Ann Inst Stat Math (2013) 65:639–665
DOI 10.1007/s10463-012-0387-4

Empirical likelihood semiparametric nonlinear regression analysis for longitudinal data with responses missing at random

Nian-Sheng Tang · Pu-Ying Zhao

Received: 9 April 2012 / Revised: 10 September 2012 / Published online: 18 November 2012

Abstract This paper develops the empirical likelihood (EL) inference on parameters and baseline function in a semiparametric nonlinear regression model for longitudinal data in the presence of missing response variables. We propose two EL-based ratio statistics for regression coefficients by introducing the working covariance matrix and a residual-adjusted EL ratio statistic for baseline function. We establish asymptotic properties of the EL estimators for regression coefficients and baseline function. Simulation studies are used to investigate the finite sample performance of our proposed EL methodologies. An AIDS clinical trial data set is used to illustrate our proposed methodologies.

Keywords Empirical likelihood · Imputation · Longitudinal data · Missing at random · Semiparametric nonlinear regression model

1 Introduction

Longitudinal data are often encountered in economical, psychological, biomedical, behavioral, educational and social research. In longitudinal studies, subjects are observed repeatedly over time and responses of interest are recorded together with covariates. Semiparametric regression models are often employed to fit various longitudinal data because the parametric part provides an interpretable data summary and the nonparametric functions provide flexibility to all the data to decide some unknown or uncertain components such as the shape of the mean response over time. Various statistical methods have been developed to estimate the regression coefficients and smoothing functions in a semiparametric regression model in past years. For example,

N.-S. Tang (✉) · P.-Y. Zhao
Department of Statistics, Yunnan University, Kunming 650091, People's Republic of China
e-mail: nstang@ynu.edu.cn

see Green (1987), Zeger and Diggle (1994), Lin and Carroll (2001), Ruppert et al. (2003) and Fan and Li (2004). However, nonlinear relations among the covariates are important for developing more reasonable and meaningful models, see Bates and Watts (1988). Recently, semiparametric nonlinear regression models have received considerable attention, for example, see Zhu et al. (2000), Li and Nie (2008), and Wang and Ke (2009). These existing theories and methods have been developed under the assumption that responses or covariates in semiparametric nonlinear regression models are not subject to missingness. Hence, this paper aimed to develop an inference procedure for regression coefficients and smoothing functions in semiparametric nonlinear regression models with missing responses at random.

Since missing data are often encountered in various fields, such as surveys, clinical trials and longitudinal studies (Little and Rubin 2002) due to some potential reasons such as study drop-out or study subjects' refusal to answer some items on a questionnaire or failing to attend a scheduled clinic visit, various methods have been developed to analyse semiparametric regression models with missing data. For example, see Yi and Cook (2002), Shardell and Miller (2008), Chen et al. (2008). Particularly, EL inference for semiparametric regression models with missing data has received a lot of attention in recent years because it is especially useful for constructing confidence intervals or regions of parameters of interest in the considered models. For example, see Wang et al. (2004), Liang et al. (2007), Liang and Qin (2008), Xue and Xue (2011). Also, nonlinear regression models with responses missing at random were studied in recent years, for example, see Müller (2009) and Ciuperca (2011). However, it is more challenging to deal with semiparametric nonlinear regression models for longitudinal data with missing responses at random due to nonlinearity of unknown regression coefficients and the within-subject correlation. Moreover, there is little work done on the development of the EL method for semiparametric nonlinear regression models for longitudinal data with missing responses at random.

The aim of this paper was to develop a general EL inference procedure for parameters and baseline function using the complete-case data set or the imputed values in a semiparametric nonlinear regression model for longitudinal data with responses missing at random. In our proposed methods, the value for a missing response is imputed using the inverse-probability weighted imputed method, and the within-subject correlation structure is considered by introducing the working covariance matrix into the proposed auxiliary random vectors. Particularly, to avoid selecting the optimal bandwidth and the so-called "curse of dimensionality" in estimating selection probability function via the kernel method, we employ a logistic regression model, which is widely used in missing data analysis (see Ibrahim et al. 2001; Lee and Tang 2006; Chen and Zhong 2010), to evaluate estimation of the selection probability function by maximizing the corresponding likelihood function of the given logistic model. Our proposed EL method has the following features: (1) the EL ratio statistic on β follows asymptotically the central Chi-squared distribution, which can be directly used to construct confidence regions of the parameters without any extra Monte Carlo approximation needed when our proposed EL method is not used; (2) unlike normal-approximation-based (NA-based) method for constructing confidence region on β, a consistent estimator of the asymptotic covariance matrix is not needed; (3) our empirical results show that the EL-based method has advantage over the NA-based in terms of the coverage probability and the interval width; and (4) our proposed theoretical results are new since other literature only considered nonlinear models with responses missing at random (Ciuperca 2011) or semiparametric linear regression models with responses missing at random or within-subject independence structure. We here extend the EL inference for semiparametric regression models with missing responses at random to semiparametric nonlinear regression models for longitudinal data with missing responses at random by incorporating the within-subject correlation into the constructed auxiliary vectors. We systematically investigate the asymptotic properties of the maximum EL estimators (MELEs) under this new setting.

The rest of the paper is organized as follows: Section 2 outlines the formulations of two ELs for β based on the complete-case data and the inverse probability weighted imputation technique. We propose a calibrated method for constructing EL ratios and an imputation estimator for $g(t)$ in Sect. 2. In Sect. 3, we establish the asymptotic properties of the proposed three EL ratio functions and their corresponding EL estimators. Numerical illustrations including two simulation studies and a real example are presented to compare the finite sample performance of the proposed methods in Sect. 4. Some concluding remarks are given in Sect. 5. Technical details are presented in the Appendix.

2 Methods

2.1 Model and notation

Consider a data set from n independent subjects. For the ith subject, we suppose that Y_{ij} is the observed value of a scale response variable at time T_{ij}, and X_{ij} is the corresponding $p \times 1$ covariate vector for $i = 1, \ldots, n$, $j = 1, \ldots, n_i$. Under the abovementioned assumption, a semiparametric nonlinear regression model can be written as

$$Y_{ij} = f(X_{ij}; \beta) + g(T_{ij}) + \varepsilon_{ij} \tag{1}$$

for $i = 1, \ldots, n,\ \ j = 1, \ldots, n_i$. Here $f(X; \beta)$ is a twice continuously differentiable function with respect to β (a p-dimensional unknown parameter) but is nonlinear with respect to β; $g(\cdot)$ is an unknown regression function defined on the interval [0, 1]. The time points T_{ij} are known design points. We assume that ε_{ij} satisfies $E(\varepsilon_{ij}|X_{ij}, T_{ij}) = 0$, and $\varepsilon_1, \ldots, \varepsilon_n$ are mutually independent with zero mean and the positive definite covariance matrix Σ_i, i.e. $\text{var}(\varepsilon_i) = \Sigma_i$, where $\varepsilon_i = (\varepsilon_{i1}, \ldots, \varepsilon_{in_i})^{\mathrm{T}}$ for $i = 1, \ldots, n$.

Throughout this paper, we assume that Y_{ij}'s are subject to missingness and X_{ij}'s and T_{ij}'s are completely observed. Let $\delta_{ij} = 0$ if Y_{ij} is missing and $\delta_{ij} = 1$ if Y_{ij} is observed. Generally, the missing components may vary across different subjects. Here we assume that Y_{ij} is missing at random (MAR), i.e. δ_{ij} and Y_{ij} are conditionally independent given X_{ij} and T_{ij}: $P(\delta_{ij} = 1|X_{ij}, Y_{ij}, T_{ij}) = P(\delta_{ij} = 1|X_{ij}, T_{ij}) \triangleq p(X_{ij}, T_{ij})$. It is assumed that δ_{ij} is independent of δ_{ik} for any $j \neq k$. Without loss of generality, we also assume that T_{ij}'s are all scaled into the interval [0, 1].

For simplicity, we consider the following missingness data mechanism model:

$$P(\delta_{ij}=1|X_{ij},Y_{ij},T_{ij})=p(X_{ij},T_{ij};\gamma)=\frac{\exp\{\gamma_0+\gamma_1^{\mathrm{T}}X_{ij}+\gamma_2T_{ij}\}}{1+\exp\{\gamma_0+\gamma_1^{\mathrm{T}}X_{ij}+\gamma_2T_{ij}\}}, \quad (2)$$

where $\gamma_1=(\gamma_{11},\ldots,\gamma_{1q})^{\mathrm{T}}$, γ_0 is a constant term and $\gamma=(\gamma_0,\gamma_1^{\mathrm{T}},\gamma_2)^{\mathrm{T}}$. The logistic regression model (2) is a widely used model in many missing data literature, for example, see Ibrahim et al. (2001) and Lee and Tang (2006) and among others. In fact, model (2) can be also relaxed by assuming a more complicated interaction/quadratic covariates parametric model or a nonparametric model or an exponential tilting model for missingness data mechanism as done in many missing data literature, for instance, see Liang et al. (2007), Wang et al. (2004), Kim and Yu (2011) and among others. Also, model (2) can be regarded as a first-order approximation to nonparametric function $p(x,t)$ and it can avoid selecting the optimal bandwidth and the so-called "curse-of-dimensionality" in estimating selection probability via the kernel method.

Parameter γ can be estimated by maximizing the following binary likelihood:

$$L(\gamma)=\prod_{i=1}^{n}\prod_{j=1}^{n_i}p(X_{ij},T_{ij};\gamma)^{\delta_{ij}}(1-p(X_{ij},T_{ij};\gamma))^{1-\delta_{ij}}.$$

The re-weighted least squares iterative algorithm can be used to obtain consistent estimator $\hat{\gamma}$ of unknown parameter γ.

2.2 MELE of β with the complete-case data

To delete the incomplete cases, we pre-multiply (1) by the observation indicator δ_{ij}, which yields $\delta_{ij}Y_{ij}=\delta_{ij}f(X_{ij};\beta)+\delta_{ij}g(T_{ij})+\delta_{ij}\varepsilon_{ij}$. It follows from the above assumptions that $E(\delta_{ij}Y_{ij}|T_{ij}=t)=E(\delta_{ij}f(X_{ij};\beta)|T_{ij}=t)+E(\delta_{ij}|T_{ij}=t)g(t)$. Let $g_2^{\mathrm{C}}(t)=E(\delta_{ij}Y_{ij}|T_{ij}=t)/E(\delta_{ij}|T_{ij}=t)$ and $g_1^{\mathrm{C}}(t;\beta)=E(\delta_{ij}f(X_{ij};\beta)|T_{ij}=t)/E(\delta_{ij}|T_{ij}=t)$. Then, we obtain $g(t)=g_2^{\mathrm{C}}(t)-g_1^{\mathrm{C}}(t;\beta)$. The kernel estimators of $g_1^{\mathrm{C}}(t;\beta)$ and $g_2^{\mathrm{C}}(t)$ are

$$\hat{g}_1^{\mathrm{C}}(t;\beta)=\sum_{i=1}^{n}\sum_{j=1}^{n_i}W_{ij}^{\mathrm{C}}(t)f(X_{ij};\beta)\ \text{ and }\ \hat{g}_2^{\mathrm{C}}(t)=\sum_{i=1}^{n}\sum_{j=1}^{n_i}W_{ij}^{\mathrm{C}}(t)Y_{ij}, \quad (3)$$

respectively, where $W_{ij}^{\mathrm{C}}(t)=\delta_{ij}K_h(T_{ij}-t)/\{\sum_{k=1}^{n}\sum_{l=1}^{n_k}\delta_{kl}K_h(T_{kl}-t)\}$ is the kernel weight function, $K_h(t)=K(t/h)$ in which $K(u)$ is a kernel function on the real line, $h=h_n$ is a positive smoothing bandwidth sequence such that $h_n\to 0$ and $nh_n\to\infty$ as $n\to\infty$. It is easily shown that $\hat{g}_1^{\mathrm{C}}(t;\beta)$ and $\hat{g}_2^{\mathrm{C}}(t)$ are the consistent estimators of $g_1^{\mathrm{C}}(t;\beta)$ and $g_2^{\mathrm{C}}(t)$, respectively, and $\hat{g}(t)=\hat{g}_2^{\mathrm{C}}(t)-\hat{g}_1^{\mathrm{C}}(t;\beta)$ is also a consistent estimator of $g(t)$.

Let $\tilde{y}_{ij}=Y_{ij}-\sum_{k=1}^{n}\sum_{l=1}^{n_k}W_{kl}^{\mathrm{C}}(T_{ij})Y_{kl}$, $\tilde{f}_{ij}(\beta)=f(X_{ij};\beta)-\sum_{k=1}^{n}\sum_{l=1}^{n_k}W_{kl}^{\mathrm{C}}(T_{ij})f(X_{kl};\beta)$, $\tilde{d}_{ij}(\beta)=d_{ij}(\beta)-\sum_{k=1}^{n}\sum_{l=1}^{n_k}W_{kl}^{\mathrm{C}}(T_{ij})d_{kl}(\beta)$ with $d_{ij}(\beta)=\partial f(X_{ij};\beta)/\partial\beta$

for $j = 1, \ldots, n_i$, $\tilde{y}_i = (\tilde{y}_{i1}, \ldots, \tilde{y}_{in_i})^{\mathrm{T}}$, $\tilde{f}_i(\beta) = (\tilde{f}_{i1}(\beta), \ldots, \tilde{f}_{in_i}(\beta))^{\mathrm{T}}$, $\Delta_i = \mathrm{diag}(\delta_{i1}, \ldots, \delta_{in_i})$ and $D_i(\beta) = (\tilde{d}_{i1}(\beta), \ldots, \tilde{d}_{in_i}(\beta))^{\mathrm{T}}$ for $i = 1, \ldots, n$. To develop the EL procedure for β, we consider the following auxiliary random vectors:

$$Z_{i1}(\beta) = D_i^{\mathrm{T}}(\beta)\Delta_i V_i^{-1}\Delta_i(\tilde{y}_i - \tilde{f}_i(\beta)), \quad i = 1, \ldots, n, \tag{4}$$

where V_i is an arbitrarily specified working covariance matrix. If $V_i = I$ (a $n_i \times n_i$ identity matrix), the observations within the same subject are independent; if V_i is the true covariance matrix of n_i observations for the ith subject, the within-subject correlation structures for the longitudinal data are considered. When the working covariance matrix V_i is unknown, we should first use the method of moments (e.g., see Lin and Carroll 2001) to estimate it and then discuss statistical inference on β based on estimator of V_i. For example, V_i can be estimated by $n^{-1}\sum_{i=1}^n \tilde{e}_i\tilde{e}_i^{\mathrm{T}}$, where $\tilde{e}_i = \tilde{y}_i^o - \tilde{f}_i^o(\hat{\beta})$, $\tilde{y}_i^o = (\tilde{y}_{i1}^o, \ldots, \tilde{y}_{in_i}^o)^{\mathrm{T}}$, $\tilde{f}_i^o(\hat{\beta}) = (\tilde{f}_{i1}^o(\hat{\beta}), \ldots, \tilde{f}_{in_i}^o(\hat{\beta}))^{\mathrm{T}}$, $\tilde{y}_{ij}^o = Y_{ij}^o - \sum_{k=1}^n\sum_{l=1}^{n_k} W_{kl}(T_{ij})Y_{kl}^o$ with $Y_{ij}^o = \delta_{ij}Y_{ij} + (1-\delta_{ij})(f(X_{ij}, \hat{\beta}) + \hat{g}(T_{ij}))$, $\tilde{f}_{ij}^o(\hat{\beta}) = f(X_{ij}; \hat{\beta}) - \sum_{k=1}^n\sum_{l=1}^{n_k} W_{kl}(T_{ij})f(X_{kl}; \hat{\beta})$, and $\hat{\beta}$ is obtained by solving the following equation: $n^{-1}\sum_{i=1}^n Z_{i1}(\beta) = 0$ with $V_i = I$ in Eq. (4).

