Brazilian Journal of Probability and Statistics
2018, Vol. 32, No. 1, 86–116
https://doi.org/10.1214/16-BJPS333

Nonlinear measurement errors models subject to partial linear additive distortion

Jun Zhang, Nanguang Zhou, Qian Chen and Tianyue Chu

Shenzhen University

Abstract. We study nonlinear regression models when the response and predictors are unobservable and distorted in a multiplicative fashion by partial linear additive models (PLAM) of some observed confounding variables. After approximating the additive nonparametric components in the PLAM via polynomial splines and calibrating the unobserved response and unobserved predictors, we develop a semi-parametric profile nonlinear least squares procedure to estimate the parameters of interest. The resulting estimators are shown to be asymptotically normal. To construct confidence intervals for the parameters of interest, an empirical likelihood-based statistic is proposed to improve the accuracy of the associated normal approximation. We also show that the empirical likelihood statistic is asymptotically chi-squared. Moreover, a test procedure based on the empirical process is proposed to check whether the parametric regression model is adequate or not. A wild bootstrap procedure is proposed to compute p-values. Simulation studies are conducted to examine the performance of the estimation and testing procedures. The methods are applied to re-analyze real data from a diabetes study for an illustration.

1 Introduction

Measurement error is common in many disciplines, such as economics, health science and medical research, due to improper instrument calibration or many other reasons. It is known that simply ignoring the errors can cause estimation bias. Therefore, it requires particular care to eliminate such bias when estimating target parameters to accurately detecting the relationship among variables. Research on classical errors-in-variables have been widely studied in the last two decades, Carroll et al. (2006) gave comprehensive surveys on linear and nonlinear measurement errors models for a variety of such real-world examples. In this paper, we consider a situation both the response and predictors are unobservable and distorted by the multiplicative effects of some observable confounding variables in

Key words and phrases. Bootstrap approximation, confounding variables, covariate-adjusted regression, distorting function, empirical likelihood, empirical process, measurement errors models, model checking, multiplicative effect, regression spline.

Received April 2015; accepted August 2016.

the following covariate-adjusted setting:

$$\begin{cases} Y = f(\boldsymbol{X}, \boldsymbol{\beta}) + \varepsilon, \\ \tilde{Y} = \phi(\boldsymbol{U})Y, \\ \tilde{\boldsymbol{X}} = \psi(\boldsymbol{U})\boldsymbol{X}, \end{cases} \tag{1.1}$$

where $f(\cdot,\cdot)$ is a known continuous nonlinear function, $\boldsymbol{\beta}$ is an unknown $q \times 1$ parameter vector on a compact parameter space $\Theta_{\boldsymbol{\beta}} \subset \mathbb{R}^q$. Y is the unobservable response, $\boldsymbol{X} = (X_1, X_2, \dots, X_p)^\tau$ is the unobservable continuous predictor vector (the superscript τ denotes the transpose operator throughout this paper), while $\tilde{Y}$ and $\tilde{\boldsymbol{X}}$ are observed and distorted response and predictors, $\psi(\cdot)$ is a $p \times p$ diagonal matrix $\mathrm{diag}(\psi_1(\cdot), \dots, \psi_p(\cdot))$, where $\phi(\cdot)$ and $\psi_r(\cdot)$ are unknown contaminating smooth functions of an observed confounding vector $\boldsymbol{U} = (U_1, \dots, U_d)^\tau$. The diagonal form of $\psi(\cdot)$ indicates that the confounding vector $\boldsymbol{U}$ distorts each component of the unobserved predictors $\boldsymbol{X}$ in a multiplicative fashion. The confounding vector $\boldsymbol{U}$ and model error ε are independent of $(\boldsymbol{X}, Y)$.

This type of measurement error models is revealed by Şentürk and Müller (2005, 2006) for analyze the data from some biomedical and health-related studies, in which confounding variables such as the body mass index, height or weight usually have multiplicative effects on the primary response and predictor variables. For example, Şentürk and Nguyen (2009) suggested the covariate "body mass index" (BMI) to be a potential confounding variable. Kaysen et al. (2002) also treat BMI as the confounder on hemodialysis patients and they further realized that the fibrinogen level and serum transferrin level should be divided by BMI to eliminate the potential bias possibly caused by BMI. In fact, no one knows the exact relationship between the confounder and primary variables, and the way of naively dividing BMI may not be appropriate and possibly lead to biased estimators of the parameters. To avoid the model misspecified, Şentürk and Müller (2005, 2006) introduced a flexible multiplicative adjustment by using unknown smooth distorting functions $\phi(\cdot)$, $\psi_r(\cdot)$ for the confounding variable $\boldsymbol{U}$.

Recently, there are two main estimation procedures for the distortion measurement errors, namely, the transformation method proposed by Şentürk and Müller (2005, 2006, 2009), and the direct plug-in method proposed by Cui et al. (2009). The readers can refer to the following work on the development of these two methods, see for example, Zhang, Zhu and Liang (2012), Zhang, Gai and Wu (2013), Zhang et al. (2014) for the multivariate confounders, Li, Lin and Cui (2010), Zhang et al. (2013) for some semi-parametric models, Zhang, Feng and Zhou (2014) for an efficient estimator of the correlation coefficient between two variables, Li et al. (2014) for the variable selection by employing the smoothly clipped absolute deviation penalty (Fan and Li (2001), Liang and Li (2009, SCAD)), and Zhang, Li and Feng (2015) for the problem of model checking on a parametric regression model by using a residuals based empirical process based test statistic. Delaigle, Hall and Wen-Xin (2016) obtained a fundamental work of nonparametric estimation of a regression curve when the data are observed with multiplicative distortion. Zhang et al. (2016) considered the estimation and hypothesis testing problems for the partial linear regression models under the distortion measurement error setting, and the proposed estimation procedure is designed to accommodate undistorted as well as distorted variables.

