CrossMark

Empirical likelihood for heteroscedastic partially linear single-index models with growing dimensional data

Jianglin Fang[1] · **Wanrong Liu[2]** · **Xuewen Lu[3]**

Abstract In this paper, we propose a new approach to the empirical likelihood inference for the parameters in heteroscedastic partially linear single-index models. In the growing dimensional setting, it is proved that estimators based on semiparametric efficient score have the asymptotic consistency, and the limit distribution of the empirical log-likelihood ratio statistic for parameters $(\beta^\top, \theta^\top)^\top$ is a normal distribution. Furthermore, we show that the empirical log-likelihood ratio based on the subvector of β is an asymptotic chi-square random variable, which can be used to construct the confidence interval or region for the subvector of β. The proposed method can naturally be applied to deal with pure single-index models and partially linear models with high-dimensional data. The performance of the proposed method is illustrated via a real data application and numerical simulations.

1 Introduction

Consider the partially linear single-index model

$$Y_i = X_i^\top \beta + g\left(Z_i^\top \theta\right) + \varepsilon_i, \quad (i = 1, \ldots, n), \tag{1}$$

✉ Jianglin Fang
  fangjlin@foxmail.com

[1] College of Science, Hunan Institute of Engineering, Xiangtan 411104, Hunan, China

[2] College of Mathematics and Computer Science, Hunan Normal University, Changsha 410081, China

[3] Department of Mathematics and Statistics, University of Calgary, Calgary, AB T2N 1N4, Canada

where $X_i \in \mathbb{R}^p$ and $Z_i \in \mathbb{R}^r$ are random variables with dimensions p and r respectively, Y_i is a response variable, β and θ are unknown parameters, $g(\cdot)$ is an unknown function, ε_i's are independent random errors with $E(\varepsilon_i|X_i, Z_i) = 0$ and $Var(\varepsilon_i|X_i, Z_i) = v(X_i, Z_i) > 0$, $v(X, Z)$ is a function of (X, Z) representing possible heteroscedasticity. For identifiability, we assume that the first component of θ is 1, and we use Z_{-1} to denote the last $r - 1$ components of Z. An interesting problem is statistical inference about $(\beta^\top, \theta^\top)^\top$ in partially linear single-index models, whereas all the other unknown components of model (1), such as the unknown function $g(\cdot)$ and the unspecified function $v(X, Z)$, are termed nuisance parameters. Model (1) was proposed by Carroll et al. (1997), and is a natural extension of the partially linear model and the single-index model (See Engle et al. 1986; Xia et al. 1999). The parametric component $X_i^\top \beta$ provides a simple summary of covariate effects which are of the main scientific interest. The index $Z_i^\top \theta$ enables us to simplify the treatment of the multiple auxiliary variables, and the smooth baseline component $g(\cdot)$ enriches model flexibility. Since the partially linear single-index model was introduced, it has been broadly and deeply studied by many authors in various disciplines. For example, Yu and Ruppert (2002) proposed a penalized spline estimation procedure; Xia et al. (2002) integrated the dimension reduction idea and minimum average variance estimation; Xia and Härdle (2006) showed semiparametric estimation of partially linear single-index models; Lai and Wang (2014) studied semiparametric efficient estimation with responses missing at random. Although, such as Yu and Ruppert (2002), they did not explicitly make an equal variance assumption for ε, they did not account for the heteroscedasticity of model (1). Therefore, these estimators are not efficient when heteroscedasticity is present.

In practice, many variables are introduced to reduce possible modeling biases. In the early days, the value for the number of parameters p is in the ranges 10–500, and sample size n is in the ranges 100–10,000. From problems in X-ray crystallography, Huber (1973) noted that in a variable selection context the number of parameters is often large and should be modeled as p_n, which tends to ∞. Afterwards, with the development of technology and huge investment in various forms of data gathering, Donohn (2000) demonstrated with web term-document data, gene expression data and consumer financial history data, large sample sizes with high dimensions are important characteristics. He also found that even in a classical setting such as the Framingham heart study, the sample size is as large as $n = 25,000$ and the dimension is $p = 100$, which can be modeled as $p = O(n^{1/3})$ or $p = O(n^{1/2})$. Recently, high-dimensional data becomes more and more popular in many areas, such as financial and statistical applications, hyperspectral imagery, internet portals, high-throughput genomic data analysis and other areas of computational biology; see, e.g., Bai and Saranadasa (1996), Ledoit and Wolf (2002), Hjort et al. (2009), Zhang et al. (2012) and Ma and Zhu (2013). Zhang et al. (2012) integrated the dimension reduction idea and variable selection in partially linear single-index models with high-dimensional covariates, Ma and Zhu (2013) proposed efficient estimators for heteroscedastic partially linear single-index models allowing high-dimensional covariates. For model (1), although they allowed high-dimensional covariates, the dimension of covariates is fixed, and can not tend to infinity as the sample size $n \to \infty$.

The method of empirical likelihood introduced in Owen ([1988](#), [1990](#), [1991](#), [2001](#)) might be useful for the purpose of making inference for model ([1](#)). To the best of our knowledge, there is not much literature on this model by using empirical likelihood method, although it has been successfully applied to various models, see, e.g., linear models (Owen [1991](#)), generalized linear models (Kolaczyk [1994](#)), partially linear models (Shi and Lau [2000](#)), heteroscedastic partially linear models (Lu [2009](#)), single-index models (Xue and Zhu [2006](#)), partially linear single-index models (Zhu and Xue [2006](#)), general estimating equations (Qin and Lawless [1994](#)), quantile estimation (Chen and Hall [1993](#)), errors-in-covariables models (Wang and Rao [2002](#)), Cox regression models(Qin and Jing [2001](#)), additive risk models (Lu and Qi [2004](#)). Empirical likelihood has also been applied to some high-dimensional problems and its asymptotic behavior under the setting where n and p both tend to infinity has also been carefully studied. Hjort et al. ([2009](#)) derived the limit distribution of the EL ratio statistic based on p-dimensional estimating equations when $p \to \infty$ with n at the rate $p = o\left(n^{1/3}\right)$; Chen et al. ([2009](#)) improved upon the rate restriction in Hjort et al. ([2009](#)) and established a nondegenerate limit distribution of the EL ratio statistic, allowing $p = o\left(n^{1/2}\right)$ under suitable regularity conditions.

Zhu and Xue ([2006](#)) investigated likelihood confidence regions in a partially linear single-index model. In their paper, empirical likelihood was constructed from the components of a semiparametric inefficient estimating equation. We think that it might be more informative if we use semiparametric efficient score to construct the empirical likelihood. Furthermore, Zhu and Xue ([2006](#)) assumed that dimensions p and r for β and θ are fixed. Their results may not be valid when dimensions $p \to \infty$ and $r \to \infty$, as $n \to \infty$. Motivated by the empirical likelihood method for high-dimensional data in Hjort et al. ([2009](#)), in this paper, we propose a new approach to the empirical likelihood inference about $(\beta^\top, \theta^\top)^\top$ for heteroscedastic partially linear single-index models with high-dimensional data based on the semiparametric efficient score, where dimensions $p \to \infty$ and $r \to \infty$, as $n \to \infty$. We will show that the limit distribution of the empirical log-likelihood ratio for β is a normal distribution. Furthermore, we will show that the empirical log-likelihood ratio based on a k-dimensional ($k < p$) subvector of β is an asymptotically standard chi-square random variable, which can be used to construct confidence intervals or regions for the k-dimensional subvector of β.

The remainder of this paper is organized as follows. In Sect. [2](#), we introduce the empirical likelihood method for the inference of heteroscedastic partially linear single-index models and present our main results. The pure single-index model and the partially linear model, as the special examples, are discussed in Sect. [3](#). In Sect. [4](#), we report the results from simulation studies, and a real data example is presented in Sect. [5](#). Finally, the technical proofs of the main results are given in the "Appendix".

2 Methodology and main results

Firstly, we introduce the efficient estimation method of Ma and Zhu ([2013](#)) for the parameter $(\beta^\top, \theta^\top)^\top$ in the heteroscedastic partially linear single-index model ([1](#)),

and propose the empirical likelihood method for $(\beta^\top, \theta^\top)^\top$. To estimate $(\beta^\top, \theta^\top)^\top$, Ma and Zhu (2013) reviewed the following class of weighted estimating equations:

$$
\frac{1}{\sqrt{n}} \sum_{i=1}^{n} w\left(X_i, Z_i\right) \left\{ Y_i - X_i^\top \beta - f\left(Z_i^\top \theta\right) \right\} \left[a\left(X_i, Z_i\right) \right. \\
\left. - \tilde{E}\left\{ a\left(X, Z\right) | Z_i^\top \theta \right) \right\} \right] = 0, \tag{2}
$$

where $f(Z_i^\top \theta)$ is a given function of $Z_i^\top \theta$ which may or may not equal $g(Z_i^\top \theta)$, $\tilde{E}\{\cdot | Z_i^\top \theta)\}$ denotes a function of $Z_i^\top \theta$ that may or may not be the true $E\{\cdot | Z_i^\top \theta)\}$, $w_i = w(X_i, Z_i) = \text{var}(\varepsilon_i | X_i, Z_i)^{-1}$ and $a(\cdot, \cdot) \in \mathbb{R}^{p+r-1}$ is an arbitrary function of X and Z. They pointed out that consistent estimator $(\hat{\beta}^\top, \hat{\theta}^\top)^\top$ can be obtained by (2) if w_i and $f(\cdot)$ can be consistently estimated or is completely known. However, the doubly robustness and efficiency property of $(\hat{\beta}^\top, \hat{\theta}^\top)^\top$ is lost. In order to develop estimators for $(\beta^\top, \theta^\top)^\top$ with the doubly robustness and efficiency property, Ma and Zhu (2013) proposed an improved class of weighted estimation equations:

$$
\begin{cases}
\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \tilde{\varepsilon}_i \hat{w}\left(X_i, Z_i\right) \left[X_i - \dfrac{\hat{E}\{\hat{w}(X,Z)X | Z_i^\top \theta\}}{\hat{E}\{\hat{w}(X,Z) | Z_i^\top \theta\}} \right] = 0, \\
\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \tilde{\varepsilon}_i \hat{w}\left(X_i, Z_i\right) \hat{g}'\left(Z_i^\top \theta\right) \left[Z_{-1,i} - \dfrac{\hat{E}\{\hat{w}(X,Z)Z_{-1} | Z_i^\top \theta\}}{\hat{E}\{\hat{w}(X,Z) | Z_i^\top \theta\}} \right] = 0,
\end{cases} \tag{3}
$$

where $\tilde{\varepsilon}_i = Y_i - X_i^\top \beta - \hat{g}(Z_i^\top \theta)$, $w_i = w(X_i, Z_i) = E(\varepsilon_i^2 | X_i, Z_i)^{-1}$, $\hat{g}(Z_i^\top \theta)$, $\hat{g}'(Z_i^\top \theta)$, $\hat{E}\{\hat{w}(X, Z) | Z_i^\top \theta\}$, $\hat{E}\{\hat{w}(X, Z)X | Z_i^\top \theta\}$ and $\hat{E}\{\hat{w}(X, Z)Z_{-1} | Z_i^\top \theta\}$ are the nonparametric estimators via kernel estimation. We assume that $\eta_i = \eta(X_i, Z_i)$ is a low dimensional variable such that $\text{var}(\varepsilon_i | X_i, Z_i) = \text{var}(\varepsilon_i | \eta_i)$ and η has a known form. For example, η can be $Z^\top \theta$ or $X^\top \beta$, which means that the error variance depends on the covariates through $Z^\top \theta$ or $X^\top \beta$ only. Certainly, it can also be a combination of these two or can have any other form. Let $K_h(\cdot) = h^{-1} K(\cdot / h)$, where $K_h(\cdot)$ is a kernel function with bandwidth $h \to 0$. For bandwidths h_1, h_2 and h_3, we set

$$
\hat{g}\left(Z_i^\top \theta\right) = \sum_{j \neq i} K_{h_1}\left(Z_i^\top \theta - Z_j^\top \theta\right)\left(Y_j - X_j^\top \beta\right) / \sum_{j \neq i} K_{h_1}\left(Z_i^\top \theta - Z_j^\top \theta\right),
$$