Without loss of generality, we assume that V_i is known in this paper. It can be shown from MAR assumption that $E(Z_{i1}(\beta)) = 0$ when β is the true parameter. Thus, the true parameter β can be estimated from the completely observed data using the following estimating equations: $E\{H(\beta)\} = 0$, where $H(\beta) = n^{-1}\sum_{i=1}^n Z_{i1}(\beta)$, which shows that estimate (say $\hat{\beta}_M$) of parameter β can be obtained by using the following iterative formula:

$$\beta^{(k+1)} = \beta^{(k)} + \left\{\sum_{i=1}^n D_i^{\mathrm{T}}(\beta)\Delta_i V_i^{-1}\Delta_i D_i(\beta)\right\}^{-1} \times \left\{\sum_{i=1}^n D_i^{\mathrm{T}}(\beta)\Delta_i V_i^{-1}\Delta_i(\tilde{y}_i - \tilde{f}_i(\beta))\right\}, \quad k = 0, 1, \ldots,$$

where $\beta^{(k+1)}$ is the value of β at the kth iteration, and $D_i(\beta)$ and $\tilde{f}_i(\beta)$ are evaluated at $\beta^{(k)}$. Here $\hat{\beta}_M$ is referred to as the generalized least squares estimator (GLSE). It is easily seen from the above iterative formula that when the rank of $\sum_{i=1}^n D_i^{\mathrm{T}}(\beta)\Delta_i V_i^{-1}\Delta_i D_i(\beta)$ is less than p, it is impossible to implement the above iterative procedure. The EL method of Owen (2001) is a very powerful nonparametric method for making inference on β based on the estimating equation $E\{H(\beta)\} = 0$ and it has many advantages over NA-based method (Owen 2001). For example, it has better small sample performance than NA-based approach, and EL-based confidence regions are range preserving and transformation respecting and the regularity conditions for EL-based method are weak and natural. The EL method has become increasingly common in recent years and has been used widely in many applied areas (Wang et al. 2004; Liang and Qin 2008; Ciuperca 2011). Hence, an alternative EL approach is developed to obtain estimator of parameter β and construct the confidence interval of β based on estimating equations $E\{H(\beta)\} = 0$ as follows:

Let p_i be the probability weight allocated to $Z_{i1}(\beta)$ such that $\sum_{i=1}^n p_i = 1$ and $p_i \geq 0$ for each i. The EL for β based on $H(\beta)$ can be defined as

$$L_n(\beta) = \sup\left\{\prod_{i=1}^n p_i \mid p_i \geq 0, \sum_{i=1}^n p_i = 1, \sum_{i=1}^n p_i Z_{i1}(\beta) = 0\right\}.$$

Using the Lagrange multiplier method, the optimal value of p_i is $\hat{p}_i = n^{-1}\{1 + \lambda_{n1}^{\mathrm{T}}(\beta)Z_{i1}(\beta)\}^{-1}$, where $\lambda_{n1}(\beta)$ (an $p \times 1$ vector) is the Lagrange multiplier and satisfies $Q_{n1}(\beta, \lambda_{n1}) = n^{-1}\sum_{i=1}^n Z_{i1}(\beta)/\{1 + \lambda_{n1}^{\mathrm{T}}(\beta)Z_{i1}(\beta)\} = 0$. Then, the log empirical likelihood ratio function (LELRF) for β with the complete-case data is

$$\ell_c(\beta) = -2\log\left\{\prod_{i=1}^n (n\hat{p}_i)\right\} = 2\sum_{i=1}^n \log\{1 + \lambda_{n1}^{\mathrm{T}}(\beta)Z_{i1}(\beta)\}. \tag{5}$$

Maximizing $-\ell_c(\beta)$ yields the MELE of β, denoted by $\hat{\beta}_c$. Under some regular conditions, $\hat{\beta}_c$ can be obtained by simultaneously solving the following two equations: $Q_{n1}(\beta, \lambda_{n1}) = 0$ and $Q_{n2}(\beta, \lambda_{n1}) = n^{-1}\sum_{i=1}^n \lambda_{n1}^{\mathrm{T}}(\beta) \times \partial_\beta Z_{i1}(\beta)/\{1 + \lambda_{n1}^{\mathrm{T}}(\beta)Z_{i1}(\beta)\} = 0$, where ∂_β represents taking partial derivative with respect to β. An estimator of $g(t)$ with the complete-case data is $\hat{g}_{\mathrm{C}}(t) = \hat{g}_2^{\mathrm{C}}(t) - \hat{g}_1^{\mathrm{C}}(t; \hat{\beta}_c)$.

2.3 MELE of β with the imputed values

Clearly, the above-presented EL with the complete-case data do not completely use all the information contained in the data set $\{(X_{ij}, Y_{ij}, T_{ij}, \delta_{ij}) : i = 1, \ldots, n, j = 1, \ldots, n_i\}$. In particular, when the proportion of missing responses is large, statistical inference such as estimator of parameter β and its confidence region based on $\ell_c(\beta)$ may lead to unreasonable conclusions. To overcome the above-mentioned shortcomings, the imputation method is here employed to deal with missing values of responses in model (1). Inspired by linear regression imputation (Yates 1933), we impute $\tilde{y}_{ij}$ by $\widetilde{f}_{ij}(\hat{\beta}_c))$ if Y_{ij} is missing and obtain the imputed values of $\tilde{y}_{ij}$ by $\tilde{y}_{ij}^* = \delta_{ij}\tilde{y}_{ij}/p(X_{ij}, T_{ij}) + (1 - \delta_{ij}/p(X_{ij}, T_{ij}))\widetilde{f}_{ij}(\hat{\beta}_c))$. In this case, when V_i is unknown, V_i can be estimated by $n^{-1}\sum_{i=1}^n \tilde{\epsilon}_i\tilde{\epsilon}_i^{\mathrm{T}}$, where $\tilde{\epsilon}_i = \tilde{y}_i^* - \tilde{f}_i(\hat{\beta}_c)$, $\tilde{y}_i^* = (\tilde{y}_{i1}^*, \ldots, \tilde{y}_{in_i}^*)^{\mathrm{T}}$ and $\tilde{f}_i(\hat{\beta}_c) = (\tilde{f}_{i1}(\hat{\beta}_c), \ldots, \tilde{f}_{in_i}(\hat{\beta}_c))^{\mathrm{T}}$ for $i = 1, \ldots, n$. Then, we introduce the following auxiliary random vectors:

$$Z_{i2}(\beta) = D_i^{\mathrm{T}}(\beta)V_i^{-1}(\tilde{y}_i^* - \tilde{f}_i(\beta)), \quad i = 1, \ldots, n.$$

The empirical log-likelihood for β based on the imputed values can be defined as

$$\ell_I(\beta) = -2\max\left\{\sum_{i=1}^n \log(np_i) \mid p_i \geq 0, \sum_{i=1}^n p_i = 1, \sum_{i=1}^n p_i Z_{i2}(\beta) = 0\right\}.$$

Clearly, $\ell_I(\beta)$ is more reasonable than the empirical log-likelihood $\ell_c(\beta)$ because it fully explores the information contained in the data set. Then, the LELRF for β is $\ell_I(\beta) = 2\sum_{i=1}^n \log\{1+\lambda_{n2}^{\mathrm{T}} Z_{i2}(\beta)\}$, where λ_{n2} is the Lagrange multiplier and satisfies $M_{n1}(\beta, \lambda_{n2}) = n^{-1}\sum_{i=1}^n Z_{i2}(\beta)/\{1+\lambda_{n2}^{\mathrm{T}} Z_{i2}(\beta)\} = 0$. Maximizing $-\ell_I(\beta)$ leads to the MELE of β, denoted by $\hat{\beta}_I$. Under some regular conditions, $\hat{\beta}_I$ can be obtained by simultaneously solving the two equations: $M_{n1}(\beta, \lambda_{n2}) = 0$ and $M_{n2}(\beta, \lambda_{n2}) = n^{-1}\sum_{i=1}^n \lambda_{n2}^{\mathrm{T}} \partial_\beta Z_{i2}(\beta)/\{1+\lambda_{n2}^{\mathrm{T}} Z_{i2}(\beta)\} = 0$. And an estimator of $g(t)$ with the imputed values of missing responses is $\hat{g}_{\mathrm{I}}(t) = \hat{g}_2^{\mathrm{C}}(t) - \hat{g}_1^{\mathrm{C}}(t; \hat{\beta}_I)$.

2.4 Maximum residual-adjusted EL estimator for $g(t)$

By Eq. (3), $g(t)$ can be estimated by $\hat{g}_C(t) = \sum_{i=1}^n \sum_{j=1}^{n_i} W_{ij}^{\mathrm{C}}(t)(Y_{ij} - f(X_{ij}; \hat{\beta}_c)) \triangleq \hat{g}_2^{\mathrm{C}}(t) - \hat{g}_1^{\mathrm{C}}(t; \hat{\beta}_c)$. Then, it follows from Eq. (1) that for any $t \in [0, 1]$, we have

$$\hat{g}_C(t) - g(t) = \sum_{i=1}^n \sum_{j=1}^{n_i} W_{ij}^{\mathrm{C}}(t)\{\varepsilon_{ij} + f(X_{ij}; \beta) - f(X_{ij}; \hat{\beta}_c) + g(T_{ij})\} - g(t).$$

It follows from $\sum_{i=1}^n \sum_{j=1}^{n_i} W_{ij}^{\mathrm{C}}(t)\{g(T_{ij}) - g(t)\} = O_p(h^2)$ that the LELRF for $g(t)$ constructed from $\hat{g}_C(t)$ is not asymptotically distributed as a Chi-squared distribution. To overcome the above-mentioned difficulties, a modified estimator of $g(t)$ is defined by $\hat{g}_{\mathrm{MC}}(t) = \sum_{i=1}^n \sum_{j=1}^{n_i} W_{ij}^{\mathrm{C}}(t)\{Y_{ij} - f(X_{ij}; \hat{\beta}_c) - (\hat{g}_C(T_{ij}) - \hat{g}_C(t))\}$. Then, we can define the following auxiliary random variables $\hat{\eta}_{iR}(g(t)) = \sum_{j=1}^{n_i} \delta_{ij}\{Y_{ij} - f(X_{ij}; \hat{\beta}_c) - g(t) - (\hat{g}_C(T_{ij}) - \hat{g}_C(t))\}K_h(T_{ij} - t)$ for $i = 1, \ldots, n$. A residual-adjusted EL for $g(t)$ can be defined as

$$\ell_R(g(t)) = -2\max\left\{\sum_{i=1}^n \log(np_i) | p_i \geq 0, \sum_{i=1}^n p_i = 1, \sum_{i=1}^n p_i \hat{\eta}_{iR}(g(t)) = 0\right\}.$$

The LELRF for $g(t)$ with the complete-case data is $\ell_R(g(t)) = 2\sum_{i=1}^n \log\{1 + \lambda_{n3}\hat{\eta}_{iR}(g(t))\}$, where λ_{n3} satisfies $S_{n1}(g(t), \lambda_{n3}) = n^{-1}\sum_{i=1}^n \hat{\eta}_{iR}(g(t))/\{1 + \lambda_{n3}\hat{\eta}_{iR}(g(t))\} = 0$. Maximizing $-\ell_R(g(t))$ results in the maximum residual-adjusted EL estimator of $g(t)$, denoted as $\hat{g}(t)$.

2.5 Imputation estimator for $g(t)$

All the above-presented estimators for $g(t)$ are obtained from the complete-case data set and do not sufficiently use the information contained in the data set, which may yield bias estimator of $g(t)$. Motivated by the imputation method for missing responses given in Sect. 2.3, we propose an imputation estimator for $g(t)$ as follows:

Let $\tilde{Y}_{ij}^I = \delta_{ij}Y_{ij}/p(X_{ij}, T_{ij}) + (1 - \delta_{ij}/p(X_{ij}, T_{ij}))(f(X_{ij}; \beta) + g(T_{ij}))$. Under MAR assumption, it can be shown that $E(\tilde{Y}_{ij}^I | X_{ij}, T_{ij}) = f(X_{ij}; \beta) + g(T_{ij})$, which implies $\tilde{Y}_{ij}^I = f(X_{ij}; \beta) + g(T_{ij}) + \epsilon_{ij}$ for $i = 1, \ldots, n$ and $j = 1, \ldots, n_i$, where

$E(\epsilon_{ij}|X_{ij}, T_{ij}) = 0$. Let $g_1(t; \beta) = E\{f(X_{ij}; \beta)|T_{ij} = t\}$ and $g_2(t) = E\{\tilde{Y}_{ij}^I|T_{ij} = t\}$, which implies that $g(t) = g_2(t) - g_1(t; \beta)$. The kernel estimators of $g_1(t; \beta)$ and $g_2(t)$ are

$$\hat{g}_1^{\mathrm{IP}}(t; \beta) = \sum_{i=1}^{n}\sum_{j=1}^{n_i} W_{ij}(t) f(X_{ij}; \beta) \ \text{ and } \ \hat{g}_2^{\mathrm{IP}}(t) = \sum_{i=1}^{n}\sum_{j=1}^{n_i} W_{ij}(t)\tilde{Y}_{ij}^I, \tag{6}$$

respectively, where $W_{ij}(t) = K_b(T_{ij} - t)/\sum_{k=1}^n \sum_{l=1}^{n_k} K_b(T_{kl} - t)$ is a kernel weight function. Under some regular conditions, we can show that $\hat{g}_1^{\mathrm{IP}}(t; \beta)$ and $\hat{g}_2^{\mathrm{IP}}(t)$ are the consistent estimators of $g_1(t; \beta)$ and $g_2(t)$, respectively, and $\hat{g}^{\mathrm{IP}}(t) = \hat{g}_2^{\mathrm{IP}}(t) - \hat{g}_1^{\mathrm{IP}}(t; \beta)$ is a consistent estimator of $g(t)$. Unfortunately, $\tilde{Y}_{ij}^I$ contains unknown parameter β and nonparametric function $g(T_{ij})$. A natural idea for solving this problem is to replace these unknown quantities by their corresponding estimators. Here, using $\hat{\beta}_I$ (defined in Sect. 2.3) and $\hat{g}(t)$ (defined in Sect. 2.4) to replace β and $g(t)$ in $\hat{g}_1^{\mathrm{IP}}(t; \beta)$ and $\hat{g}_2^{\mathrm{IP}}(t)$ leads to a new estimator of $g(t)$, which is given by $\hat{g}^{\mathrm{MIP}}(t) = \hat{g}_2^{\mathrm{MIP}}(t) - \hat{g}_1^{\mathrm{IP}}(t; \hat{\beta}_I)$, where $\hat{g}_2^{\mathrm{MIP}}(t) = \sum_{i=1}^n \sum_{j=1}^{n_i} W_{ij}(t)\tilde{Y}_{ij}^{\mathrm{MIP}}$ with $\tilde{Y}_{ij}^{\mathrm{MIP}} = \delta_{ij} Y_{ij}/p(X_{ij}, T_{ij}) + (1 - \delta_{ij}/p(X_{ij}, T_{ij}))(f(X_{ij}; \hat{\beta}_I) + \hat{g}(T_{ij}))$.