In this paper, we consider when more than one confounding variables affect the response and predictors, that is, a d-dimensional confounding vector $\boldsymbol{U}$. Note that the distorting functions $\phi(\cdot)$ and $\psi(\cdot)$ are unknown, the "curse of dimensionality" problem occurs if one uses the d-dimensional smooth models: $\phi(\boldsymbol{U}) = \phi(U_1, \dots, U_d)$, $\psi_r(\boldsymbol{U}) = \psi_r(U_1, \dots, U_d)$. To overcome the problem, Nguyen and Şentürk (2008) modeled the distortions by single-index models (SIM) and considered to model the primary variables as a linear regression model. Nguyen and Şentürk (2008) proposed a hybrid backfitting algorithm to simultaneously estimate unknown single-index parameters and varying coefficient functions, and derived final estimators of the parameters with a weighted-average of the estimated coefficient functions. Zhang, Zhu and Liang (2012) extended the single-index distortions to the nonlinear regression models, and proposed to use the estimating function method (EFM) proposed by Cui, Härdle and Zhu (2011) to estimate the single-index parameters, and the direct plug-in procedure was used to give a nonlinear least squares estimators of $\boldsymbol{\beta}$. Recently, Zhang et al. (2014) proposed to use additive models (AM) as a competitor for the distortion functions, because of the unique feature of interpretability and flexibility of the additive models. For the estimations and application of additive models, the readers can refer the following literature Hastie and Tibshirani (1990), Huang (1999, 2003), Li (2000), Li and Ruppert (2008), Liang et al. (2008), Opsomer and Ruppert (1997), Stone (1985).

As claimed in Stone (1985), the AM and partial linear additive models (PLAM) can also achieve dimension reduction as SIM does, and, be possibly greater estimation accuracy than fully nonparametric estimation. Moreover, SIM, AM, and PLAM are nonnested each other. Each of the models can relax some strong assumptions of standard parametric models. Therefore, in most cases, at most one of these models can be correctly specified in a given application, and an analyst should choose among the models. In Zhang, Zhu and Liang (2012), Zhang et al. (2014), they modeled distorting functions as SIM and AM, respectively. As we indicated above, the nonnesting phenomenon of SIM and AM, PLAM motivates us to model distorting functions as PLAM in this paper. For the case of additive components PLAM, the spline approximation method will be used to approximate unknown distorting functions, then the profile least-square technique can be used to obtain estimators of parameters. It is also of interest to investigate the asymptotic properties of the parameters when we use the PLAM distortion functions. In this paper, we investigate PLAM distortion phenomenon in the context of multiple distorting measurement error setting. To the best of our knowledge, this kind of error-prone distortion is new and the findings contribute to the literature on the distortion measurement error modeling.

For the parameters estimation, we first calibrate distorted response and predictors by using spline approximations for PLAM, and investigate the asymptotic normality for the parameter in PLAM. Then, we estimate the parameter $\boldsymbol{\beta}$ by profile nonlinear least squares procedure and establish asymptotic normality. The linear regression models, a special case of nonlinear models, are discussed as well. To construct the confidence intervals of the parameter $\boldsymbol{\beta}$, an empirical likelihood based statistic is proposed. For the model checking on the adequacy of the parametric models, we also develop a lack-of-fit test for model (1.1). We developed a Cramér-von Mises (CvM) test proposed by Stute, González Manteiga and Presedo Quindimil (1998a), and we further present the bootstrap approximation for calculating critical values.

The paper is organized as follows. In Section 2, we propose an estimation procedure for the partial linear additive distortions, associated asymptotic results. We also develop an empirical log-likelihood ratio statistic for the parameter $\boldsymbol{\beta}$, and show that the empirical likelihood statistic has an asymptotic chi-squared distribution. In Section 3, we propose a Cramér-von Mises (CvM) test to investigate the model checking. In Section 4, we conduct simulation studies to examine the performance of the proposed methods. In Section 5, an analysis of a very low birth weight infants dataset is presented. All the technical proofs of the asymptotic results are given in the Appendix.

2 Estimation procedure for semiparametric additive distorting functions

For real data analysis, a partial linear additive structure of distorting functions is valid, i.e., partial linear models (PLMs) (Härdle and Liang (2007), Härdle, Liang and Gao (2000), Heckman (1986), Speckman (1988)), additive partial linear models (APLMs) (Li (2000), Liang et al. (2008), Liang et al. (2009b)). The PLMs and APLMs combine both linear components and non-parametric components, so that they balance model interpretability and flexibility better than the additive models. Thus, it is desirable to determine the linear components in the additive models. Chen, Liang and Wang (2011) proposed a test procedure based on the square of the differences of the fitted models under null and alternative hypothesis, and the authors further proposed a bootstrap procedure to obtain the critical values for approximating the true underlying finite-sample distributions of the test statistic. Under the distorting measurement error setting considered in this paper, one can easily use the Chen, Liang and Wang's (2011) test procedure to investigate the linear components of every additive distorting function in model (1.1). Without loss of generality, the partial linear additive distorting functions can be written as

$$\begin{cases} \phi(\boldsymbol{U}) = \alpha_0 + \zeta_{0,1}(U_1 - E(U_1)) + \cdots + \zeta_{0,d_1}(U_{d_1} - E(U_{d_1})) \\ \qquad\quad + \phi_{d_1+1}(U_{d_1+1}) + \cdots + \phi_d(U_d), \\ \psi_r(\boldsymbol{U}) = \alpha_r + \zeta_{r,1}(U_1 - E(U_1)) + \cdots + \zeta_{r,d_1}(U_{d_1} - E(U_{d_1})) \\ \qquad\quad + \psi_{r,d_1+1}(U_{d_1+1}) + \cdots + \psi_{r,d}(U_d). \end{cases} \tag{2.1}$$

In model (2.1), we consider a simple case in which the first d_1 components of $\boldsymbol{U}$, $U_1, \ldots, U_{d_1}$ are all in the linear part of distorting functions $\phi(\cdot)$ and $\psi_r(\cdot)$'s, $r = 1, \ldots, p$ and the rest $d - d_1$ components of $\boldsymbol{U}$ are all involved in the additive structure. A more general setting of model (2.1) can be determined by this test procedure proposed by Chen, Liang and Wang (2011) in practice. Here we use the model (2.1) to illustrate our methodology and theoretical results.