$$
\hat{g}'\left(Z_i^\top \theta\right) = h_1^{-1} \left\{ \sum_{j \neq i} K_{h_1}'\left(Z_i^\top \theta - Z_j^\top \theta\right)\left(Y_j - X_j^\top \beta\right) \sum_{j \neq i} K_{h_1}\left(Z_i^\top \theta - Z_j^\top \theta\right) \right. \\
\left. - \sum_{j \neq i} K_{h_1}\left(Z_i^\top \theta - Z_j^\top \theta\right)\left(Y_j - X_j^\top \beta\right) \right. \\
\left. \times \sum_{j \neq i} K_{h_1}'\left(Z_i^\top \theta - Z_j^\top \theta\right) \right\} \bigg/ \left\{ \sum_{j \neq i} K_{h_1}\left(Z_i^\top \theta - Z_j^\top \theta\right) \right\}^2,
$$

$$
\hat{w}\left(X_i, Z_i\right) = \sum_{j \neq i} K_{h_2}\left(\eta_i - \eta_j\right) / \sum_{j \neq i} K_{h_2}\left(\eta_i - \eta_j\right) e_j^2,
$$

$$\hat{E}\left\{\hat{w}\left(X, Z\right) | Z_i^\top \theta\right\}$$
$$= \sum_{j \neq i} K_{h_3}\left(Z_i^\top \theta - Z_j^\top \theta\right) \hat{w}\left(X_j, Z_j\right) / \sum_{j \neq i} K_{h_3}\left(Z_i^\top \theta - Z_j^\top \theta\right),$$
$$\hat{E}\left\{\hat{w}\left(X, Z\right) X | Z_i^\top \theta\right\}$$
$$= \sum_{j \neq i} K_{h_3}\left(Z_i^\top \theta - Z_j^\top \theta\right) \hat{w}\left(X_j, Z_j\right) X_j / \sum_{j \neq i} K_{h_3}\left(Z_i^\top \theta - Z_j^\top \theta\right),$$
$$\hat{E}\left\{\hat{w}\left(X, Z\right) Z_{-1,i} | Z_i^\top \theta\right\}$$
$$= \sum_{j \neq i} K_{h_3}\left(Z_i^\top \theta - Z_j^\top \theta\right) \hat{w}\left(X_j, Z_j\right) Z_{-1,j} / \sum_{j \neq i} K_{h_3}\left(Z_i^\top \theta - Z_j^\top \theta\right).$$

The algorithm in calculating the estimators of $(\beta^\top, \theta^\top)^\top$ can be found in Ma and Zhu (2013).

Ma and Zhu (2013) proved that, under some mild conditions, the semiparametric efficient score is

$$S_{eff} = w\varepsilon\left(X^\top - \frac{E(wX^\top | Z^\top \theta)}{E(w | Z^\top \theta)}, g'(Z^\top \theta)\left\{Z_{-1}^\top - \frac{E\left(wZ_{-1}^\top | Z^\top \theta\right)}{E(w | Z^\top \theta)}\right\}\right)^\top, \quad (4)$$

and estimators $(\hat{\beta}^\top, \hat{\theta}^\top)^\top$ by solving (3) are doubly robust and efficient. In Ma and Zhu (2013), although they allowed high-dimensional covariates, the dimension of covariates is fixed, and can not tend to infinity as the sample size $n \to \infty$.

We first extend the fixed-dimensional results in Ma and Zhu (2013) to cases with diverging dimensionality, i.e., $p, r \to \infty$ as $n \to \infty$. Let $B_n \in \mathbb{R}^{(p+r-1) \times q}$ with fixed q. $B_n^\top S_{eff}$ represents a projection of the diverging dimensional vector S_{eff} to a fixed dimension q.

Definition 1 Assume that S_{eff} is a diverging dimensional semiparametric score, and $B_n^\top S_{eff}$ represents a projection of the diverging dimensional vector S_{eff} to a fixed dimension q. We say that the diverging dimensional semiparametric score is a semiparametric efficient score if the fixed dimensional projection $B_n^\top S_{eff}$ of S_{eff} is a semiparametric efficient score.

Similar to Proposition 1 of Ma and Zhu (2013), we can prove that, for any fixed q, the projection $B_n^\top S_{eff}$ of the diverging dimensional semiparametric score S_{eff} is a semiparametric efficient score. Therefore, according to Definition 1, we say that the diverging dimensional semiparametric score S_{eff} is a semiparametric efficient score.

The following theorem shows the theoretical properties of high-dimensional estimator based on the semiparametric score.

Theorem 2.1 *Let* $(\beta_0^\top, \theta_0^\top)^\top$ *be the true value of parameter vector* $(\beta^\top, \theta^\top)^\top$, *under Assumptions* 1–9, *then*

$$\sqrt{n} A V^{1/2}\left\{\left(\hat{\beta}^\top, \hat{\theta}^\top\right)^\top - \left(\beta_0^\top, \theta_0^\top\right)^\top\right\} \xrightarrow{L} N\left(0, G\right),$$

where $\overset{L}{\rightarrow}$ stands for convergence in distribution, $A_1 \in R^{q_1 \times p}$ and $A_2 \in R^{q_2 \times r}$ are $q_1 \times p$ and $q_2 \times r$ matrixes respectively, $AA^\top \rightarrow G$ and G is a $(q_1 + q_2) \times (q_1 + q_2)$ matrix with fixed q_1 and q_2,

$$A = \begin{pmatrix} A_1 & 0 \\ 0 & A_2 \end{pmatrix}, \quad V = \begin{pmatrix} V_{11} & V_{12} \\ V_{21} & V_{22} \end{pmatrix},$$

$$V_{11} = E\left\{ wXX^\top - \frac{E(wX|Z^\top\theta)E(wX^\top|Z^\top\theta)}{E(w|Z^\top\theta)} \right\},$$

$$V_{12} = E\left[g'(Z^\top\theta)\left\{ wXZ_{-1}^\top - \frac{E(wX|Z^\top\theta)E(wZ_{-1}^\top|Z^\top\theta)}{E(w|Z^\top\theta)} \right\} \right],$$

$$V_{21} = E\left[g'(Z^\top\theta)\left\{ wZ_{-1}X^\top - \frac{E(wZ_{-1}|Z^\top\theta)E(wX^\top|Z^\top\theta)}{E(w|Z^\top\theta)} \right\} \right],$$

$$V_{22} = E\left[g'(Z^\top\theta)^2\left\{ wZ_{-1}Z_{-1}^\top - \frac{E(wZ_{-1}|Z^\top\theta)E(wZ_{-1}^\top|Z^\top\theta)}{E(w|Z^\top\theta)} \right\} \right].$$

Theorem 2.1 extends the fixed-dimensional results in Ma and Zhu (2013) to cases with diverging dimensionality. Furthermore, in Theorem 2.1, A represents a projection of the diverging dimensional vector to a fixed dimension $q_1 + q_2$, and the limiting distribution of the projected vector of $\{(\hat{\beta}^\top, \hat{\theta}^\top)^\top - (\beta_0^\top, \theta_0^\top)^\top\}$ is a multivariate normal distribution. This theorem provides the consistency and normality of the projected vector of the estimator $(\hat{\beta}^\top, \hat{\theta}^\top)^\top$ for heteroscedastic partially linear single-index models.

Next, we introduce an auxiliary random vector by using the semiparametric efficient score. Let

$$\xi_i(\beta, \theta) = w_i\varepsilon_i \left(X_i^\top - \frac{E\left(wX^\top|Z_i^\top\theta\right)}{E(w|Z_i^\top\theta)}, g'\left(Z_i^\top\theta\right)\left\{ Z_{-1,i}^\top - \frac{E\left(wZ_{-1}^\top|Z_i^\top\theta\right)}{E(w|Z_i^\top\theta)} \right\} \right)^\top.$$

Note that $E\{\xi_i(\beta, \theta)\} = 0$ for $i = 1, \ldots, n$, if $(\beta^\top, \theta^\top)^\top$ is the true parameter. According to this fact, we apply the empirical likelihood method of Owen (1988, 1990) to make inference about $(\beta^\top, \theta^\top)^\top$. Let $\pi = (\pi_1, \cdots, \pi_n)$ be a probability vector, satisfying $\sum_{i=1}^{n} \pi_i = 1$ $\pi_i \geq 0$ for $i = 1, \ldots, n$. The traditional empirical likelihood function for $(\beta^\top, \theta^\top)^\top$ is defined as follows:

$$L(\beta, \theta) = \sup\left\{ \prod_{i=1}^{n}(n\pi_i) : \sum_{i=1}^{n}\pi_i = 1, \pi_i \geq 0, \sum_{i=1}^{n}\pi_i\xi_i(\beta, \theta) = 0 \right\}. \tag{5}$$

Because (5) contains unknown functions $w(X_i, Z_i)$, $g(Z_i^\top\theta)$, $E\{w(X, Z)|Z_i^\top\theta\}$, $g'(Z_i^\top\theta)$, $E\{w(X, Z)X|Z_i^\top\theta\}$ and $E\{w(X, Z)Z_{-1}|Z_i^\top\theta\}$, it cannot be used directly to make inference on $(\beta^\top, \theta^\top)^\top$. To solve this problem, a natural method is to replace those by their estimators $\hat{w}(X_i, Z_i)$, $\hat{g}(Z_i^\top\theta)$, $\hat{g}'(Z_i^\top\theta)$, $\hat{E}\{\hat{w}(X, Z)|Z_i^\top\theta\}$,

$\hat{E}\{\hat{w}(X, Z)X|Z_i^\top\theta\}$ and $\hat{E}\{\hat{w}(X, Z)Z_{-1}|Z_i^\top\theta\}$ given above. Define an estimated empirical likelihood function for $(\beta^\top, \theta^\top)^\top$ as

$$\tilde{L}(\beta, \theta) = \sup\left\{\prod_{i=1}^n (n\pi_i) : \sum_{i=1}^n \pi_i = 1, \pi_i \geq 0, \sum_{i=1}^n \pi_i \hat{\xi}_i(\beta, \theta) = 0\right\}, \quad (6)$$

where

$$\hat{\xi}_i(\beta, \theta) = \hat{w}_i\tilde{\varepsilon}_i\left(X_i^\top - \frac{\hat{E}\left(\hat{w}X^\top|Z_i^\top\theta\right)}{\hat{E}\left(\hat{w}|Z_i^\top\theta\right)}, \hat{g}'\left(Z_i^\top\theta\right)\left\{Z_{-1,i}^\top - \frac{\hat{E}\left(\hat{w}Z_{-1}^\top|Z_i^\top\theta\right)}{\hat{E}\left(\hat{w}|Z_i^\top\theta\right)}\right\}\right)^\top,$$

and $\tilde{\varepsilon}_i = Y_i - X_i^\top\beta - \hat{g}(Z_i^\top\theta)$. The estimated empirical log-likelihood ratio is

$$\tilde{l}(\beta, \theta) = -2\left\{\log\{\tilde{L}(\beta, \theta)\} - n\log(n)\right\}. \quad (7)$$

According to Tsao (2004), when the dimension $p + r$ is moderately large but fixed, the distribution of empirical likelihood ratio $\tilde{l}(\beta, \theta)$ has an atom at infinity for fixed sample size n: the probability of $\tilde{l}(\beta, \theta) = \infty$ is nonzero. Based on the paper of Tsao (2004), the dimension $p + r$ and the sample size n increase at the same rate such that $p/n \geq 0.5$, the probability of $\tilde{l}(\beta, \theta) = \infty$ converges to 1 since the probability of $(\beta^\top, \theta^\top)^\top$ being contained in the convex hull of the sample converges to 0. These reveal the effects of the dimension $p + r$ on the empirical likelihood from another perspective. In this paper, we analyze the empirical likelihood for heteroscedastic partially linear single-index models with high-dimensional data, which p, r and n increase at the mild rate in order to ensure the empirical likelihood ratio $\tilde{l}(\beta, \theta)$ having definition.