3 Asymptotic properties

Here, we assume that function $\partial f(X_{ij}; \beta)/\partial\beta$ can be written as

$$\frac{\partial f(X_{ij}; \beta)}{\partial \beta_a} = h_a(T_{ij}; \beta) + u_{ija}(\beta), \quad i = 1, \ldots, n,\ j = 1, \ldots, n_i,\ a = 1, \ldots, p,$$

where $h_a(T_{ij}; \beta) = E(\partial f(X_{ij}; \beta)/\partial\beta_a|T_{ij})$. Then, we have $\tilde{d}_{ij}(\beta) = u_{ija}(\beta) + \check{h}_a(T_{ij}; \beta)$, where $\check{h}_a(T_{ij}; \beta) = h_a(T_{ij}; \beta) - \hat{h}_a(T_{ij}; \beta)$ with $\hat{h}_a(T_{ij}; \beta) = \sum_{k=1}^n \sum_{l=1}^{n_k} W_{kl}^{\mathrm{C}}(T_{ij})\partial f(X_{kl}; \beta)/\partial\beta_a$.

Based on the above-mentioned notation, we consider asymptotic distributions of the LELRFs $\ell_l(\beta)$ and the estimators $\hat{\beta}_l$ ($l = c, I$) for parameter β presented in Sects. 2.2 and 2.3.

Theorem 1 *Suppose that the conditions (A1)–(A11) given in the Appendix hold. If β is the true parameter, then $\ell_l(\beta) \xrightarrow{\mathcal{L}} \chi_p^2$ for $l = c$ and I, where χ_p^2 is the Chi-squared distribution with p degrees of freedom, and $\xrightarrow{\mathcal{L}}$ denotes the convergence in distribution.*

Let $\chi_{p,\alpha}^2$ be the upper α-percentile of the central Chi-squared distribution with p degrees of freedom for $0 < \alpha < 1$. It follows from Theorem 1 that the approximate $100(1-\alpha)$ % EL confidence region (ELCR) for β can be obtained by $\{\beta : \ell_l(\beta) \leq \chi_{p,\alpha}^2\}$ for $l = c$ and I.

Theorem 2 *Suppose that the conditions (A1)–(A11) given in the Appendix hold. If β is the true parameter, then we have*

$$\sqrt{n}(\hat{\beta}_k - \beta) \xrightarrow{\mathcal{L}} N(0, \Xi_k^{-1}\Lambda_k\Xi_k^{-1}) \ \text{ for } \ k = c, I,$$

where $\Lambda_c = \lim_{n\to\infty} n^{-1}\sum_{i=1}^n u_i^{\mathrm{T}}\Delta_i V_i^{-1}\Delta_i\Sigma_i\Delta_i V_i^{-1}\Delta_i u_i$, $\Xi_c = \lim_{n\to\infty} n^{-1}\sum_{i=1}^n u_i^{\mathrm{T}}\Delta_i V_i^{-1}\Delta_i u_i$, $\Lambda_I = \lim_{n\to\infty} n^{-1}\sum_{i=1}^n u_i^{\mathrm{T}} V_i^{-1}\tilde{\Delta}_i\Sigma_i\tilde{\Delta}_i V_i^{-1}u_i$, $\tilde{\Delta}_i = \mathrm{diag}(\delta_{i1}/P(X_{i1},T_{i1}),\ldots,\delta_{in_i}/P(X_{in_i},T_{in_i}))$, $\Xi_I = \lim_{n\to\infty} n^{-1}\sum_{i=1}^n u_i^{\mathrm{T}} V_i^{-1}\tilde{\Delta}_i u_i$, $u_i = (u_{i1},\ldots,u_{in_i})^{\mathrm{T}}$ *with* $u_{ij} = (u_{ij1},\ldots,u_{ijp})^{\mathrm{T}}$.

Let $\hat{\Omega}_k = \hat{\Xi}_k^{-1}\hat{\Lambda}_k\hat{\Xi}_k^{-1}$, *where* $\hat{\Lambda}_k = n^{-1}\sum_{i=1}^n Z_{ik}(\hat{\beta})Z_{ik}^{\mathrm{T}}(\hat{\beta})$ *and* $\hat{\Xi}_k = n^{-1}\sum_{i=1}^n\{\partial Z_{ik}(\beta)/\partial\beta\}_{\beta=\hat{\beta}_k}$ *for* $k = c$ *and* I. *It is easily shown that* $\hat{\Omega}_k$ *is the consistent estimator of* $\Xi_k^{-1}\Lambda_k\Xi_k^{-1}$ *for* $k = c$ *and* I. *Then, it follows from Theorem 2 that* $\sqrt{n}\hat{\Omega}_k^{-1/2}(\hat{\beta}_k-\beta)\xrightarrow{\mathcal{L}} N(0, I_p)$, *which yields* $n(\hat{\beta}_k-\beta)^{\mathrm{T}}\hat{\Omega}_k^{-1}(\hat{\beta}_k-\beta)\xrightarrow{\mathcal{L}}\chi_p^2$, *where* I_p *is the* $p\times p$ *identity matrix. Therefore, the approximate* $100(1-\alpha)$ % *ELCR for* β *can be constructed by* $\{\beta : n(\hat{\beta}_k-\beta)^{\mathrm{T}}\hat{\Omega}_k^{-1}(\hat{\beta}_k-\beta)\le\chi_{p,\alpha}^2\}$ *for* $k = c$ *and* I.

Theorem 3 *Suppose that the conditions* (*A1*)–(*A11*) *given in the Appendix hold and the kernel function* $K(\cdot)$ *is twice continuously differentiable on* $[0, 1]$. *If* $g(t_0)$ *is the true value of the baseline function* $g(t)$, *then we have* $\ell_R(g(t_0))\xrightarrow{\mathcal{L}}\chi_1^2$.

By Theorem 3, an approximate $100(1-\alpha)$ % *pointwise EL confidence interval (CI) for* $g(t_0)$ *can be constructed by* $\{g(t_0) : \hat{\ell}(g(t_0))\le\chi_{1,\alpha}^2\}$.

Theorem 4 *Suppose that the conditions* (*A1*)–(*A11*) *in the Appendix hold. Then, we have*

$$\sqrt{Nh}\{\hat{g}(t_0)-g(t_0)\}-b(t_0)\{q(t_0)\kappa(t_0)\}^{-1}\xrightarrow{\mathcal{L}} N(0,\gamma^2(t_0)),$$

where $b(t_0) = h_0^{5/2}[g'(t_0)\{q'(t_0)\kappa(t_0)+q(t_0)\kappa'(t_0)\}+\frac{1}{2}g''(t_0)q(t_0)\kappa(t_0)]\int_{-1}^1 u^2 K(u)\mathrm{d}u$, $\gamma^2(t_0) = V^2(t_0)\{q(t_0)\kappa(t_0)\}^{-2}$ *with* $V^2(t_0) = \sigma_\varepsilon^2(t_0)q(t_0)\kappa(t_0)\int_{-1}^1 K^2(u)\mathrm{d}u$, *the definitions of* $q(t_0)$ *and* $\kappa(t_0)$ *are given in Appendix, and* h_0 *is a constant defined in the condition* (*A3*) *of Appendix.*

Proposition 1 *If the condition* (*A2*) *is substituted by the condition that* $Nh^2/\log(N)\to\infty$ *and* $Nh^5\to 0$, *then the bias term* $b(t_0)$ *is asymptotically zero and* $\sqrt{Nh}\{\hat{g}(t_0)-g(t_0)\}\xrightarrow{\mathcal{L}} N(0,\gamma^2(t_0))$.

To construct the pointwise CI for $g(t_0)$ *based on the above-presented normal approximation* (*NA*), *we must first estimate* $b(t_0)$ *and* $\gamma^2(t_0)$. *It is easily shown from* $\int uK(u)\mathrm{d}u = 0$ *and* $h\to 0$ *that* $\sqrt{N/h}E\{\delta(g(T)-g(t_0))K_h(T-t_0)\} = b(t_0)+o_p(1)$, *which implies that a consistent estimator of* $b(t_0)$ *can be expressed as*

$$\hat{b}(t_0) = (Nh)^{-1/2}\sum_{i=1}^n\sum_{j=1}^{n_i}\delta_{ij}\{\hat{g}(T_{ij})-\hat{g}(t_0)\}K_h(T_{ij}-t_0).$$

Similar to Xue and Xue (2011), we can estimate $\gamma^2(t_0)$ *by* $\hat{\gamma}^2(t_0) = \hat{V}^2(t_0)\{\hat{q}(t_0)\hat{\kappa}(t_0)\}^{-2}$, *where* $\hat{\kappa}(t_0) = (Nh)^{-1}\sum_{i=1}^n\sum_{j=1}^{n_i}K_h(T_{ij}-t_0)$, $\hat{q}(t_0) = (Nh)^{-1}\sum_{i=1}^n\sum_{j=1}^{n_i}K_h(T_{ij}-t_0)\delta_{ij}/\hat{\kappa}(t_0)$ *and* $\hat{V}(t_0) = (Nh)^{-1}\sum_{i=1}^n\hat{\eta}_{iE}(g(t_0))$ *with* $\hat{\eta}_{iE}(g(t)) = \sum_{j=1}^{n_i}\delta_{ij}\{Y_{ij}-f(X_{ij};\hat{\beta})-g(t)\}K_h(T_{ij}-t)$. *Then, it follows from Theorem 4 that*

$\hat{\gamma}^{-1}(t_0)[\sqrt{Nh}\{\hat{g}(t_0)-g(t_0)\}-\hat{b}(t_0)\{\hat{q}(t_0)\hat{\kappa}(t_0)\}^{-1}] \xrightarrow{\mathcal{L}} N(0,1)$. *Thus, an approximate* $100(1-\alpha)$ % *CI for* $g(t_0)$ *is given by*

$$\hat{g}(t_0) - (Nh)^{-1/2}\hat{b}(t_0)\{\hat{q}(t_0)\hat{\kappa}(t_0)\}^{-1} \pm z_{\alpha/2}(Nh)^{-1/2}\hat{\gamma}(t_0),$$

where $z_{\alpha/2}$ *is the upper* $\alpha/2$ *percentile of the standard normal distribution, "−" and "+" correspond to the lower limit and the upper limit of the confidence interval, respectively.*

Proposition 2 *If the condition presented in Proposition 1 holds, the approximation* $100(1-\alpha)$ % *CI for* $g(t_0)$ *can be expressed as* $\hat{g}(t_0) \pm z_{\alpha/2}(Nh)^{-1/2}\hat{\gamma}(t_0)$.

Theorem 5 *Suppose that the conditions (A1)–(A11) given in the Appendix hold. Then, we have* $\hat{g}^{\mathrm{MIP}}(t) - g(t) = O_p((nh)^{-\frac{1}{2}} + (nb)^{-\frac{1}{2}} + b + h)$. *In particular, if* $h = O(n^{-1/3})$ *and* $b = O(n^{-1/3})$, *we have* $\hat{g}^{\mathrm{MIP}}(t) - g(t) = O_p(n^{-1/3})$.

Theorem 5 shows that $\hat{g}^{\mathrm{MIP}}(t)$ *attains the optimal convergence rate of nonparametric kernel regression estimator when* $h = O(n^{-1/3})$ *and* $b = O(n^{-1/3})$ *(Stone 1980).*

4 Numerical examples

4.1 Simulation studies

(1) One-dimensional case

In the simulation study, the data set $\{Y_{ij} : i = 1, \ldots, n, j = 1, \ldots, n_i\}$ was generated from the following semiparametric nonlinear model: $Y_{ij} = \exp(X_{ij}\beta) + \cos(4\pi T_{ij}) + \varepsilon_{ij}$ with the true value of parameter β being $\beta = 1.5$. To generate Y_{ij}, we independently simulated X_{ij} and the time point T_{ij} from the uniform distribution $U(0,1)$ and then generated ε_{ij} via $\varepsilon_{ij} = e_i + v_{ij}$ in which e_i and v_{ij} were independently generated from $N(0, \sigma_e^2)$ and $N(0, \sigma_v^2)$ with the true values of parameters σ_e^2 and σ_v^2 being $\sigma_e^2 = 1.0$ and $\sigma_v^2 = 1.0$. This structure for generating ε_{ij} ensures dependence among the repeated measurements Y_{ij} for each subject i because $\mathrm{cov}(\varepsilon_{ij}, \varepsilon_{ik}) = \sigma_e^2$ and the correlation coefficient between Y_{ij} and Y_{ik} is $\sigma_e^2/(\sigma_e^2 + \sigma_v^2)$ for $j \neq k$. For simplicity, we consider the balanced design, i.e. $n_1 = \cdots = n_n = J$. To create the missing data for responses Y_{ij}, we consider the following four cases for the selection probability function $p(x, t; \gamma) = \exp(\gamma_0 + \gamma_1 x + \gamma_2 t)/(1 + \exp(\gamma_0 + \gamma_1 x + \gamma_2 t))$ with $\gamma = (\gamma_0, \gamma_1, \gamma_2)$ specified by (1) $\gamma = (1.85, 0.02, 0.05)$, (2) $\gamma = (1.0, 0.5, 0.05)$, (3) $\gamma = (1.0, 0.001, 0.012)$ and (4) $\gamma = (0.4, 0.01, 0.02)$. Clearly, the considered missing data mechanism is MAR. For each given case of the selection probability $p(x, t; \gamma)$, the missing data Y_{ij}'s were created via the following steps: (a) we first generated a random number τ from the uniform distribution $U(0,1)$ and then (b) the observation Y_{ij} was missing if $\tau \leq 1 - p(X_{ij}, T_{ij}; \gamma)$ and we set $\delta_{ij} = 0$, and $\delta_{ij} = 1$ otherwise. In evaluating MELE and CI for β and estimating the parametric function $g(t) = \cos(4\pi t)$, we took the kernel function to be the Gaussian kernel $K(u) = (2\pi)^{-1/2}\exp(-u^2/2)$ and set the bandwidths h and b to be $n^{-1/5}$; we use the reweighted least squares iterative algorithm to estimate the parameter γ. We considered the following three different kinds of working covariance matrices in the simulation study, that is, we took $V = I_J$ (working independence), $V = \Sigma_i$ (true covariance matrix) and $V = \tilde{V}_i$ (estimator of V), where $\tilde{V}_i$ is evaluated using the formulae introduced in Sects. 2.2 and 2.3.

