left(\int_{L_i}^{R_i} \phi(\boldsymbol{Y}_i \big| \boldsymbol{W}_i, L_i, R_i, t; \boldsymbol{\theta}) dF(t) \right) \right\}. \end{aligned}$$

Here $\hat{\boldsymbol{\theta}}_n$ is the value of $\mathbb{Q}_n$ assessed at the estimated distribution $\hat{F}_n$, i.e., $\hat{\boldsymbol{\theta}}_n = \mathbb{Q}_n(\hat{F}_n)$. If F_0 is known, the true parameter $\boldsymbol{\theta}_0$ can be estimated from Model (2). Let $\tilde{\boldsymbol{\theta}}_n$ be the estimate under F_0, i.e., $\tilde{\boldsymbol{\theta}}_n = \mathbb{Q}_n(F_0)$. It follows from the standard maximum likelihood theory that $\tilde{\boldsymbol{\theta}}_n$ is a consistent and asymptotic normal estimator of $\boldsymbol{\theta}_0$.

When F_0 is unknown, we approach the problem as follows: Suppose $\hat{F}_n$ is a consistent estimator of F_0, and $\mathbb{Q}_n$ is a smooth functional, then $\hat{\boldsymbol{\theta}}_n = \mathbb{Q}_n(\hat{F}_n)$ is potentially asymptotically equivalent to $\tilde{\boldsymbol{\theta}}_n = \mathbb{Q}_n(F_0)$, so it is possibly a consistent estimator of $\boldsymbol{\theta}_0$. While the idea behind the approach is simple, a rigorous study of the asymptotic properties of $\hat{\boldsymbol{\theta}}_n$ is much involved, because of the extra variability associated with the estimation of $\hat{F}_n$. Here, the $n^{1/3}$ rate of convergence of $\hat{F}_n$ (Groeneboom and Wellner, 1992), also complicates the study of the asymptotic distribution of $\hat{\boldsymbol{\theta}}_n$.

To study the asymptotic properties of $\hat{\boldsymbol{\theta}}_n$, we use techniques from the empirical process theory. The following regularity conditions are sufficient to justify the forthcoming theorem on the asymptotic properties of $\hat{\boldsymbol{\theta}}_n$.

1. Regularity conditions on the interval-censored data

   F1: There exist constants $\tau_1 < \tau_2 < \infty$ such that the support of the density function of the anchoring-event time T is contained in $[\tau_1, \tau_2]$.

   F2: The anchoring event time T is conditionally independent of $\boldsymbol{W}$, given L and R. The censoring interval $(L, R]$ is independent of T.

   F3: The support of F_0 is included in the union of the supports of the CDF of L and the CDF of R. And F_0 uninformative to $\boldsymbol{\theta}$.

   F4: There exists a constant c such that $P\big(F_0(R) - F_0(L) > c\big) = 1$.

   F5: The sum of density functions of L and R, $f_L + f_R$, is strictly positive in $[\tau_1, \tau_2]$.

   F6: The joint density function of (L, T, R), is twice differentiable in $[\tau_1, \tau_2]$. In particular, f_L and f_R are differentiable and uniformly bounded in $[\tau_1, \tau_2]$.

   F7: The density function of T is twice differentiable.

2. Regularity conditions on the longitudinal model when F_0 is known.

   Let $\nabla^k_{\boldsymbol{\theta}}$ denote the differential operator of all kth order partial derivatives with respect to the vector variable $\boldsymbol{\theta}$. Let $d = \dim(\boldsymbol{\theta})$ be the dimension of $\boldsymbol{\theta}$. The model parameter space $\boldsymbol{\Theta}$ is a subset of $\mathbb{R}^d$ such that:

   M1: $\int_L^R \phi(\boldsymbol{Y} \big| \boldsymbol{W}, L, R, t; \boldsymbol{\theta}_1) dF_0(t) \neq \int_L^R \phi(\boldsymbol{Y} \big| \boldsymbol{W}, L, R, t; \boldsymbol{\theta}_2) dF_0(t)$ for any $\boldsymbol{\theta}_1 \neq \boldsymbol{\theta}_2$.

   M2: The true parameter $\boldsymbol{\theta}_0$ is an inner point of $\boldsymbol{\Theta}$.

   M3: The support of $\int_L^R \phi(\boldsymbol{Y} \big| \boldsymbol{W}, L, R, t; \boldsymbol{\theta}) dF_0(t)$ does not dependent on $\boldsymbol{\theta} \in \boldsymbol{\Theta}$.

M4: The conditional density function $\phi = \phi(\boldsymbol{Y}\big|\boldsymbol{W}, L, R, T; \boldsymbol{\theta})$ is continuous. The third order partial derivative $\nabla^3_{\boldsymbol{\theta}}(\phi)$ exists and is continuous. Both ϕ and its partial derivative function $\boldsymbol{u} = \nabla_{\boldsymbol{\theta}}(\phi)$ have continuous partial derivatives with respect to T.

M5: Let P be the probability measure associated with $(\boldsymbol{Y}, \boldsymbol{W}, L, R)$, then

$$\nabla_{\boldsymbol{\theta}}\left[\mathrm{P}\left[\log\left(\int \phi dF_0\right)\right]\right] = \mathrm{P}\left[\nabla_{\boldsymbol{\theta}}\left[\log\left(\int \phi dF_0\right)\right]\right]$$

$$\nabla^2_{\boldsymbol{\theta}}\left[\mathrm{P}\left[\log\left(\int \phi dF_0\right)\right]\right] = \mathrm{P}\left[\nabla^2_{\boldsymbol{\theta}}\left[\log\left(\int \phi dF_0\right)\right]\right]$$

3. The random vector $(\boldsymbol{Y}, \boldsymbol{W}, L, R)$ is bounded with probability 1.

Theorem 2.1. *Under the stated regularity conditions, the model estimate $\hat{\boldsymbol{\theta}}_n$ of Model (3) is consistent and asymptotically normally distributed. More precisely, let* P *and* $\mathbb{P}_n$ *be the respective probability measure and empirical probability measure associated with* $(\boldsymbol{Y}, \boldsymbol{W}, L, R)$*,* $\boldsymbol{\theta}_0$ *the true parameter,* $\boldsymbol{u} = \nabla_{\boldsymbol{\theta}}(\phi)$ *the gradient of* ϕ *with respect to* $\boldsymbol{\theta}$*,* $\boldsymbol{U}_{\boldsymbol{\theta}_0, F_0} = \left[\int_L^R \boldsymbol{u} dF_0 \Big/ \int_L^R \phi dF_0\right]_{\boldsymbol{\theta}=\boldsymbol{\theta}_0}$ *the score function and* $\boldsymbol{A} = \left[\mathrm{P}\big(\nabla_{\boldsymbol{\theta}}(\boldsymbol{U}_{\boldsymbol{\theta}, F_0})\big)\right]_{\boldsymbol{\theta}=\boldsymbol{\theta}_0}$ *the Hessian matrix when* F_0 *is known, and* $\boldsymbol{\Phi} = \boldsymbol{\Phi}(L, R)$ *the multidimensional function that has mean zero and uniquely solves the following integral equation system*

$$\int_{L<t\le R} \boldsymbol{\Phi}(L, R) d\mathrm{P} = \int_{S_t} \left[\left(\int_L^R \phi dF_0\right)^{-2} \left(\boldsymbol{u}\int_L^R \phi dF_0 - \phi \int_L^R \boldsymbol{u} dF_0\right)\right]_{\boldsymbol{\theta}=\boldsymbol{\theta}_0} d\mathrm{P}_t,$$

where S_t *denotes the domain of* $(\boldsymbol{Y}, \boldsymbol{W}, L, R)$ *given value* $T = t$*, and* P_t *the conditional measure of* P *when restricted to* S_t*. Then*

$$\sqrt{n}\big(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0\big) = -\boldsymbol{A}^{-1} \cdot \sqrt{n}\mathbb{P}_n\big(\boldsymbol{U}_{\boldsymbol{\theta}_0, F_0} + \boldsymbol{\Phi}\big) + \boldsymbol{o}_p(1). \tag{4}$$

In particular, $\sqrt{n}(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0) \xrightarrow{P} \boldsymbol{N}(\mathbf{0}, \boldsymbol{\Sigma})$*, with the asymptotic variance* $\boldsymbol{\Sigma}$ *given by*

$$\boldsymbol{\Sigma} \;=\; \boldsymbol{A}^{-1} + \boldsymbol{A}^{-1}\mathrm{P}\left(\boldsymbol{U}_{\boldsymbol{\theta}_0,F_0}\boldsymbol{\Phi}^t + \boldsymbol{\Phi}\boldsymbol{U}^t_{\boldsymbol{\theta}_0,F_0} + \boldsymbol{\Phi}^{\otimes 2}\right)\boldsymbol{A}^{-1},$$

where $\boldsymbol{M}^{\otimes 2}$ *denotes* $\boldsymbol{M}\boldsymbol{M}^t$ *for a matrix* $\boldsymbol{M}$*.*

We also establish a more general asymptotic normality theorem for semiparametric Z-estimators in the Appendix. To prove Theorem 2.1, one only has to verify the conditions for the general theorem.

A few remarks are in order for Theorem 2.1.

Remark 1. Parallel to the classical result of Gong and Samaniego (1981) on parametric maximum pseudo-likelihood estimation, Theorem 2.1 shows that the asymptotic variance $\boldsymbol{\Sigma}$ for the semiparametric maximum pseudolikelihood estimation can be decomposed into two components, with the first component being the asymptotic variance in estimating the model parameters when the nuisance infinite dimensional parameter F_0 is known, and the other component the extra variability associated with the estimation of F_0.

Remark 2. The regularity conditions are mild and they pose no extra restrictions in most applications. The first set of conditions are usually assumed in order to guarantee that the values of smooth functionals on estimated CDF of the interval-censored event times have good properties (Geskus and Groeneboom, 1999). The second set of conditions are the usual regularity conditions assumed in the maximum likelihood theory. The third condition is generally satisfied in practice. It means that, as long as the data do not contain substantial amount of extreme observations, the parameter estimate is asymptotically normally distributed as described in Theorem 2.1.

Remark 3. The asymptotic variance $\boldsymbol{\Sigma}$ has a complicated expression. Equation 4 in Theorem 2.1 showed that the proposed estimator $\hat{\boldsymbol{\theta}}_n$ is asymptotically linear and normally distributed. By Mammen (2012, Theorem 1, Chapter 1), $\mathbf{\Sigma}$ can be consistently estimated from the following bootstrap resampling method. For a data set containing n subjects, we draw bootstrap samples that contain n subjects, each of which is drawn from the original data set with equal weight and with replacement. The parameter estimate is obtained using the bootstrap samples. We independently repeat this procedure to obtain B estimates $\hat{\boldsymbol{\theta}}_n^{(b)}, b = 1, \cdots, B$, where B is a pre-specified number. The sample variance matrix of $\left\{\hat{\boldsymbol{\theta}}_n^{(b)} : b = 1, \ldots, B\right\}$ is a consistent estimate of $\mathbf{\Sigma}$.

3 Linear mixed-effects models: A case study

As a more concrete example, we used the general method described in Section 2 to study parameter estimation in linear mixed-effects models with interval-censored anchoring events.