The distorting functions ϕ_s's and $\psi_{r,s}$'s are unknown smooth functions. To ensure identifiability for the additive model (2.1), one need to assume that

$$\begin{gathered} E\{\phi_s(U_s)\} = 0, \qquad E\{\psi_{r,s}(U_s)\} = 0, \\ s = d_1 + 1, \ldots, d, \qquad r = 1, \ldots, p. \end{gathered} \tag{2.2}$$

Suggested by Şentürk and Müller (2005, 2006, 2009), the following identifiability conditions are also essential for the distorting functions $\phi(\boldsymbol{U})$ and $\psi_r(\boldsymbol{U})$:

$$E\{\phi(\boldsymbol{U})\} = 1, \qquad E\{\psi_r(\boldsymbol{U})\} = 1, \qquad r = 1, \ldots, p. \tag{2.3}$$

Identifiability conditions (2.2) and (2.3) entail that $\alpha_0 = 1$, $\alpha_r = 1$. An equivalent model for (2.1) is

$$\begin{cases} \phi(\boldsymbol{U}) = \dot{\alpha}_0 + \dot{\zeta}_{0,1}U_1 + \cdots + \dot{\zeta}_{0,d_1}U_{d_1} + \phi_{d_1+1}(U_{d_1+1}) \\ \qquad\quad + \cdots + \phi_d(U_d), \\ \psi_r(\boldsymbol{U}) = \dot{\alpha}_r + \dot{\zeta}_{r,1}U_1 + \cdots + \dot{\zeta}_{r,d_1}U_{d_1} \\ \qquad\quad + \psi_{r,d_1+1}(U_{d_1+1}) + \cdots + \psi_{r,d}(U_d). \end{cases} \tag{2.4}$$

For the model (2.4), besides the identifiability conditions $E\{\phi_s(U_s)\} = 0$ and $E\{\psi_l(U_l)\} = 0$, the identifiability conditions (2.3) entail that $\dot{\alpha}_0 + \sum_{m=1}^{d_1} \dot{\zeta}_{0,m} E(U_m) = 1$, $\dot{\alpha}_r + \sum_{m=1}^{d_1} \dot{\zeta}_{r,m} E(U_m) = 1$. We use model (2.1) instead of model (2.4) for simplicity.

Without loss of generality, suppose the covariate U_s is distributed on a compact interval [0, 1] for $s = d_1 + 1, \ldots, d$. Under the smoothness assumptions given in the Appendix, ϕ^o and ψ_r^o can be well approximated by spline functions. Let $\mathcal{Q}_n$ denote polynomial splines on [0, 1] of degree $\rho \geq 1$. A knot sequence for these polynomial splines with J_n interior knots is denoted by $k_{-\rho} = \cdots = k_{-1} = k_0 = 0 < k_1 < \cdots < k_{J_n} < 1 = k_{J_n+1} = \cdots = k_{J_n+\rho+1}$, where J_n increases with the increases of sample size n. The exact order of J_n is presented in condition (C4) in the Appendix. Moreover, $\mathcal{Q}_n$ consists of functions $g(\cdot)$ which satisfy:

- $g(\cdot)$ is a polynomial of degree ρ on each of the subintervals $[k_t, k_{t+1})$, $t = 0, 1, \ldots, J_n - 1$, and the last subinterval is $[k_{J_n}, 1]$;
- $\rho \geq 2$, $g(\cdot)$ is a $(\rho - 1)$-times continuously differentiable function on [0, 1].

For simplicity, we use equally spaced knots. Let $h = \frac{1}{J_n+1}$ be the distance between two consecutive knots.

2.1 Estimation for additive partial linear distorting functions

By the independence of $\boldsymbol{U}$ and $(Y, \boldsymbol{X})$, we know that $E(\tilde{Y}|\boldsymbol{U}) = \phi(\boldsymbol{U})E(Y)$, $E(\tilde{X}_r|\boldsymbol{U}) = \psi_r(\boldsymbol{U})E(X_r)$. Moreover, the identifiability condition (2.3) entails that $E(\tilde{Y}) = E(Y)$, $E(\tilde{X}_r) = E(X_r)$. As such,

$$E\left\{\frac{\tilde{Y}}{E(\tilde{Y})}\Big|\boldsymbol{U}\right\} = \phi(\boldsymbol{U}), \qquad E\left\{\frac{\tilde{X}_r}{E(\tilde{X}_r)}\Big|\boldsymbol{U}\right\} = \psi_r(\boldsymbol{U}). \tag{2.5}$$

Recalling the distorting functions ϕ, ψ_r's in (2.1) have a partial linear additive mean structure, a natural way of estimating ϕ and ψ_r's is to find functions ϕ^o, $\psi_r^o \in \mathcal{Q}_n$ and values of α_0, α_r that minimize the least squares objective functions

$$\begin{cases} \sum_{i=1}^{n}\Bigg\{\tilde{Y}_i/\bar{\tilde{Y}} - \alpha_0 - \sum_{l=1}^{d_1}\zeta_{0,l}(U_{l,i} - \bar{U}_l) \\ \qquad - \vec{\phi}_{d_1+1}(U_{d_1+1,i}) - \cdots - \vec{\phi}_d(U_{d,i})\Bigg\}^2, \\ \sum_{i=1}^{n}\Bigg\{\tilde{X}_{r,i}/\bar{\tilde{X}}_r - \alpha_r - \sum_{l=1}^{d_1}\zeta_{r,l}(U_{l,i} \\ \qquad - \bar{U}_l) - \vec{\psi}_{r,d_1+1}(U_{d_1+1,i}) - \cdots - \vec{\psi}_{r,d}(U_{d,i})\Bigg\}^2, \\ \qquad \vec{\phi}_s \in \mathcal{Q}_n, \vec{\psi}_{r,s} \in \mathcal{Q}_n, r = 1, \ldots, p, s = d_1 + 1, \ldots, d, \end{cases} \tag{2.6}$$

where $\bar{\tilde{Y}} = \frac{1}{n}\sum_{i=1}^{n}\tilde{Y}_i$, $\bar{\tilde{X}}_r = \frac{1}{n}\sum_{i=1}^{n}\tilde{X}_{r,i}$, $r = 1, \ldots, p$, $\bar{U}_l = \frac{1}{n}\sum_{i=1}^{n}U_{l,i}$, $l = d_1 + 1, \ldots, d$ are the moment estimators of $E(\tilde{Y})$, $E(\tilde{X}_r)$'s and $E(U_l)$'s.