For each of the above-specified four cases for γ, we independently simulated 500 random samples of incomplete data set $\{(X_{ij}, Y_{ij}, T_{ij}, \delta_{ij}) : i = 1,\ldots,n, j = 1,\ldots,J\}$ with $n = 50$ and 100 and $J = 4$. The mean response rates for the above given four cases were roughly $E[p(X,T;\gamma)] \approx 90.07$, 83.47, 79.87 and 70.10 %, respectively. Results are reported in Table 1 in which 'Bias' is the absolute difference between the true value and the mean of 500

Table 1 Bias, RMS, coverage probability and average length of β under different missing functions $P(X,T)$ and sample sizes when nominal level is 0.95 and $p = 1$

Methods	$n = 50$						$n = 100$					
	CEL			IEL			CEL			IEL		
	I	Σ_i	$\tilde{V}_i$	I	Σ_i	$\tilde{V}_i$	I	Σ_i	$\tilde{V}_i$	I	Σ_i	$\tilde{V}_i$
Case 1												
Bias	0.003	0.003	0.004	0.003	0.003	0.004	0.003	0.000	0.000	0.002	0.000	0.000
RMS	0.087	0.0710	0.073	0.087	0.073	0.076	0.062	0.050	0.051	0.062	0.051	0.052
NACP	0.922	0.938	0.920	0.946	0.938	0.908	0.932	0.934	0.926	0.966	0.944	0.926
NAAL	0.319	0.266	0.256	0.356	0.271	0.262	0.226	0.186	0.182	0.251	0.190	0.187
ELCP	0.922	0.936	0.918	0.922	0.932	0.912	0.930	0.940	0.930	0.930	0.946	0.940
ELAL	0.325	0.267	0.256	0.325	0.273	0.263	0.225	0.184	0.180	0.225	0.188	0.184
Case 2												
Bias	0.003	0.003	0.004	0.003	0.003	0.004	0.002	0.000	0.000	0.002	0.001	0.001
RMS	0.091	0.074	0.077	0.091	0.079	0.083	0.063	0.052	0.053	0.063	0.053	0.055
NACP	0.912	0.940	0.926	0.974	0.934	0.920	0.946	0.936	0.928	0.980	0.944	0.924
NAAL	0.331	0.278	0.269	0.395	0.287	0.279	0.233	0.194	0.191	0.277	0.201	0.198
ELCP	0.920	0.936	0.924	0.922	0.924	0.918	0.946	0.932	0.932	0.944	0.944	0.930
ELAL	0.337	0.280	0.270	0.336	0.290	0.282	0.232	0.192	0.188	0.232	0.199	0.196
Case 3												
Bias	0.004	0.005	0.006	0.004	0.004	0.006	0.003	0.001	0.001	0.003	0.000	0.000
RMS	0.092	0.078	0.081	0.092	0.082	0.087	0.064	0.053	0.055	0.064	0.055	0.057
NACP	0.924	0.942	0.918	0.978	0.930	0.916	0.944	0.932	0.934	0.982	0.944	0.930
NAAL	0.340	0.288	0.280	0.429	0.298	0.292	0.239	0.201	0.198	0.300	0.209	0.207
ELCP	0.924	0.940	0.916	0.928	0.926	0.916	0.944	0.936	0.932	0.942	0.948	0.940
ELAL	0.347	0.291	0.282	0.344	0.302	0.295	0.239	0.199	0.196	0.238	0.208	0.205
Case 4												
Bias	0.002	0.002	0.003	0.001	0.001	0.002	0.002	0.002	0.002	0.002	0.001	0.000
RMS	0.099	0.084	0.087	0.099	0.092	0.098	0.067	0.058	0.060	0.067	0.061	0.063
NACP	0.922	0.936	0.922	0.984	0.922	0.914	0.944	0.942	0.934	0.998	0.944	0.938
NAAL	0.361	0.314	0.307	0.521	0.328	0.326	0.255	0.219	0.216	0.364	0.230	0.229
ELCP	0.918	0.940	0.928	0.920	0.930	0.908	0.944	0.942	0.938	0.946	0.954	0.942
ELAL	0.369	0.318	0.310	0.365	0.335	0.331	0.255	0.217	0.215	0.253	0.230	0.229

estimates, and 'RMS' is the root mean square between 500 estimates and its true value; 'CEL' and 'IEL' represent the EL methods with the complete-case data and the imputed values for missing responses, respectively; 'NACP' and 'ELCP' denote coverage probabilities of NA-based and EL-based CIs for β with 95 % confidence level, respectively; 'NAAL' and 'ELAL' denote average lengths (AL) of NA-based and EL-based CIs for β with 95 % confidence level, respectively.

From Table 1, we have following observations: (1) the CEL method has shorter interval length than the IEL method; (2) the EL-based method produces shorter interval length but larger coverage probability than the NA-based method; (3) the coverage probabilities for our considered EL-based CI and NA-based CI are close to the prespecified nominal level when the sample size is large or the average proportion of missing data is small; (4) the widths for the EL-based CI and the NA-based CI decrease as sample size n increases for every fixed selection probability function; (5) the average length depends on the selection probability function, namely, the average length increases as the missing rate increases; (6) the EL-based estimate for β is reasonably accurate under different cases for the selection probability function and all considered sample sizes including small sample case; and (7) the values of Bias and RMS via the true working covariance matrix are smaller than the other two cases, whilst the method via the estimated working covariance matrix performs better than the method with the identity working covariance matrix; the CI via the estimated working covariance matrix outperforms the CI via the identity and true working covariance matrices in terms of the length of CI. These results show that increasing n or reducing missing rate can improve the accuracy of estimators.

To investigate the performance of the constructed pointwise CIs for $g(t)$, we compute the 95 % confidence bands of $g(t)$ with 400 simulation runs via the residual-adjusted-EL-based method (see Sect. 2.4) and NA-based method (see Theorem 4) under the first case for the selection probability function $p(x, t; \gamma)$. Results for $n = 100$ are presented in Fig. 1, which indicates that the proposed EL-based method

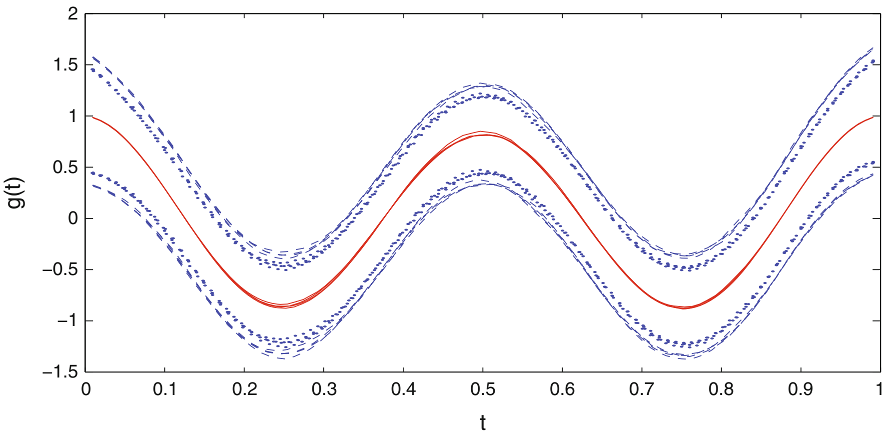

Fig. 1 95 % Confidence bands for $g(t)$ based on EL (*dashed curves*) and NA (*dotted curves*) with $n = 100$ and $p = 1$ for the first case of the selection probability function. The *solid curve* represents the real curve of $g(t)$

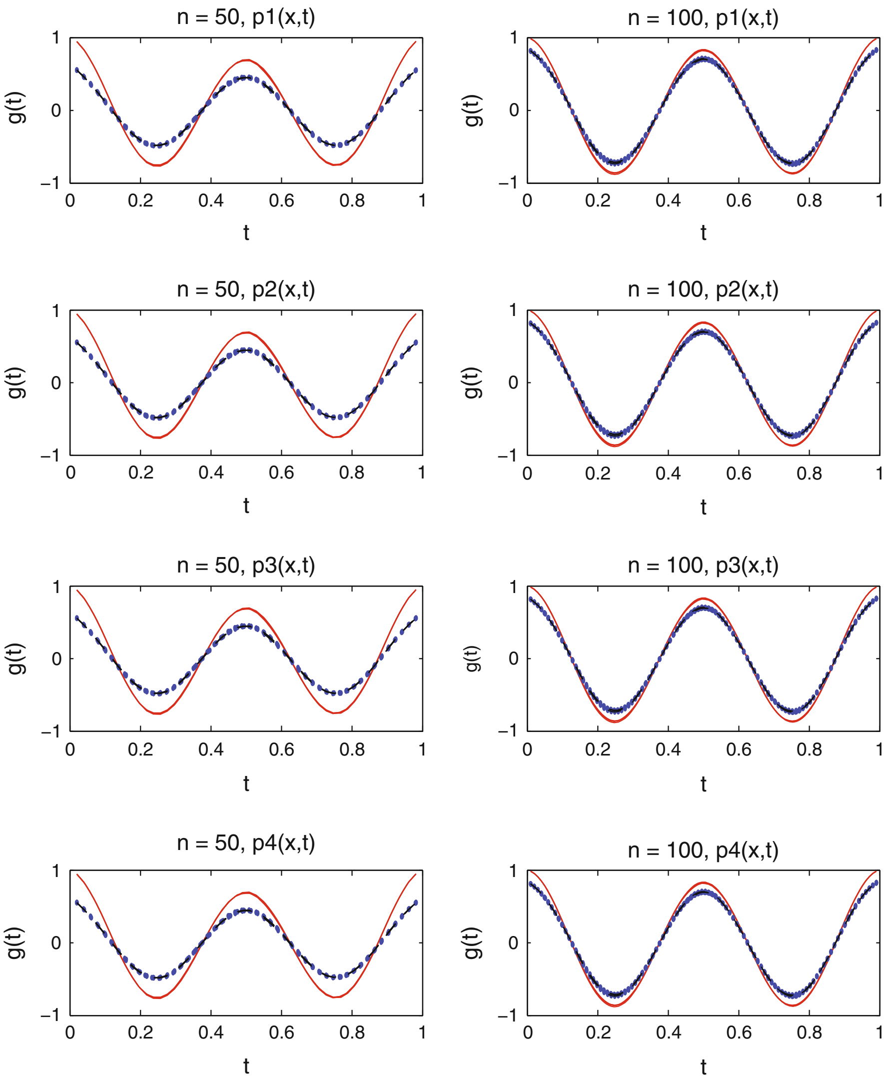

Fig. 2 Simulated curves of $\hat{g}_C(t)$, $\hat{g}^{[\mathrm{MIP}]}(t)$ for four missing functions $P(x,t)$ and two sample sizes when $p = 1$. The *red solid line* represents the true curve of $g(t)$, the *dotted curve* represents the estimated curve $\hat{g}_C(t)$ based on CEL method, the *dashed lines* are $\hat{g}^{[\mathrm{MIP}]}(t)$, respectively

behaves satisfactorily. Although the NA-based method gives a slightly narrower confidence band than the EL-based method, the latter does not require consistent estimator for the asymptotic variance, it is much easier to implement than the NA-based method.

To investigate the accuracy of the proposed $\hat{g}_C(t)$ and $\hat{g}^{\mathrm{MIP}}(t)$ for $g(t)$ under different missing cases for $p(x,t;\gamma)$ and sample sizes, we compute 1,000 simulated values of $\hat{g}_C(t)$ (see Sect. 2.2) and $\hat{g}_n^{\mathrm{MIP}}(t)$. Figure 2 presents their corresponding simulated curves on the inner points against the true curve of $g(t)$. Figure 2 shows that our proposed estimated curves are rather close to the true one in general.

(2) Two-dimensional case

In this simulation study, we consider the following two-dimensional semiparametric nonlinear model for longitudinal data $Y_{ij} = \exp\{X_{1ij}\beta_1 + X_{2ij}\beta_2\} + \cos(4\pi T_{ij}) + \varepsilon_{ij}$. Here, X_{1ij}, X_{2ij} and T_{ij} were independently generated from the uniform distribution $U(0, 1)$, ε_{ij} was generated by $\varepsilon_{ij} = e_i + v_{ij}$ in which e_i and v_{ij} were independently generated from $N(0, \sigma_e^2)$ and $N(0, \sigma_v^2)$ with the true values of σ_e^2 and σ_v^2 being $\sigma_e^2 = \sigma_v^2 = 1.0$, leading to a correlation structure for $y_i = (y_{i1}, \ldots, y_{in_i})^{\mathrm{T}}$. Then, Y_{ij}'s were generated from the above specified two-dimensional semiparametric nonlinear model with the true value of β being $\beta = (\beta_1, \beta_2)^{\mathrm{T}} = (1.0, 0.5)^{\mathrm{T}}$. We set the number of repeated measures n_i to be the same, say m. The selection probability $p(x, t; \gamma) = \exp(\gamma_0 + \gamma_1 x + \gamma_2 t)/(1 + \exp(\gamma_0 + \gamma_1 x + \gamma_2 t))$ with $\gamma = (\gamma_0, \gamma_1, \gamma_2)$ is taken to be (1) $\gamma = (1, 0.5, 0.5, 0.05)$ and (2) $\gamma = (0.4, 0.01, 0.02)$. For each of two cases, the missing data are created as done in the one-dimensional case. In evaluating EL estimates and confidence regions for β and estimating $g(t) = \cos(4\pi t)$, we took the same kernel function and bandwidth h as done in the one-dimensional case; we also used the reweighted least squared iterative algorithm to obtain estimate of parameter γ. For each case, we independently generated 500 random samples of incomplete data set $\{(X_{1ij}, X_{2ij}, Y_{ij}, T_{ij}, \delta_{ij}) : i = 1, \ldots, n, j = 1, \ldots, m\}$ with $n = 50$ and 100 and $m = 4$. The mean response rates for two cases are $E[p(X, T)] \approx 86.44$ and 70.23 %, respectively. Based on the generated 500 data sets for each given selection probability function $p(x, t; \gamma)$, we computed the values of Bias and RMS, and the coverage probabilities and interval lengths for the 95 % CIs of β_1 and β_2 via the EL-based method and the NA-based method under $n = 50$ and 100 with $m = 4$. Here, a grid search algorithm was used to evaluate the EL-based CIs for β_1 and β_2 via the following steps: (i) arbitrarily give two intervals which, respectively, contain the true values $\beta_1 = 1.0$ and $\beta_2 = 0.5$; (ii) given a search step length, we evaluated the LELRF $\ell_l(\beta)$ ($l = c, I$) at each search point belonging to the given interval and found the gridpoint $\hat{\beta}_0 = (\hat{\beta}_{10}, \hat{\beta}_{20})$ such that $\ell_l(\hat{\beta}_0) \leq \chi^2_{2,\alpha}$, which indicates that $\hat{\beta}_0$ is just the upper or lower bound of the EL-based CI. Results are presented in Table 2. Examination of Table 2 shows that (1) MELEs and the 95 % CIs for β_1 and β_2 are rather accurate; (2) the efficiency of MELE can be improved by considering the within-group correlation structure; and (3) the CP of the CEL method with true covariance matrix is closer to the prespecified confidence level than that of the CEL method with estimated covariance matrix when sample size is small (e.g., $n = 50$), but the CEL method with true covariance matrix becomes more conservative than that with estimated covariance matrix when sample size is large (e.g., $n = 100$) whose main reason is that the missing rate corresponding to $n = 100$ is higher than that corresponding to $n = 50$.