3.1 Linear mixed-effects models with interval-censored anchoring events

As before, we let $\boldsymbol{Y}$ be the longitudinal outcome, $\boldsymbol{W}$ the covariates, and $(L, R]$ the time interval that brackets the unobserved anchoring event time T. We consider a linear mixed-effects model as follows:

$$\boldsymbol{Y}\big|(\boldsymbol{W}, L, R, T, \boldsymbol{r}) \sim \boldsymbol{N}\left(\boldsymbol{X\beta} + \boldsymbol{Zr}, \sigma^2\boldsymbol{I}\right), \quad \boldsymbol{r} \sim \boldsymbol{N}(\mathbf{0}, \boldsymbol{G}),$$

where $\boldsymbol{G}$ is the fixed but unknown covariance matrix of the random effects $\boldsymbol{r}$, $\boldsymbol{X} = \boldsymbol{X}(\boldsymbol{W}, L, R, T)$ and $\boldsymbol{Z} = \boldsymbol{Z}(\boldsymbol{W}, L, R, T)$ are respectively the design matrices for the fixed and random effects. Entries of $\boldsymbol{X}$ and $\boldsymbol{Z}$ are functions of $\boldsymbol{W}$, L, R and T. The parameter vector of interest is $\boldsymbol{\theta} = (\boldsymbol{\beta}, \sigma^2, \boldsymbol{G})$.

To estimate the true parameter $\boldsymbol{\theta}_0$ using the semiparametric pseudolikelihood method proposed in Section 2.1, we first obtain the NPMLE $\hat{F}_n$ of F_0 by using the interval-censored data $(L, R]$. Assuming that $\hat{F}_n$ has jumps p_j at time s_j, $j = 1, \cdots, k$, we can write the pseudolikelihood of $\boldsymbol{Y}\big|(\boldsymbol{W}, L, R, L < T \le R)$ (2) as

$$\sum_{L<s_j\le R} \frac{p_j}{\sum\limits_{L<s_k\le R} p_k} |\boldsymbol{V}(s_j)|^{-1/2} \exp\left(-\frac{1}{2}\big(\boldsymbol{Y}-\boldsymbol{X}(s_j)\boldsymbol{\beta}\big)^t \boldsymbol{V}(s_j)^{-1}\big(\boldsymbol{Y}-\boldsymbol{X}(s_j)\boldsymbol{\beta}\big)\right),$$

where $\boldsymbol{X}(s_j) = \boldsymbol{X}|_{T=s_j}$, $\boldsymbol{Z}(s_j) = \boldsymbol{Z}|_{T=s_j}$ and $\boldsymbol{V}(s_j) = \sigma^2\boldsymbol{I} + \boldsymbol{Z}(s_j)\boldsymbol{G}\boldsymbol{Z}(s_j)^t$.

Given a random sample $\Big\{\boldsymbol{D}_i = (\boldsymbol{Y}_i, \boldsymbol{W}_i, L_i, R_i) : i = 1, \cdots, n\Big\}$, for the i-th subject and index j such that $L_i < s_j \le R_i$, we write $\boldsymbol{X}_{ij} = \boldsymbol{X}(\boldsymbol{W}_i, L_i, R_i, s_j)$, $\boldsymbol{Z}_{ij} = \boldsymbol{Z}_i(\boldsymbol{W}_i, L_i, R_i, s_j)$, $\boldsymbol{Vec}_{ij} = \boldsymbol{Y}_i - \boldsymbol{X}_{ij}\boldsymbol{\beta}$, $\boldsymbol{V}_{ij} = \sigma^2\boldsymbol{I} + \boldsymbol{Z}_{ij}\boldsymbol{G}\boldsymbol{Z}_{ij}^t$, and $p_{ij} = p_j \Big/ \sum\limits_{L_i<s_k\le R_i} p_k$.

Under this abbreviated notation, we write the log pseudolikelihood for the observed data as

$$\mathcal{L}_n^{pl}(\boldsymbol{\theta}) \propto \sum_{i=1}^{n} \log\left[\sum_{L_i<s_j\le R_i} p_{ij}|\boldsymbol{V}_{ij}|^{-1/2} \exp\left(-\frac{1}{2}\boldsymbol{Vec}_{ij}^t \boldsymbol{V}_{ij}^{-1} \boldsymbol{Vec}_{ij}\right)\right]. \tag{5}$$

The parameter estimate $\hat{\boldsymbol{\theta}}_n$ is the maximizer of the above function $\mathcal{L}_n^{pl}(\boldsymbol{\theta})$.

3.2 Computation

Although the proposed semiparametric pseudolikelihood estimation has reduced computation burden than a joint estimation of F and $\boldsymbol{\theta}$, maximizing the function $\mathcal{L}_n^{pl}(\boldsymbol{\theta})$ is still not an easy task, because it has a complicated structure. The commonly used computation algorithms for fitting traditional mixed-effects models, namely the profile likelihood method and the restricted maximum likelihood method, do not seem to be easily applicable to maximize $\mathcal{L}_n^{pl}(\boldsymbol{\theta})$. Therefore, we propose a hybrid computation algorithm to maximize $\mathcal{L}_n^{pl}(\boldsymbol{\theta})$, combining the Fisher-Scoring (FS) algorithm with an EM-algorithm. The hybrid algorithm is more robust than the FS-algorithm, and converges faster than the EM-algorithm.

The following notation is defined for description of the computation algorithm. For a positive definite matrix $\boldsymbol{G}$, there exists a unique lower triangular matrix $\boldsymbol{A}$ with positive diagonal entries such that $\boldsymbol{G} = \boldsymbol{A}\boldsymbol{A}^t$. We reparameterize $\boldsymbol{G}$ with $\boldsymbol{A}$ for the computational advantage that the boundary condition is easier to check, because $\boldsymbol{G}$ is positive definite if and only if $\boldsymbol{A}$ has positive diagonal entries. To simplify the notation, for any parameter value $\boldsymbol{\theta} = (\boldsymbol{\beta}, \sigma^2, \boldsymbol{A})$, interval $(L_i, R_i]$, and index j such that $L_i < s_j \leq R_i$, we let $\tilde{p}_{ij}(\boldsymbol{\theta})$ denote the quantity $\tilde{p}_{ij}(\boldsymbol{\theta}) = p_{ij}|\boldsymbol{V}_{ij}|^{-1/2} \exp\left(-\frac{1}{2}\boldsymbol{Vec}_{ij}^t \cdot \boldsymbol{V}_{ij}^{-1} \cdot \boldsymbol{Vec}_{ij}\right)$, and let $p_{ij}(\boldsymbol{\theta})$ denote the quantity $p_{ij}(\boldsymbol{\theta}) = \tilde{p}_{ij}(\boldsymbol{\theta}) \Big/ \sum_{L_i < s_k \leq R_i} \tilde{p}_{ik}(\boldsymbol{\theta})$. Let $\boldsymbol{X}_i$ and $\boldsymbol{Z}_i$ denote the functions of T, defined as $\boldsymbol{X}_i(T) = \boldsymbol{X}(\boldsymbol{W}_i, L_i, R_i, T)$ and $\boldsymbol{Z}_i(T) = \boldsymbol{Z}(\boldsymbol{W}_i, L_i, R_i, T)$.

The score function $\boldsymbol{U}(\boldsymbol{\theta})$ can be expressed as $\boldsymbol{U}(\boldsymbol{\theta}) = \frac{\partial}{\partial \boldsymbol{\theta}}\left(\mathcal{L}_n^{pl}(\boldsymbol{\theta})\right) = \sum_{i=1}^{n} \boldsymbol{U}_i(\boldsymbol{\theta})$, where $\boldsymbol{U}_i(\boldsymbol{\theta})$ is the score function computed from the i-th subject. Using vector calculus (Wand, 2002), we compute the component functions of $\boldsymbol{U}_i(\boldsymbol{\theta})$ as follows:

$$
\begin{aligned}
\boldsymbol{U}_i(\boldsymbol{\beta}) &= \sum_{L_i < s_j \leq R_i} \boldsymbol{X}_{ij}^t \boldsymbol{V}_{ij}^{-1} \boldsymbol{Vec}_{ij} \cdot p_{ij}(\boldsymbol{\theta}), \\
U_i(\sigma^2) &= \sum_{L_i < s_j \leq R_i} \tfrac{1}{2} Tr\left((\boldsymbol{V}_{ij}^{-1}\boldsymbol{Vec}_{ij})^{\otimes 2} - \boldsymbol{V}_{ij}^{-1}\right) \cdot p_{ij}(\boldsymbol{\theta}), \\
U_i(a_{pq}) &= \sum_{L_i < s_j \leq R_i} \boldsymbol{E}_p^t \boldsymbol{Z}_{ij}^t \left((\boldsymbol{V}_{ij}^{-1}\boldsymbol{Vec}_{ij})^{\otimes 2} - \boldsymbol{V}_{ij}^{-1}\right) \boldsymbol{Z}_{ij} \boldsymbol{A} \boldsymbol{E}_q \cdot p_{ij}(\boldsymbol{\theta}),
\end{aligned}
$$

where a_{pq} is the (p, q)-th entry of $\boldsymbol{A}$ and $p \geq q$, and $\boldsymbol{E}_k$ is the column vector with all entries being zero and the kth entry being 1. With the above formulae, we use the FS-algorithm with a step-halving line search strategy.

To implement the FS-algorithm, a good initial value is essential. To ensure a good initial value, we start with an EM-algorithm. The derivation of the E-step and M-step is lengthy and algebraically heavy. We provide the essential details in the online supplementary material. Given the current parameter estimate $\boldsymbol{\theta}^{(k)} = \big(\boldsymbol{\beta}(k), \sigma^2(k), \boldsymbol{G}(k)\big)$, the EM-algorithm computes the next estimate $\boldsymbol{\theta}^{(k+1)} = \big(\boldsymbol{\beta}(k+1), \sigma^2(k+1), \boldsymbol{G}(k+1)\big)$ as

$$\begin{aligned}
&\boldsymbol{\beta}(k+1) = \boldsymbol{\beta}(k) + \sigma^2(k) \cdot \arg\min_{\boldsymbol{\Delta}} \sum_{i=1}^{n} \boldsymbol{B}_i(k) \\
&\sigma^2(k+1) = \sigma^2(k) + \sigma^4(k) \left(\sum_{i=1}^{n} q_i\right)^{-1} \left\{\min_{\boldsymbol{\Delta}} \sum_{i=1}^{n} \boldsymbol{B}_i(k) - \sum_{i=1}^{n} E_{T_i(k)}\left[Tr\left(\boldsymbol{V}_i(k)^{-1}\right)\right]\right\} \\
&\boldsymbol{G}(k+1) = \boldsymbol{G}(k) + \tfrac{1}{n}\boldsymbol{G}(k) \left(\sum_{i=1}^{n} E_{T_i(k)}\left[\boldsymbol{Z}_i^t \left(\left(\boldsymbol{V}_i(k)^{-1}\boldsymbol{Vec}_i(k)\right)^{\otimes 2} - \boldsymbol{V}_i(k)^{-1}\right)\boldsymbol{Z}_i\right]\right)\boldsymbol{G}(k)
\end{aligned}$$

where $\boldsymbol{B}_i(k) = E_{T_i(k)}\left[\left(\boldsymbol{V}_i(k)^{-1}\boldsymbol{Vec}_i(k) - \boldsymbol{X}_i\boldsymbol{\Delta}\right)^t\right]^{\otimes 2}$ is a function of $\boldsymbol{\Delta}$; $E_{T_i(k)}$ denotes the expectation with respect to the random variable $T_i(k)$, which has density $p_{ij}(\boldsymbol{\theta}^{(k)})$ at $s_j \in (L_i, R_i]$ and 0 elsewhere; $\boldsymbol{V}_i(k) = \sigma^2(k)\boldsymbol{I} + \boldsymbol{Z}_i\boldsymbol{G}(k)\boldsymbol{Z}_i^t$ and $\boldsymbol{Vec}_i(k) = \boldsymbol{Y}_i - \boldsymbol{X}_i\boldsymbol{\beta}(k)$ are functions of $T = T_i(k)$; and q_i is the number of observations for the i-th subject. Note that $\sum_{i=1}^{n} \boldsymbol{B}_i(k)$ is a quadratic function of $\boldsymbol{\Delta}$. So minimizing $\sum_{i=1}^{n} \boldsymbol{B}_i(k)$ over $\boldsymbol{\Delta}$ is easily accomplishable. The above formula for $\boldsymbol{\theta}^{(k+1)}$ does not guarantee $\sigma^2(k+1)$ or $\boldsymbol{G}(k+1)$ to be non-negatively definite. The step-halving line search strategy is built in the algorithm to guarantee that $\boldsymbol{\theta}^{(k+1)}$ stays inside $\boldsymbol{\Theta}$.