For the sth covariate U_s, let $b_{j,s}(u_s)$ denote the B-spline basis function of degree ρ. For any ϕ^o, $\psi_r^o \in \mathcal{Q}_n$, we can use B-spline basis function of degree ρ to approximate $\sum_{s=d_1+1}^{d}\phi_s(u_s)$ and $\sum_{s=d_1+1}^{d}\psi_{r,s}(u_s)$ by

$$\begin{aligned} \sum_{s=d_1+1}^{d}\phi_s(u_s) &\approx \boldsymbol{\gamma}_{0,d-d_1}^{\tau}\mathbf{b}_{d-d_1}(\boldsymbol{u}_{d-d_1}), \\ \sum_{s=d_1+1}^{d}\psi_{rs}(u_s) &\approx \boldsymbol{\gamma}_{r,d-d_1}^{\tau}\mathbf{b}_{d-d_1}(\boldsymbol{u}_{d-d_1}), \end{aligned} \tag{2.7}$$

where $\boldsymbol{u}_{d-d_1} = (u_{d_1+1}, \dots, u_d)^\tau$, $\mathbf{b}_{d-d_1}(\boldsymbol{u}_{d-d_1}) = \{b_{j,s}(u_s), j = -\rho, \dots, J_n, s = d_1 + 1, \dots, d\}^\tau$ are the spline basis functions and $\boldsymbol{\gamma}_{0,d-d_1} = \{\gamma_{0,j,s}, j = -\rho, \dots, J_n, s = d_1 + 1, \dots, d\}$, $\boldsymbol{\gamma}_{r,d-d_1} = \{\gamma_{r,j,s}, j = -\rho, \dots, J_n, s = d_1 + 1, \dots, d\}$, $r = 1, \dots, p$.

Let $\boldsymbol{\zeta}_0 = (\zeta_{01}, \dots, \zeta_{0d_1})^\tau$, $\boldsymbol{\zeta}_r = (\zeta_{r1}, \dots, \zeta_{rd_1})^\tau$. The estimation procedure of $(\alpha_0, \boldsymbol{\zeta}_0^\tau, \boldsymbol{\gamma}_{r,d-d_1}^\tau)^\tau$, $(\alpha_r, \boldsymbol{\zeta}_r^\tau, \boldsymbol{\gamma}_{r,d-d_1}^\tau)^\tau$ can be obtained by finding α_0^*, α_r^*, $\boldsymbol{\zeta}_0^*, \boldsymbol{\zeta}_r^*, \boldsymbol{\gamma}_{0,d-d_1}^*, \boldsymbol{\gamma}_{r,d-d_1}^*$ which minimize

$$\begin{cases} \sum_{i=1}^n \{\tilde{Y}_i/\bar{\tilde{Y}} - \alpha_0^* - \boldsymbol{\zeta}_0^{*\tau}(\boldsymbol{U}_{i,d_1} - \bar{\boldsymbol{U}}_{d_1}) + \cdots \\ \qquad + \boldsymbol{\gamma}_{0,d-d_1}^{*\tau}\mathbf{b}_{d-d_1}(\boldsymbol{U}_{i,d-d_1})\}^2, \\ \sum_{i=1}^n \{\tilde{X}_{ri}/\bar{\tilde{X}}_r - \alpha_r^* - \boldsymbol{\zeta}_r^{*\tau}(\boldsymbol{U}_{i,d_1} - \bar{\boldsymbol{U}}_{d_1}) + \cdots \\ \qquad + \boldsymbol{\gamma}_{r,d-d_1}^{*\tau}\mathbf{b}_{d-d_1}(\boldsymbol{U}_{i,d-d_1})\}^2, \end{cases} \tag{2.8}$$

where $\boldsymbol{U}_{i,d_1} = (U_{i,1}, \dots, U_{i,d_1})^\tau$, $\boldsymbol{U}_{i,d-d_1} = (U_{i,d_1+1}, \dots, U_{i,d})^\tau$ and $\bar{\boldsymbol{U}}_{d_1} = (n^{-1}\sum_{i=1}^n U_{i,1}, \dots, n^{-1}\sum_{i=1}^n U_{i,d_1})^\tau$.

Note that $E(\phi_r(U_s)) = 0$, $E(\psi_{r,s}(U_s)) = 0$, the centered spline estimators of each component function are

$$\hat{\phi}_s^*(u_s) = \sum_{j=-\rho}^{J_n} \hat{\gamma}_{0,j,s} b_{j,s}(u_s) - \frac{1}{n}\sum_{i=1}^n \sum_{j=-\rho}^{J_n} \hat{\gamma}_{0,j,s} b_{j,s}(U_{i,s}), \tag{2.9}$$

$$\hat{\psi}_{r,s}^*(u_s) = \sum_{j=-\rho}^{J_n} \hat{\gamma}_{r,j,s} b_{j,s}(u_s) - \frac{1}{n}\sum_{i=1}^n \sum_{j=-\rho}^{J_n} \hat{\gamma}_{r,j,s} b_{j,s}(U_{i,s}), \quad r = 1, \dots, p. \tag{2.10}$$

Now we further define a minimizing function which is mathematically equivalent to (2.8). For $s = d_1 + 1, \dots, d$ and $j = -\rho + 1, \dots, J_n$, let $b_{j,s}^*(u_s) = b_{j,s}(u_s) - \frac{\|b_{j,s}\|_{2s}}{\|b_{j-1,s}\|_{2s}} b_{j-1,s}(u_s)$, where $\|b_{j,s}\|_{2s}$ is defined as

$$\|b_{j,s}\|_{2s} = \{E[b_{j,s}^2(U_s)]\}^{1/2} = \left\{\int_0^1 b_{j,s}^2(u_s) f_s(u_s)\, du_s\right\}^{1/2},$$

where $f_s(u_s)$ is the density function of U_s. Define the standardized version of spline basis as

$$B_{j,s}(u_s) = \frac{b_{j,s}^*(u_s)}{\|b_{j,s}^*\|_{2s}}. \tag{2.11}$$

It is worth to mention that the minimization problem in (2.8) is equivalent to finding those $\alpha_0, \alpha_r, \boldsymbol{\zeta}_0, \boldsymbol{\zeta}_r, \boldsymbol{\gamma}_{0,d-d_1}, \boldsymbol{\gamma}_{r,d-d_1}$ which minimize

$$\begin{cases} \sum_{i=1}^{n}\{\tilde{Y}_i/\bar{\tilde{Y}} - \alpha_0 - \boldsymbol{\zeta}_0^\tau(\boldsymbol{U}_{i,d_1} - \bar{\boldsymbol{U}}_{d_1}) - \boldsymbol{\gamma}_{0,d-d_1}^\tau \boldsymbol{B}_{d-d_1}(\boldsymbol{U}_{i,d-d_1})\}^2, \\ \sum_{i=1}^{n}\{\tilde{X}_{ri}/\bar{\tilde{X}}_r - \alpha_r - \boldsymbol{\zeta}_r^\tau(\boldsymbol{U}_{i,d_1} - \bar{\boldsymbol{U}}_{d_1}) - \boldsymbol{\gamma}_{r,d-d_1}^\tau \boldsymbol{B}_{d-d_1}(\boldsymbol{U}_{i,d-d_1})\}^2, \end{cases} \tag{2.12}$$

where $\boldsymbol{B}_{d-d_1}(\boldsymbol{u}_{d-d_1}) = \{B_{j,s}(u_s), j = -\rho+1, \ldots, J_n, s = d_1+1, \ldots, d\}^\tau$.