By using the Lagrange multiplier method, $\{\pi_i\}_{i=1}^n$ in (6) are

$$\pi_i = \frac{1}{n}\frac{1}{1 + \lambda^\top\hat{\xi}_i(\beta, \theta)},$$

with the restriction on λ that is

$$\frac{1}{n}\sum_{i=1}^n \frac{\hat{\xi}_i(\beta, \theta)}{1 + \lambda^\top\hat{\xi}_i(\beta, \theta)} = 0. \quad (8)$$

Therefore, the estimated empirical log-likelihood ratio function for $(\beta^\top, \theta^\top)^\top$ defined in (7) is given by

$$\tilde{l}(\beta, \theta) = -2\left\{\log\{\tilde{L}(\beta, \theta)\} - n\log(n)\right\} = 2\sum_{i=1}^n \log\left\{1 + \lambda^\top\hat{\xi}_i(\beta, \theta)\right\}. \quad (9)$$

The following theorem gives the asymptotic distribution of $\tilde{l}(\beta, \theta)$.

Theorem 2.2 *Let* $(\beta_0^\top, \theta_0^\top)^\top$ *be the true value of parameter vector* $(\beta^\top, \theta^\top)^\top$, *under Assumptions* 1–9, $\tilde{l}(\beta_0, \theta_0)$ *has an asymptotic standard normal distribution, i.e.,*

$$\left\{2(p + r - 1)\right\}^{-1/2}\left\{\tilde{l}(\beta_0, \theta_0) - (p + r - 1)\right\} \xrightarrow{L} N(0, 1).$$

Therefore, Theorem 2.2 can be used to test the hypothesis

$$H_0 : \left(\beta^\top, \theta^\top\right)^\top = \left(\beta_0^\top, \theta_0^\top\right)^\top \quad vs \quad H_1 : \left(\beta^\top, \theta^\top\right)^\top \neq \left(\beta_0^\top, \theta_0^\top\right)^\top,$$

and on the other hand, it also can be used to construct confidence regions for $(\beta^\top, \theta^\top)^\top$. Let

$$I_\alpha(\beta, \theta) = \left\{(\beta^\top, \theta^\top)^\top : \left|\tilde{l}(\beta, \theta) - (p + r - 1)\right| \leq z_{\alpha/2}\sqrt{2(p + r - 1)}\right\},$$

where $z_{\alpha/2}$ is the upper $\alpha/2-$quantile of the standard normal distribution, then by Theorem 2.2, $I_\alpha(\beta, \theta)$ gives an approximate confidence region for $(\beta^\top, \theta^\top)^\top$ with asymptotically correct coverage probability $1 - \alpha$, i.e.,

$$P((\beta^\top, \theta^\top)^\top \in I_\alpha(\beta, \theta)) = 1 - \alpha + o(1).$$

Remark 1 Theorem 2.2 gives the asymptotic distribution of $\tilde{l}(\beta, \theta)$ with diverging dimensionality, i.e., $p, r \to \infty$ as $n \to \infty$. When p and r are fixed and do not diverge with n, $\tilde{l}(\beta_0, \theta_0)$ is asymptotically a standard chi-square random variable, and it is easy to see from the proof of Theorem 2.2 in "Appendix" that

$$\tilde{l}(\beta_0, \theta_0) \xrightarrow{L} \chi^2_{p+r-1},$$

where χ^2_k denotes the chi-square distribution with $(p + r - 1)$ degrees of freedom.

In practice, it is rarely the case that confidence regions for the entire parameter vector will be sought, because as soon as dimension $p > 3$, it is hard to visualize or represent the confidence region. Usually one will only be interested in a one or two-dimensional subvector of parameter. Assume $(\beta^\top, \theta^\top)^\top = (\beta^{(1)\top}, \beta^{(2)\top}, \theta^\top)^\top$, where $\beta^{(1)}$ is k-dimensional (k is fixed and does not diverge with n) for which the confidence interval/region is to be constructed. The true parameter vector can be partitioned as $(\beta_0^\top, \theta_0^\top)^\top = (\beta_0^{(1)\top}, \beta_0^{(2)\top}, \theta_0^\top)^\top$, and X_i and Z_i can be similarly partitioned.

We can obtain, with fixed $\beta^{(1)}$, estimators for the rest of the parameters by solving (3), denoted by $\hat{\beta}^{(2)}$ and $\hat{\theta}$. Define

$$\hat{\tilde{\xi}}_i(\beta^{(1)}) = \hat{w}_i \left\{Y_i - X_i^{(1)\top}\beta^{(1)} - X_i^{(2)\top}\hat{\beta}^{(2)}\right.$$
$$\left. - \hat{g}\left(Z_i^\top\hat{\theta}\right)\right\}\left(X_i^{(1)} - \frac{\hat{E}\left(\hat{w}X^{(1)}|Z_i^\top\hat{\theta}\right)}{\hat{E}\left(\hat{w}|Z_i^\top\hat{\theta}\right)}\right)$$

The estimated empirical log-likelihood for $\beta^{(1)}$ is given by

$$\tilde{l}(\beta^{(1)}) = 2 \sum_{i=1}^{n} \log \left\{ 1 + \lambda^{(1)\top} \hat{\tilde{\xi}}_i (\beta^{(1)}) \right\}. \tag{10}$$

Theorem 2.3 *Under the same assumptions as in Theorem* 2.2,

$$\tilde{l}\left(\beta_0^{(1)}\right) \xrightarrow{L} \chi_k^2.$$

Based on Theorem 2.3, a confidence interval or region for $\beta^{(1)}$ can be easily constructed. For any given $0 < \alpha < 1$, there exists c_α such that $P(\chi_k^2 > c_\alpha) = \alpha$, then

$$I_\alpha(\beta^{(1)}) = \left\{ \beta^{(1)} \in R^k : \tilde{l}(\beta^{(1)}) \leq c_\alpha \right\}$$

is the confidence interval or region of the $\beta^{(1)}$ with asymptotically correct coverage probability $1 - \alpha$.

Remark 2 Based on Tsao (2004), in high dimensional setting, the dimension will have non-ignorable effect on coverage probabilities of empirical likelihood confidence regions. Therefore, coverage probabilities of empirical likelihood confidence regions may be less than the nominal significance level. We hope to consider how to improve coverage probabilities of empirical likelihood confidence regions under the high dimensional situations in future communications.

3 Two special cases: partially linear models and single-index models

In this section, we present the empirical likelihood inference for β or θ of two special cases in model (1). First we consider the heteroscedastic partially linear model with a diverging number parameters, meaning that single-index component in model (1) does not contain parameters, and the model can be written as

$$Y_i = X_i^\top \beta + g(Z_i) + \varepsilon_i, \quad (i = 1, \ldots, n). \tag{11}$$

The weighted estimating equations for model (11) become

$$n^{-1/2} \sum_{i=1}^{n} \hat{w}(X_i, Z_i) \left(Y_i - X_i^\top \beta - \hat{g}(Z_i) \right) \left[X_i - \frac{\hat{E}\{\hat{w}(X, Z)X|Z_i\}}{\hat{E}\{\hat{w}(X, Z)|Z_i\}} \right] = 0. \tag{12}$$

The estimator $\hat{\beta}$ can be obtained by solving semiparametric efficient score equations (12). Rewrite the auxiliary random vector

$$\hat{\xi}_i(\beta) = \hat{w}_i \left\{ Y_i - X_i^\top \beta - \hat{g}(Z_i) \right\} \left\{ X_i - \frac{\hat{E}(\hat{w}X|Z_i)}{\hat{E}(\hat{w}|Z_i)} \right\}.$$

Therefore, the estimated empirical log-likelihood ratio function for β is

$$\tilde{l}_1(\beta) = 2 \sum_{i=1}^{n} \log \left\{ 1 + \lambda^\top \hat{\xi}_i(\beta) \right\}. \tag{13}$$

Assume $\beta = (\beta^{(1)\top}, \beta^{(2)\top})^\top$, where $\beta^{(1)}$ is a k-dimensional (k is fixed and does not diverge with n) component of β. The true parameter vector can be partitioned as $\beta_0 = (\beta_0^{(1)\top}, \beta_0^{(2)\top})^\top$, and X_i can be similarly partitioned. We can obtain, with fixed $\beta^{(1)}$, estimators for the rest of the parameters by solving (12), denoted by $\hat{\beta}^{(2)}$. Define

$$\tilde{\hat{\xi}}_i(\beta^{(1)}) = \hat{w}_i \left\{ Y_i - X_i^{(1)\top} \beta^{(1)} - X_i^{(2)\top} \hat{\beta}^{(2)} - \hat{g}(Z_i) \right\} \left\{ X_i^{(1)} - \frac{\hat{E}(\hat{w}X^{(1)}|Z_i)}{\hat{E}(\hat{w}|Z_i)} \right\}.$$

The estimated empirical log-likelihood for $\beta^{(1)}$ is given by

$$\tilde{\hat{l}}(\beta^{(1)}) = 2 \sum_{i=1}^{n} \log \left\{ 1 + \lambda^{(1)\top} \tilde{\hat{\xi}}_i(\beta^{(1)}) \right\}. \tag{14}$$

In high-dimensional setting, there are some theoretical properties for model (11) as follows.

Theorem 3.1 *Let β_0 be the true value of parameter vector β, under Assumptions 1–9, then*

$$(1) \quad \sqrt{n} A_n V_1^{-1/2} (\hat{\beta} - \beta_0) \xrightarrow{L} N(0, G),$$
$$(2) \quad (2p)^{-1/2} \left\{ \tilde{l}_1(\beta_0) - p \right\} \xrightarrow{L} N(0, 1),$$
$$(3) \quad \tilde{\hat{l}}(\beta_0^{(1)}) \xrightarrow{L} \chi_k^2,$$

where $A_n \in R^{q \times p}$ is a $q \times p$ matrix such that $A_n A_n^\top \to G$ and G is a $q \times q$ matrix with fixed q, and

$$V_1 = \left[E \left\{ wXX^\top - \frac{E(wX|Z)E(wX|Z)^\top}{E(w|Z)} \right\} \right]^{-1}.$$

Theorem 3.1(1) extends the fixed-dimensional results in Ma et al. (2006) to cases with diverging dimensionality. Theorem 3.1(2) can be used to construct empirical likelihood confidence regions for β. A $(1 - \alpha)100\%$ confidence region for β is given by

$$I_\alpha(\beta) = \left\{ \beta : \left| \tilde{l}_1(\beta) - p \right| \leq z_{\alpha/2} \sqrt{2p} \right\}.$$

When p is fixed and does not diverge with n, $\tilde{l}_1(\beta_0)$ has the asymptotic chi-square distribution, and this corresponds to the empirical likelihood method considered by Lu (2009) for heteroscedastic partially linear models. Theorem 3.1(3) can be used to construct the confidence interval or region of subvector of parameter β. Based on Theorem 3.1(3), for $0 < \alpha < 1$, the confidence interval or region for $\beta^{(1)}$ can be given by

$$\tilde{I}_\alpha(\beta^{(1)}) = \left\{ \beta^{(1)} | \tilde{\tilde{l}}(\beta^{(1)}) \le \chi_k^2(\alpha) \right\}.$$

For high-dimensional data, we now consider the heteroscedastic pure single-index model

$$Y_i = g\left(Z_i^\top \theta\right) + \varepsilon_i, \quad (i = 1, \ldots, n). \tag{15}$$

It means that there is no linear component in model (1). The weighted estimating equations can be written as

$$n^{-1/2} \sum_{i=1}^n \hat{w} \left\{ Y_i - \hat{g}\left(Z_i^\top \theta\right) \right\} \hat{g}'\left(Z_i^\top \tilde{\theta}\right) \left[Z_{-1,i} - \frac{\hat{E}\left\{ \hat{w} Z_{-1} | Z_i^\top \tilde{\theta} \right\}}{\hat{E}\left\{ \hat{w} | Z_i^\top \tilde{\theta} \right\}} \right] = 0, \tag{16}$$

where $\tilde{\theta}$ is the solution of the following equations

$$n^{-1/2} \sum_{i=1}^n \hat{w} \left\{ Y_i - f\left(Z_i^\top \theta\right) \right\} \left\{ a(Z_i) - \tilde{E}\left\{ a(Z) | Z_i^\top \theta \right\} \right\} = 0.$$