Regardless of initial values, the FS-algorithm could still fail to converge because of the quality of approximating the Hessian matrix, especially when the sample size is small. To overcome this algorithmic difficulty, we propose the following hybrid approach: For the current parameter estimate $\boldsymbol{\theta}^{(k)}$, we compute a temporary parameter estimate $\widetilde{\boldsymbol{\theta}}^{(k+1)}$ using the FS-algorithm. If $\mathcal{L}_n^{pl}\left(\widetilde{\boldsymbol{\theta}}^{(k+1)}\right) \geq \mathcal{L}_n^{pl}\left(\boldsymbol{\theta}^{(k)}\right)$, the updated parameter estimate $\boldsymbol{\theta}^{(k+1)}$ is set to be $\widetilde{\boldsymbol{\theta}}^{(k+1)}$. Otherwise, $\boldsymbol{\theta}^{(k+1)}$ is obtained by running the EM-algorithm for N iterations, where $N \geq 2$ is a pre-specified number. In other words, this hybrid algorithm attempts to use the FS-algorithm to accelerate the EM, while also keeping the FS-iterations in the right track of increasing the likelihood with the assistance of the EM.

The performance of this algorithm was tested in the simulation study and in a real data analysis reported in the following sections. In all these applications, the hybrid algorithm produced algorithmically convergent series of updated parameter estimates.

4 Simulation studies

Simulation studies were conducted to investigate the performance of the proposed model in finite-sample situations. Two sample sizes were considered: $n = 200$, and 400. To evaluate the impact of the interval lengths, for each given sample size n, we simulated two scenarios: (1) average censoring interval length $l = 1$; and (2) average censoring interval length $l = 2$.

For a given sample size n and the average interval length l, we generated a total of 1500 simulated data sets as follows: For the ith subject, the true anchoring-event time T_i was independently generated from a Weibull distribution with shape parameter 80 and scale parameter 12. For each non-overlapping time window $(kl, (k+1)l]$, where $k = 0, 1, 2, \cdots$, a uniformly distributed screening time was generated. The censoring interval $(L_i, R_i]$ was identified as the adjacent screening times that bracket T_i, i.e., $L_i < T_i \leq R_i$. To allow covariates in the proposed method, we also simulated a binary covariate X_{1i} with equal probability $P(X_{1i} = 0) = P(X_{2i} = 1) = 1/2$, and a continuous covariate X_{2i} that was $N(0, 1)$ distributed. The observations at the two endpoints of the censoring interval, $Y_{L,i}$ and $Y_{R,i}$, were then generated from the following linear mixed-effects model

$$\begin{cases} Y_{L,i} = \lambda + \beta_1 X_{1i} + \beta_2 X_{2i} + \alpha(L_i - T_i) + \lambda_i + \alpha_i(L_i - T_i) + \epsilon_{L,i} \\ Y_{R,i} = \lambda + \beta_1 X_{1i} + \beta_2 X_{2i} + \beta(R_i - T_i) + \lambda_i + \beta_i(R_i - T_i) + \epsilon_{R,i} \end{cases} \tag{6}$$

where $(\lambda, \beta_1, \beta_2, \alpha, \beta)$ were the parameters for the population fixed effects, $(\lambda_i, \alpha_i, \beta_i)$ were the subject-specific random deviations, and $\epsilon_{L,i}$ and $\epsilon_{R,i}$ were the independent error terms. The true values of the parameters were $\lambda = 50, \beta_1 = -2, \beta_2 = -3, \alpha = 5, \beta = 8$. The random effects $(\lambda_i, \alpha_i, \beta_i)$ were generated from $\boldsymbol{N}(\mathbf{0}, \boldsymbol{G})$, and the error terms $(\epsilon_{L,i}, \epsilon_{R,i})$ were generated from $\boldsymbol{N}(\mathbf{0}, diag(\sigma^2, \sigma^2))$, where $\sigma^2 = 2.25$ and

$$\boldsymbol{G} = \begin{pmatrix} 9 & 1 & -1 \\ 1 & 3 & -1 \\ -1 & -1 & 4 \end{pmatrix}.$$

This data setting mimics the PGS study that is illustrated in the next section.

For each simulated data set, three models were fitted. First, Model (6) was fitted using the proposed functional estimation procedure, with standard errors of estimators estimated using bootstrap method with 50 resamples. The second model that we fitted used the midpoint of the censoring interval to impute the unobserved T_i, i.e., set $T_i = \frac{1}{2}(L_i + R_i)$, as done in practice (Shankar et al., 2005). Finally, we fitted Model (6), using the true anchoring event time T_i. In real world applications, this final model is unrealistic because of unknown T. In the simulation study, we simply use the model with true T values to benchmark the performance of the proposed method.

We note that the standard algorithm for computing the maximum likelihood estimates for longitudinal normal data with midpoint imputation failed to converge in about 20% of the simulated data sets. For each sample size n and average censoring interval length l, we summarized the simulation result from the first 1000 data sets that provided numerically convergent parameter estimates from all of the three methods. The percentage of bias in parameter estimates (% Bias), Monte-Carlo standard deviation (M-C SD), average bootstrap standard error (Av. SE) and coverage probability of the 95% Wald-confidence intervals (95% CP) are reported in Tables 1 and 2.

##Insert Table 1 here##

##Insert Table 2 here##

The simulation results in Tables 1 and 2 showed that the proposed method had an excellent finite-sample performance. Estimation bias was very small and virtually ignorable when the sample size was large (less than 1%). Both bias and Monte-Carlo standard deviation decreased with sample size. The bootstrap standard error were very close to Monte-Carlo standard deviation, which justified the use of bootstrap method in estimating the standard errors. The empirical coverage probabilities of the 95% Wald-confidence intervals were all close to the nominal level. The result clearly validates the large sample theory derived in Theorem 2.1.

For a fixed sample size n, the simulation study showed an improved model performance with wider censoring intervals, which are expected in view of the regularity condition (F4). When the censoring interval is narrow and n is not sufficiently large, the NPMLE $\hat{F}_n$ may not be a satisfactory estimate of F_0, thus leading to a decreased performance of the proposed method. In such situations, a larger sample size is generally required to have an asymptotic normal distribution for the estimated model parameters.

Not surprisingly, the midpoint imputation method did not fare well at all, with drastically larger biases, especially in α and β, the main parameters of interest. The bias remained substantial even when the sample size was increased to 400. At the same time, Monte-Carlo standard deviations almost tripled in comparison with that of the proposed method. Interestingly, however, estimates for the parameters that are not associated with the anchoring time T, such as $\lambda, \beta_1, \beta_2$, the biases were less severely affected.

As expected, the method using the true anchoring event time outperformed all its competitors. But it is important to note that the proposed method also had practically ignorable bias and comparable efficiency in parameter estimation. To empirically evaluate the relative efficiency, we computed the ratios of the Monte-Carlo standard deviations of the parameter estimates between the proposed method and the model under the known true anchoring event times; see Tabel 3. The Monte-Carlo standard deviations of the parameter estimates in the proposed method were 1.4%-12.3% larger, indicating only a mild loss of efficiency compared to the ideal situation of knowing T values.

Insert Tabel 3 here

To investigate the robustness of the normality assumption on the error terms in the proposed method, we repeated the previous simulation with a change in the error terms $\epsilon_{L,i}$ and $\epsilon_{R,i}$ in Model (6). We simulated errors from a mixture of normal distributions

$$Z_1 Z_2 + (1 - Z_1) Z_3,$$

where Z_1 was a binary random variable with equal probability $P(Z_1 = 0) = P(Z_1 = 1) = 1/2$, $Z_2 \sim N(-1, 1^2)$, and $Z_3 \sim N(1, 0.5^2)$. The simulation result was provided in the online supplementary material. It generally showed patterns similar to those reported in Tables 1, 2, and 3, which indicated that the normal assumption on the error terms in the proposed method was generally robust for larger samples.

5 Analysis of weight change around the PGS

In a longitudinal study of pubertal growth and blood pressure regulation, school children aged from 5 to 17 were recruited for longitudinal assessment of somatic growth and blood pressure. At each assessment, somatic growth measures were taken and recorded. The detailed study protocol was described in Tu et al. (2009, 2014). For each study participant, the investigators identified the interval that showed the greatest rate of height increase and used it as the interval containing the PGS (Shankar et al., 2005).

An overarching objective of this research was to quantify the weight changes around the time of PGS, with the goal of improving the existing understanding of the adolescent growth. We focused on growth rates around of the time of PGS, because they are thought to set the trajectory of adolescent development into the adulthood. The outcome of interest in this particular analysis was weight, which is one of the the primary markers of body development. Specifically, we attempted to compare: (1) the pre and post-PGS rates of weight increase, (2) the average weights at the time of PGS between races, and (3) the race difference in weight increase rates around the time of PGS. Because of the known differences in pubertal growth patterns between boys and girls, analyses often proceed in sex-specific groups. The current analysis was based on data from 188 boys. Because the focus of the current analysis was the quantification of local rates of skeletal changes around the PGS, we used only data collected at the two ends of the peak growth intervals.

The peak growth intervals, i.e., the censoring intervals containing the unobserved PGS, are presented in the left panel of Figure 1. Weights measured at the endpoints of the censoring intervals are depicted in the right panel of Figure 1. The pre and post-PGS weight measures from the same individual are connected by a line segment.

##Insert Figure 1 here##

To analyze, we considered the following piece-wise linear mixed-effects model with random intercepts and random post-PGS growth rates. We did not include random pre-PGS slopes because existing literature suggests that heterogeneity in rate of weight increase started at the PGS.

$$\begin{cases} Y_{L,i} = \lambda_1 + \alpha_1(L_i - T_i) + \big(\lambda_2 + \alpha_2(L_i - T_i)\big) * I_w + \lambda_i + \epsilon_{L,i} \\ Y_{R,i} = \lambda_1 + \beta_1(R_i - T_i) + \big(\lambda_2 + \beta_2(R_i - T_i)\big) * I_w + \lambda_i + \beta_i(R_i - T_i) + \epsilon_{R,i} \end{cases} \tag{7}$$

where $(L_i, R_i]$ was the censoring interval for the i-th subject; T_i was the unobserved PGS time; $Y_{L,i}$ and $Y_{R,i}$ were the respective weights measured at L_i and R_i; I_w was an indicator for being whites; λ_1 was the average weight at PGS in non-whites; and α_1 and β_1 were the respectively average pre and post-PGS weight growth rates in non-whites. Similarly, λ_2, α_2 and β_2 respectively represented the differences in average value, pre and post-PGS weight growth rates between non-whites and whites. Here λ_i and β_i were the random intercept and slope; and $\epsilon_{L,i}$ and $\epsilon_{R,i}$ were the random errors.