We computed the 95 % confidence band of $g(t)$ with 400 simulation runs via the EL-based method and the NA-based method for the first case of the selection probability function. Results for $n = 100$ were shown in Fig. 3, which implies that our proposed EL-based method behaves satisfactorily. In addition, Fig. 4 displayed the simulated curves on the inner points against the true curve of $g(t)$ based on 1, 000 simulated values of $\hat{g}_C(t)$ and $\hat{g}_n^{\mathrm{MIP}}(t)$ under different missing functions $p(x, t)$ and sample sizes, which shows that the same findings are observed as in Fig. 2.

Table 2 Bias, RMS, coverage probability and average length of β under different missing functions $P(X, T)$ and sample size when nominal level is 0.95 and $p = 2$

Methods	$n = 50$						$n = 100$					
	CEL			IEL			CEL			IEL		
	I	Σ_i	$\tilde{V}_i$	I	Σ_i	$\tilde{V}_i$	I	Σ_i	$\tilde{V}_i$	I	Σ_i	$\tilde{V}_i$
Estimate of β_1 *with* $p_1(x, t)$												
Bias	0.001	0.004	0.004	0.002	0.003	0.002	0.001	0.001	0.000	0.001	0.000	0.001
RMS	0.121	0.100	0.105	0.122	0.103	0.110	0.088	0.072	0.072	0.088	0.076	0.078
NACP	0.946	0.868	0.922	0.944	0.948	0.924	0.938	0.876	0.944	0.942	0.936	0.926
NAAL	0.477	0.326	0.386	0.478	0.414	0.403	0.339	0.227	0.275	0.337	0.289	0.284
ELCP	0.937	0.953	0.930	0.945	0.945	0.928	0.957	0.965	0.943	0.949	0.967	0.947
ELAL	0.280	0.228	0.217	0.279	0.236	0.226	0.194	0.158	0.154	0.193	0.161	0.158
Estimate of β_1 with $p_2(x, t)$												
Bias	0.001	0.002	0.003	0.001	0.004	0.005	0.006	0.003	0.003	0.006	0.003	0.003
RMS	0.134	0.115	0.119	0.134	0.123	0.132	0.099	0.083	0.085	0.099	0.088	0.092
NACP	0.950	0.884	0.916	0.944	0.946	0.928	0.950	0.864	0.932	0.950	0.940	0.940
NAAL	0.534	0.375	0.449	0.529	0.490	0.486	0.378	0.261	0.319	0.375	0.342	0.339
ELCP	0.941	0.932	0.915	0.930	0.939	0.909	0.945	0.963	0.949	0.947	0.943	0.926
ELAL	0.319	0.267	0.261	0.319	0.287	0.282	0.215	0.183	0.180	0.213	0.191	0.188
Estimate of β_2 with $p_1(x, t)$												
Bias	0.004	0.009	0.008	0.003	0.007	0.005	0.007	0.004	0.004	0.006	0.003	0.002
RMS	0.141	0.112	0.118	0.141	0.116	0.124	0.099	0.079	0.081	0.099	0.083	0.086
NACP	0.944	0.892	0.944	0.942	0.952	0.926	0.948	0.902	0.946	0.952	0.954	0.944
NAAL	0.539	0.372	0.438	0.538	0.469	0.455	0.381	0.255	0.309	0.379	0.324	0.319
ELCP	0.937	0.953	0.930	0.945	0.945	0.928	0.957	0.965	0.943	0.949	0.967	0.947
ELAL	0.280	0.228	0.217	0.279	0.236	0.226	0.194	0.158	0.154	0.193	0.161	0.158
Estimate of β_2 with $p_2(x, t)$												
Bias	0.009	0.010	0.009	0.009	0.012	0.011	0.004	0.004	0.003	0.004	0.002	0.001
RMS	0.160	0.135	0.141	0.160	0.140	0.152	0.113	0.093	0.095	0.112	0.102	0.106
NACP	0.938	0.892	0.924	0.942	0.944	0.938	0.936	0.868	0.928	0.944	0.928	0.924
NAAL	0.604	0.426	0.509	0.600	0.556	0.551	0.426	0.294	0.358	0.421	0.383	0.381
ELCP	0.941	0.932	0.915	0.930	0.939	0.909	0.945	0.963	0.949	0.947	0.943	0.926
ELAL	0.319	0.267	0.261	0.319	0.287	0.282	0.215	0.183	0.180	0.213	0.191	0.188

4.2 A real example

A longitudinal data set from the pediatric AIDS clinical trial group ACTG 315 study was used to illustrate our proposed methodologies. In an AIDS clinical trial, plasma HIV RNA copies (viral load) and CD4+ cell counts were two important surrogate markers for evaluating antiviral therapies (Saag et al. 1996; Mellors et al. 1996). Clinical investigators' main purpose is to study their relationship during antiviral treatment. In this study, viral load and CD4+ cell counts from 46 patients were measured on treatment days $t = 0, 2, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 25, 27, \ldots, 175$,

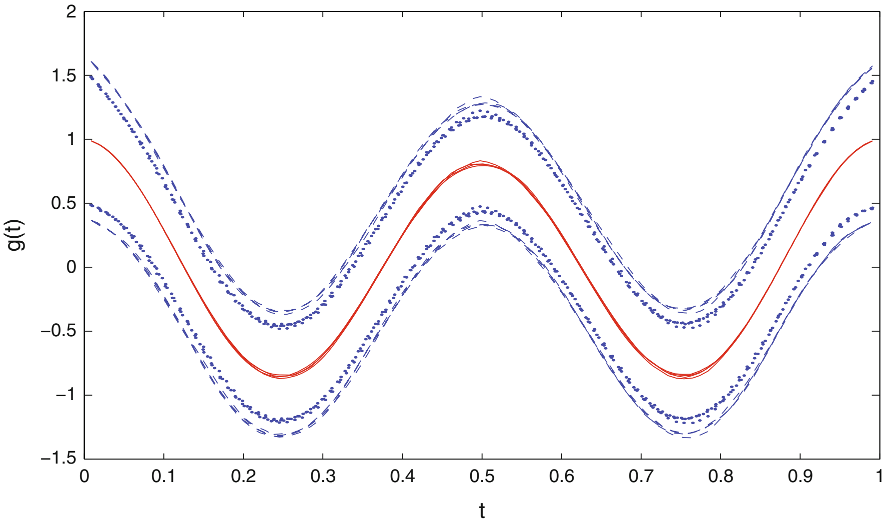

Fig. 3 95 % confidence bands for $g(t)$ based on EL (*dashed curves*) and NA (*dotted curves*) with $n = 100$ and $p = 2$ for the first case of the selection probability function. The *solid curve* represents the real curve of $g(t)$

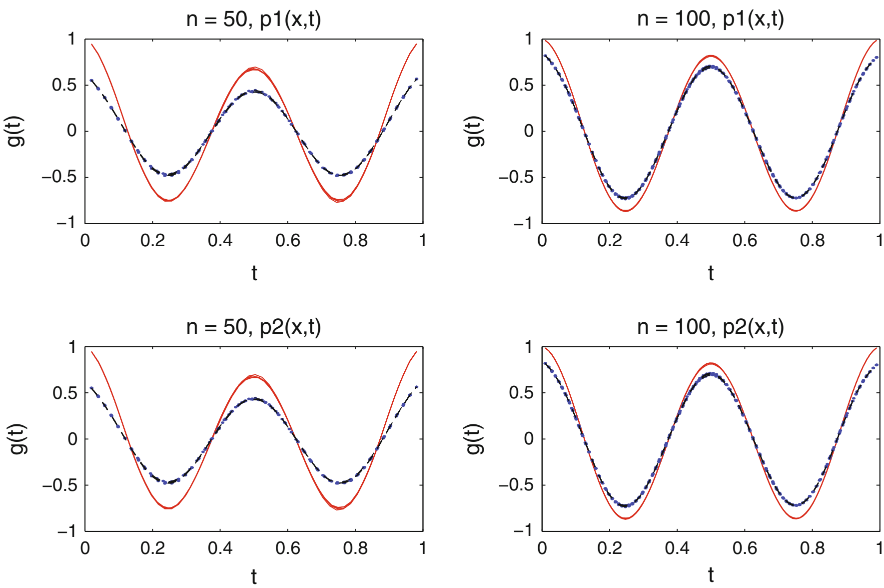

Fig. 4 Simulated curves of $\hat{g}_C(t)$, $\hat{g}^{[\mathrm{MIP}]}(t)$ with two missing functions $P(x, t)$ and two sample sizes when $p = 2$. The *red solid line* represents the true curve of $g(t)$, the *dotted curve* represents the estimated curve $\hat{g}_C(t)$ based on CEL method, and the *dashed line* represents the estimated curves $\hat{g}^{[\mathrm{MIP}]}(t)$, respectively

182, 196 after initiation of an antiviral therapy, and 361 complete pairs of viral load and CD4+ cell count were obtained. The number of the measured time points on individual patients ranges from 4 to 8. The data set has even been analysed by Liang et al. (2003) and Xue and Xue (2011). The preceding studies in

Liang et al. (2003) and Xue and Xue (2011) suggested that viral load depends linearity on CD4 cell count but nonlinearly on treatment time; however, the scatterplot between viral load and CD4 cell count shows that there is no rigorous linearity between viral load and CD4 cell count. Therefore, here we used the following semiparametric nonlinear model to formulate the relationship between viral load and CD4 cell count: $Y_{ij} = \exp(X_{ij}\beta) + g(T_{ij}) + \varepsilon_{ij}$, where Y_{ij} and X_{ij} are the viral load and the CD4+ cell count for subject i at treatment time T_{ij}, respectively. To illustrate the application of our proposed methodologies, we created missing data via the following selection probability function: $p(x, t; \gamma) = \exp(\gamma_0+\gamma_1 x+\gamma_2 t)/(1+\exp(\gamma_0+\gamma_1 x+\gamma_2 t))$ with $\gamma = (\gamma_0, \gamma_1, \gamma_2) = (0.4, 0.05, 0.1)$. Based on this selection probability function and the assumption that Y_{i1} was always observed, the missing data for Y_{ij} were created with the following steps: (a) we generated a random number τ from the uniform distribution $U(0, 1)$, (b) Y_{ij} was missing if $\tau \le p(X_{ij}, T_{ij}; \gamma)$ for $i = 1, \ldots, 46$, $j = 1, \ldots, n_i$. The corresponding missing proportion is roughly 15 %. As commonly done in AIDS clinical trials, we used $\log_{10}$ scale in viral load and 100^{-1} scale in CD4 cell counts to stabilize the variance and computational algorithms.

In the real example analysis, we took the kernel function to be $K(u) = (2\pi)^{-1/2}\exp(-u^2/2)$, the bandwidth h to be $h = 24.35$ (Xue and Xue 2011) and used the reweighted least squares iterative algorithm to obtain estimate of parameter γ in the selection probability function. Based on the above-given kernel function and bandwidth, we computed estimate for β and its corresponding 95 % EL-based and NA-based CIs. Estimate of β is $\hat{\beta}_I = -0.5713$, which indicated that the CD4+ cell counts have a negative effect on viral load during antiviral treatments; this result is consistent with that given in Liang et al. (2003) and Xue and Xue (2011). The 95 % EL-based and NA-based (NA-based) CIs for β are $(-0.7200, -0.4620)$ and $(-0.6964, -0.4462)$, respectively. In addition, we evaluated the 95 % EL-based

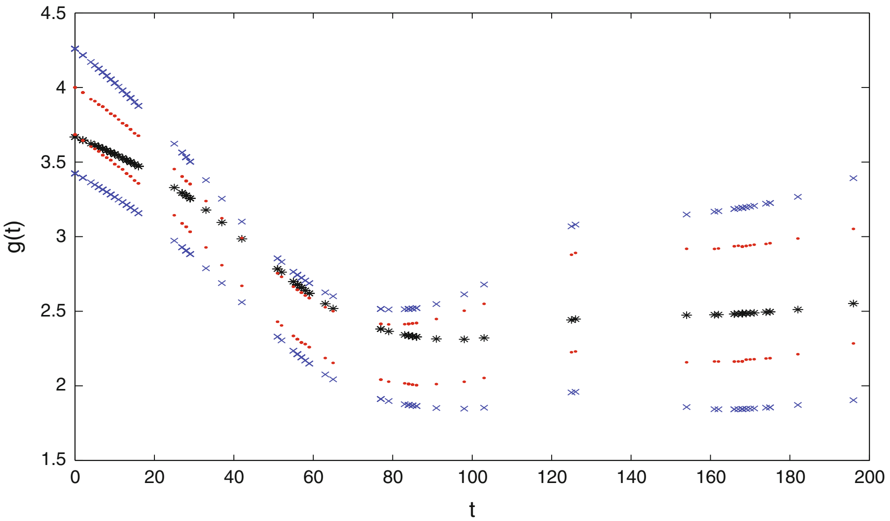

Fig. 5 95 % confidence bands for $g(t)$ based on EL (the *dotted curve*) and NA (the "*x-symbol*" *curve*) in the real example. The *star curve* represents the estimated curve of $g(t)$

(see Sect. 2.4) and NA-based CIs for $g(t)$ (see Theorem 4). The corresponding results were reported in Fig. 5, which shows that (1) the viral load RNA levels rapidly decrease after initial antiviral treatment, then rebound a bit little and finally become nearly flat, (2) the EL-based method gives a narrower band than the NA-based method; these results were consistent with those given in Xue and Xue (2011).

5 Conclusions

By introducing the working covariance matrix into the auxiliary random vector, we develop an EL-based inference procedure for a semiparametric nonlinear regression model for longitudinal data with response missing at random. Two MELEs for unknown parameter β in our considered semiparametric nonlinear regression models were presented on the basis of the complete-case data and the imputed values of missing responses. Also, a maximum residual-adjusted EL estimator and an imputation estimator for the smoothing functions were proposed. We systematically investigate the asymptotic properties of the MELEs under this new setting. Our main contribution is that (1) our considered model is more general than nonlinear regression model and semiparametric regression model with response missing at random, which indicates that our proposed theoretical results are new; (2) the working covariance matrix is introduced to accommodate for the within-subject correlation, which can be used to improve the efficiency of MELE; and (3) we proved that our constructed EL ratio statistic for β follows asymptotically the central Chi-squared distribution, which can be directly used to construct confidence regions of parameters in our considered semiparametric nonlinear regression model without any extra Monte Carlo approximation needed when our proposed EL method is not used. We extended the EL inference procedure for semiparametric regression models with missing response at random to semiparametric nonlinear regression models for longitudinal data with missing response at random by incorporating the within-subject correlation into the constructed auxiliary vectors.