The estimator $\hat{\theta}$ can be got by solving semiparametric efficient score equations (16). Similarly, redefine the auxiliary random vector

$$\hat{\xi}_i(\theta) = \hat{w}_i \left\{ Y_i - \hat{g}\left(Z_i^\top \theta\right) \right\} \hat{g}'\left(Z_i^\top \theta\right) \left\{ Z_{-1,i} - \frac{\hat{E}\left(\hat{w} Z_{-1} | Z_i^\top \theta\right)}{\hat{E}\left(\hat{w} | Z_i^\top \theta\right)} \right\}.$$

and the estimated empirical log-likelihood ratio is

$$\tilde{l}_2(\theta) = 2 \sum_{i=1}^n \log \left\{ 1 + \lambda^\top \hat{\xi}_i(\theta) \right\}.$$

Theorem 3.2 *Under Assumptions 1–9, if θ_0 is the true parameter value, then*

(1) $\sqrt{n} A_n V_2^{-1/2}(\hat{\theta}_{-1} - \theta_{0,-1}) \xrightarrow{L} N(0, G),$

(2) $\{2(r-1)\}^{-1/2} \left\{ \tilde{l}_2(\theta_0) - (r-1) \right\} \xrightarrow{L} N(0, 1),$

where $A_n \in R^{q \times (r-1)}$ is a $q \times (r-1)$ matrix such that $A_n A_n^{\top} \to G$ and G is a $q \times q$ matrix with fixed q, and

$$V_2 = E\left[g'(Z^{\top}\theta)^2 \left\{ w Z_{-1} Z_{-1}^{\top} - \frac{E(w Z_{-1}|Z^{\top}\theta) E(w Z_{-1}^{\top}|Z^{\top}\theta)}{E(w|Z^{\top}\theta)} \right\} \right].$$

The theoretical properties of $\hat{\theta}$ is given by Theorem 3.2(1). When r is fixed, $\tilde{l}_2(\theta_0) \xrightarrow{L} \chi^2_{r-1}$. According to Theorem 3.2(2), a $(1-\alpha)100\%$ empirical likelihood confidence region for θ is

$$I_{\alpha}(\theta) = \left\{ \theta : \left| \tilde{l}_2(\theta) - (r-1) \right| \leq z_{\alpha/2}\sqrt{2(r-1)} \right\}.$$

4 Simulation studies

Firstly, we describe how to approach the optimization problems posed by the empirical likelihood method. Due to the nonconvexity, computing the empirical likelihood ratio is nontrivial. We can use the following algorithm to obtain the estimated empirical likelihood ratio defined by (7).

Step 1 Obtain $\hat{g}(Z_i^{\top}\theta), \hat{g}'(Z_i^{\top}\theta), \hat{w}(X_i, Z_i), \hat{E}\{\hat{w}(X, Z)|Z_i^{\top}\theta\}, \hat{E}\{\hat{w}(X, Z)X|Z_i^{\top}\theta\}$ and $\hat{E}\{\hat{w}(X, Z)Z_{-1}|Z_i^{\top}\theta\}$ described above by using the fixed values of $(\beta^{\top}, \theta^{\top})^{\top}$.
Step 2 Obtain the auxiliary random vector $\hat{\hat{\xi}}_i(\beta, \theta)$.
Step 3 Obtain $\hat{\lambda}$ by using the Newtons method to minimize (9) over λ.
Step 4 Obtain $\tilde{l}(\beta, \theta)$ defined by (9) based on $\hat{\hat{\xi}}_i(\beta, \theta)$ and $\hat{\lambda}$.

Next, we present some Monte Carlo experiments to compare the finite sample performance of EL with normal approximation that is based on the doubly robust and efficient method (DRE) in Ma and Zhu (2013). Throughout these simulations, we use the Epanechnikov kernel $K(t) = 3/4(1-t^2)_+$ in all the nonparametric regression procedures and use the cross validation method to choose the optimal bandwidth h_{opt} satisfying Assumption 6. We consider the heteroscedastic partially single-index model (1). In our simulations, $(X_1, \ldots, X_p, Z_1, \ldots, Z_r)^{\top}$ are generated from a multivariate normal distribution with mean 0 and covariance matrix $(\sigma_{ij})_{(p+r)\times(p+r)}$ where $\sigma_{ij} = 0.5^{|i-j|}$, and Y are generated from a normal distribution with mean $X^{\top}\beta + \exp(Z^{\top}\theta)$, and variance function $|Z^{\top}\theta|$. Furthermore, $\beta = (1, 0.5, -1, 1, 0.5, -1, \ldots)^{\top}$ and $\theta = (1, 0.25, -0.25, 0.25, 0, \ldots, 0)^{\top}$. We use the dimensions $p = 10, 20, 30$ and the corresponding dimension $r = 10$.

In the first simulation, we consider the confidence interval for β_1 constructed by using empirical likelihood and results for other parameters are similar and not presented here. For comparison, we also construct the confidence interval for β_1 by using the doubly robust and efficient method in Ma and Zhu (2013). In each case we repeat the simulation 1000 times, and the nominal confidence level $1-\alpha$ is taken to 0.95. The results are presented in Table 1. In addition, the column of time in Table 1 denote the computing cost of the simulation. For example, the value of 2.6 s in Table 1 means the

Table 1 The coverage probability (CP) and average length (AL) for β_1

(p,r)	n	EL			DRE		
		CP	AL	Time (s)	CP	AL	Time (s)
(10,10)	200	0.939	0.351	2.6	0.927	0.367	4.2
	400	0.947	0.318	7.3	0.941	0.342	9.7
	600	0.953	0.282	14.5	0.948	0.324	15.4
(20,10)	400	0.932	0.423	15.4	0.916	0.451	17.1
	600	0.941	0.376	19.7	0.932	0.409	20.6
	800	0.946	0.354	22.1	0.937	0.385	24.3
(30,10)	600	0.915	0.447	23.9	0.903	0.483	27.5
	800	0.926	0.419	27.8	0.912	0.458	31.8
	900	0.940	0.385	31.3	0.931	0.426	35.2

We compare confidence intervals based on EL method with intervals based on the doubly robust and efficient method (DRE)

average computing cost of each simulation, and the total computing cost of repeating the simulation 1000 times is $2.6 \times 1000 = 2600(s)$. This is also the case in Tables 2 and 3.

From Table 1, we have the following results:

(1) We can see that the coverage probabilities (CP) for β_1 based on the EL and the doubly robust and efficient method (DRE) method increase as the sample size n increases, and the coverage probabilities appear to be close to the nominal levels especially with moderate sample size. Certainly, the coverage probabilities based on the empirical likelihood can not be up to the nominal levels when the dimension p grows with the sample size n. According to Tsao (2004), the dimension will have non-ignorable effect on coverage probabilities of empirical likelihood confidence regions. Some methods may be used to improve coverage probabilities of empirical likelihood confidence regions/intervals under the high dimensional situation, such as Bartlett adjustment. How to improve it under the high dimensional situations will be investigated in our future communications.

(2) The confidence intervals based on the empirical likelihood method consistently have better coverage probabilities than the intervals based on DRE, and the average length (AL) of the intervals based on the empirical likelihood is slightly shorter than that of the intervals based on DRE.

(3) The average computing cost of the empirical likelihood method is less than the DRE method from Table 1. However, It must have a lot to do with computer configuration, and high performance computers can improve the computing cost. The outcomes in Table 1 are based on a normal computer.

In the second simulation, we further consider confidence regions for (β_1, β_2) constructed by EL and DRE where we show the coverage probability of the constructed regions. The results are presented in Table 2. Finally, we consider confidence regions for $(\beta^\top, \theta^\top)^\top$ constructed by using EL where we show the coverage of the constructed regions. For comparison, we also construct confidence regions for $(\beta^\top, \theta^\top)^\top$ by using

Table 2 Comparison of coverage probability for (β_1, β_2) between EL method and DRE method

p	n	$1-\alpha = 0.90$				$1-\alpha = 0.95$			
		EL	Time (s)	DRE	Time (s)	EL	Time (s)	DRE	Time (s)
(10,10)	200	0.871	3.5	0.857	4.9	0.919	3.7	0.908	5.0
	400	0.887	8.2	0.874	10.3	0.932	8.1	0.916	10.1
	600	0.901	15.4	0.889	16.1	0.947	15.3	0.929	16.2
(20,10)	400	0.859	16.3	0.847	18.5	0.908	16.6	0.897	18.8
	600	0.881	20.6	0.872	21.4	0.923	20.5	0.912	21.9
	800	0.893	24.1	0.881	25.8	0.935	24.1	0.924	25.4
(30,10)	600	0.862	24.9	0.856	28.2	0.903	25.1	0.891	28.6
	800	0.875	28.7	0.867	32.6	0.920	28.8	0.903	32.3
	900	0.884	32.4	0.873	35.7	0.931	32.2	0.915	35.6

Table 3 Comparison of coverage probability for $(\beta^\top, \theta^\top)^\top$ between EL method and DRE method

p	n	$1-\alpha = 0.90$				$1-\alpha = 0.95$			
		EL	Time (s)	DRE	Time (s)	EL	Time (s)	DRE	Time (s)
(10,10)	800	0.866	18.7	0.854	20.3	0.921	18.5	0.914	20.1
	1000	0.873	24.2	0.861	25.1	0.927	24.3	0.919	25.4
	1200	0.885	29.5	0.876	31.4	0.934	29.7	0.925	31.5
(20,10)	1000	0.852	32.4	0.843	34.6	0.905	32.3	0.892	34.4
	1200	0.867	37.9	0.859	38.2	0.913	38.1	0.901	38.3
	1400	0.882	41.6	0.871	43.5	0.928	41.8	0.917	43.8
(30,10)	1200	0.843	43.8	0.837	45.7	0.882	44.1	0.879	45.6
	1400	0.861	48.1	0.852	49.4	0.910	48.2	0.892	50.2
	1600	0.875	54.3	0.864	55.8	0.921	54.6	0.913	55.7

DRE. In each case we repeat the simulation 1000 times. The results are presented in Table 3. Observing Tables 2 and 3, the message is similar as before and we can find that EL consistently achieves slightly higher coverage probabilities than DRE.

5 Real data application

We further illustrate our proposed method by applying the heteroscedastic partially linear single index model to data from AIDS Clinical Trials Group Protocol 175 (ACTG175) which has been analyzed by Hammer et al. (1996), Davidian et al. (2005) and Lai and Wang (2014). CD4 is a co-receptor that assists the T cell receptor (TCR) with an antigen-presenting cell, and many HIV clinical trials focus on comparing treatment effects on CD4 count after a specified period. ACTG175 data concludes four antiretroviral regimens [zidovudine (ZDV) monotherapy, ZDV+didanosine (ddI), ZDV+zalcitabine, ddI] and randomizes 2139 patients to four antiretroviral regimens

Table 4 95% confidence intervals based on EL and DRE, and estimators for $(\beta^\top, \theta^\top)^\top$

	$(\hat{\beta}^\top, \hat{\theta}^\top)^\top$	Confidence intervals	
		EL	DRE
hemo	0.0044	$[-0.0507, 0.0932]$	$[-0.3218, 0.3305]$
homo	0.0054	$[-0.0526, 0.0103]$	$[-0.1784, 0.1892]$
drugs	0.0010	$[-0.0543, 0.0597]$	$[-0.2795, 0.2814]$
karnof	0.0001	$[-0.0482, 0.0504]$	$[-0.0136, 0.0138]$
race	-0.0019	$[-0.0685, 0.0306]$	$[-0.1890, 0.1851]$
gender	-0.0039	$[-0.0881, 0.0109]$	$[-0.2274, 0.2196]$
str2	-0.0089	$[-0.0138, -0.0405]$	$[-0.1697, 0.1520]$
symptom	-0.0088	$[-0.0137, -0.0394]$	$[-0.2290, 0.2115]$
age	0.5687	$[0.2687, 0.9687]$	$[-0.6460, 0.7597]$
wtkg	-0.5925	$[-0.9925, -0.2925]$	$[-1.1214, 1.0029]$
cd80	0.0447	$[-0.1553, 0.3947]$	$[-0.5243, 0.5332]$
cd820	-0.0989	$[-0.1489, 0.1511]$	$[-0.5509, 0.5311]$
cd40	-0.4832	$[-0.8332, -0.1832]$	$[-0.6623, 0.5657]$
cd420	0.3701	$[0.0701, 0.5701]$	$[-0.5681, 0.6421]$

in equal proportions. The findings of ACTG 175 indicate that antiretroviral therapy could reduce the risk in people with intermediate stage HIV disease and no symptoms.