Similarly to (2.9) and (2.10), we can also defined the centered spline estimators of every component function $\hat{\phi}_s$ and $\hat{\psi}_{r,s}$, $s = d_1+1, \ldots, d$ and $r = 1, \ldots, p$. As noted in Wang et al. (2011), Zhang and Liang (2011), the basis $\{b_{j,s}(u_s), j = -\rho, \ldots, J_n, s = d_1+1, \ldots, d\}^\tau$ is used for data analytic implementation in practice, and the mathematical equivalent expression (2.11) is convenient for asymptotic derivation. Similarly, the estimators $\hat{\alpha}_0^*, \hat{\alpha}_r^*, \hat{\boldsymbol{\zeta}}_0^*, \hat{\boldsymbol{\zeta}}_r^*, \hat{\boldsymbol{\gamma}}_{0,d-d_1}^*, \hat{\boldsymbol{\gamma}}_{r,d-d_1}^*$ from (2.8) can be used for data analysis in practice, and $\hat{\alpha}_0, \hat{\alpha}_r, \hat{\boldsymbol{\zeta}}_0, \hat{\boldsymbol{\zeta}}_r, \hat{\boldsymbol{\gamma}}_{0,d-d_1}, \hat{\boldsymbol{\gamma}}_{r,d-d_1}$ from (2.12) is used for mathematical asymptotic derivation.

2.2 Estimating parameter $\boldsymbol{\beta}$

After obtaining estimators $\hat{\phi}$ and $\hat{\psi}_r$'s, denoted as

$$\hat{\phi}(\boldsymbol{U}_i) = \hat{\alpha}_0 + \hat{\boldsymbol{\zeta}}_0^\tau(\boldsymbol{U}_{i,d_1} - \bar{\boldsymbol{U}}_{d_1}) + \hat{\boldsymbol{\gamma}}_{0,d-d_1}^\tau \boldsymbol{B}_{d-d_1}(\boldsymbol{U}_{i,d-d_1}),$$
$$\hat{\psi}_r(\boldsymbol{U}_i) = \hat{\alpha}_r + \hat{\boldsymbol{\zeta}}_r^\tau(\boldsymbol{U}_{i,d_1} - \bar{\boldsymbol{U}}_{d_1}) + \hat{\boldsymbol{\gamma}}_{r,d-d_1}^\tau \boldsymbol{B}_{d-d_1}(\boldsymbol{U}_{i,d-d_1}).$$

The unobservable response and predictors $\{Y_i, \boldsymbol{X}_i\}$, $i = 1, \ldots, n$ can be obtained as

$$\hat{Y}_i = \frac{\tilde{Y}_i}{\hat{\phi}(\boldsymbol{U}_i)}, \qquad \hat{X}_{ri} = \frac{\tilde{X}_{ri}}{\hat{\psi}_r(\boldsymbol{U}_i)}, \qquad r = 1, \ldots, p. \tag{2.13}$$

As a result, the nonlinear least squares estimator $\hat{\boldsymbol{\beta}}$ is obtained by solving

$$\sum_{i=1}^{n}\{\hat{Y}_i - f(\hat{\boldsymbol{X}}_i, \boldsymbol{\beta})\}\frac{\partial f(\hat{\boldsymbol{X}}_i, \boldsymbol{\beta})}{\partial \boldsymbol{\beta}_r} = 0, \qquad r = 1, \ldots, p, \tag{2.14}$$

where $\hat{\boldsymbol{X}}_i = (\hat{X}_{1i}, \ldots, \hat{X}_{pi})^\tau$ and $\hat{Y}_i$ are given in (2.13), $\frac{\partial f(\hat{\boldsymbol{X}}_i, \boldsymbol{\beta})}{\partial \boldsymbol{\beta}_r}$ is the partial derivative of f with respect to $\boldsymbol{\beta}_r$. When (2.14) has no closed form solution, one may solve these equations by the Newton–Raphson iterative method.

2.3 Asymptotic results

In this subsection, we investigate the asymptotic properties for the estimators proposed in Section 2.1 and 2.2. In what follows, define $A^{\otimes 2} = AA^\tau$ for any vector or matrix A. Moreover, let $\boldsymbol{e}=(1,\mathbf{0}_{1\times d_1})^\tau$, $\check{\boldsymbol{U}}_{d_1}=\boldsymbol{U}_{d_1}-E(\boldsymbol{U}_{d_1}|\boldsymbol{U}_{d-d_1})$, $\boldsymbol{\Delta}=E[(1,\check{\boldsymbol{U}}_{d_1}^\tau)^\tau]^{\otimes 2}$, $\boldsymbol{\Theta}=E\{[\boldsymbol{U}_{d_1}-E(\boldsymbol{U}_{d_1})][\tilde{Y}-E(\tilde{Y})]\}$ and $\boldsymbol{\Sigma}_{d_1}=\mathrm{Cov}(\boldsymbol{U}_{d_1})$. For any function $g(\cdot)$ satisfying $Eg^2(\boldsymbol{U})<\infty$, we further define

$$\boldsymbol{\Delta}_g=E\big[(1,\check{\boldsymbol{U}}_{d_1}^\tau)^\tau g(\boldsymbol{U})\big]^{\otimes 2},\qquad \boldsymbol{\Xi}_g=E\big[(1,\check{\boldsymbol{U}}_{d_1}^\tau)^\tau g(\boldsymbol{U})\big],$$
$$\boldsymbol{A}_g=E\big\{(1,\check{\boldsymbol{U}}_{d_1}^\tau)^\tau g(\boldsymbol{U})[Y-E(Y)][\tilde{Y}-E(\tilde{Y})]\big\},$$
$$\boldsymbol{B}_g=E\big\{(1,\check{\boldsymbol{U}}_{d_1}^\tau)^\tau[\boldsymbol{U}_{d_1}-E(\boldsymbol{U}_{d_1})]^\tau[Y-E(Y)]g(\boldsymbol{U})\big\},$$