All regularity conditions were satisfied in this data application. For example, Condition (F2) was satisfied, as Figure 1 did not show substantial differences in PGS distributions between whites and non-whites. So we applied the proposed functional estimation method to fit Model (7). The parameter estimates $\hat{\boldsymbol{\theta}}_n = (\hat{\lambda}_{n,1}, \hat{\lambda}_{n,2}, \hat{\alpha}_{n,1}, \hat{\beta}_{n,1}, \hat{\alpha}_{n,2}, \hat{\beta}_{n,2})$ are summarized in Table 4.

Insert Table 4 here##

The following covariance matrix $\hat{\boldsymbol{\Sigma}}_n$ of the parameter estimates was estimated using bootstrap resampling method with 100 resamples.

$$\hat{\boldsymbol{\Sigma}}_n = \begin{bmatrix} 3.613 & -3.409 & 0.278 & 0.769 & -0.329 & -0.659 \\ -3.409 & 3.968 & -0.117 & -0.879 & 0.334 & 0.733 \\ 0.278 & -0.117 & 0.658 & -0.442 & -0.671 & 0.491 \\ 0.769 & -0.879 & -0.442 & 1.009 & 0.450 & -1.004 \\ -0.329 & 0.334 & -0.671 & 0.450 & 0.942 & -0.741 \\ -0.659 & 0.733 & 0.491 & -1.004 & -0.741 & 1.577 \end{bmatrix}.$$

We then proceeded to make inferences on parameters of interest along the lines laid out by Theorem 2.1. First, the pre and post-PGS rates of weight increase can be compared by testing hypothesis that $\alpha_1 = \beta_1$ and $\alpha_2 = \beta_2$ in all ethnic groups. Using contrast

$$\boldsymbol{M} = \begin{pmatrix} 0 & 0 & 1 & -1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & -1 \end{pmatrix},$$

the test statistic

$$T_n = (\boldsymbol{M}\hat{\boldsymbol{\theta}}_n)^t(\boldsymbol{M}\hat{\boldsymbol{\Sigma}}_n\boldsymbol{M}^t)^{-1}(\boldsymbol{M}\hat{\boldsymbol{\theta}}_n)$$

was approximately $\chi^2(2)$-distributed under the null hypothesis. The observed value of T_n was calculated to be 10.150, which resulted in a p-value of 0.006 for a χ^2-test. So we had evidence that rate of weight growth was greater in the post-PGS period.

To compare the average weight at PGS between individuals in different ethnic groups, we simply tested the hypothesis that $\lambda_2 = 0$. Under the null hypothesis, the Z-score of $\hat{\lambda}_2$ was calculated as $-6.927/1.992 = -3.477$, which resulted in a p-value of 0.0005 for the two-sided Z-test. So we concluded that whites and ethnic minority children had different weights at PGS, with whites had lower weights.

Finally, to compare the rates of weight increase around PGS between individuals of different ethnic groups, we tested the hypothesis $\alpha_2 = \beta_2 = 0$. To implement, we calculated the value of the statistic $(\hat{\alpha}_{n,2}, \hat{\beta}_{n,2}) \cdot \hat{\boldsymbol{\Sigma}}_n^{\prime -1} \cdot (\hat{\alpha}_{n,2}, \hat{\beta}_{n,2})^t$, which was approximated by a $\chi^2(2)$-distribution under the null hypothesis. Here matrix $\hat{\boldsymbol{\Sigma}}_n^{\prime}$ was the covariance matrix of $\hat{\alpha}_{n,2}$ and $\hat{\beta}_{n,2}$, which was a submatrix of $\hat{\boldsymbol{\Sigma}}_n$. The observed value of this statistic was calculated to be 5.877, which led to a p-value of 0.053. So we concluded that there was some indication that whites had slower rate of weight gain around PGS, although the difference did not reach the threshold of 0.05 to be statistically significant. We note that all findings were consistent with the observed data shown in Figure 1.

Through this real data example, we demonstrated the operations of parameter estimation and statistical inference using the new method, which handled interval-censored PGS times nicely in this application. While the findings were largely consistent with the existing theory of human growth (Hall, 2006), no studies to the best of our knowledge have actually quantified the growth parameters around PGS, because of the traditional longitudinal models' inability to accommodate the unobserved anchoring event times in the absence of strong parametric assumptions.

6 Discussion

In the past quarter of century, we have witnessed a remarkable growth of methodology in longitudinal data analysis. The development, in many ways, has not only enhanced statisticians' analytical toolbox, but also influenced the way that scientists approach their investigations. Among other things, quantitative depiction of temporal changes has becoming more commonplace and generally accepted, with important discoveries being made along the way. If the raison d'être of longitudinal data analysis is indeed characterization of changes in the response of interest over time, as some have convincingly argued (Fitzmaurice and Ravichandran, 2008), a necessary but often implicit requirement is a well-defined study timeline. In many research settings, such as clinical trials, the requirement is automatically satisfied because of the use of explicitly specified starting points. But there are many studies whose timelines are anchored by unobserved events. Such situations are especially abundant in clinical investigation. Examples include unobserved oncogenic onset and puberty growth spurt; the latter was described in the current paper. In the absence of directly observed anchoring events, study timelines become undefined and all measurements lose their temporal references, thus rendering the longitudinal models inoperable or uninterpretable. Considering how common the situation is, it is surprising that the problem has not caught more attention earlier.

In this research, we took the initial steps towards solving the problem for a general class of longitudinal models. Here, the data likelihood is essentially a function of the model parameters $\boldsymbol{\theta} \in \boldsymbol{\Theta}$ and a nuisance parameter F (CDF of the interval-censored anchoring-event time). While it is straightforward to formulate the model and likelihood function, jointly estimating $\boldsymbol{\theta}$ and F is a computationally daunting, if not impossible task, unless F is specified parametrically. In this research, we present a solution that is conceptually simple and computationally straightforward: By expressing the longitudinal model parameter estimators $\hat{\boldsymbol{\theta}}_n$ as stochastic functionals of F, we were able to achieve valid parameter estimates via a semiparametric pseudolikelihood method, built upon the well studied NPMLE of F from interval-censored observations.

Along this line, we established asymptotic normality for the functional estimates in a general class of longitudinal models without imputing the anchoring event times or making parametric assumptions for their distribution. To illustrate its use, we presented a case study of the frequently used linear mixed-effects models, along with an efficient computational algorithm. We also provided comprehensive numerical evidence in support of its good finite-sample performance, even in presence of model misspecification. We explicitly stated the sufficient conditions for the main theoretical results. These conditions are mild and are generally satisfied by most observational longitudinal studies. Although our numerical example depicts a situation where there are only two observation time points, the method is readily extendable to studies with multiple follow-up assessments.

By studying analytical methods in longitudinal studies with less well defined timelines, we hope to broaden the scope of application for the existing longitudinal models so that they could better meet the need of scientific investigation. But considering the levels of maturity of the existing longitudinal models, this work is still an initial attempt towards the goal of a more complete solution. There are certainly important questions that remain to be addressed. Among them are the incorporation of functional analysis of longitudinal outcomes and covariate-dependent anchoring-event time distributions. In a situation where the anchoring-event time T depends on a set of covariates $\boldsymbol{W}$, there are a number of semiparametric models for interval-censored survival data (Huang and Wellner, 1997) that can be used to estimate the distribution function of T given $\boldsymbol{W}$, $\hat{F}_n(t|\boldsymbol{W})$. All of the technical steps presented in this paper can similarly proceed by replacing $\hat{F}_n$ with $\hat{F}_n(t|\boldsymbol{W})$. Although the notation will be more involved in the method development and the empirical process arguments, the main theoretical result will hold with mild regularity conditions. New computational algorithms, however, will be needed for model fitting.

Notwithstanding these limitations, we put forward a carefully justified method for analyzing longitudinal data with interval-censored anchoring events. The method helps to address the scientific need for estimating the change rates around an unobserved anchoring event, a need that would be difficult to meet with the traditional longitudinal data analysis.

Acknowledgement

The authors appreciate the editor and two anonymous reviewers for their helpful comments and suggestions.

References

G. M. Fitzmaurice and C. Ravichandran. A primer in longitudinal data analysis. *Circulation*, 118(19):2005–2010, 2008.

R. Geskus and P. Groeneboom. Asymptotically optimal estimation of smooth functionals for interval censoring, case 2. *The Annals of Statistics*, 27(2):627–674, 1999.

G. Gong and F. J. Samaniego. Pseudo maximum likelihood estimation: theory and applications. *The Annals of Statistics*, pages 861–869, 1981.

P. Groeneboom and J. A. Wellner. *Information bounds and nonparametric maximum likelihood estimation*, volume 19. Springer Science & Business Media, 1992.

S. S. Hall. *Size Matters: How Height Affects the Health, Happiness, and Success of Boys–and the Men They Become*. Houghton Mifflin Harcourt, 2006.

J. Huang and J. Wellner. Interval-censored survival data: a review of recent progress. In *Proceedings of the First Seattle Symposium in Biostatistics. Lecture Notes in Statistics*. Springer-Verlag, 1997.

S. Kong, B. Nan, J. D. Kalbfleisch, R. Saran, and R. Hirth. Conditional modeling of longitudinal data with terminal event. *Journal of the American Statistical Association*, pages 1–12, 2017.

E. Mammen. *When does bootstrap work?: asymptotic results and simulations*, volume 77. Springer Science & Business Media, 2012.

L. F. Robinson, T. D. Wager, and M. A. Lindquist. Change point estimation in multi-subject fmri studies. *Neuroimage*, 49(2):1581–1592, 2010.

R. R. Shankar, G. J. Eckert, C. Saha, W. Tu, and J. H. Pratt. The change in blood pressure during pubertal growth. *The Journal of Clinical Endocrinology & Metabolism*, 90(1):163–167, 2005.

J. M. Tanner and R. H. Whitehouse. Clinical longitudinal standards for height, weight, height velocity, weight velocity, and stages of puberty. *Archives of disease in childhood*, 51(3):170–179, 1976.

W. Tu, G. J. Eckert, C. Saha, and J. H. Pratt. Synchronization of adolescent blood pressure and pubertal somatic growth. *The Journal of Clinical Endocrinology & Metabolism*, 94(12):5019–5022, 2009.

W. Tu, G. J. Eckert, T. S. Hannon, H. Liu, L. M. Pratt, M. A. Wagner, L. A. DiMeglio, J. Jung, and J. H. Pratt. Racial differences in sensitivity of blood pressure to aldosterone. *Hypertension*, 63(6):1212–1218, 2014.

A. van den Hout, G. Muniz-Terrera, and F. E. Matthews. Change point models for cognitive tests using semi-parametric maximum likelihood. *Computational statistics & data analysis*, 57(1):684–698, 2013.

A. van der Vaart. On differentiable functionals. *The Annals of Statistics*, 19(1):178–204, 1991.

A. van der Vaart. *Asymptotic Statistics*. Cambridge University Press, Cambridge, 1998.

A. van der Vaart and J. A. Wellner. *Weak Convergence and Empirical Processes with Application to Statistics*. Springer Verlage, 1996.