In this paper, we will build a heteroscedastic partially linear single-index model for a subject's response with ZDV monotherapy. There are 532 patients with CD4 cell counts between 200 and 500 per cubic millimeter. The response Y is the CD4 cell count at 96 ± 5 weeks (CD496). Because some CD496 are missing in the ACTG175 data, we consider a subset of the ACTG175 data representing 321 samples from subjects with CD496. The predictor variables are age, wtkg, hemo, homo, drugs, karnof, race, gender, str2, symptom, CD40, CD420, CD80 and CD820. We divide the discrete variables (hemo, homo, drugs, karnof , race, gender, str2, symptom) into the linear part and the standardized continuous variables (age,wtkg, CD80, CD820, CD40, CD420) into the single-index part. Furthermore, we standardize the response variable CD496, and build the following heteroscedastic partially linear single-index model

$$Y = X^\top \beta + g\left(Z^\top \theta\right) + \varepsilon,$$

where $X^\top$ = (hemo, homo, drugs, karnof , race, gender, str2, symptom), $Z^\top$ = (age,wtkg, CD80, CD820, CD40,CD420), Y = CD496 and $g(\cdot)$ is an unknown function. According to the plot of estimated errors in Lai and Wang (2014), we can find that ε is heteroscedastic. Using the method proposed in Sect. 2, we can obtain the estimators and confidence intervals for the parameters, and the results are summarized in Table 4. It is seen from Table 4 that the confidence intervals based on the empirical likelihood method are somewhat different from those of the doubly robust and efficient method. The doubly robust and efficient method gives larger intervals

and imposes symmetry on the confidence intervals. Lai and Wang (2014) analyzed the ACTG175 data by using the variable selection method. They found that the effect of CD40 and CD420 were significant, and the corresponding variables were included in the final fitted model. From Table 4, we find that the confidence intervals based on the empirical likelihood method of CD40 and CD420 do not cover the zero point. However, the confidence intervals based on the doubly robust and efficient method of CD40 and CD420 cover the zero point, which mean that the effect of CD40 and CD420 are insignificant. In addition, according to the empirical likelihood method, the coefficient of antiretroviral history (str2), symptomatic indicator (symptom) and weight (wtkg) are significantly negative and the coefficient of age is significantly positive. These factors play an important role for four different antiretroviral regimens. We think that those findings corroborate the results in Lai and Wang (2014) very well.

Appendix

To prove the main theorems, we need to give the following set of conditions.

Assumption 1 Let $\text{Var}(X_i) = \Sigma_{xi}$ and $\text{Var}(Z_i) = \Sigma_{zi}$, the eigenvalues of Σ_{xi} and Σ_{zi} satisfy $C_1 \leq \gamma_1(\Sigma_{xi}) \leq \cdots \leq \gamma_p(\Sigma_{xi}) \leq C_2$ and $C_1 \leq \gamma_1(\Sigma_{zi}) \leq \cdots \leq \gamma_r(\Sigma_{zi}) \leq C_2$ for some constants $0 < C_1 < C_2$, for $i = 1 \cdots n$. There is a constant $\delta > 0$ such that $E(\varepsilon^{4+\delta}|X, Z) < \infty$.

Assumption 2 There are $v(\cdot)$, $\eta = \eta(X, Z)$, such that $E(\varepsilon^2|X, Z) = v(\eta)$, $0 < C_1 < v(\cdot) < C_2 < \infty$ for some constants $0 < C_1 < C_2$, and The eigenvalues of $\text{Var}(X_i|\eta(X_i, Z_i))$ are bounded away from zero and infinity.

Assumption 3 There exists $v_1(X, Z)$ such that

$$\left| \frac{\partial^2 E(X|Z^\top \theta)}{\partial \theta_i \partial \theta_j} \right|, \left| \frac{\partial^2 E(Z|Z^\top \theta)}{\partial \theta_i \partial \theta_j} \right|, \left| \frac{\partial^2 E(w|Z^\top \theta)}{\partial \theta_i \partial \theta_j} \right|, \left| \frac{\partial^2 E(wZ|Z^\top \theta)}{\partial \theta_i \partial \theta_j} \right|,$$

$$\left| \frac{\partial^2 E(wX|Z^\top \theta)}{\partial \theta_i \partial \theta_j} \right| < v_1(X, Z), Ev_1^2 < \infty, (i, j = 2, \ldots, r).$$

Further there exists $v_2(X, Z)$ such that

$$\left| \frac{\partial^3 \eta(X, Z)}{\partial \gamma_i \partial \gamma_j \partial \gamma_l} \right| < v_2(X, Z), Ev_2^2 < \infty, (i, j, l = 1, \ldots, p + r),$$

where $(X^\top, Z^\top)^\top = (\gamma_1, \ldots, \gamma_{p+r})^\top$. Further there exists $v_3(X, Z)$ such that

$$\left| \frac{\partial^4 g(Z^\top \theta)}{\partial \theta_i \partial \theta_j \partial \theta_k \partial \theta_l} \right|, \left| \frac{\partial^4 v(\eta)}{\partial \eta_{i_1} \partial \eta_{j_1} \partial \eta_{k_1} \partial \eta_{l_1}} \right| < v_3(X, Z), E v_3^2 < \infty,$$

where the dimension of η is p_1, and $i, j, k, l = 2, \ldots, r, i_1, j_1, k_1, l_1 = 1, \ldots, p_1$.

Assumption 4 Assume that the random variable η and $Z^\top \theta$ have densities $f_\eta(\eta)$ and $f_{Z^\top \theta}(Z^\top \theta)$, satisfying $0 < \inf f_\eta(\eta) \leq \sup f_\eta(\eta) < \infty$ and $0 < \inf f_{Z^\top \theta}(Z^\top \theta) \leq \sup f_{Z^\top \theta}(Z^\top \theta) < \infty$. Further there exists $v_4(X, Z)$ such that

$$\left| \frac{\partial^2 f_{Z^\top \theta}(Z^\top \theta))}{\partial \theta_i \partial \theta_j} \right|, \left| \frac{\partial^2 f_\eta(\eta)}{\partial \eta_k \partial \eta_l} \right|$$
$$< v_4(X, Z), E v_4^2 < \infty, (i, j = 2, \ldots, p; \; k, l = 1, \ldots, p_1).$$

Assumption 5 The kernel function $K_h(\cdot)$ is symmetric and its derivative is continuous with compact support contained in $[-1, 1]$.

Assumption 6 The bandwidths h_i satisfy $\log^2(n)/(nh_i) \to 0$ for $i = 1, 2, 3$. In addition, $nh_1^4 \to \infty$, $nh_1^8 \to 0$, $h_1^4 \log^2(n)/h_i \to 0$ and $\log^4(n)/(nh_1 h_i) \to 0$ for $i = 1, 2, 3$, $h_2 = O(n^{-1/5})$ and $h_3 = O(n^{-1/5})$.

Assumption 7 $p, r \to \infty$, $pn^{-1/5} \to 0$, $rn^{-1/5} \to 0$, as $n \to \infty$.

Assumption 8 $E\|X\|^4 < \infty$, $E\|Z\|^4 < \infty$, $E\|\varepsilon X\|^4 < \infty$, $E\|\varepsilon Z\|^4 < \infty$ and $E|\varepsilon|^4 < \infty$.

Assumption 9 Let

$$\xi_n(\beta, \theta) = w\varepsilon \left[X^\top - \frac{E(wX^\top | Z^\top \theta)}{E(w | Z^\top \theta)}, g'(Z^\top \theta) \left\{ Z^\top - \frac{E(wZ^\top | Z^\top \theta)}{E(w | Z^\top \theta)} \right\} \right]^\top,$$

and $\xi_{nl}(\beta, \theta)$ be the l-th component of $\xi_n(\beta, \theta)$, $l = 1, \ldots, p, p + 2, \ldots, p + r$. As $n \to \infty$, there is a positive constant C such that, $E(\|\xi_n(\beta, \theta)/\sqrt{p}\|^4) < C$, $E(\|XX^\top\|^4) < C$, $E(\|XZ^\top\|^4) < \infty$ and $E(\|ZX^\top\|^4) < C$.

Assumptions 1–6 ensure the function $g(Z_i^\top \theta)$, $g'(Z_i^\top \theta)$, $w(X_i, Z_i)$, $E\{\hat{w}(X, Z) | Z_i^\top \theta\}$, $E\{\hat{w}(X, Z)X | Z_i^\top \theta\}$ and $E\{\hat{w}(X, Z)Z_{-1} | Z_i^\top \theta\}$ are estimated with retained precision and the nonparametric estimation does not affect the asymptotic result of the estimated empirical likelihood ratio, i.e., the estimated empirical likelihood ratio $\tilde{L}(\beta, \theta)$ has the same asymptotic distribution as the ordinary empirical likelihood ratio $L(\beta, \theta)$. Furthermore, Assumptions 1–6 ensure the existence of the estimator $(\hat{\beta}^\top, \hat{\theta}^\top)^\top$ for parameters $(\beta^\top, \theta^\top)^\top$. Assumption 7 is a technical condition, and Assumption 8 ensures that there exists an asymptotic variance for the estimator of the growing parameters $(\beta^\top, \theta^\top)^\top$. Assumption 9 controls the tail probability behavior of the estimating equation. Because establishing the asymptotic theoretical results for

empirical likelihood approach under the situation with diverging dimensionality on covariates is very challenging, these conditions are not the weakest possible and the bounds in the stochastic analysis are conservative. This is also the case in Ma and Zhu (2013), these strong conditions facilitate technical derivations.

Let $\tilde{l}(\lambda, \beta, \theta) = n^{-1} \sum_{i=1}^{n} \log \left\{1 + \lambda^\top \hat{\xi}_i(\beta, \theta)\right\}$, $\bar{\hat{\xi}}(\beta, \theta) = n^{-1} \sum_{i=1}^{n} \hat{\xi}_i(\beta, \theta)$, $a_n = O_p\{(p/n)^{1/2}\}$ and C will denote a generic positive constant that may be different in different uses throughout the "Appendix". In addition, we use the Frobenius norm of a matrix A, defined as $\|A\| = \{\mathrm{tr}(A^\top A)\}^{\frac{1}{2}}$, where $\mathrm{tr}(A)$ denotes the trace of matrix A.