Define $\boldsymbol{\vartheta}_0=(\alpha_0,\boldsymbol{\zeta}_0^\tau)^\tau$, $\boldsymbol{\vartheta}_r=(\alpha_r,\boldsymbol{\zeta}_r^\tau)^\tau$. For the estimators $\hat{\boldsymbol{\vartheta}}_0=(\hat{\alpha}_0,\hat{\boldsymbol{\zeta}}_0^\tau)^\tau$ and $\hat{\boldsymbol{\vartheta}}_r=(\hat{\alpha}_r,\hat{\boldsymbol{\zeta}}_r^\tau)^\tau$, we have the following asymptotic results.

Theorem 1. *Under the conditions* (C1)–(C8) *in the Appendix, we have that*

$$\sqrt{n}(\hat{\boldsymbol{\vartheta}}_0-\boldsymbol{\vartheta}_0)\xrightarrow{L}N\big(\mathbf{0},\boldsymbol{\Delta}^{-1}\boldsymbol{\Sigma}_{\boldsymbol{\vartheta}_0}\boldsymbol{\Delta}^{-1}\big),$$
$$\sqrt{n}(\hat{\boldsymbol{\vartheta}}_r-\boldsymbol{\vartheta}_r)\xrightarrow{L}N\big(\mathbf{0},\boldsymbol{\Delta}^{-1}\boldsymbol{\Sigma}_{\boldsymbol{\vartheta}_r}\boldsymbol{\Delta}^{-1}\big),$$

where

$$\boldsymbol{\Sigma}_{\boldsymbol{\vartheta}_0}=\frac{\mathrm{Var}(Y)}{[E(Y)]^2}\boldsymbol{\Delta}_\phi+\frac{\mathrm{Var}(\tilde{Y})}{[E(Y)]^2}\boldsymbol{\Xi}_\phi^{\otimes 2}+\frac{\boldsymbol{B}_\phi\boldsymbol{\zeta}_0\boldsymbol{e}^\tau+\boldsymbol{e}\boldsymbol{\zeta}_0^\tau\boldsymbol{B}_\phi^\tau}{E(Y)}-\frac{\boldsymbol{A}_\phi\boldsymbol{\Xi}_\phi^\tau+\boldsymbol{\Xi}_\phi\boldsymbol{A}_\phi^\tau}{[E(Y)]^2}$$
$$-\frac{\boldsymbol{\Xi}_\phi\boldsymbol{\Theta}^\tau\boldsymbol{\zeta}_0\boldsymbol{e}^\tau+\boldsymbol{e}\boldsymbol{\zeta}_0^\tau\boldsymbol{\Theta}\boldsymbol{\Xi}_\phi^\tau}{E(Y)}+\boldsymbol{e}^{\otimes 2}\zeta_0^\tau\boldsymbol{\Sigma}_{d_1}\zeta_0,$$

$$\boldsymbol{\Sigma}_{\boldsymbol{\vartheta}_r}=\frac{\mathrm{Var}(X_r)}{[E(X_r)]^2}\boldsymbol{\Delta}_{\psi_r}+\frac{\mathrm{Var}(\tilde{X}_r)}{[E(X_r)]^2}\boldsymbol{\Xi}_{\psi_r}^{\otimes 2}+\frac{\boldsymbol{B}_{\psi_r}\boldsymbol{\zeta}_r\boldsymbol{e}^\tau+\boldsymbol{e}\boldsymbol{\zeta}_r^\tau\boldsymbol{B}_{\psi_r}^\tau}{E(X_r)}$$
$$-\frac{\boldsymbol{A}_{\psi_r}\boldsymbol{\Xi}_{\psi_r}^\tau+\boldsymbol{\Xi}_{\psi_r}\boldsymbol{A}_{\psi_r}^\tau}{[E(X_r)]^2}-\frac{\boldsymbol{\Xi}_{\psi_r}\boldsymbol{\Theta}^\tau\boldsymbol{\zeta}_r\boldsymbol{e}^\tau+\boldsymbol{e}\boldsymbol{\zeta}_r^\tau\boldsymbol{\Theta}\boldsymbol{\Xi}_{\psi_r}^\tau}{E(X_r)}+\boldsymbol{e}^{\otimes 2}\zeta_r^\tau\boldsymbol{\Sigma}_{d_1}\zeta_r.$$

For the nonlinear least squares estimator $\hat{\boldsymbol{\beta}}$, we have the following asymptotic expression. Let $f'_{\boldsymbol{\beta}}(\boldsymbol{X},\boldsymbol{\beta})=(f'_1(\boldsymbol{X},\boldsymbol{\beta}),\ldots,f'_q(\boldsymbol{X},\boldsymbol{\beta}))^\tau$, $f'_r(\boldsymbol{x},\boldsymbol{\beta})=\frac{\partial f(\boldsymbol{x},\boldsymbol{\beta})}{\partial\boldsymbol{\beta}_r}$, and $\boldsymbol{\Gamma}_{\boldsymbol{\beta}}=E\{f_{\boldsymbol{\beta}}^{'\otimes 2}(\boldsymbol{X},\boldsymbol{\beta})\}$.