Appendix

For convenience and simplicity, let c denote a positive constant which may represent a different value at different cases throughout this paper. Denote $q(t) = E(\delta|T = t)$ and assume that variable T has the probability density function $\kappa(t)$. Denote $N = \sum_{i=1}^{n} n_i$ and suppose $n = O(N)$. The following conditions are required for results given in Theorems 1–5:

- (A1) The selection probability function $p(x, t)$ and the X-density function $\Gamma(x)$ have bounded partial derivatives up to order s with $s \geq 2$.
- (A2) Let $S(\gamma)$ be the score function of the partial likelihood $L(\gamma)$ for parameter $\gamma = (\gamma_0, \gamma_1^{\mathrm{T}}, \gamma_2)^{\mathrm{T}}$ defined in Sect. 2.1 and γ^* be in the interior of compact set Υ. We assume $\mathrm{var}(\mathrm{S}(\gamma))$ is a finite and positive definite matrix, and $E(\partial S(\gamma)/\partial\gamma|_{\gamma=\gamma^*})$ exists and is invertible. The missing propensity $p(X_{ij}, T_{ij}; \gamma) > c_0 > 0$ for all $i \in \{1, \ldots, n\}$ and $j \in \{1, \ldots, n_i\}$.

(A3) The bandwidth satisfies $h = h_0 N^{-1/5}$ for some constant $h_0 > 0$, and $b = b_0 N^{-1/5}$ for some constant $b_0 > 0$.
(A4) The kernel function $K(\cdot)$ is a symmetric and bounded probability density function with support $[-1, 1]$.
(A5) For each design, points $\{T_{ij} : i = 1, \ldots, n, j = 1, \ldots, n_i\}$ are assumed to be independent and identically distributed from a super-population density $\kappa(t)$. Both $q(t)$ and $\kappa(t)$ have continuous and bounded derivatives on (0,1) and are bounded away from zero and infinity on [0,1].
(A6) The residuals ε_{ij} and u_{ij} are independent of each other, and ε_{ij} and u_{ij} are, respectively, independent of $\varepsilon_{i'j}$ and $u_{i'j}$ for any $i \neq i'$. Further, we assume that $E|\varepsilon_{ij}|^{4+r} < \infty$, $\max_{1\leq i\leq n} \|u_{ij}\| = o_p\{n^{\frac{2+r}{2(4+r)}}(\log n)^{-1}\}$ for some $r > 0$.
(A7) The matrices Λ_i and Ξ_i $(i = c, I)$ defined in Theorem 2 are positive definite.
(A8) The functions $g(t)$ and $h(t)$ are twice continuously differentiable on (0,1).
(A9) The function $f(X; \beta)$ is continuous with respect to β in a compact set Θ.
(A10) There exit two positive constants c_1 and c_2 such that

$$0 \leq \min_{1\leq i\leq n} \lambda_{i1} \leq \min_{1\leq i\leq n} \lambda_{in_i} \leq c_2 < \infty,$$

where λ_{i1} and λ_{in_i} denote the smallest and largest eigenvalues of Σ_i, respectively.
(A11) There exit two positive constants c_3 and c_4 such that

$$0 \leq \min_{1\leq i\leq n} \lambda'_{i1} \leq \min_{1\leq i\leq n} \lambda'_{in_i} \leq c_2 < \infty,$$

where λ'_{i1} and λ'_{in_i} denote the smallest and largest eigenvalues of V_i, respectively.

Condition (A1) is the standard assumption for nonparametric regression problem. $p(x, t)$ being bounded away from zero in condition (A1) indicates that data cannot be missing with probability 1 anywhere in the domain of the (X, T)-variable. Condition (A2) is a regular condition for consistence of MLE for parameter γ in the selection probability. Smoothing conditions (A4), (A5) and (A8) are the standard conditions for nonparametric problems. Conditions (A6) and (A7) are necessary for asymptotic normality. Condition (A3) gives the rate of the optimal bandwidth for estimating $g(t)$ and ensures that undersmoothing $\hat{g}(t)$ is not needed so that we can use the data-driven approach to select the optimal bandwidth. Condition (A9) is a regular condition for the general nonlinear models (Wu 1981). Conditions (A10) and (A11) are widely used in longitudinal data analysis.

To complete Proofs of Theorems 1–5, the following Lemmas are needed:

Lemma 1 *Suppose that the conditions* (*A1*)–(*A11*) *hold. Then, for any constants a and b with* $0 < a < b < 1$, *we have*

$$\sup_{a\le t\le b} E\{|\hat{g}_{1n}^{\mathrm{C}}(T_{ij};\beta)-g_1^{\mathrm{C}}(T_{ij};\beta)|^2|T_{ij}=t\}=O((nh)^{-1}+h^4),$$
$$\sup_{a\le t\le b} E\{|\hat{g}_{2n}^{\mathrm{C}}(T_{ij})-g_2^{\mathrm{C}}(T_{ij})|^2|T_{ij}=t\}=O((nh)^{-1}+h^4),$$
$$\sup_{a\le t\le b} E\{|\hat{h}(T_{ij};\beta)-h(T_{ij};\beta)|^2|T_{ij}=t\}=O((nh)^{-1}+h^4).$$

Proof For simplicity, we only prove the second equation. The other two equations can be similarly proved. According to the inequality $(A+B)^2 \le 2A^2+2B^2$ for any constants A and B and $\sum_{k=1}^n\sum_{l=1}^{n_i} W_{kl}^{\mathrm{C}}(T_{ij})=1$, we can prove that $E\{|\hat{g}_{2n}^{\mathrm{C}}(T_{ij})-g_2^{\mathrm{C}}(T_{ij})|^2|T_{ij}=t\}\le I_1(t)+I_2(t)$, where $I_1(t)=2E\{|\sum_{k=1}^n\sum_{l=1}^{n_k} W_{kl}^{\mathrm{C}}(T_{ij})(Y_{kl}-g_2^{\mathrm{C}}(T_{kl}))|^2|T_{ij}=t\}$ and $I_2(t)=2E\{|\sum_{k=1}^n\sum_{l=1}^{n_k} W_{kl}^{\mathrm{C}}(T_{ij})(g_2^{\mathrm{C}}(T_{kl})-g_2^{\mathrm{C}}(T_{ij}))|^2|T_{ij}=t\}$.

We first prove that $\sup_{a\le t\le b} I_2(t)=O(n^{-1}h+h^4)$. Let $q(t)=E(\delta|T=t)$, $m(t)=q(t)\kappa(t)$ and $\hat{m}(t)=(nh)^{-1}\sum_{i=1}^n\sum_{j=1}^{n_i}\delta_{ij}\ K_h(T_{ij}-t)$. Following the standard procedure in a nonparametric regression, it can be shown that $\max_{a\le t\le b}|\hat{m}(t)-m(t)|=O(n^{-1/5})$ a.s.. Hence, it follows from condition (A4) that there are two positive constants c_1 and c_2 such that $\min_{0\le t\le 1} m(t)\ge c_1$ and $\min_{0\le t\le 1}\hat{m}(t)\ge c_2$ a.s.. Let $\psi_{kl}(T_{ij})=K_h(T_{kl}-T_{ij})\delta_{kl}\{g_2^{\mathrm{C}}(T_{kl})-g_2^{\mathrm{C}}(T_{ij})\}$. Then, by conditions (A3), (A4) and (A7), we have $\max_{a\le t\le b}|E\{\psi_{kl}(T_{ij})|T_{ij}=t\}|=O(h^3)$ and $\max_{a\le t\le b}|E\{\psi_{kl}^2(T_{ij})|T_{ij}=t\}|=O(h^3)$. Based on these results, it is easy to show that $I_2(t)\le cn^{-1}h+ch^4$.

Again, it is easy to show that $E\{\delta_{kl}(Y_{kl}-g_2^{\mathrm{C}}(T_{kl}))\}=0$. Then, we can obtain that $I_1(t)\le c(nh)^{-1}$. Combining the above inequalities finishes the proof of the second equation. □

Lemma 2 *Suppose that the conditions* (*A1*)–(*A11*) *hold. Then, we have*

$$n^{-1/2}\sum_{i=1}^n Z_{i1}(\beta)\xrightarrow{\mathcal{L}} N(0,\Lambda_c),\quad n^{-1/2}\sum_{i=1}^n Z_{i2}(\beta)\xrightarrow{\mathcal{L}} N(0,\Lambda_I),$$

where Λ_c *and* Λ_I *are defined in Theorem 2.*

Proof Let $\breve{g}(T_{ij})=g(T_{ij})-\hat{g}(T_{ij})=g(T_{ij})-\hat{g}_{2n}^{\mathrm{C}}(T_{ij})+\hat{g}_{1n}^{\mathrm{C}}(T_{ij};\beta)$. Denote σ_i^{kl} be the (k,l)th component of V_i^{-1}. Then, we have $n^{-1/2}\sum_{i=1}^n Z_{i1}(\beta)\triangleq U_1+U_2+U_3+U_4$, where $U_1=n^{-1/2}\sum_{i=1}^n\sum_{k=1}^{n_i}\sum_{l=1}^{n_i}\{\delta_{ik}\delta_{il}u_{ik}\sigma_i^{kl}\varepsilon_{il}\}$, $U_2=n^{-1/2}\sum_{i=1}^n\sum_{k=1}^{n_i}\sum_{l=1}^{n_i}\{\delta_{ik}\delta_{il}\sigma_i^{kl}\breve{h}(T_{ik},\beta)\varepsilon_{il}\}$, $U_3=n^{-1/2}\sum_{i=1}^n\sum_{k=1}^{n_i}\sum_{l=1}^{n_i}\{\delta_{ik}\delta_{il}\sigma_i^{kl}u_{ik}\breve{g}(T_{il})\}$, $U_4=n^{-1/2}\sum_{i=1}^n\sum_{k=1}^{n_i}\sum_{l=1}^{n_i}\{\delta_{ik}\delta_{il}\sigma_i^{kl}\breve{h}(T_{ik},\beta)\breve{g}(T_{il})\}$.

We first prove $U_k=o_p(1)$ for $k=2,3,4$. It follows from Lemma 1 that $E\|U_2\|^2\le c\{(nh)^{-1}+h^4\}\to 0$. Similarly, we obtain $E\|U_3\|^2\le c\{(nh)^{-1}+h^4\}\to 0$. By Lemma 1 and the Cauchy–Schwarz inequality, we can obtain $E\|U_3\|\le c\sqrt{n}\{(nh)^{-1}+h^4\}\to 0$. Based on the above equations, we can prove that $U_j\xrightarrow{P} 0$ for $j=2,3$ and 4. These results show that we only need to prove $U_1\xrightarrow{\mathcal{L}} N(0,\Lambda_1)$ to show that Lemma 2 holds. It is easy to show that $\mathrm{var}(U_1)=\Lambda_c$ because U_1 is a sum of independent random variables. Thus, we only need to check whether U_1 satisfies condition of the Cramer–Wold Theorem and the Lindeberg–Feller condition. For any $\alpha \in R^p$ and $\varepsilon > 0$, let $L_n \triangleq \sum_{i=1}^n \mathrm{var}\{\sum_{k=1}^{n_i}\sum_{l=1}^{n_i}\alpha'\delta_{ik}\delta_{il}u_{ik}\sigma_i^{kl}\varepsilon_{il}\} = O(n)$ and $I(\cdot)$ be an indicator function. Then, we can show

$$g_n(\varepsilon) = \frac{1}{L_n}\sum_{i=1}^n E\left\{I\left(\sum_{k=1}^{n_i}\sum_{l=1}^{n_i}\alpha'\delta_{ik}\delta_{il}u_{ik}(\beta)\sigma_i^{kl}\varepsilon_{il} \geq \varepsilon\sqrt{L_n}\right)\right.$$
$$\left.\times\left(\sum_{k=1}^{n_i}\sum_{l=1}^{n_i}\alpha'\delta_{ik}\delta_{il}u_{ik}(\beta)\sigma_i^{kl}\varepsilon_{il}\right)^2\right\} \to 0.$$

Therefore, it follows from the Cramer–Wold Theorem and Lindeberg–Feller Theorem that $n^{-1/2}\sum_{i=1}^n Z_{i1}(\beta) \xrightarrow{\mathcal{L}} N(0, \Lambda_c)$, where $\Lambda_c = \lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^n u_i^{\mathrm{T}}\Delta_i V_i^{-1}\Delta_i\Sigma_i\Delta_i V_i^{-1}\Delta_i u_i$.

Denote $P_{ij}(\hat{\gamma}) = P(X_{ij}, T_{ij}; \hat{\gamma})$, where $\hat{\gamma}$ is a consistent estimator of $\gamma = (\gamma_0, \gamma_1^{\mathrm{T}}, \gamma_2)^{\mathrm{T}}$. Since $\delta_{ij}/P_{ij}(\hat{\gamma}) = \{\delta_{ij}P_{ij}^{-1}(\gamma)\}\{1 - P_{ij}^{'}(\gamma)(\hat{\gamma}-\gamma)/P_{ij}(\gamma) + o_p(n^{-1/2})\}$ and $\tilde{y}_{ij}^* - \widetilde{f}_{ij}(\beta) = \delta_{ij}P_{ij}^{-1}(\hat{\gamma})\{\tilde{y}_{ij} - \widetilde{f}_{ij}(\beta)\} + \left\{1-\delta_{ij}/P_{ij}(\hat{\gamma})\right\}\{\widetilde{f}_{ij}(\hat{\beta}) - \widetilde{f}_{ij}(\beta)\}$, we can obtain

$$\begin{aligned}
&n^{-1/2}\sum_{i=1}^n Z_{i2}(\beta)\\
&\quad= \frac{1}{\sqrt{n}}\sum_{i=1}^n\sum_{k=1}^{n_i}\sum_{l=1}^{n_i}\left\{\{u_{ik}(\beta)+\check{h}(T_{ik},\beta)\}\sigma_i^{kl}\frac{\delta_{il}}{P_{il}(\gamma)}\{\varepsilon_{il}+\check{g}(T_{il})\}\{1+o_p(1)\}\right\}\\
&\qquad+\frac{1}{n}\sum_{i=1}^n\sum_{k=1}^{n_i}\sum_{l=1}^{n_i}\left\{\tilde{d}_{ik}(\beta)\sigma_i^{kl}\{1-\frac{\delta_{il}}{P_{il}(\gamma)}\{1+o_p(1)\}\}\tilde{d}_{il}(\beta)\right\}\sqrt{n}(\hat{\beta}-\beta).\\
&\quad\triangleq J_1 + J_2.
\end{aligned}$$

For J_1, we have

$$\begin{aligned}
J_1 &= \frac{1}{\sqrt{n}}\sum_{i=1}^n\sum_{k=1}^{n_i}\sum_{l=1}^{n_i}\left\{u_{ik}(\beta)\sigma_i^{kl}\frac{\delta_{il}}{P_{il}(\gamma)}\varepsilon_{il}\{1+o_p(1)\}\right\}\\
&\quad+\frac{1}{\sqrt{n}}\sum_{i=1}^n\sum_{k=1}^{n_i}\sum_{l=1}^{n_i}\left\{u_{ik}(\beta)\sigma_i^{kl}\frac{\delta_{il}}{P_{il}(\gamma)}\check{g}(T_{il})\{1+o_p(1)\}\right\}\\
&\quad+\frac{1}{\sqrt{n}}\sum_{i=1}^n\sum_{k=1}^{n_i}\sum_{l=1}^{n_i}\left\{\check{h}(T_{ik},\beta)\sigma_i^{kl}\frac{\delta_{il}}{P_{il}(\gamma)}\varepsilon_{il}\{1+o_p(1)\}\right\}\\
&\quad+\frac{1}{\sqrt{n}}\sum_{i=1}^n\sum_{k=1}^{n_i}\sum_{l=1}^{n_i}\left\{\check{h}(T_{ik},\beta)\sigma_i^{kl}\frac{\delta_{il}}{P_{il}(\gamma)}\check{g}(T_{il})\{1+o_p(1)\}\right\}\\
&\triangleq J_{11}+J_{12}+J_{13}+J_{14}.
\end{aligned}$$

Since $\sqrt{n}J_{11}$ is sum of i.i.d random variables, it follows from the Central Limit Theorem that $J_{11} \xrightarrow{\mathcal{L}} N(0, \Lambda_I)$, where $\Lambda_I = \lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^n u_i^{\mathrm{T}}(\beta)V_i^{-1}\tilde{\Delta}_i\Sigma_i\tilde{\Delta}_iV_i^{-1}u_i(\beta)$. Similarly, for $J_{1k}(k = 2, 3, 4)$, we can prove that $J_{1k} = o_p(1)$ for $k = 2, 3, 4$. Under the MAR assumption and the fact that $\sqrt{n}(\hat{\beta} - \beta) = O_p(1)$, we can show that $J_2 = o_p(1)$. Combining the above equations yields that $n^{-1/2}\sum_{i=1}^n Z_{i2}(\beta) \xrightarrow{\mathcal{L}} N(0, \Lambda_I)$.