Proof of Theorem 2.1

Proof We first expand

$$
\begin{aligned}
0 &= \frac{1}{\sqrt{n}} A_2 \sum_{i=1}^{n} \left\{Y_i - X_i^\top \hat{\beta} + \hat{g}'\left(Z_i^\top \hat{\theta}\right)\right\} \hat{w}_i \hat{g}'\left(Z_i^\top \hat{\theta}\right) \left\{Z_i - \frac{\hat{E}\left(\hat{w} Z | Z_i^\top \hat{\theta}\right)}{\hat{E}\left(\hat{w} | Z_i^\top \hat{\theta}\right)}\right\} \\
&= \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \hat{w}_i \hat{g}'\left(Z_i^\top \hat{\theta}\right) A_2 \left\{Z_i - \frac{E\left(w Z | Z_i^\top \theta_0\right)}{E\left(w | Z_i^\top \theta_0\right)}\right\} X_i^\top \left(\beta_0 - \hat{\beta}\right) \\
&\quad + \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \hat{w}_i \hat{g}'\left(Z_i^\top \hat{\theta}\right) A_2 \left\{\frac{E\left(w Z | Z_i^\top \theta_0\right)}{E\left(w | Z_i^\top \theta_0\right)} - \frac{\hat{E}\left(\hat{w} Z | Z_i^\top \hat{\theta}\right)}{\hat{E}\left(\hat{w} | Z_i^\top \hat{\theta}\right)}\right\} X_i^\top \left(\beta_0 - \hat{\beta}\right) \\
&\quad + \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left\{g\left(Z_i^\top \theta_0\right) - \hat{g}\left(Z_i^\top \theta_0\right)\right\} \hat{w}_i \hat{g}'\left(Z_i^\top \hat{\theta}\right) A_2 \left\{Z_i - \frac{E\left(w Z | Z_i^\top \theta_0\right)}{E\left(w | Z_i^\top \theta_0\right)}\right\} \\
&\quad + \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left\{g\left(Z_i^\top \theta_0\right) - \hat{g}\left(Z_i^\top \theta_0\right)\right\} \hat{w}_i \hat{g}'\left(Z_i^\top \hat{\theta}\right) A_2 \left\{\frac{E\left(w Z | Z_i^\top \theta_0\right)}{E\left(w | Z_i^\top \theta_0\right)}\right. \\
&\quad \left. - \frac{\hat{E}\left(\hat{w} Z | Z_i^\top \hat{\theta}\right)}{\hat{E}\left(\hat{w} | Z_i^\top \hat{\theta}\right)}\right\} \\
&\quad + \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left\{g\left(Z_i^\top \theta_0\right) - \hat{g}\left(Z_i^\top \hat{\theta}\right)\right\} \hat{w}_i \hat{g}'\left(Z_i^\top \hat{\theta}\right) A_2 \left\{Z_i - \frac{\hat{E}\left(\hat{w} Z | Z_i^\top \hat{\theta}\right)}{\hat{E}\left(\hat{w} | Z_i^\top \hat{\theta}\right)}\right\} \\
&\quad + \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \varepsilon_i \hat{w}_i \hat{g}'\left(Z_i^\top \hat{\theta}\right) A_2 \left\{Z_i - \frac{\hat{E}\left(\hat{w} Z | Z_i^\top \hat{\theta}\right)}{\hat{E}\left(\hat{w} | Z_i^\top \hat{\theta}\right)}\right\}.
\end{aligned}
\tag{17}
$$

Similar to the proof of Proposition 2 in Ma and Zhu (2013), we can obtain from the second equation in (3) that

$$
A_2 E \left[w g'\left(Z^\top \theta_0\right) \left\{Z - \frac{E\left(w Z | Z^\top \theta_0\right)}{E\left(w | Z^\top \theta_0\right)}\right\} X^\top\right] \sqrt{n}\left(\hat{\beta} - \beta_0\right)
$$

$$+A_2 E\left[w\{g'\left(Z^\top\theta_0\right)\}^2\left\{Z-\frac{E\left(wZ|Z^\top\theta_0\right)}{E\left(w|Z^\top\theta_0\right)}\right\}Z^\top\right]\sqrt{n}\left(\hat{\theta}-\theta_0\right)$$

$$=\frac{1}{\sqrt{n}}\sum_{i=1}^n\varepsilon_i w_i g'\left(Z_i^\top\theta_0\right)A_2\left\{Z_i-\frac{E\left(wZ|Z_i^\top\theta_0\right)}{E\left(w|Z_i^\top\theta_0\right)}\right\}+o_p\left(1\right). \qquad (18)$$

Similarly, from the first equation in (3), we have that

$$A_1 E\left[w\left\{X-\frac{E\left(wX|Z^\top\theta_0\right)}{E\left(w|Z^\top\theta_0\right)}\right\}X^\top\right]\sqrt{n}\left(\hat{\beta}-\beta_0\right)$$

$$+A_1 E\left[wg'\left(Z^\top\theta_0\right)\left\{X-\frac{E\left(wX|Z^\top\theta_0\right)}{E\left(w|Z^\top\theta_0\right)}\right\}Z^\top\right]\sqrt{n}\left(\hat{\theta}-\theta_0\right)$$

$$=\frac{1}{\sqrt{n}}\sum_{i=1}^n\varepsilon_i w_i A_1\left\{X_i-\frac{E\left(wX|Z_i^\top\theta_0\right)}{E\left(w|Z_i^\top\theta_0\right)}\right\}+o_p\left(1\right). \qquad (19)$$

Combining (17) and (18) implies that

$$AV^{1/2}\begin{pmatrix}\hat{\beta}-\beta_0\\\hat{\theta}-\theta_0\end{pmatrix}=AV^{-1/2}\left(\begin{array}{c}\frac{1}{\sqrt{n}}\sum_{i=1}^n\varepsilon_i w_i\left\{X_i-\frac{E(wX|Z_i^\top\theta_0)}{E(w|Z_i^\top\theta_0)}\right\}\\\frac{1}{\sqrt{n}}\sum_{i=1}^n\varepsilon_i w_i g'\left(Z_i^\top\theta_0\right)\left\{Z_i-\frac{E(wZ|Z_i^\top\theta_0)}{E(w|Z_i^\top\theta_0)}\right\}\end{array}\right)+o_p(1).$$

Applying the Lindeberg–Feller central limit theorem, we can establish

$$\sqrt{n}AV^{1/2}\left\{\left(\hat{\beta}^\top,\hat{\theta}^\top\right)^\top-\left(\beta_0^\top,\theta_0^\top\right)^\top\right\}\to N(0,G)$$

in distribution, and the proof of Theorem 2.1 is completed. □

Next, we present the following lemmas before proving Theorem 2.2.

Lemma 5.1 *Under Assumptions of Theorem* 2.2, $\max_{1\leq i\leq n}\|\hat{\xi}_i(\beta,\theta)\|=o_p(n^{1/4}\sqrt{p})$ *and* $\max_{1\leq i\leq n}|\lambda^\top\hat{\xi}_i(\beta,\theta)|=o_p(1)$ *for all* $\lambda=O_p(a_n)$.

Proof From Assumptions 8 and 9, for any $\epsilon>0$,

$$P\left\{\max_{1\leq i\leq n}\|\xi_i\left(\beta,\theta\right)\|\leq n^{1/4}\sqrt{p}\epsilon\right\}\leq\sum_{i=1}^n P\left\{\|\xi_i\left(\beta,\theta\right)\|\leq n^{1/4}\sqrt{p}\epsilon\right\}$$

$$\leq\frac{1}{np^2\epsilon^4}\sum_{i=1}^n E\|\xi_i\left(\beta,\theta\right)\|^4$$

$$=\frac{1}{\epsilon^k}E\|\xi_1\left(\beta,\theta\right)/\sqrt{p}\|^4. \qquad (20)$$

By Cauchy–Schwarz inequality, $\|\xi_1(\beta,\theta)/\sqrt{p}\|^4 \le 1/p \sum_{l=1}^{p+r} |\xi_{1l}(\beta,\theta)|^4$, where $\xi_{1l}(\beta,\theta)$ are the lth component of $\xi_1(\beta,\theta)$. According to (20), we have

$$\max_{1 \le i \le n} \|\xi_i(\beta,\theta)\| = o_p\left(n^{1/4}\sqrt{p}\right).$$

Similar to the proof of (17) and (18) above, it is easy to check that

$$\|\hat{\bar{\xi}}_i(\beta,\theta)\| = \|\xi_i(\beta,\theta)\| + O_p(p).$$

Then, by Assumption 7, we have

$$\|\hat{\bar{\xi}}_i(\beta,\theta)\| = o_p\left(n^{1/4}\sqrt{p}\right) + O_p(p) = o_p\left(n^{1/4}\sqrt{p}\right),$$

and for all $\lambda = O_p(a_n)$,

$$\max_{1 \le i \le n} |\lambda^\top \hat{\xi}_i(\beta,\theta)| = o_p(1).$$

The proof of Lemma 5.1 is completed.   □

Lemma 5.2 *Under Assumptions of Theorem* 2.2, $\|S_n - V\| = O_p(p/\sqrt{n})$, *where* $S_n = 1/n \sum_{i=1}^n \hat{\xi}_i(\beta,\theta)\hat{\xi}_i(\beta,\theta)^\top$.

Proof Similar to the proof of Lemma 5.4 in Chen et al. (2009), we have $tr\{(S_n - V)^{\otimes 2}\} = O_p(p^2/n)$. Therefore, by the definition of Frobenius norm, $\|S_n - V\| = \{tr[(S_n - V)^\top(S_n - V)]\}^{1/2} = O_p(p/\sqrt{n})$.   □

Lemma 5.3 *Under Assumptions of Theorem* 2.2, $\|\lambda\| = O_p(a_n)$, *where λ is the root of* (8).

Proof According to (8), $\lambda \in \mathbb{R}^{p+r}$ satisfies

$$0 = \frac{1}{n} \sum_{i=1}^n \frac{\hat{\bar{\xi}}_i(\beta,\theta)}{1 + \lambda^\top \hat{\xi}_i(\beta,\theta)} =: \psi(\lambda).$$

Let $\lambda = \rho\alpha$, where $\rho \ge 0, \alpha \in \mathbb{R}^{p+r}$ and $\|\alpha\| = 1$. Substituting $1/(1+\lambda^\top \hat{\xi}_i(\beta,\theta)) = 1 - \lambda^\top \hat{\xi}_i(\beta,\theta)/(1 + \lambda^\top \hat{\xi}_i(\beta,\theta))$ into $\alpha^\top \psi(\lambda) = 0$, we have

$$|\alpha^\top \hat{\bar{\xi}}_i(\beta,\theta)| \ge \frac{\rho}{1 + \rho \max_{1 \le i \le n} \|\xi_i(\beta,\theta)\|} \alpha^\top S_n \alpha,$$

where $S_n = \frac{1}{n} \sum_{i=1}^n \hat{\xi}_i(\beta,\theta)\hat{\xi}_i(\beta,\theta)^\top$. Because of

$$0 < 1 + \lambda^\top \hat{\xi}_i(\beta,\theta) \le 1 + \rho \max_{1 \le i \le n} \|\xi_i(\beta,\theta)\|,$$

we have

$$\rho[\alpha^\top S_n \alpha - \alpha^\top \hat{\xi}_i(\beta, \theta) \max_{1 \le i \le n} \|\xi_i(\beta, \theta)\|] \le \left| \alpha^\top \bar{\hat{\xi}}_i(\beta, \theta) \right|. \tag{21}$$

Because $|\alpha^\top \bar{\hat{\xi}}_i(\beta, \theta)| \le \|\bar{\hat{\xi}}_i(\beta, \theta)\| = O_p(\sqrt{p/n})$ and Lemma 5.1, then

$$\max_{1 \le i \le n} \|\xi_i(\beta, \theta)\| \left| \alpha^\top \bar{\hat{\xi}}_i(\beta, \theta) \right| = o_p(1). \tag{22}$$

By combining (21) and (22), we have

$$|\rho[\alpha^\top S_n \alpha + o_{p(1)}]| = O_p(\sqrt{p/n}).$$

According to Lemma 5.2, for a constant $C_1 > 0$, $P(\alpha^\top S_n \alpha \ge \frac{1}{2} C_1) \to 1$ as $n \to \infty$. Hence, $\rho = O_p(\sqrt{p/n})$, that is $\|\lambda\| = \rho = O_p(\sqrt{p/n})$, and the proof of Lemma 5.3 is completed. $\qquad\square$

Lemma 5.4 *Under Assumptions of Theorem* 2.2, *as* $n \to \infty$,

$$\{2(p+r-1)\}^{-1} \left\{ \left(\frac{1}{\sqrt{n}} \sum_{i=1}^n \hat{\xi}_i(\beta, \theta) \right)^\top V^{-1} \left(\frac{1}{\sqrt{n}} \sum_{i=1}^n \hat{\xi}_i(\beta, \theta) \right) \right.$$
$$\left. - (p+r-1) \right\} \xrightarrow{L} N(0, 1).$$

Proof The proof entails applying the martingale central limit theorem as given in Hall and Hyde (1980), and is omitted. $\qquad\square$