Theorem 2. *Under the conditions of Theorem 1, we have that*

$$\sqrt{n}(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta})$$
$$=\frac{1}{\sqrt{n}}\sum_{i=1}^n\boldsymbol{\Gamma}_{\boldsymbol{\beta}}^{-1}\varepsilon_i f'_{\boldsymbol{\beta}}(\boldsymbol{X}_i,\boldsymbol{\beta})$$
$$-\frac{1}{\sqrt{n}}\sum_{i=1}^n\boldsymbol{\Gamma}_{\boldsymbol{\beta}}^{-1}\frac{Y_i f'_{\boldsymbol{\beta}}(\boldsymbol{X}_i,\boldsymbol{\beta})}{\phi(\boldsymbol{U}_i)}\big(1,(\boldsymbol{U}_{i,d_1}-E\boldsymbol{U}_{d_1})^\tau\big)(\hat{\boldsymbol{\vartheta}}_0-\boldsymbol{\vartheta}_0)$$

$$
\begin{aligned}
&+\frac{1}{\sqrt{n}}\sum_{l=1}^{p}\sum_{i=1}^{n}\boldsymbol{\Gamma}_{\boldsymbol{\beta}}^{-1}f'_{\boldsymbol{\beta}}(\boldsymbol{X}_i,\boldsymbol{\beta})\\
&\times\frac{X_{li}f'_{x_l}(\boldsymbol{X}_i,\boldsymbol{\beta})}{\psi_l(\boldsymbol{U}_i)}\big(1,(\boldsymbol{U}_{i,d_1}-E\boldsymbol{U}_{d_1})^{\tau}\big)(\hat{\boldsymbol{\vartheta}}_l-\boldsymbol{\vartheta}_l)\\
&-\sqrt{n}\boldsymbol{\Gamma}_{\boldsymbol{\beta}}^{-1}(\bar{\boldsymbol{U}}_{d_1}-E\boldsymbol{U}_{d_1})^{\tau}\Bigg\{\boldsymbol{\zeta}_0E\bigg\{\frac{Yf'_{\boldsymbol{\beta}}(\boldsymbol{X},\boldsymbol{\beta})}{\phi(\boldsymbol{U})}\bigg\}\\
&-\sum_{l=1}^{p}\boldsymbol{\zeta}_lE\bigg\{\frac{X_lf'_{\boldsymbol{\beta}}(\boldsymbol{X},\boldsymbol{\beta})f'_{x_l}(\boldsymbol{X},\boldsymbol{\beta})}{\psi_l(\boldsymbol{U})}\bigg\}\Bigg\}\\
&+o_P(1).
\end{aligned}
\tag{2.15}
$$

If the model (1.1) is in fact a linear regression model: $Y=\boldsymbol{\beta}_{\mathrm{LS}}^{\tau}\boldsymbol{X}^o+\varepsilon$, where $\boldsymbol{X}^o=(1,\boldsymbol{X}^{\tau})^{\tau}$, $\boldsymbol{\beta}_{\mathrm{LS}}=(\boldsymbol{\beta}_0,\boldsymbol{\beta}_1,\dots,\boldsymbol{\beta}_p)^{\tau}$, $\boldsymbol{\beta}_0$ represents the intercept. The estimation equation (2.14) becomes the classical least squares estimate. In this context, denote the estimator obtained from (2.14) as $\hat{\boldsymbol{\beta}}_{\mathrm{LS}}$, we have the following asymptotic expression.

Theorem 3. *Let $\boldsymbol{\Omega}=E[\{\boldsymbol{X}^o\}^{\otimes 2}]$ and $\psi_0(\boldsymbol{U})\equiv 1$. Under the conditions of Theorem 1, we have that*

$$
\begin{aligned}
&\sqrt{n}(\hat{\boldsymbol{\beta}}_{\mathrm{LS}}-\boldsymbol{\beta}_{\mathrm{LS}})\\
&\quad=\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\boldsymbol{\Omega}^{-1}\varepsilon_i\boldsymbol{X}_i^o\\
&\quad-\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\boldsymbol{\Omega}^{-1}\frac{Y_i\boldsymbol{X}_i^o}{\phi(\boldsymbol{U}_i)}\big(1,(\boldsymbol{U}_{i,d_1}-E\boldsymbol{U}_{d_1})^{\tau}\big)(\hat{\boldsymbol{\vartheta}}_0-\boldsymbol{\vartheta}_0)\\
&\quad+\frac{1}{\sqrt{n}}\sum_{l=1}^{p}\sum_{i=1}^{n}\boldsymbol{\Omega}^{-1}\boldsymbol{X}_i^o\frac{X_{li}\boldsymbol{\beta}_l}{\psi_l(\boldsymbol{U}_i)}\big(1,(\boldsymbol{U}_{i,d_1}-E\boldsymbol{U}_{d_1})^{\tau}\big)(\hat{\boldsymbol{\vartheta}}_l-\boldsymbol{\vartheta}_l)\\
&\quad-\boldsymbol{\Omega}^{-1}\sqrt{n}(\bar{\boldsymbol{U}}_{d_1}-E\boldsymbol{U}_{d_1})^{\tau}\Bigg\{\boldsymbol{\zeta}_0E\bigg\{\frac{Y\boldsymbol{X}^o}{\phi(\boldsymbol{U})}\bigg\}-\sum_{l=0}^{p}\boldsymbol{\zeta}_lE\bigg\{\frac{X_l\boldsymbol{X}^o\boldsymbol{\beta}_l}{\psi_l(\boldsymbol{U})}\bigg\}\Bigg\}\\
&\quad+o_P(1).
\end{aligned}
\tag{2.16}
$$

2.4 Empirical likelihood based inference

We can use the asymptotic results of Theorem 1 and the asymptotic expressions (A.15) and (2.16) to present the asymptotic covariance matrices of $\hat{\boldsymbol{\beta}}$ and $\hat{\boldsymbol{\beta}}_{\mathrm{LS}}$. For example, the confidence regions and intervals for $\hat{\boldsymbol{\beta}}$ got from normal approximation is, $I_{\alpha,\mathrm{NOR}}=\{\boldsymbol{\beta}':n(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}')^{\tau}\hat{\boldsymbol{\Sigma}}_{\boldsymbol{\beta}}(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}')\le c_{\alpha}\}$, where $\hat{\boldsymbol{\Sigma}}_{\boldsymbol{\beta}}$ is the direct plug-in estimator of the asymptotic covariance matrix of $\hat{\boldsymbol{\beta}}$. As revealed by Cui et al. (2009), Li, Lin and Cui (2010), Zhang, Feng and Zhou (2014), Zhang et al. (2014), the direct plug-in estimator for $\hat{\boldsymbol{\Sigma}}_{\boldsymbol{\beta}}$ may not be appropriate, as one need to estimate so many complicated terms in the asymptotic variance and their finite-sample behaviors may not perform well. Although the estimator can be shown to be consistent under some mild assumptions. As such, the empirical likelihood (EL) method (Owen (2001), Qin and Lawless (1994)) or the bootstrap procedure (Zhang, Feng and Zhou (2014), Zhang, Li and Feng (2015), Zhang et al. (2014)) is a preferable choice.