A. W. van der Vaart, J. A. W. Wellner, et al. Empirical processes indexed by estimated functions. In *Asymptotics: particles, processes and inverse problems*, pages 234–252. Institute of Mathematical Statistics, 2007.

M. Wand. Vector differential calculus in statistics. *The American Statistician*, 56(1): 55–62, 2002.

Y. Zhang and M. Jamshidian. On algorithms for the nonparametric maximum likelihood estimator of the failure function with censored data. *Journal of Computational and Graphical Statistics*, 13(1):123–140, 2004.

Y. Zhang, G. Cheng, and W. Tu. Robust nonparametric estimation of monotone regression functions with interval-censored observations. *Biometrics*, 72:720–730, 2016.

Correspondence

Ying Zhang, University of Nebraska Medical Center, Omaha, NE 68198-4375, USA. Email: ying.zhang@unmc.edu

APPENDIX

Proof of Theorem 2.1. We need a general theorem on the asymptotic normality of semiparametric Z-estimators and a technical lemma, whose proof is provided in the online supplementary material. To state the general theorem, let P be the probability measure associated with observed data $\boldsymbol{D}$ and $\mathbb{P}_n$, the empirical measure associated with n independent and identically distributed copies of $\boldsymbol{D}$. For any P-measurable function f, the integrals $\int f d\mathrm{P}$ and $\int f d\mathbb{P}_n$ are respectively denoted as $\mathrm{P}(f)$ and $\mathbb{P}_n(f)$.

Let P be the probability measure associated with observed data $\boldsymbol{D}$ and $\mathbb{P}_n$, the empirical measure associated with n independent and identically distributed copies of $\boldsymbol{D}$. For any P-measurable function f, the integrals $\int f d\mathrm{P}$ and $\int f d\mathbb{P}_n$ are respectively denoted as $\mathrm{P}(f)$ and $\mathbb{P}_n(f)$.

We consider semiparametric model that satisfies

$$\mathrm{P}\big(\boldsymbol{\psi}(\boldsymbol{D}, F_0, \boldsymbol{\theta}_0)\big) = \mathbf{0}$$

where $\quad = \boldsymbol{\psi}(\boldsymbol{D}, F, \boldsymbol{\theta})$ is a d-dimensional estimating function for $\boldsymbol{\theta} \in \boldsymbol{\Theta} \subset \mathbb{R}^d$ given $F \in \mathcal{F}$, a class of functions, and $\boldsymbol{\theta}_0 \in \boldsymbol{\Theta}$ and $F_0 \in \mathcal{F}$ are the true model parameters. When F_0 is unknown, as long as a consistent estimator $\hat{F}_n$ of F_0 can be obtained from the data, one can obtain a Z-estimator of $\boldsymbol{\theta}_0$ by solving the estimating equation

$$\mathbb{P}_n\big(\quad (\boldsymbol{D}, \hat{F}_n, \boldsymbol{\theta})\big) = \mathbf{0},$$

for $\boldsymbol{\theta}$, denoted by $\hat{\boldsymbol{\theta}}_n$.

The following theorem provides sufficient conditions for $\sqrt{n}(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0)$ to converge in distribution. For convenience of presentation, we define a mapping $\boldsymbol{\Psi} : \boldsymbol{\Theta} \times \mathcal{F} \to \mathbb{R}$ by setting $\boldsymbol{\Psi}(\boldsymbol{\theta}, F) = \mathrm{P}\big(\boldsymbol{\psi}(\boldsymbol{D}, F, \boldsymbol{\theta})\big)$ for any $(\boldsymbol{\theta}, F) \in \boldsymbol{\Theta} \times \mathcal{F}$. Let $\boldsymbol{\Psi}_n$ be the empirical version of $\boldsymbol{\Psi}$, i.e., $\boldsymbol{\Psi}_n(\boldsymbol{\theta}, F) = \mathbb{P}_n\big(\boldsymbol{\psi}(\boldsymbol{D}, F, \boldsymbol{\theta})\big)$.

Theorem 1. *Suppose $\boldsymbol{\theta}_0$ satisfies $\boldsymbol{\Psi}(\boldsymbol{\theta}_0, F_0) = \mathbf{0}$. Let $\hat{\boldsymbol{\theta}}_n$ be a solution of $\boldsymbol{\Psi}_n(\boldsymbol{\theta}, \hat{F}_n) = \mathbf{0}$, where $\hat{F}_n$ is an estimate of F_0 from the sample. If the following conditions hold,*

- *T1. $\boldsymbol{\theta}_0$ is an inner point of $\boldsymbol{\Theta}$. The function $\boldsymbol{\theta} \mapsto \boldsymbol{\Psi}(\boldsymbol{\theta}, F_0)$ has continuous second order derivatives in a neighborhood of $\boldsymbol{\theta}_0$ and the matrix $\boldsymbol{A} = \nabla_{\boldsymbol{\theta}} \boldsymbol{\Psi}(\boldsymbol{\theta}_0, F_0)$ is nonsingular;*
- *T2. $\hat{\boldsymbol{\theta}}_n \xrightarrow{P} \boldsymbol{\theta}_0$;*
- *T3. $\sqrt{n}\boldsymbol{\Psi}_n(\boldsymbol{\theta}_0, \hat{F}_n) \xrightarrow{D} \boldsymbol{Z}$ for some random vector $\boldsymbol{Z}$;*
- *T4. $\Big(1 + \sqrt{n}\|\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0\|\Big)^{-1} \Big\|\sqrt{n}\big(\boldsymbol{\Psi}(\hat{\boldsymbol{\theta}}_n, F_0) + \boldsymbol{\Psi}_n(\boldsymbol{\theta}_0, \hat{F}_n)\big)\Big\| \xrightarrow{P} 0$,*

then $\sqrt{n}(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0) \xrightarrow{D} -\boldsymbol{A}^{-1}\boldsymbol{Z}$.

Condition $T1$ in Theorem 1 is the general regularity condition for parametric models when F_0 is known, which is usually satisfied if the estimating function is a smooth function of $\boldsymbol{\theta}$. Methods for verifying Conditions $T2$ and $T3$ depend on the specific model setting, and usually require more efforts with empirical process theory. Condition $T4$ is essentially equivalent to the general result given by van der Vaart et al. (2007) for studying the asymptotics in pseudolikelihood estimation methods. This condition is, however, easier to verify than van der Vaart-Wellner's condition for proving Theorem 2.1 with the following lemma, which stipulates a set of sufficient conditions that justify Condition $T4$.

Lemma 2. *Let $\boldsymbol{\Theta}$ be a compact set that contains $\boldsymbol{\theta}_0$ as an inner point. Let $\|\cdot\|_\infty$ be the supremum norm on $\mathcal{F}$. Assume that*

- *L1. For any $F \in \mathcal{F}$, the Stieltjes-Lebesgue measure dF exists and is supported in a finite closed interval $[\tau_1, \tau_2]$, where the constants $\tau_1 < \tau_2$ do not depend on F.*

L2. $\|\hat{F}_n - F_0\|_\infty = o_p(n^{-1/4})$.

L3. $\sqrt{n}\Big(\boldsymbol{\Psi}_n(\hat{\boldsymbol{\theta}}_n, F) - \boldsymbol{\Psi}(\hat{\boldsymbol{\theta}}_n, F)\Big) - \sqrt{n}\Big(\boldsymbol{\Psi}_n(\boldsymbol{\theta}_0, F) - \boldsymbol{\Psi}(\boldsymbol{\theta}_0, F)\Big) = \boldsymbol{o}_p(1 + \sqrt{n}\|\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0\|)$, *uniformly over the class* $\mathcal{F}$.

L4. $\boldsymbol{\Psi}(\boldsymbol{\theta}, \hat{F}_n) - \boldsymbol{\Psi}(\boldsymbol{\theta}, F_0) = \int \kappa(\boldsymbol{\theta}, t) d(\hat{F}_n(t) - F_0(t)) + \boldsymbol{O}_p(\|\hat{F}_n - F_0\|_\infty^2)$ *uniformly over* $\boldsymbol{\Theta}$*, where* $\kappa(\boldsymbol{\theta}, t) \in C^1\Big(\boldsymbol{\Theta} \times [\tau_1, \tau_2]\Big)$*, the set of functions on* $\boldsymbol{\Theta} \times [\tau_1, \tau_2]$ *that have continuous first-order derivatives.*

Then Condition T4 of Theorem 1 is satisfied.

In the remaining part of the Appendix, we apply Theorem 1 to study the asymptotic distribution of the proposed model parameter estimates, under the regularity conditions stated in Section 2.2. Let K denote a constant, whose value may differ from place to place.

Recall that $\phi = \phi(\boldsymbol{Y}|\boldsymbol{W}, L, R, T, \boldsymbol{\theta})$ denote the conditional density of $\boldsymbol{Y}$ given $\boldsymbol{W}, L, R$ and the anchoring-event time T, and $\boldsymbol{u} = \nabla_{\boldsymbol{\theta}}\phi$ its partial derivative with respect to $\boldsymbol{\theta}$. Let H, P and $\mathbb{P}_n$ denote, respectively, the density function, the probability measure, and the corresponding empirical measure of $\boldsymbol{D} = (\boldsymbol{Y}, \boldsymbol{W}, L, R)$. By the regularity condition 3 , the probability measure P has a compact support, which we denote by Ω. To apply the general result in Theorem 1, we define the following multivariate random functional,

$$\psi(\boldsymbol{D}, F, \boldsymbol{\theta}) = \frac{\int_L^R \boldsymbol{u}(\boldsymbol{Y}, \boldsymbol{W}, L, R, t; \boldsymbol{\theta})\, dF(t)}{\int_L^R \phi(\boldsymbol{Y}, \boldsymbol{W}, L, R, t; \boldsymbol{\theta})\, dF(t)} = \frac{\int_L^R \boldsymbol{u}\, dF}{\int_L^R \phi\, dF}$$

which is the score function of the parameter $\boldsymbol{\theta}$ based on the conditional likelihood of $\boldsymbol{Y}|(\boldsymbol{W}, L, R, L \le T < R)$, assuming the CDF of T is $F \in \mathcal{F}$. By the second set of regularity conditions, the usual MLE theory implies that the true parameter $\boldsymbol{\theta}_0$ solves the estimating equation

$$\mathrm{P}\big(\psi(\boldsymbol{D}, F_0, \boldsymbol{\theta})\big) = \int \frac{\int_L^R \boldsymbol{u}\, dF_0}{\int_L^R \phi\, dF_0}\, d\mathrm{P} = \boldsymbol{0}.$$

From the regularity conditions (M2) and (M4), there exists a closed ball $\boldsymbol{\Theta}$ with radius $K > 0$ and with center $\boldsymbol{\theta}_0$, over which the function $\mathrm{P}\big(\psi(\boldsymbol{D}, F_0, \boldsymbol{\theta})\big)$ is C^3 (i.e., having continuous third-order derivatives), and strictly convex with respect to $\boldsymbol{\theta}$. Let $\mathcal{F}_\delta$ denote the class of all CDFs supported in $[\tau_1, \tau_2]$ (the finite interval from the regularity condition F1), whose $\|\cdot\|_\infty$-distances from F_0 are less than a small number δ. For the sake of notational convenience, we let $\boldsymbol{U}_{\boldsymbol{\theta},F}$ denote the following random variable on Ω, where $(\boldsymbol{\theta}, F) \in \boldsymbol{\Theta} \times \mathcal{F}_\delta$.

$$\boldsymbol{U}_{\boldsymbol{\theta},F}(\boldsymbol{D}) = \psi(\boldsymbol{D}, F, \boldsymbol{\theta}), \quad \text{where } \boldsymbol{D} \in \Omega.$$

Define an empirical process $\boldsymbol{\Psi}_n$ by setting $\boldsymbol{\Psi}_n(\boldsymbol{\theta}, F) = \mathbb{P}_n \boldsymbol{U}_{\boldsymbol{\theta},F}$, and the corresponding functional $\boldsymbol{\Psi}$ by setting $\boldsymbol{\Psi}(\boldsymbol{\theta}, F) = \mathrm{P}\boldsymbol{U}_{\boldsymbol{\theta},F}$, where $(\boldsymbol{\theta}, F) \in \boldsymbol{\Theta} \times \mathcal{F}_\delta$.