Lemma 5.5 *Under Assumptions of Theorem* 2.2,

$$\left\{ \frac{1}{\sqrt{n}} \sum_{i=1}^n \hat{\xi}_i(\beta, \theta) \right\}^\top \left(S_n^{-1} - V^{-1} \right) \left\{ \frac{1}{\sqrt{n}} \sum_{i=1}^n \hat{\xi}_i(\beta, \theta) \right\} = o_p(\sqrt{p}).$$

Proof Let $D_n = V^{-1/2} S_n V^{-1/2} - I_{p+r}$, where I_{p+r} is the $p+r$ dimensional identity matrix.

$$S_n^{-1} - V^{-1} = V^{-1/2} \left(V^{1/2} S_n^{-1} V^{1/2} - I_{p+r} \right) V^{-1/2}$$
$$= V^{-1/2} \left\{ -D_n + D_n^2 + D_n^2 \left(V^{1/2} S_n^{-1} V^{1/2} - I_{p+r} \right) \right\} V^{-1/2}.$$

It is easy to check that

$$tr(S_n - V) = tr\left(V^{1/2} \left(V^{-1/2} S_n V^{-1/2} - I_{p+r} \right) V^{1/2} \right)$$
$$= tr(D_n V D_n V) \ge \gamma_1^2(V) tr\left(D_n^2 \right),$$

where $\gamma_1(V)$ is the smallest eigenvalue of V. Similar to the proof of Lemma 5.4 in Chen et al. (2009), we have

$$tr\left(D_n^2\right) \le tr\left\{(S_n - V)^2\right\} = O_p\left(p^2/n\right).$$

Then

$$
\begin{aligned}
tr\left(S_n^{-1} - V^{-1}\right)^2 &\le 2tr\left\{V^{-2}\left(-D_n + D_n^2\right)^2\right\} + 2tr\left\{D_n^4\left(S_n^{-1} - V^{-1}\right)^2\right\} \\
&\le 2tr\left\{V^{-2}\left(-D_n + D_n^2\right)^2\right\} \\
&\quad + 2\left\{tr\left(D_n^2\right)\right\}^2 tr\left\{\left(S_n^{-1} - V^{-1}\right)^2\right\} \\
&= 2tr\left\{V^{-2}\left(-D_n + D_n^2\right)^2\right\} + o_p\left(tr\left\{\left(S_n^{-1} - V^{-1}\right)^2\right\}\right) \\
&= o_p\left(p^2/n\right).
\end{aligned}
$$

Because $\|\frac{1}{n}\sum_{i=1}^n \hat{\xi}_i(\beta, \theta)\| = O_p(\sqrt{p/n})$, we can obtain

$$
\left\{\tfrac{1}{\sqrt{n}}\sum_{i=1}^n \hat{\xi}_i(\beta, \theta)\right\}^\top \left(S_n^{-1} - V^{-1}\right)\left\{\tfrac{1}{\sqrt{n}}\sum_{i=1}^n \hat{\xi}_i(\beta, \theta)\right\}
$$

$$
\le n\|\tfrac{1}{n}\sum_{i=1}^n \hat{\xi}_i(\beta, \theta)\|^2\sqrt{tr\left(S_n^{-1} - V^{-1}\right)^2} = o_p\left(\sqrt{p}\right).
$$

$\square$

Proof of Theorem 2.2

Proof Put $W_i = \lambda^\top\hat{\xi}_i(\beta, \theta), i = 1, \ldots, n$. By expanding Eq. (8), we obtain

$$
0 = \sum_{i=1}^n \frac{\hat{\xi}_i(\beta, \theta)}{1 + \lambda^\top\hat{\xi}_i(\beta, \theta)} = \sum_{i=1}^n \hat{\xi}_i(\beta, \theta) - \sum_{i=1}^n \left\{\hat{\xi}_i(\beta, \theta)\hat{\xi}_i(\beta, \theta)^\top\right\}\lambda + R_n, \quad (23)
$$

where $R_n = \sum_{i=1}^n \frac{\hat{\xi}_i(\beta,\theta)(\lambda^\top\hat{\xi}_i(\beta,\theta))^2}{(1+\vartheta_i)^3}$ and $|\vartheta_i| \le |\lambda^\top\hat{\xi}_i(\beta, \theta)|$. By Lemma 5.1, we have $\max_{1\le i\le n}|\vartheta_i| = o_p(1)$. Hence $R_n = R_{n1}\{1 + o_p(1)\}$, where

$$
R_{n1} = \sum_{i=1}^n \hat{\xi}_i(\beta, \theta)\left(\lambda^\top\hat{\xi}_i(\beta, \theta)\right)^2.
$$

Apply Lemmas 5.1 and 5.3, we obtain

$$
\|n^{-1}R_n\| \le C\|\lambda\|^2 \max_{1\le i\le n}\|\hat{\xi}_i(\beta, \theta)\|n^{-1}\sum_{i=1}^n \|\hat{\xi}_i(\beta, \theta)\|^2 = o_p(a_n). \quad (24)
$$

By (23), we have

$$\lambda = \left\{ \sum_{i=1}^{n} \hat{\xi}_i(\beta, \theta) \hat{\xi}_i(\beta, \theta)^\top \right\}^{-1} \sum_{i=1}^{n} \hat{\xi}_i(\beta, \theta) + \left\{ \sum_{i=1}^{n} \hat{\xi}_i(\beta, \theta) \hat{\xi}_i(\beta, \theta)^\top \right\}^{-1} R_n.$$

Applying Taylor's expansion, for some ζ_i such that $|\zeta_i| \le |\lambda^\top \hat{\xi}_i(\beta, \theta)|$, we obtain

$$\log\left(1 + \lambda^\top \hat{\xi}_i(\beta, \theta)\right) = \lambda^\top \hat{\xi}_i(\beta, \theta) - \frac{\left\{ \lambda^\top \hat{\xi}_i(\beta, \theta) \right\}^2}{2} + \frac{\left\{ \lambda^\top \hat{\xi}_i(\beta, \theta) \right\}^3}{3(1+\zeta_i)^4}.$$

Therefore,

$$
\begin{aligned}
\tilde{l}(\beta, \theta) &= \left(\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \hat{\xi}_i(\beta, \theta) \right)^\top \left\{ \frac{1}{n} \sum_{i=1}^{n} \hat{\xi}_i(\beta, \theta) \hat{\xi}_i(\beta, \theta)^\top \right\}^{-1} \left(\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \hat{\xi}_i(\beta, \theta) \right) \\
&\quad - \frac{1}{n} R_n^\top \left\{ \frac{1}{n} \sum_{i=1}^{n} \hat{\xi}_i(\beta, \theta) \hat{\xi}_i(\beta, \theta)^\top \right\}^{-1} R_n + \sum_{i=1}^{n} \frac{2 \left\{ \lambda^\top \hat{\xi}_i(\beta, \theta) \right\}^3}{3(1+\zeta_i)^4} \\
&= \left(\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \hat{\xi}_i(\beta, \theta) \right)^\top V^{-1} \left(\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \hat{\xi}_i(\beta, \theta) \right) \\
&\quad + \left(\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \hat{\xi}_i(\beta, \theta) \right)^\top \left[\left\{ \frac{1}{n} \sum_{i=1}^{n} \hat{\xi}_i(\beta, \theta) \hat{\xi}_i(\beta, \theta)^\top \right\}^{-1} - V^{-1} \right] \\
&\quad \times \left(\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \hat{\xi}_i(\beta, \theta) \right) - \frac{1}{n} R_n^\top \left\{ \frac{1}{n} \sum_{i=1}^{n} \hat{\xi}_i(\beta, \theta) \hat{\xi}_i(\beta, \theta)^\top \right\}^{-1} R_n \\
&\quad + \frac{2}{3} \sum_{i=1}^{n} \left\{ \lambda^\top \hat{\xi}_i(\beta, \theta) \right\}^3 \left\{ 1 + o_p(1) \right\}.
\end{aligned}
\tag{25}
$$

By Lemma 5.5, we have

$$
\begin{aligned}
&\left(\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \hat{\xi}_i(\beta, \theta) \right)^\top \left[\left\{ \frac{1}{n} \sum_{i=1}^{n} \hat{\xi}_i(\beta, \theta) \hat{\xi}_i(\beta, \theta)^\top \right\}^{-1} - V^{-1} \right] \\
&\quad \times \left(\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \hat{\xi}_i(\beta, \theta) \right) = o_p(1).
\end{aligned}
\tag{26}
$$

By Lemmas 5.1–5.3 and (24), we can obtain

$$\frac{1}{n} R_n^\top \left\{ \frac{1}{n} \sum_{i=1}^{n} \hat{\xi}_i(\beta, \theta) \hat{\xi}_i(\beta, \theta)^\top \right\}^{-1} R_n = o_p(1),
\tag{27}$$

and

$$\frac{2}{3} \sum_{i=1}^{n} \left\{ \lambda^\top \hat{\xi}_i \left(\beta, \theta\right) \right\}^3 \left\{ 1 + o_p\left(1\right) \right\} = o_p\left(\sqrt{p}\right).$$

(28)

It follows from (25)–(28) that

$$\tilde{l}\left(\beta, \theta\right) = \left(\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \hat{\xi}_i\left(\beta, \theta\right)\right)^\top V^{-1} \left(\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \hat{\xi}_i\left(\beta, \theta\right)\right) + o_p\left(\sqrt{p}\right).$$

Hence the theorem follows from Lemmas 5.4 and 5.5, and the proof of Theorem 2.2 is completed.   □

Proof of Theorem 2.3

Proof We first prove that $\max_{1 \le i \le n} \|\hat{\tilde{\xi}}_i(\beta^{(1)})\| = o_p(n^{1/2})$. It can be shown that

$$\begin{aligned}
\hat{\tilde{\xi}}_i\left(\beta\right) &= \hat{w}_i \left\{ Y_i - X_i^{(1)\top}\beta^{(1)} - X_i^{(2)\top}\hat{\beta}^{(2)} - \hat{g}\left(Z_i^\top\hat{\theta}\right) \right\} \left\{ X_i^{(1)} - \frac{\hat{E}\left(\hat{w}X^{(1)}|Z_i^\top\hat{\theta}\right)}{\hat{E}\left(\hat{w}|Z_i^\top\hat{\theta}\right)} \right\} \\
&= \left\{ w_i \left(1 + o_p\left(1\right)\right) \right\} \left\{ \varepsilon_i + X_i^\top\left(\beta - \hat{\beta}\right) + X_i^{(1)\top}\left(\hat{\beta}^{(1)} - \beta^{(1)}\right) + \left(g\left(Z_i^\top\hat{\theta}\right)\right. \right. \\
&\quad \left. -\hat{g}\left(Z_i^\top\hat{\theta}\right)\right) \right\} \\
&\quad \times \left\{ \left(X_i^{(1)} - \frac{E\left(wX^{(1)}|Z_i^\top\hat{\theta}\right)}{E\left(w|Z_i^\top\hat{\theta}\right)}\right) + \left(\frac{E\left(wX^{(1)}|Z_i^\top\hat{\theta}\right)}{E\left(w|Z^\top\hat{\theta}\right)} - \frac{\hat{E}\left(\hat{w}X^{(1)}|Z_i^\top\hat{\theta}\right)}{\hat{E}\left(\hat{w}|Z_i^\top\hat{\theta}\right)}\right) \right\} \\
&= M_{i1} + M_{i2} + M_{i3} + M_{i4} + M_{i5} + M_{i6} + M_{i7} + M_{i8},
\end{aligned}$$