Lemma 3 *Suppose that the conditions (A1)–(A11) hold. Then, we have*

$$\begin{aligned}
&\frac{1}{n}\sum_{i=1}^n Z_{i1}(\beta)Z_{i1}^{\mathrm{T}}(\beta) \xrightarrow{P} \Lambda_c, \quad \frac{1}{n}\sum_{i=1}^n \frac{\partial Z_{i1}(\beta)}{\partial\beta} \xrightarrow{P} \Xi_c,\\
&\frac{1}{n}\sum_{i=1}^n Z_{i2}(\beta)Z_{i2}^{\mathrm{T}}(\beta) \xrightarrow{P} \Lambda_I, \quad \frac{1}{n}\sum_{i=1}^n \frac{\partial Z_{i2}(\beta)}{\partial\beta} \xrightarrow{P} \Xi_I,
\end{aligned}$$

where $\Xi_c = -\lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^n u_i^{\mathrm{T}}\Delta_iV_i^{-1}\Delta_iu_i$, *and* $\Xi_I = -\lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^n u_i^{\mathrm{T}}V_i^{-1}\tilde{\Delta}_iu_i$.

Proof Let $V_{i1} = \sum_{k=1}^{n_i}\sum_{l=1}^{n_i}\delta_{ik}\delta_{il}u_{ik}\sigma_i^{kl}\varepsilon_{il}$ and $V_{i2} = \sum_{k=1}^{n_i}\sum_{l=1}^{n_i}\delta_{ik}\delta_{il}\sigma_i^{kl}\{\check{h}(T_{ik},\beta)\varepsilon_{il}+u_{ik}\check{g}(T_{il})+\check{h}(T_{ik},\beta)\check{g}(T_{il})\}$, where $\check{h}(\cdot;\cdot)$ and $\check{g}(\cdot)$ are defined in proof of Lemma 2. Then, it follows from the definition of $Z_{i1}(\beta)$ that $Z_{i1}(\beta) = V_{i1} + V_{i2}$, which leads to

$$\begin{aligned}
\frac{1}{n}\sum_{i=1}^n Z_{i1}(\beta)Z_{i1}^{\mathrm{T}}(\beta) &= \frac{1}{n}\sum_{i=1}^n V_{i1}V_{i1}^{\mathrm{T}} + \frac{1}{n}\sum_{i=1}^n V_{i2}V_{i2}^{\mathrm{T}} + \frac{1}{n}\sum_{i=1}^n V_{i1}V_{i2}^{\mathrm{T}} + \frac{1}{n}\sum_{i=1}^n V_{i2}V_{i1}^{\mathrm{T}}\\
&\triangleq H_1 + H_2 + H_3 + H_4.
\end{aligned}$$

Using Laws of Large Number, we can obtain $H_1 \xrightarrow{P} \Lambda_c$. Next, we study the asymptotic properties of H_v for $v = 2, 3$ and 4. We first study asymptotic property of H_2. Let $H_{2,rs}$ be the (r, s)th component of H_2, and $V_{2i,r}$ be the rth component of V_{i2}. By the Canchy–Schwarz inequality, we have $\|H_{2,rs}\| \le (n^{-1}\sum_{i=1}^n V_{2i,r}^2)^{\frac{1}{2}}(n^{-1}\sum_{i=1}^n V_{2i,s}^2)^{\frac{1}{2}}$. It follows from Lemma 1 that $\frac{1}{n}\sum_{i=1}^n V_{2i,r}^2 \xrightarrow{P} 0$, which indicates $H_2 \xrightarrow{P} 0$. Similarly, we can show that $H_3 \xrightarrow{P} 0$ and $H_4 \xrightarrow{P} 0$. Therefore, combining the above results yields $\frac{1}{n}\sum_{i=1}^n Z_{i1}(\beta)Z_{i1}^{\mathrm{T}}(\beta) \xrightarrow{P} \Lambda_c$.

Again, by the definition of $Z_{i1}(\beta)$, it is easy to show that

$$\begin{aligned}
\frac{1}{n}\sum_{i=1}^n \frac{\partial Z_{i1}(\beta)}{\partial\beta} &= \frac{1}{n}\sum_{i=1}^n\sum_{k=1}^{n_i}\sum_{l=1}^{n_i}\left\{\delta_{ik}\delta_{il}\sigma_i^{kl}\frac{\partial\tilde{d}_{ik}(\beta)}{\partial\beta^{\mathrm{T}}}\varepsilon_{il}\right\}\\
&\quad+\frac{1}{n}\sum_{i=1}^n\sum_{k=1}^{n_i}\sum_{l=1}^{n_i}\left\{\delta_{ik}\delta_{il}\sigma_i^{kl}\frac{\partial\tilde{d}_{ik}(\beta)}{\partial\beta^{\mathrm{T}}}\check{g}(T_{il})\right\}\\
&\quad-\frac{1}{n}\sum_{i=1}^n\sum_{k=1}^{n_i}\sum_{l=1}^{n_i}\left\{\delta_{ik}\delta_{il}\sigma_i^{kl}u_{ik}u_{il}^{\mathrm{T}}\right\}
\end{aligned}$$

$$
\begin{aligned}
&-\frac{1}{n}\sum_{i=1}^{n}\sum_{k=1}^{n_i}\sum_{l=1}^{n_i}\left\{\delta_{ik}\delta_{il}\sigma_i^{kl}u_{ik}\check{h}^{\mathrm{T}}(T_{il},\beta)\right\}\\
&-\frac{1}{n}\sum_{i=1}^{n}\sum_{k=1}^{n_i}\sum_{l=1}^{n_i}\left\{\delta_{ik}\delta_{il}\sigma_i^{kl}\check{h}(T_{il},\beta)u_{ik}^{\mathrm{T}}\right\}\\
&-\frac{1}{n}\sum_{i=1}^{n}\sum_{k=1}^{n_i}\sum_{l=1}^{n_i}\left\{\delta_{ik}\delta_{il}\sigma_i^{kl}\check{h}(T_{ik},\beta)\check{h}^{\mathrm{T}}(T_{il},\beta)\right\}\\
&\triangleq M_1+M_2+M_3+M_4+M_5+M_6.
\end{aligned}
$$

By the Law of Large Number, we obtain $M_1 \xrightarrow{P} 0$ and $M_3 \xrightarrow{P} \Xi_c$. It follows from Lemma 1 that $M_v \xrightarrow{P} 0$ for $v = 2, 4, 5, 6$. Combining the above equations yields $\frac{1}{n}\sum_{i=1}^{n}\frac{\partial Z_{i1}(\beta)}{\partial\beta} \xrightarrow{P} \Xi_c$. Similarly, we can show that other two equations also hold. □

Proof of Theorem 1 Let $\ell_l(\beta) = 2\sum_{i=1}^{n}\log(1+\lambda_{nl}^{\mathrm{T}}(\beta)Z_{il}(\beta)) \triangleq 2\sum_{i=1}^{n}\log(1+r_{il})$, where $r_{il} = \lambda_{nl}^{\mathrm{T}}(\beta)Z_{il}(\beta)$ for $l = c$ and I. Taking Taylor expansion of $\ell_l(\beta)$ at $r_{il} = 0$ yields

$$
\begin{aligned}
\ell_l(\beta) &= 2\sum_{i=1}^{n}(r_{il}-\frac{1}{2}r_{il}^2+\eta_{il}) = 2n\lambda_{nl}^{\mathrm{T}}\left\{\frac{1}{n}\sum_{i=1}^{n}Z_{il}(\beta)\right\} - n\lambda_{nl}^{\mathrm{T}}S_l\lambda_{nl}+2\sum_{i=1}^{n}\eta_{il}\\
&= n\left\{\frac{1}{n}\sum_{i=1}^{n}Z_{il}(\beta)\right\}^{\mathrm{T}}S_l^{-1}\left\{\frac{1}{n}\sum_{i=1}^{n}Z_{il}(\beta)\right\} - n\xi_{nl}^{\mathrm{T}}S_l^{-1}\xi_{nl}+2\sum_{i=1}^{n}\eta_{il},
\end{aligned}
$$

where $\xi_{nl} = n^{-1}\sum_{i=1}^{n}Z_{il}(\beta)r_{il}^2/(1+r_{il}) = O_p(n^{-\frac{1}{2}})$, $S_l = \frac{1}{n}\sum_{i=1}^{n}Z_{il}(\beta)Z_{il}^{\mathrm{T}}(\beta)$ and η_{il} is the remainder term with respect to r_{il} for $l = c$ and I.

From Lemmas 2 and 3, we obtain $n\{\frac{1}{n}\sum_{i=1}^{n}Z_{il}(\beta)\}^{\mathrm{T}}S_l^{-1}\{\frac{1}{n}\sum_{i=1}^{n}Z_{il}(\beta)\} \xrightarrow{\mathcal{L}} \chi_p^2$ as $n \to \infty$. It follows from the definitions of ξ_{nl} and S_l and the above equations that $n\xi_{nl}^{\mathrm{T}}S_l^{-1}\xi_{nl} = no_p(n^{-\frac{1}{2}})O_p(1)o_p(n^{-\frac{1}{2}}) = o_p(1)$ and $2\sum_{i=1}^{n}\eta_{il} \le 2C\|\lambda_{nl}\|^3\sum_{i=1}^{n}\|Z_{il}(\beta)\|^3 = O_p(n^{-\frac{3}{2}})o_p(n^{\frac{3}{2}}) = o_p(1)$. Then, combining the above equations leads to $\ell_l(\beta) \xrightarrow{\mathcal{L}} \chi_p^2$ for $l = c$ and I. □

Proof of Theorem 2 Let $T_{1nl}(\beta,\lambda_{nl}) = n^{-1}\sum_{i=1}^{n}Z_{il}(\beta)/\{1+\lambda_{nl}^{\mathrm{T}}Z_{il}(\beta)\}$ and $T_{2nl}(\beta,\lambda_{nl}) = n^{-1}\sum_{i=1}^{n}\{\partial Z_{il}(\beta)/\partial\beta\}^{\mathrm{T}}\lambda_{nl}/\{1+\lambda_{nl}^{\mathrm{T}}Z_{il}(\beta)\}$ for $l = c$ and I. Then, $\hat{\beta}_l$ and $\hat{\lambda}_{nl}$ are the solutions of the following equations: $T_{1nl}(\beta,\lambda_{nl}) = 0$ and $T_{2nl}(\beta,\lambda_{nl}) = 0$. Taking Taylor expansions of $T_{1nl}(\hat{\beta}_l,\hat{\lambda}_{nl})$ and $T_{2nl}(\hat{\beta}_l,\hat{\lambda}_{nl})$ at $(\beta, 0)$ yields

$$
\begin{aligned}
0 = T_{1nl}(\hat{\beta}_l,\hat{\lambda}_{nl}) = T_{1nl}(\beta,0)+\frac{\partial T_{1nl}(\beta,0)}{\partial\beta}(\hat{\beta}_l-\beta)+\frac{\partial T_{1nl}(\beta,0)}{\partial\lambda_{nl}}\hat{\lambda}_{nl}+o_p(\sigma_{nl}),\\
0 = T_{2nl}(\hat{\beta}_l,\hat{\lambda}_{nl}) = T_{2nl}(\beta,0)+\frac{\partial T_{2nl}(\beta,0)}{\partial\beta}(\hat{\beta}_l-\beta)+\frac{\partial T_{2nl}(\beta,0)}{\partial\lambda_{nl}}\hat{\lambda}_{nl}+o_p(\sigma_{nl}),
\end{aligned}
$$

which leads to

$$\begin{pmatrix} \hat{\lambda}_{nl} \\ \hat{\beta}_l - \beta \end{pmatrix} = S_{nl}^{-1} \begin{pmatrix} -T_{1nl}(\beta, 0) + o_p(\sigma_{nl}) \\ o_p(\sigma_{nl}) \end{pmatrix},$$

where $\sigma_{nl} = \|\hat{\beta}_l - \beta\| + \|\hat{\lambda}_{nl}\|$ and

$$\begin{aligned} S_{nl} &= \begin{pmatrix} \frac{\partial T_{1nl}(\beta,0)}{\partial \lambda_{nl}} & \frac{\partial T_{1nl}(\beta,0)}{\partial \beta} \\ \frac{\partial T_{2nl}(\beta,0)}{\partial \lambda_{nl}} & \frac{\partial T_{2nl}(\beta,0)}{\partial \beta} \end{pmatrix} \\ &= \begin{pmatrix} -\frac{1}{n}\sum_{i=1}^{n} Z_{il}(\beta) Z_{il}^{\mathrm{T}}(\beta) & \frac{1}{n}\sum_{i=1}^{n} \frac{\partial Z_{il}(\beta)}{\partial \beta^{\mathrm{T}}} \\ \frac{1}{n}\sum_{i=1}^{n} \frac{\partial Z_{il}(\beta)}{\partial \beta^{\mathrm{T}}} & 0 \end{pmatrix} \xrightarrow{P} \begin{pmatrix} S_{l11} & S_{l12} \\ S_{l21} & 0 \end{pmatrix}. \end{aligned}$$

Then, we have

$$S_{nl}^{-1} \xrightarrow{P} \begin{pmatrix} S_{l11}^{-1} + S_{l11}^{-1} S_{l12} S_{l22.1}^{-1} S_{l22} S_{l11}^{-1} & -S_{l11}^{-1} S_{l12} S_{l22.1}^{-1} \\ -S_{l22.1}^{-1} S_{l21} S_{l11}^{-1} & S_{l22.1}^{-1} \end{pmatrix},$$

where $S_{l22.1} = -S_{l21} S_{l11}^{-1} S_{l12}$. It follows from Lemma 3 that $T_{1nl}(\beta, 0) = \frac{1}{n}\sum_{i=1}^{n} Z_{il}(\beta) = O_p(n^{-\frac{1}{2}})$ and $\|\lambda_{nl}\| = O_p(n^{-\frac{1}{2}})$, which indicates that $\sigma_{nl} = \|\hat{\beta}_l - \beta\| + \|\hat{\lambda}_{nl}\| = o_p(n^{-\frac{1}{2}})$. Combining the above equations leads to $\sqrt{n}(\hat{\beta}_l - \beta) = S_{l22.1}^{-1} S_{l21} S_{l11}^{-1} \sqrt{n}\ T_{1nl}(\beta, 0) + o_p(1)$ for $l = c$ and I. Then, it follows from the above equations and Lemmas 2 and 3 that $\sqrt{n}(\hat{\beta}_l - \beta) \xrightarrow{\mathcal{L}} N(0, \Xi_l^{-1} \Lambda_l \Xi_l^{-1})$ for $l = c$ and I. □

Proof of Theorem 3 Using Taylor expansion as done in Theorem 1, we obtain $\hat{\ell}(g(t_0)) = \left(\sum_{i=1}^{n} \hat{\eta}_{iR}(g(t_0))\right)^2 / \sum_{i=1}^{n} \hat{\eta}_{iR}^2(g(t_0)) + o_p(1)$. By the definition of $\hat{\eta}_{iR}(g(t))$, it is easy to obtain that

$$\begin{aligned} \frac{1}{\sqrt{Nh}} \sum_{i=1}^{n} \hat{\eta}_{iR}(g(t_0)) &= \left\{ \frac{1}{\sqrt{Nh}} \sum_{i=1}^{n} \hat{\eta}_{iE}(g(t_0) - b(t_0)) \right\} - [\hat{b}(t_0) - b(t_0)], \\ \sum_{i=1}^{n} \hat{\eta}_{iR}^2(g(t_0)) &= \sum_{i=1}^{n} \hat{\eta}_{iE}^2(g(t_0)) - 2\sum_{i=1}^{n} \hat{\eta}_{iE}^2(g(t_0))\hat{\varphi}_i(t_0) + \sum_{i=1}^{n} \hat{\varphi}_i^2(t_0), \end{aligned}$$

where $\hat{\varphi}_i(t_0) = \sum_{j=1}^{n_i}\{\hat{g}_n^{\mathrm{C}}(T_{ij}) - \hat{g}_n^{\mathrm{C}}(t_0)\}\delta_{ij} K_n(T_{ij} - t_0)$.