To prove Theorem 2.1, we first need to study the properties of several classes of functions that are relevant to the problem.

Let $G = G(\boldsymbol{D}, t; \boldsymbol{\theta})$, which could be either the function ϕ or a component function of $\boldsymbol{u}$ in our application. Let $\mathcal{G}_{\boldsymbol{\Theta},\mathcal{F}_\delta}$ be the following induced class of function on Ω indexed by $\boldsymbol{\Theta} \times \mathcal{F}_\delta$.

$$\mathcal{G}_{\boldsymbol{\Theta},\mathcal{F}_\delta} = \left\{ G_{\boldsymbol{\theta},F} = \int_L^R G\, dF = \int_L^R G(\boldsymbol{D}, t; \boldsymbol{\theta})\, dF(t) :\ F \in \mathcal{F}_\delta, \boldsymbol{\theta} \in \boldsymbol{\Theta} \right\}.$$

We evaluate $N_{[]}(\epsilon, \mathcal{G}_{\boldsymbol{\Theta},\mathcal{F}_\delta}, L^2(\mathrm{P}))$, the $L^2(\mathrm{P})$-norm ϵ-bracketing number of $\mathcal{G}_{\boldsymbol{\Theta},\mathcal{F}_\delta}$ with respect to the probability measure P.

By Theorem 2.7.5 of van der Vaart and Wellner (1996), the family $\mathcal{F}_\delta$ can be covered by N_ϵ number of ϵ-brackets in L^2-norm $\|\cdot\|_2$ with respect to the Borel measure with

$N_\epsilon \leq \exp(\frac{K}{\epsilon})$. In other words, there exist pairs of measurable functions

$$\left\{\left(F_i^-(t), F_i^+(t)\right) : i = 1, \cdots N_\epsilon\right\}$$

such that there exists for any $F \in \mathcal{F}_\delta$ a bracket $\left(F_i^-(t), F_i^+(t)\right)$ satisfying $F_i^-(t) \leq F(t) \leq F_i^+(t)$ and $\| F_i^-(t) - F_i^+(t) \|_2 < \epsilon$. We assume that each bracket contains at least one function F in $\mathcal{F}_\delta$. Otherwise, such a bracket should be removed and results in fewer ϵ-brackets. We can also require that $0 \leq F_i^+ \leq 1$ and $0 \leq F_i^- \leq 1$. It is obvious that there are no more than $\left(\frac{K}{\epsilon}\right)^d$ solid hypercubes $\{Q_1, Q_2, \cdots, Q_K\}$, whose union covers $\boldsymbol{\Theta}$ and whose sides have lengths ϵ.

For any hypercube Q_j and any $t \in [\tau_1, \tau_2]$, define the following functions $S_{j,t}^-$, $S_{j,t}^+$, $S_{j,t}^{'-}$ and $S_{j,t}^{'+}$ of $\boldsymbol{D} \in \Omega$.

$$\begin{aligned} S_{j,t}^-(\boldsymbol{D}) &= \min_{\boldsymbol{\theta} \in Q_j} G(\boldsymbol{D}, t; \boldsymbol{\theta}), S_{j,t}^+(\boldsymbol{D}) = \max_{\boldsymbol{\theta} \in Q_j} G(\boldsymbol{D}, t; \boldsymbol{\theta}); \\ S_{j,t}^{'-}(\boldsymbol{D}) &= \min_{\boldsymbol{\theta} \in Q_j} \frac{\partial G}{\partial t}(\boldsymbol{D}, t; \boldsymbol{\theta}), S_{j,t}^{'+}(\boldsymbol{D}) = \max_{\boldsymbol{\theta} \in Q_j} \frac{\partial G}{\partial t}(\boldsymbol{D}, t; \boldsymbol{\theta}). \end{aligned}$$

By the regularity condition (M4) on the smoothness of ϕ and $\boldsymbol{u}$, and by the regularity condition 3, both G and $\frac{\partial G}{\partial t}$ are continuous on the compact set $\Omega \times \boldsymbol{\Theta}$, and hence absolutely continuous. So $|S_{j,t}^+ - S_{j,t}^-| \leq K\epsilon$ and $|S_{j,t}^{'+} - S_{j,t}^{'-}| \leq K\epsilon$ for all j and t, where the value of K does not depend on j or t. For any bracket $\left(F_i^-(t), F_i^+(t)\right)$ and hypercube Q_j, we define the following functions of $\boldsymbol{D} \in \Omega$.

$$\begin{aligned} G_{ij}^-(\boldsymbol{D}) = S_{j,R}^-(\boldsymbol{D}) \cdot \Big(F_i^-(R) \cdot 1(S_{j,R}^-(\boldsymbol{D}) > 0) + F_i^+(R) \cdot 1(S_{j,R}^-(\boldsymbol{D}) \leq 0)\Big) \\ -S_{j,L}^+(\boldsymbol{D}) \cdot \Big(F_i^+(L) \cdot 1(S_{j,L}^+(\boldsymbol{D}) > 0) + F_i^-(L) \cdot 1(S_{j,L}^+(\boldsymbol{D}) \leq 0)\Big) \\ - \int_L^R S_{j,t}^{'+}(\boldsymbol{D}) \cdot \Big(F_i^+(t) \cdot 1(S_{j,t}^{'+}(\boldsymbol{D}) > 0) + F_i^-(t) \cdot 1(S_{j,t}^{'+}(\boldsymbol{D}) \leq 0)\Big) dt \end{aligned}$$

$$
\begin{aligned}
G_{ij}^{+}(\boldsymbol{D}) = S_{j,R}^{+}(\boldsymbol{D}) \cdot \Big(F_i^{+}(R) \cdot 1(S_{j,R}^{+}(\boldsymbol{D}) > 0) + F_i^{-}(R) \cdot 1(S_{j,R}^{+}(\boldsymbol{D}) \leq 0) \Big) \\
- S_{j,L}^{-}(\boldsymbol{D}) \cdot \Big(F_i^{-}(L) \cdot 1(S_{j,L}^{-}(\boldsymbol{D}) > 0) + F_i^{+}(L) \cdot 1(S_{j,L}^{-}(\boldsymbol{D}) \leq 0) \Big) \\
- \int_L^R S_{j,t}^{\prime -}(\boldsymbol{D}) \cdot \Big(F_i^{-}(t) \cdot 1(S_{j,t}^{\prime -}(\boldsymbol{D}) > 0) + F_i^{+}(t) \cdot 1(S_{j,t}^{\prime -}(\boldsymbol{D}) \leq 0) \Big) dt.
\end{aligned}
$$

Although the expressions of G_{ij}^{-} and G_{ij}^{+} are complicate, it is easy to see that the summands of these functions bracket the summands of the following integral in order.

$$
G_{\boldsymbol{\theta},F} = \int_L^R G\, dF = G\big|_{t=R} \cdot F(R) - G\big|_{t=L} \cdot F(L) - \int_L^R \frac{\partial G}{\partial t} \cdot F\, dt,
$$

where $F_i^{-} \leq F \leq F_i^{+}$ and $\boldsymbol{\theta} \in Q_j$. So we have $G_{ij}^{-} \leq G_{\boldsymbol{\theta},F} \leq G_{ij}^{+}$. In other words, the set of brackets $\big\{ \big(G_{ij}^{-}, G_{ij}^{+}\big) : i, j \big\}$ covers $G_{\boldsymbol{\theta},F}$. Let $\|\cdot\|_{2,\mathrm{P}}$ denote the $L^2(\mathrm{P})$-norm with respect to the probability measure P. It is shown in the online supplementary material that the $\|\cdot\|_{2,\mathrm{P}}$-length of the bracket $\big(G_{ij}^{-}, G_{ij}^{+}\big)$ satisfies

$$
\|G_{ij}^{+} - G_{ij}^{-}\|_{2,\mathrm{P}} \leq K\epsilon. \tag{8}
$$

In summary, we found a total of $(K/\epsilon)^d\, N_\epsilon$ brackets for $\mathcal{G}_{\boldsymbol{\Theta},\mathcal{F}_\delta}$, each of length $\leq K\epsilon$, where K is independent on ϵ. So $N_{[]}(\epsilon, \mathcal{G}_{\boldsymbol{\Theta},\mathcal{F}_\delta}, L^2(\mathrm{P}))$, the $L^2(\mathrm{P})$-norm ϵ-bracketing number for $\mathcal{G}_{\boldsymbol{\Theta},\mathcal{F}_\delta}$, is bounded by $(K/\epsilon)^d \exp(K/\epsilon)$. Then it follows that

$$
\begin{aligned}
J_{[]}\big(\epsilon, \mathcal{G}_{\boldsymbol{\Theta},\mathcal{F}_\delta}, L^2(\mathrm{P})\big) &= \int_0^1 \sqrt{\log\big(N_{[]}(\epsilon, \mathcal{G}_{\boldsymbol{\Theta},\mathcal{F}_\delta}, L^2(\mathrm{P}))\big)}\, d\epsilon \\
&\leq \int_0^1 \sqrt{\frac{K}{\epsilon} - K\log(\epsilon)}\, d\epsilon < \infty.
\end{aligned}
$$

By Donsker's Thoerem (van der Vaart (1998), page 270), $\mathcal{G}_{\boldsymbol{\Theta},\mathcal{F}_\delta}$ is a P-Donsker class.