(29)

where

$$M_{i1} = \left\{ w_i\left(1 + o_p\left(1\right)\right) \right\} \varepsilon_i \left\{ X_i^{(1)} - \frac{E\left(wX^{(1)}|Z_i^\top\hat{\theta}\right)}{E\left(w|Z_i^\top\hat{\theta}\right)} \right\},$$

$$M_{i2} = \left\{ w_i\left(1 + o_p\left(1\right)\right) \right\} \varepsilon_i \left\{ \frac{E\left(wX^{(1)}|Z_i^\top\hat{\theta}\right)}{E\left(w|Z^\top\hat{\theta}\right)} - \frac{\hat{E}\left(\hat{w}X^{(1)}|Z_i^\top\hat{\theta}\right)}{\hat{E}\left(\hat{w}|Z^\top\hat{\theta}\right)} \right\},$$

$$M_{i3} = \left\{ w_i\left(1 + o_p\left(1\right)\right) \right\} X_i^\top\left(\beta - \hat{\beta}\right) \left\{ X_i^{(1)} - \frac{E\left(wX^{(1)}|Z_i^\top\hat{\theta}\right)}{E\left(w|Z_i^\top\hat{\theta}\right)} \right\},$$

$$M_{i4} = \left\{ w_i\left(1 + o_p\left(1\right)\right) \right\} X_i^\top\left(\beta - \hat{\beta}\right) \left\{ \frac{E\left(wX^{(1)}|Z_i^\top\hat{\theta}\right)}{E\left(w|Z^\top\hat{\theta}\right)} - \frac{\hat{E}\left(\hat{w}X^{(1)}|Z_i^\top\hat{\theta}\right)}{\hat{E}\left(\hat{w}|Z^\top\hat{\theta}\right)} \right\},$$

$$M_{i5} = \{w_i \left(1 + o_p(1)\right)\} X_i^{(1)\top} \left(\hat{\beta}^{(1)} - \beta^{(1)}\right) \left\{ X_i^{(1)} - \frac{E\left(wX^{(1)}|Z_i^\top \hat{\theta}\right)}{E\left(w|Z_i^\top \hat{\theta}\right)} \right\},$$

$$M_{i6} = \{w_i \left(1 + o_p(1)\right)\} X_i^{(1)\top} \left(\hat{\beta}^{(1)}\right.$$
$$\left. -\beta^{(1)}\right) \left\{ \frac{E\left(wX^{(1)}|Z_i^\top \hat{\theta}\right)}{E\left(w|Z_i \hat{\theta}\right)} - \frac{\hat{E}\left(\hat{w}X^{(1)}|Z_i^\top \hat{\theta}\right)}{\hat{E}\left(\hat{w}|Z^\top \hat{\theta}\right)} \right\},$$

$$M_{i7} = \{w_i \left(1 + o_p(1)\right)\} \left(g\left(z_i^\top \hat{\theta}\right) - \hat{g}\left(Z_i^\top \hat{\theta}\right)\right) \left\{ X_i^{(1)} - \frac{E\left(wX^{(1)}|Z_i^\top \hat{\theta}\right)}{E\left(w|Z_i^\top \hat{\theta}\right)} \right\},$$

$$M_{i8} = \{w_i \left(1 + o_p(1)\right)\} \left(g\left(z_i^\top \hat{\theta}\right)\right.$$
$$\left. -\hat{g}\left(Z_i^\top \hat{\theta}\right)\right) \left\{ \frac{E\left(wX^{(1)}|Z_i^\top \hat{\theta}\right)}{E\left(w|Z_i^\top \hat{\theta}\right)} - \frac{\hat{E}\left(\hat{w}X^{(1)}|Z_i^\top \hat{\theta}\right)}{\hat{E}\left(\hat{w}|Z^\top \hat{\theta}\right)} \right\}.$$

By (29), we can obtain that

$$\max_{1\le i \le n} \|\hat{\tilde{\xi}}_i(\beta^{(1)})\| \le \max_{1\le i \le n} \|M_{i1}\| + \max_{1\le i \le n} \|M_{i2}\| + \max_{1\le i \le n} \|M_{i3}\| + \max_{1\le i \le n} \|M_{i4}\|$$
$$+ \max_{1\le i \le n} \|M_{i5}\| + \max_{1\le i \le n} \|M_{i6}\| + \max_{1\le i \le n} \|M_{i7}\| + \max_{1\le i \le n} \|M_{i8}\|.$$

Similar to the proof of Proposition 2 in Ma and Zhu (2013), it is easy to show that

$$\max_{1\le i \le n} \|M_{l1}\| = o_p(n^{1/2}), \quad l = 1, \dots, 8.$$

Therefore, we have $\max_{1\le i \le n} \|\hat{\tilde{\xi}}_i(\beta^{(1)})\| = o_p(n^{1/2})$. In addition, from the proof of Theorem 3.1 in Li and Wang (2003), as $n \to \infty$, we can also show that

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \hat{\tilde{\xi}}_i\left(\beta^{(1)}\right) \xrightarrow{L} N\left(0, V_1\left(\beta^{(1)}\right)\right), \tag{30}$$

$$\frac{1}{n} \sum_{i=1}^{n} \hat{\tilde{\xi}}_i\left(\beta^{(1)}\right) \hat{\tilde{\xi}}_i^\top\left(\beta^{(1)}\right) \xrightarrow{p} V_1\left(\beta^{(1)}\right), \tag{31}$$

where

$$V_1\left(\beta^{(1)}\right) = E\left\{ wX^{(1)}X^{(1)\top} - \frac{E\left(wX^{(1)}|Z^\top \hat{\theta}\right) E\left(wX^{(1)T}|Z^\top \hat{\theta}\right)}{E\left(w|Z^\top \hat{\theta}\right)} \right\}$$

and $\xrightarrow{p}$ stands for convergence in probability. By $\max_{1 \le i \le n} \|\hat{\tilde{\xi}}_i(\beta^{(1)})\| = o_p(n^{1/2})$ and Talor expansion to (10), we can obtain that

$$\tilde{l}\left(\beta^{(1)}\right) = 2\sum_{i=1}^{n} \lambda^{(1)\top}\hat{\tilde{\xi}}_i\left(\beta^{(1)}\right) - \sum_{i=1}^{n}\left\{\lambda^{(1)\top}\hat{\tilde{\xi}}_i\left(\beta^{(1)}\right)\right\}^2 + o_p(1). \qquad (32)$$

Similar to the proof of Theorem 17 in Owen (1990), we have

$$\sum_{i=1}^{n}\left\{\lambda^{(1)\top}\hat{\tilde{\xi}}_i\left(\beta^{(1)}\right)\right\}^2 = \sum_{i=1}^{n}\lambda^{(1)\top}\hat{\tilde{\xi}}_i\left(\beta^{(1)}\right) + o_p(1), \qquad (33)$$

$$\lambda^{(1)} = \left\{\sum_{i=1}^{n}\hat{\tilde{\xi}}_i\left(\beta^{(1)}\right)\hat{\tilde{\xi}}_i\left(\beta^{(1)}\right)^\top\right\}^{-1}\sum_{i=1}^{n}\hat{\tilde{\xi}}_i\left(\beta^{(1)}\right) + o_p\left(n^{-1/2}\right). \qquad (34)$$

Combining (32)–(34) implies that

$$\tilde{l}\left(\beta^{(1)}\right)$$
$$= \left\{\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\hat{\tilde{\xi}}_i\left(\beta^{(1)}\right)\right\}^\top\left\{\frac{1}{n}\sum_{i=1}^{n}\hat{\tilde{\xi}}_i\left(\beta^{(1)}\right)\hat{\tilde{\xi}}_i^\top\left(\beta^{(1)}\right)\right\}^{-1}\left\{\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\hat{\tilde{\xi}}_i\left(\beta^{(1)}\right)\right\}$$
$$+ o_p(1).$$

Therefore, together with (30) and (31), we can show that $\tilde{l}(\beta^{(1)}) \xrightarrow{L} \chi_k^2$, and the proof is completed. $\qquad\qquad\square$

The partially linear model or the single-index model is a special case of the partially linear single-index model. We can prove Theorems 3.1 and 3.2 by using the same arguments in the proofs of Theorems 2.1–2.3, hence their proofs are omitted.

References

Bai Z, Saranadasa H (1996) Effect of high dimension: by an example of a two sample problem. Stat Sin 6:311–329

Carroll R, Fan J, Gijbels I, Wand M (1997) Generalized partially linear single-index models. J Am Stat Assoc 92:477–489

Chen S, Hall F (1993) Smoothed empirical likelihood confidence intervals for quantiles. Ann Stat 21:1166–1181

Chen S, Peng L, Qin Y (2009) Effects of data dimension on empirical likelihood. Biometrika 96:712–722

Davidian M, Tsiatis A, Leon S (2005) Semiparametric estimation of treatment effect in a pretest-posttest study with missing data. Stat Sin 20:261–301

Donohn D (2000) High-dimensional data analysis: high-dimensional data analysis: the curses and blessings of dimensionality. Aide-memoire of a lecture at AMS conference on math challenges of the 21st century

Engle R, Granger C, Rise J, Weiss A (1986) Semiparametric estimates of the relation between weather and electricity sales. J Am Stat Assoc 81:310–320

Hall P, Hyde C (1980) Martingale central limit theory and its applications. Academic Press, New York

Hammer S, Katzenstein D, Hughes M et al (1996) For the AIDS clinical trials group study 175 study team: a trial comparing nucleoside monotherapy with combination therapy in HIV-infected adults with CD4 cell counts from 200 to 500 per cubic millimeter. New Engl J Med 20:1081–1089

Hjort H, Mckeague I, Van Keilegom I (2009) Extending the scope of empirical likelihood. Ann Stat 37:1079–1111

Huber P (1973) Robust regression: asymptotics, conjectures and Monte Carlo. Ann Stat 1:799–821

Kolaczyk E (1994) Empirical likelihood for generalized linear models. Stat Sin 4:199–218

Lai P, Wang Q (2014) Semiparametric efficient estimation for partially linear single-index models with responses missing at random. J Multivar Anal 128:33–50

Ledoit O, Wolf M (2002) Some hypothesis tests for the covariance matrix when the dimension is large compared to the sample size. Ann Stat 30:1081–1102

Li G, Wang Q (2003) Empirical likelihood regression analysis for right censored data. Stat Sin 13:51–68

Lu X (2009) Empirical likelihood for heteroscedastic partially linear models. J Multivar Anal 100:387–395

Lu X, Qi Y (2004) Empirical likelihood for the additive risk model. Probab Math Stat 24:419–431

Ma Y, Zhu L (2013) Doubly robust and efficient estimators for heteroscedastic partially linear single-index models allowing high dimensional covariates. J R Stat Soc Ser B 75:305–322

Ma Y, Chiou J, Wang N (2006) Efficient semiparametric estimator for heteroscedastic partially linear models. Biometrika 943:75–84

Owen A (1988) Empirical likelihood ratio confidence intervals for a single function. Biometrika 75:237–249

Owen A (1990) Empirical likelihood ratio confidence regions. Ann Stat 18:90–120

Owen A (1991) Empirical likelihood for linear models. Ann Stat 19:1725–1747

Owen A (2001) Empirical likelihood. Chapman and Hall, London

Qin J, Lawless J (1994) Empirical likelihood and general estimating equations. Ann Stat 22:300–325

Qin G, Jing B (2001) Empirical likelihood for Cox regression model under random censorship. Commun Stat Simul Comput 30:79–90

Shi J, Lau T (2000) Empirical likelihood for partially linear models. J Multivar Anal 72:132–148

Tsao M (2004) Bounds on coverage probabilities of the empirical likelihood ratio confidence regions. Ann Stat 32:1215–1221

Wang Q, Rao J (2002) Empirical likelihood-based inference in linear errors-in-covariables models with validation data. Biometrika 89:345–358

Xia H, Härdle W (2006) Semi-parametric estimation of partially linear single-index models. J Multivar Anal 97:1162–1184

Xia Y, Tong H, Li W (1999) On extended partially linear single-index models. Biometrika 86:831–842

Xia Y, Tong H, Li W, Zhu L (2002) An adaptive estimation of dimension reduction space. J R Stat Soc Ser B 64:363–410

Xue L, Zhu L (2006) Empirical likelihood for single-index models. J Multivar Anal 97:1295–1312

Yu Y, Ruppert D (2002) Penalized spline estimation for partially linear single-index models. J Am Stat Assoc 97:1042–1054

Zhang J, Wang T, Zhu L, Liang H (2012) A dimension reduction based approach for estimation and variable selection in partially linear single-index models with high-dimensional covariates. Electron J Stat 6:2235–2273

Zhu L, Xue L (2006) Empirical likelihood confidence regions in a partially linear single-index model. J R Stat Soc Ser B 68:549–570