According to Lemmas 4, 5 and Theorem 8 given in Xue and Xue (2011), we have

$$\begin{aligned} &\frac{1}{\sqrt{Nh}} \sum_{i=1}^{n} \hat{\eta}_{iE}(g(t_0) - b(t_0)) \xrightarrow{\mathcal{L}} N(0, V^2(t_0)), \quad \frac{1}{\sqrt{Nh}} \sum_{i=1}^{n} \hat{\eta}_{iE}^2(g(t_0)) \xrightarrow{\mathcal{P}} V^2(t_0), \\ &\hat{b}(t_0) \xrightarrow{\mathcal{P}} b(t_0), \quad \frac{1}{\sqrt{Nh}} \sum_{i=1}^{n} \hat{\varphi}_i^2(t_0) \xrightarrow{\mathcal{P}} 0, \quad \frac{1}{\sqrt{Nh}} \sum_{i=1}^{n} \hat{\eta}_{iE}^2(g(t_0))\hat{\varphi}_i(t_0) \xrightarrow{\mathcal{P}} 0. \end{aligned}$$

Combining the above equations, we prove that Theorem 3 holds. □

Proof of Theorem 4 By the definition of $\hat{\eta}_{iE}(g(t))$, we obtain

$$\sqrt{Nh}\{\hat{g}(t_0) - g(t_0)\} = \frac{\frac{1}{\sqrt{Nh}}\sum_{i=1}^{n}\hat{\eta}_{iE}(g(t_0))}{m(t_0)} + o_p(1).$$

From Lemma 4 of Xue and Xue (2011), we can show that Theorem 4 holds. □

Proof of Theorem 5 By the definition of $\hat{g}^{\mathrm{MIP}}(t)$, we have

$$\begin{aligned}\hat{g}^{\mathrm{MIP}}(t)-g(t) &= (\hat{g}_2^{\mathrm{MIP}}(t)-g_2(t)) - (\hat{g}_1^{\mathrm{MIP}}(t;\beta) - g_1(t;\beta))-(g_1(t;\hat{\beta}) - g_1(t;\beta))\\ &\quad -[\hat{g}_1^{\mathrm{MIP}}(t;\hat{\beta}) - \hat{g}_1^{\mathrm{MIP}}(t;\beta) - g_1(t;\hat{\beta}) + g_1(t;\beta)],\\ &\triangleq H_1(t) - H_2(t) - H_3(t) - H_4(t).\end{aligned}$$

Again, it follows from the definition of $\hat{g}_2^{\mathrm{MIP}}(t)$ that

$$\begin{aligned}H_1(t) &= \sum_{i=1}^{n}\sum_{j=1}^{n_i} W_{ij}(t)[\tilde{Y}_{ij}^{I} - g_2(t)] + \sum_{i=1}^{n}\sum_{j=1}^{n_i} W_{ij}(t)\left(1 - \frac{\delta_{ij}}{p(X_{ij},T_{ij})}\right)(f(X_{ij};\hat{\beta})\\ &\quad - f(X_{ij};\beta)) + \sum_{i=1}^{n}\sum_{j=1}^{n_i} W_{ij}(t)\left(1 - \frac{\delta_{ij}}{p(X_{ij},T_{ij})}\right)(\hat{g}_n^{\mathrm{C}}(T_{ij}) - g(T_{ij}))\\ &\triangleq H_{11}(t) + H_{12}(t) + H_{13}(t).\end{aligned}$$

Taking Taylor expansion of $f(X_{ij};\hat{\beta})$ at $\hat{\beta} = \beta$ yields $f(X_{ij};\hat{\beta}) = f(X_{ij};\beta) + V_{ij}(\beta)(\hat{\beta}-\beta) + o_p(\|\hat{\beta}-\beta\|)$, which leads to $g_1(t;\hat{\beta}) \approx g_1(t;\beta) + (\hat{\beta}-\beta)M(t;\beta)$ and $\hat{g}_1^{\mathrm{MIP}}(t;\hat{\beta}) \approx \hat{g}_1^{\mathrm{MIP}}(t;\beta) + (\hat{\beta}-\beta)\hat{M}(t;\beta)$, where $V_{ij}(\beta) = \partial f(X_{ij};\beta)/\partial\beta$, $M(t;\beta) = E\{V_{ij}(\beta)|T_{ij} = t\}$ and $\hat{M}(t;\beta) = \sum_{i=1}^{n}\sum_{j=1}^{n_i} W_{ij}(t)V_{ij}(\beta)$. Thus, it follows from the definitions of $H_3(t)$, $H_4(t)$ and $H_{12}(t)$ that $H_3(t) \approx (\hat{\beta}-\beta)M(t;\beta)$, $H_4(t) \approx (\hat{\beta}-\beta)(\hat{M}(t;\beta) - M(t;\beta))$, $H_{12}(t) \approx \sum_{i=1}^{n}\sum_{j=1}^{n_i} W_{ij}(t)(1 - \delta_{ij}/P(X_{ij},T_{ij}))V_{ij}(\beta)(\hat{\beta}-\beta)$. Note that $E\{\tilde{Y}_{ij}^{I}|T_{ij} = t\} = g_2(t)$ and $E\{|(1 - \delta_{ij}/p(X_{ij},T_{ij}))\partial f(X_{ij};\beta)/\partial\beta|\} = 0$ under MAR assumption. Hence, by standard kernel regression theories (Wand and Jones 1995), we have

$$\begin{aligned}&\sup_t H_{11}(t) = O_p((nb)^{-\frac{1}{2}}) + O_p(b),\quad \sup_t E[|\hat{g}_n^{\mathrm{C}}(T_{ij}) - g(T_{ij})||T_{ij} = t]\\ &\qquad\qquad = O((nh)^{-\frac{1}{2}}) + O(h),\\ &\sup_t H_2(t) = O_p((nb)^{-\frac{1}{2}}) + O_p(b),\quad \sup_t|\hat{M}(t;\beta) - M(t;\beta)|\\ &\qquad\qquad = O_p((nb)^{-\frac{1}{2}}) + O_p(b),\\ &\sum_{i=1}^{n}\sum_{j=1}^{n_i} W_{ij}(t)\left(1 - \frac{\delta_{ij}}{P(X_{ij},T_{ij})}\right)V_{ij}(\beta) = O_p(1),\\ &\qquad\sum_{i=1}^{n}\sum_{j=1}^{n_i} W_{ij}(t)\left(1 - \frac{\delta_{ij}}{P(X_{ij},T_{ij})}\right) = O_p(1).\end{aligned}$$

Hence, it follows ffrom the above equations and $\hat{\beta} - \beta = O_p(n^{-\frac{1}{2}})$ that

$$\begin{aligned}\sup_t |\hat{g}^{\mathrm{MIP}}(t) - g(t)| &= O_p((nb)^{-\frac{1}{2}}) + O_p(b) + O_p(n^{-\frac{1}{2}}) + O_p((nh)^{-\frac{1}{2}}) + O_p(h) \\ &\quad + O_p((nb)^{-\frac{1}{2}}) + O_p(b) + O_p(n^{-\frac{1}{2}}) \\ &\quad + \{O_p((nb)^{-\frac{1}{2}}) + O_p(b)\} O_p(n^{-\frac{1}{2}}) \\ &= O_p((nb)^{-\frac{1}{2}}) + O_p(b) + O_p((nh)^{-\frac{1}{2}}) + O_p(h).\end{aligned}$$

Then, we prove Theorem 5. □

Acknowledgments The authors thank two anonymous referees for their helpful comments and suggestions which have substantially improved the readability and the presentation of this paper. The research was fully supported by grants from the National Natural Science Foundation of China (10961026, 11171293), Specialized Research Fund for the Doctoral Program of Higher Education of China (No. 20115301110004) and the Natural Science Key Project of Yunnan Province (No. 2010CC003).

References

Bates, D. M., Watts, D. G. (1988). *Nonlinear Regression Analysis and Its Applications*. New York: Wiley.

Chen, Q. X., Ibrahim, J. G., Chen, M. H., Senchaudhuri, P. (2008). Theory and inference for regression models with missing responses and covariates. *Journal of Multivariate Analysis*, *99*, 1302–1331.

Chen, S. X., Zhong, P. S. (2010). ANOVA for longitudinal data with missing values. *The Annals of Statistics*, *38*, 3630–3659.

Ciuperca, G. (2011). Empirical likelihood for nonlinear models with missing responses. *Journal of Statistical Computation and Simulation*. doi:10.1080/00949655.2011.635305.

Fan, J., Li, R. Z. (2004). New estimation and model selection procedures for semiparametric modeling in longitudinal data analysis. *Journal of the American Statistical Association*, *99*, 710–723.

Green, P. J. (1987). Penalized likelihhod for general semiparametric regression models. *International Statistical Review*, *55*, 245–259.

Ibrahim, J. G., Chen, M. H., Lipsitz, S. R. (2001). Missing responses in generalized linear mixed models when the missing data mechanism is nonignorable. *Biometrika*, *88*, 551–564.

Kim, J. K., Yu, C. L. (2011). A semiparametric estimation of mean functionals with nonignorable missing data. *Journal of the American Statistical Association*, *106*, 157–165.

Lee, S. Y., Tang, N. S. (2006). Analysis of nonlinear structural equation models with nonignorable missing covariates and ordered categorical data. *Statistica Sinica*, *16*, 1117–1141.

Li, R., Nie, L. (2008). Efficient statistical inference procedures for partially nonlinear models and their applications. *Biometrics*, *64*, 904–911.

Liang, H., Wang, S. J., Carroll, R. J. (2007). Partially linear models with missing response variables and error-prone covariates. *Biometrika*, *94*, 185–198.

Liang, H., Wu, H., Carroll, R. J. (2003). The relationship between virologic responses in AIDS clinical research using mixed-effects varying-coefficient models with measurement error. *Biostatistics*, *4*, 297–312.

Liang, H., Qin, Y. S. (2008). Empirical likelihood based inference for partially linear models with missing covariates. *Australian New Zealand Journal of Statistics*, *50*, 347–359.

Lin, X., Carroll, R. J. (2001). Semiparametric regression for clustered data using generalized estimating equations. *Journal of the American Statistical Association*, *96*, 1045–1056.

Little, R. J. A., Rubin, D. B. (2002). *Statistical Analysis with Missing Data* (2nd ed.). New York: John Wiley.

Mellors, J. W., Rinaldo, C. R., Gupta, P., White, R. M., Todd, J. A., Kingsley, L. A. (1996). Prognosis in HIV-1 infection predicted by the quantity of virus in plasma. *Science*, *272*, 1167–1170.

Müller, U. U. (2009). Estimating linear functionals in nonlinear regression with responses missing at random. *The Annals of Statistics*, *37*, 2245–2277.

Owen, A. B. (2001). *Empirical Likelihood*. New York: Chapman&Hall.

Ruppert, D., Wand, M. P., Carroll, R. J. (2003). *Semiparametric Regression*. New York: Cambridge University Press.

Saag, M. S., Holodniy, M., Kuritzkes, D. R., OBrien, W. A., Coombs, R., Poscher, M. E., et al. (1996). HIV viral load markers in clinical practice. *Nature Medicine.*, *2*, 625–629.

Shardell, M., Miller, R. R. (2008). Weighted estimating equations for longitudinal studies with death and non-monotone missing time-dependent covariates and outcomes. *Statistics in Medicine*, *27*, 1008–1025.

Stone, C. J. (1980). Optimal rates of convergence for nonparametric estimators. *The Annals of Statistics*, *8*, 1348–1360.

Wand, M. P., Jones, M. C. (1995). *Kernel Smoothing*. New York: Chapman, Hall.

Wang, Y., Ke, C. (2009). Smoothing spline semiparametric nonlinear regression models. *Journal of Computational and Graphical Statistics*, *18*, 165–183.

Wang, Q. H., Lindon, O., Härdle, W. (2004). Semiparametric regression analysis with missing response at random. *Journal of the American Statistical Association*, *99*, 334–345.

Wu, C. J. F. (1981). Asymptotic theory of nonlinear least squares estimation. *The Annal of Statistics*, *9*, 501–513.

Xue, L. G., Xue, D. (2011). Empirical Likelihood for semiparametric regression models with missing reponses. *Journal of Multivariate Analysis*, *102*, 723–740.

Yates, F. (1933). The analysis of replicated experiments when the field results are incomplete. *The Empire Journal of Experimental Agriculture*, *1*, 129–142.

Yi, G. Y., Cook, R. J. (2002). Marginal Methods for Incomplete Longitudinal Data Arising in Clusters. *Journal of the American Statistical Association*, *97*, 1071–1080.

Zeger, S. L., Diggle, P. L. (1994). Semiparametric model for longitudinal data with application to CD4 Cell numbers in HIV seroconveriers. *Biometrics*, *50*, 689–699.

Zhu, Z. Y., Tang, N. S., Wei, B. C. (2000). On confidence regions of semiparametric nonlinear regression models (a geometric approach). *Acta Mathematica Scientia*, *20*, 68–75.