The above argument directly yields the conclusion that the following two classes

$$\mathcal{NUM}_{\mathbf{\Theta},\mathcal{F}_\delta} = \left\{ \int_L^R \boldsymbol{u} dF : F \in \mathcal{F}_\delta, \boldsymbol{\theta} \in \mathbf{\Theta} \right\} \quad \text{and}$$

$$\mathcal{DEN}_{\mathbf{\Theta},\mathcal{F}_\delta} = \left\{ \int_L^R \phi dF : F \in \mathcal{F}_\delta, \boldsymbol{\theta} \in \mathbf{\Theta} \right\}$$

are P-Donsker. The regularity condition 3 implies that both $\mathcal{NUM}_{\mathbf{\Theta},\mathcal{F}_\delta}$ and $\mathcal{DEN}_{\mathbf{\Theta},\mathcal{F}_\delta}$ are uniformly bounded. For a small enough $\delta > 0$, $\mathcal{DEN}_{\mathbf{\Theta},\mathcal{F}_\delta}$ is uniformly bounded away from zero. By Theorem 2.10.6 of van der Vaart and Wellner (1996), both the point-wise quotient class

$$\mathcal{U}_{\mathbf{\Theta},\mathcal{F}_\delta} = \left\{ \boldsymbol{U}_{\boldsymbol{\theta},F} = \frac{\int_L^R \boldsymbol{u} dF}{\int_L^R \phi dF} : \boldsymbol{\theta} \in \mathbf{\Theta}, F \in \mathcal{F}_\delta \right\}$$

and the smooth transformation class

$$\mathcal{M}_{\mathbf{\Theta},\mathcal{F}_\delta} = \left\{ \boldsymbol{M}_{\boldsymbol{\theta},F} = \log\left(\int_L^R \phi dF \right) : \boldsymbol{\theta} \in \mathbf{\Theta}, F \in \mathcal{F}_\delta \right\}$$

are also P-Donsker. Hence $\mathcal{M}_{\mathbf{\Theta},\mathcal{F}_\delta}$ is a Glivenko–Cantelli class as well. By Example 2.10.7 of van der Vaart and Wellner (1996), the difference class

$$\mathcal{DU}_{\mathbf{\Theta},\mathcal{F}_\delta} = \left\{ \boldsymbol{U}_{\boldsymbol{\theta}_1,F_1} - \boldsymbol{U}_{\boldsymbol{\theta}_2,F_2} \;\middle|\; \boldsymbol{U}_{\boldsymbol{\theta}_1,F_1}, \boldsymbol{U}_{\boldsymbol{\theta}_2,F_2} \in \mathcal{U}_{\mathbf{\Theta},\mathcal{F}_\delta} \right\}$$

is also P-Donsker.

Next, we need to study the properties of the functional $\mathbf{\Psi}$. For any $\boldsymbol{\theta} \in \mathbf{\Theta}$, a direct computation shows $\mathbf{\Psi}(\boldsymbol{\theta}, F) - \mathbf{\Psi}(\boldsymbol{\theta}, F_0) = (I) - (II) + (III)$, where

$$(I) = \int \left(\int_L^R \phi dF_0 \right)^{-2} \cdot \left(\int_L^R \boldsymbol{u} dF \cdot \int_L^R \phi dF_0 - \int_L^R \boldsymbol{u} dF_0 \cdot \int_L^R \phi dF \right) d\mathrm{P}$$

$$(II) = \int \left(\int_L^R \phi dF_0 \right)^{-2} \cdot \int_L^R \boldsymbol{u} d(F - F_0) \cdot \int_L^R \phi d(F - F_0) \, d\mathrm{P}$$

$$(III) = \int \left(\int_L^R \phi dF \right)^{-1} \cdot \left(\int_L^R \phi dF_0 \right)^{-2} \cdot \left(\int_L^R \boldsymbol{u} dF \cdot \int_L^R \phi d(F - F_0) \right)^2 \, d\mathrm{P}$$

The first term (I) can be further calculated as

$$(I) \quad = \quad \int \left[\int_{S_t} \left(\int_L^R \phi dF_0 \right)^{-2} \left(\boldsymbol{u} \int_L^R \phi dF_0 - \phi \int_L^R \boldsymbol{u} dF_0 \right) d\mathrm{P}_t \right] d(F - F_0),$$

where S_t denotes the domain of $\boldsymbol{D} = (\boldsymbol{Y}, \boldsymbol{W}, L, R)$ given value t, and P_t denotes the induced conditional probability measure on S_t. By the regularity condition (M4) on the smoothness of ϕ and regularity condition (F6) on H, a straightforward algebra yields that

$$\kappa(\boldsymbol{\theta}, t) = \int_{S_t} \left(\int_L^R \phi dF_0 \right)^{-2} \left(\boldsymbol{u} \int_L^R \phi dF_0 - \phi \int_L^R \boldsymbol{u} dF_0 \right) d\mathrm{P}_t \tag{9}$$

is a C^1 function on $\boldsymbol{\Theta} \times [\tau_1, \tau_2]$.

Using the regularity condition 3, the terms (II) and (III) can be controlled by

$$\begin{aligned} \big| -(II) + (III) \big| \quad &\leq \quad \left(\frac{1}{K} \max(\|\boldsymbol{u}\|) \max(\Phi) + \frac{1}{K} \max(\|\boldsymbol{u}\|) \max(\Phi)^2 \right) \\ &\qquad \times \big\| F_n - F_0 \big\|_\infty^2 \\ &= \quad K \big\| F_n - F_0 \big\|_\infty^2 . \end{aligned}$$

Since the constant K does not depend on $\boldsymbol{\theta}$, we proved that

$$\boldsymbol{\Psi}(\boldsymbol{\theta}, F_n) - \boldsymbol{\Psi}(\boldsymbol{\theta}, F_0) = \int \kappa(\boldsymbol{\theta}, t) d(F_n - F_0) + \boldsymbol{O}_p(\|F_n - F_0\|_\infty^2), \tag{10}$$

uniformly on $\boldsymbol{\Theta}$, as required in Condition L4 of Lemma 2.

We are now ready to verify the four conditions in Theorem 1 as follows.

1. Condition T1.

   This is included in the regularity condition 2.

2. Condition T2: $\hat{\boldsymbol{\theta}}_n \xrightarrow{P} \boldsymbol{\theta}_0$.

   By definition, $\hat{\boldsymbol{\theta}}_n$ maximizes $\mathbb{P}_n \boldsymbol{M}_{\boldsymbol{\theta},\hat{F}_n}$, where $\boldsymbol{M}_{\boldsymbol{\theta},\hat{F}_n} = \log\left(\int_L^R \Phi d\hat{F}_n\right)$. Since $\mathcal{M}_{\boldsymbol{\Theta},\mathcal{F}_\delta}$ is a P-Glivenko–Cantelli class containing $\boldsymbol{M}_{\boldsymbol{\theta},\hat{F}_n}$, it follows that

   $$\max_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} |\big(\mathbb{P}_n - \mathrm{P}\big)\boldsymbol{M}_{\boldsymbol{\theta},\hat{F}_n}| \xrightarrow{P} 0.$$

   Since $n^{1/3}\|\hat{F}_n - F\|_\infty \xrightarrow{P} 0$ by Groeneboom and Wellner (1992), it can be easily shown by the Dominated Convergence Theorem (DCT) that

   $$\max_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} |\mathrm{P}\boldsymbol{M}_{\boldsymbol{\theta},\hat{F}_n} - \mathrm{P}\boldsymbol{M}_{\boldsymbol{\theta},F_0}| \xrightarrow{P} 0.$$

   Hence $\max_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} |\mathbb{P}_n\boldsymbol{M}_{\boldsymbol{\theta},\hat{F}_n} - \mathrm{P}\boldsymbol{M}_{\boldsymbol{\theta},F_0}| \xrightarrow{P} 0$ as well. It follows from the regularity condition 2 that $\mathrm{P}\boldsymbol{M}_{\boldsymbol{\theta},F_0}$ is strictly convex over $\boldsymbol{\Theta}$ with local maximum at $\boldsymbol{\theta}_0$, which implies

   $$\max_{\boldsymbol{\theta}\in\boldsymbol{\Theta},\|\boldsymbol{\theta}-\boldsymbol{\theta}_0\|>\epsilon} \mathrm{P}\boldsymbol{M}_{\boldsymbol{\theta},F_0} < \mathrm{P}\boldsymbol{M}_{\boldsymbol{\theta}_0,F_0}.$$

   Therefore by Theorem 5.7 of van der Vaart (1998), there exists a $\hat{\boldsymbol{\theta}}_n \in \boldsymbol{\Theta}$ that maximizes $\mathbb{P}_n\boldsymbol{M}_{\boldsymbol{\theta},\hat{F}_n}$ and $\hat{\boldsymbol{\theta}}_n \xrightarrow{P} \boldsymbol{\theta}_0$. In particular, when n is large, $\hat{\boldsymbol{\theta}}_n$ is a maximizer inside $\boldsymbol{\Theta}$ and hence a solution of $\boldsymbol{\Psi}_n(\boldsymbol{\theta}, \hat{F}_n) = \mathbf{0}$, which is the proposed estimate of the model parameter $\boldsymbol{\theta}$.

3. Condition T3: $\sqrt{n}\boldsymbol{\Psi}_n(\boldsymbol{\theta}_0, \hat{F}_n) \xrightarrow{P} \boldsymbol{Z}$ for a zero mean normal distribution $\boldsymbol{Z}$.

   Since $\hat{F}_n$ is the NPMLE from an interval censored data satisfying the regularity condition 1, the Hellinger differentiability (Geskus and Groeneboom, 1999, Pages 631-632) of $\boldsymbol{\Psi}(\boldsymbol{\theta}_0, F)$ with respect to F at F_0, as shown in Equation (10), implies that there exists a unique zero mean random variable $\mathbf{\Phi}(L,R)$ such that

$$\sqrt{n}\mathbf{\Psi}(\boldsymbol{\theta}_0,\hat{F}_n)=\sqrt{n}\mathbf{\Psi}(\boldsymbol{\theta}_0,\hat{F}_n)-\sqrt{n}\mathbf{\Psi}(\boldsymbol{\theta}_0,F_0)=\sqrt{n}\mathbb{P}_n\big(\mathbf{\Phi}(L,R)\big)+\boldsymbol{o}_p(1)$$

by Corollary 2.1 of Geskus and Groeneboom (1999) and Theorem 3.1 of van der Vaart (1991). The function $\mathbf{\Phi}(L,R)$ is characterized as the solution to the integral equation

$$\int_{L<t\leq R}\mathbf{\Phi}(L,R)d\mathrm{P}=\kappa(\boldsymbol{\theta}_0,t),$$

where $\kappa(\boldsymbol{\theta}_0,t)$ is given in Equation (9).

On the other hand, we have

$$\begin{aligned}
\sqrt{n}\mathbf{\Psi}_n(\boldsymbol{\theta}_0,\hat{F}_n)-\sqrt{n}\mathbf{\Psi}_n(\boldsymbol{\theta}_0,F_0)&=\sqrt{n}\mathbb{P}_n\left(\boldsymbol{U}_{\boldsymbol{\theta}_0,\hat{F}_n}-\boldsymbol{U}_{\boldsymbol{\theta}_0,F_0}\right)\\
&=\sqrt{n}(\mathbb{P}_n-\mathrm{P})\left(\boldsymbol{U}_{\boldsymbol{\theta}_0,\hat{F}_n}-\boldsymbol{U}_{\boldsymbol{\theta}_0,F_0}\right)+\sqrt{n}\mathrm{P}\left(\boldsymbol{U}_{\boldsymbol{\theta}_0,\hat{F}_n}-\boldsymbol{U}_{\boldsymbol{\theta}_0,F_0}\right)\\
&=\boldsymbol{o}_p(1)+\sqrt{n}\mathbf{\Psi}(\boldsymbol{\theta}_0,\hat{F}_n).
\end{aligned}$$

where the first term is $o_p(1)$ by Corollary 2.3.12 of van der Vaart and Wellner (1996), because $\boldsymbol{U}_{\boldsymbol{\theta}_0,\hat{F}_n}-\boldsymbol{U}_{\boldsymbol{\theta}_0,F_0}$ is in the P-Donsker class $\mathcal{DU}_{\mathbf{\Theta},\mathcal{F}_\delta}$ and $n^{1/3}\|\hat{F}_n-F\|_\infty\xrightarrow{P}0$ implies $\mathrm{P}\left(\boldsymbol{U}_{\boldsymbol{\theta}_0,\hat{F}_n}-\boldsymbol{U}_{\boldsymbol{\theta}_0,F_0}\right)^2\xrightarrow{P}\mathbf{0}$ by DCT. Thus, we have shown that

$$\begin{aligned}
\sqrt{n}\mathbf{\Psi}_n(\boldsymbol{\theta}_0,\hat{F}_n)&=\sqrt{n}\mathbf{\Psi}_n(\boldsymbol{\theta}_0,F_0)+\sqrt{n}\mathbf{\Psi}(\boldsymbol{\theta}_0,\hat{F}_n)+\boldsymbol{o}_p(1)\\
&=\sqrt{n}\mathbb{P}_n\big(\boldsymbol{U}_{\boldsymbol{\theta}_0,F_0}\big)+\sqrt{n}\mathbb{P}_n\big(\mathbf{\Phi}\big)+\boldsymbol{o}_p(1).
\end{aligned}$$

In particular, we have $\sqrt{n}\mathbf{\Psi}_n(\boldsymbol{\theta}_0,\hat{F}_n)\xrightarrow{P}\boldsymbol{Z}$, where $\boldsymbol{Z}$ is the limiting distribution of $\sqrt{n}\mathbb{P}_n\big(\boldsymbol{U}_{\boldsymbol{\theta}_0,F_0}+\mathbf{\Phi}\big)$, which is normally distributed with a zero mean.

4. Condition T4.

We verify the four sufficient conditions in Lemma 2. The first two conditions follow directly from the regularity conditions, and the $n^{1/3}$-convergence rate of $\hat{F}_n$ (Groeneboom and Wellner, 1992). The last condition follows from Equation (10). It remains to check the third condition.

For any $F \in \mathcal{F}$, we have

$$\begin{aligned}\sqrt{n}\Big(\boldsymbol{\Psi}_n(\hat{\boldsymbol{\theta}}_n, F) &- \boldsymbol{\Psi}(\hat{\boldsymbol{\theta}}_n, F)\Big) - \sqrt{n}\Big(\boldsymbol{\Psi}_n(\boldsymbol{\theta}_0, F) - \boldsymbol{\Psi}(\boldsymbol{\theta}_0, F)\Big) \\ &= \sqrt{n}(\mathbb{P}_n - \mathrm{P})\left(\boldsymbol{U}_{\hat{\boldsymbol{\theta}}_n,F} - \boldsymbol{U}_{\boldsymbol{\theta}_0,F}\right).\end{aligned}$$

Notice that $\boldsymbol{U}_{\hat{\boldsymbol{\theta}}_n,F} - \boldsymbol{U}_{\boldsymbol{\theta}_0,F}$ is a member in the P-Donsker class $\mathcal{DU}_{\boldsymbol{\Theta},\mathcal{F}_\delta}$. Using the regularity condition 3, the consistency of $\hat{\boldsymbol{\theta}}_n$, and the fact that the measure dF has support in the compact set $[\tau_1, \tau_2]$, it follows by DCT that $\mathrm{P}\left(\boldsymbol{U}_{\hat{\boldsymbol{\theta}}_n,F} - \boldsymbol{U}_{\boldsymbol{\theta}_0,F}\right)^2 \xrightarrow{P} \mathbf{0}$. Hence

$$\sqrt{n}(\mathbb{P}_n - \mathrm{P})\left(\boldsymbol{U}_{\hat{\boldsymbol{\theta}}_n,F} - \boldsymbol{U}_{\boldsymbol{\theta}_0,F}\right) = \boldsymbol{o}_p(1)$$

by Corollary 2.3.12 of van der Vaart and Wellner (1996). This verifies the third condition in Lemma 2.

So Condition T4 is satisfied by Lemma 2.

Finally, we complete the proof of Theorem 2.1 by applying Theorem 1 to obtain

$$\sqrt{n}\big(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0\big) = \boldsymbol{A}^{-1} \cdot \sqrt{n}\mathbb{P}_n\big(\boldsymbol{U}_{\boldsymbol{\theta}_0,F_0} + \boldsymbol{\Phi}\big) + \boldsymbol{o}_p(1),$$

where $\boldsymbol{A}$ is the information matrix, which is also the negative of Jacobian of $\mathrm{P}\left(\boldsymbol{U}_{\boldsymbol{\theta},F_0}\right)$ at $\boldsymbol{\theta} = \boldsymbol{\theta}_0$. So we have $\sqrt{n}\big(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0\big) \xrightarrow{P} \boldsymbol{N}(\mathbf{0}, \boldsymbol{\Sigma})$, where $\boldsymbol{\Sigma} = \boldsymbol{A}^{-1}\mathrm{P}\left((\boldsymbol{U}_{\boldsymbol{\theta}_0,F_0} + \boldsymbol{\Phi})^{\otimes 2}\right)\boldsymbol{A}^{-1}$.

Since $\boldsymbol{A} = \mathrm{P}\left(\boldsymbol{U}_{\boldsymbol{\theta}_0, F_0}^{\otimes 2}\right)$, the asymptotic variance matrix $\boldsymbol{\Sigma}$ can be decomposed as

$$\boldsymbol{\Sigma} = \boldsymbol{A}^{-1} + \boldsymbol{A}^{-1}\mathrm{P}\left(\boldsymbol{U}_{\boldsymbol{\theta}_0, F_0}\boldsymbol{\Phi}^t + \boldsymbol{\Phi}\boldsymbol{U}_{\boldsymbol{\theta}_0, F_0}^t + \boldsymbol{\Phi}^{\otimes 2}\right)\boldsymbol{A}^{-1}$$

as in Theorem 2.1.

Table 1: Simulation results for Scenario 1

	sample size =200					sample size =400				
	λ	β_1	β_2	α	β	λ	β_1	β_2	α	β
Proposed Model										
% Bias	0.073	0.834	0.976	1.145	0.838	0.067	0.063	0.620	0.640	0.541
M-C SD	0.571	0.469	0.812	0.415	0.432	0.383	0.335	0.559	0.284	0.290
Av. SE	0.565	0.468	0.803	0.400	0.418	0.395	0.328	0.568	0.283	0.294
95% CP	0.947	0.942	0.935	0.921	0.923	0.951	0.943	0.945	0.940	0.948
Midpoint Imputation										
% Bias	0.437	0.710	0.784	8.076	5.234	0.373	0.341	0.351	6.482	4.100
M-C SD	0.917	0.506	0.878	1.370	1.414	0.620	0.368	0.623	0.931	0.953
Av. SE	0.895	0.502	0.871	1.316	1.359	0.631	0.358	0.621	0.930	0.961
95% CP	0.944	0.947	0.943	0.929	0.938	0.939	0.954	0.950	0.933	0.935
Known Anchoring Time										
% Bias	0.021	1.194	0.868	0.048	0.041	0.009	0.078	0.526	0.256	0.087
M-C SD	0.536	0.447	0.769	0.390	0.397	0.353	0.323	0.533	0.253	0.263
Av. SE	0.523	0.442	0.766	0.357	0.368	0.370	0.314	0.544	0.254	0.262
95% CP	0.947	0.938	0.945	0.929	0.930	0.965	0.945	0.954	0.944	0.946

Table 2: Simulation results for Scenario 2

	sample size =200					sample size =400				
	λ	β_1	β_2	α	β	λ	β_1	β_2	α	β
Proposed Model										
% Bias	0.119	0.148	0.051	0.607	0.436	0.112	0.063	1.098	0.501	0.328
M-C SD	0.612	0.496	0.840	0.285	0.312	0.420	0.338	0.597	0.202	0.212
Av. SE	0.616	0.492	0.843	0.287	0.305	0.433	0.343	0.597	0.203	0.212
95% CP	0.949	0.938	0.944	0.946	0.938	0.955	0.958	0.947	0.944	0.945
Midpoint Imputation										
% Bias	0.855	0.155	0.694	7.630	4.895	0.734	0.571	0.792	6.464	3.903
M-C SD	1.029	0.616	1.083	0.811	0.848	0.712	0.439	0.754	0.545	0.573
Av. SE	1.011	0.611	1.060	0.779	0.823	0.712	0.435	0.754	0.550	0.580
95% CP	0.928	0.948	0.950	0.916	0.916	0.917	0.953	0.948	0.915	0.927
Known Anchoring Time										
% Bias	0.005	0.198	0.017	0.001	0.126	0.021	0.056	0.915	0.081	0.098
M-C SD	0.569	0.475	0.788	0.281	0.298	0.382	0.328	0.560	0.192	0.199
Av. SE	0.556	0.458	0.794	0.269	0.280	0.393	0.325	0.563	0.191	0.198
95% CP	0.952	0.928	0.944	0.944	0.937	0.957	0.946	0.946	0.951	0.953

Table 3: Ratio of Monte Carlo standard deviations: proposed model v.s. the model knowing true anchoring event times

	sample size =200					sample size =400				
	λ	β_1	β_2	α	β	λ	β_1	β_2	α	β
Scenario (1)										
ratio	1.065	1.049	1.056	1.064	1.088	1.085	1.037	1.049	1.123	1.103
Scenario (2)										
ratio	1.076	1.044	1.066	1.014	1.047	1.099	1.030	1.066	1.052	1.065

Figure 1: Peak growth intervals and observed weight

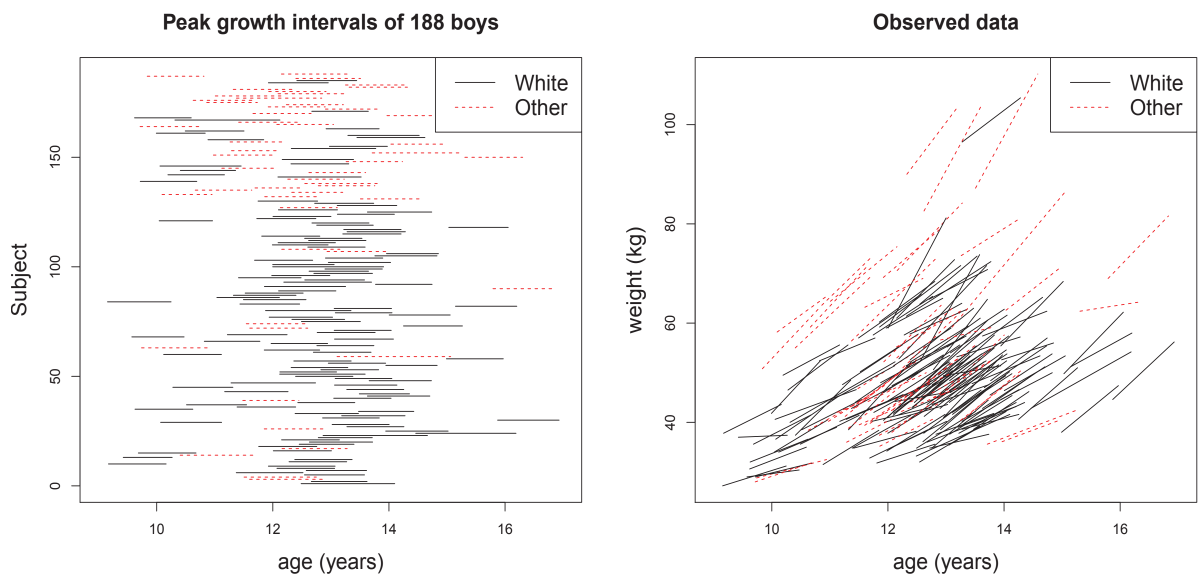

Table 4: Parameter estimates of the example data

	λ_1	λ_2	α_1	β_1	α_2	β_2
estimate	54.478	-6.927	7.526	9.730	-1.583	-0.543
se	1.901	1.992	0.811	1.005	0.971	1.256