In this paper, we suggest to use the EL method since the EL method has some attractive advantages such as: avoiding estimating asymptotic covariances, improving accuracy of coverage, implementing easily and studentizing automatically, and widely applicable. There has been many literature to discuss the EL method and its applications. We recommend the following references: Li, Lin and Zhu (2012), Lian (2012), Liang et al. (2009a), Tang and Zhao (2013a, 2013b), Wang, Li and Lin (2011),Wei and Zhu (2010), Zhang et al. (2011), Zhu et al. (2010). In the following, we make statistical inference based on the EL principle.

In previous subsections, the proposed estimation method for $\hat{\alpha}_0$, $\hat{\boldsymbol{\zeta}}_0$, $\hat{\boldsymbol{\gamma}}_{0,d-d_1}$, $\hat{\alpha}_r$, $\hat{\boldsymbol{\zeta}}_r$, $\hat{\boldsymbol{\gamma}}_{r,d-d_1}$, $r = 1, \ldots, p$ and the additive partial linear distorting functions $\hat{\phi}(\cdot)$, $\hat{\psi}_r(\cdot)$, $r = 1, \ldots, p$ are used only in the construction of the empirical likelihood based confidence interval in the following. Usually, the empirical likelihood method needs an auxiliary random vector $\wp_{n,i}(\boldsymbol{\beta}') = (\wp_{n,i}^1(\boldsymbol{\beta}'), \ldots, \wp_{n,i}^q(\boldsymbol{\beta}'))^\tau$ with the property of that $E\wp_{n,i}(\boldsymbol{\beta}') = 0$ when $\boldsymbol{\beta}' = \boldsymbol{\beta}$:

$$\wp_{n,i}^s(\boldsymbol{\beta}') = \big(Y_i - f(\boldsymbol{X}_i, \boldsymbol{\beta}')\big)\frac{\partial f(\boldsymbol{X}_i, \boldsymbol{\beta}')}{\partial \boldsymbol{\beta}_s'}.$$

Because of $\{Y_i, \boldsymbol{X}_i, i = 1, \ldots, n\}$ are unavailable, we need to estimate them by using the relationship $Y_i = \frac{\tilde{Y}_i}{\phi(\boldsymbol{U}_i)}$ and $X_{li} = \frac{\tilde{X}_{li}}{\psi_l(\boldsymbol{U}_i)}$, $l = 1, \ldots, p$. After implementing the estimation procedures proposed in Sections 2.1 and 2.2, we obtain "calibrated" variables $\hat{Y}_i = \frac{\tilde{Y}_i}{\hat{\phi}(\boldsymbol{U}_i)}$ and $\hat{X}_{li} = \frac{\tilde{X}_{li}}{\hat{\psi}_l(\boldsymbol{U}_i)}$, $l = 1, \ldots, p$.

We now define a calibrated EL principle can be applied by plugging in the estimated arguments $\{\hat{Y}_i, \hat{\boldsymbol{X}}_i\}_{i=1}^n$ into $\wp_{n,i}^s(\boldsymbol{\beta}')$:

$$\hat{l}_n(\boldsymbol{\beta}') = -2\max\left\{\sum_{i=1}^n \log(np_i) : p_i \geq 0, \sum_{i=1}^n p_i = 1, \sum_{i=1}^n p_i\hat{\wp}_{n,i}(\boldsymbol{\beta}') = 0\right\}, \qquad (2.17)$$

where $\hat{\wp}_{n,i}^s(\boldsymbol{\beta}') = (\hat{Y}_i - f(\hat{\boldsymbol{X}}_i, \boldsymbol{\beta}'))\frac{\partial f(\hat{\boldsymbol{X}}_i, \boldsymbol{\beta}')}{\partial \boldsymbol{\beta}_s'}$, $s = 1, \ldots, q$. By the Lagrange multiplier method, we have $\hat{l}_n(\boldsymbol{\beta}') = 2\sum_{i=1}^n \log\{1 + \lambda^\tau \hat{\varpi}_{n,i}(\boldsymbol{\beta}')\}$, where λ is determined by $\frac{1}{n}\sum_{i=1}^n \frac{\hat{\varpi}_{n,i}(\boldsymbol{\beta}')}{1+\lambda^\tau\hat{\varpi}_{n,i}(\boldsymbol{\beta}')} = 0$.

Theorem 4. *Suppose that conditions* (C1)–(C8) *in the Appendix hold, then* $\hat{l}_n(\boldsymbol{\beta})$ *converges to* χ_q^2, *namely, a chi-squared distribution with* q *degrees of freedom.*

Theorem 4 tells us that one can construct a confidence region of $\boldsymbol{\beta}$ by $I_{\alpha,\mathrm{EL}} = \{\boldsymbol{\beta}' : \hat{l}_n(\boldsymbol{\beta}') \le c_\alpha\}$, where c_α denotes the α quantile of the χ_q^2 distribution. As noted in Cui et al. (2009), Zhang et al. (2014), Zhang, Zhu and Liang (2012), the EL-based statistic is free of the infinite-dimensional nuisance parameters $\phi(\cdot)$, $\psi_r(\cdot)$'s. This property makes EL statistic easy to be implemented and computationally efficient.

3 Model checking

In this section, we consider the problem of model checking for the mean function $m(\boldsymbol{X}) = E(Y|\boldsymbol{X})$:

$$\mathcal{H}_0 : m(\boldsymbol{X}) = f(\boldsymbol{X}, \boldsymbol{\beta}) \qquad \text{a.s. for some } \boldsymbol{\beta}. \tag{3.1}$$

For testing (3.1), if $(\boldsymbol{X}, Y)$ can be observed directly, we can follow one of these literature: Eubank and Spiegelman (1990) for the linear models, Härdle and Mammen (1993) for comparing parametric and nonparametric fit, Stute, Thies and Zhu (1998) for a more general regression testing, and Hart (1997) for the various goodness-of-fit tests based on smoothing methods.

Our testing procedure proposed here is motivated by Stute, González Manteiga and Presedo Quindimil (1998a): an optimal test should be based on the empirical process of the regressors marked by the residuals. As $(\boldsymbol{X}_i, Y_i)$ cannot be obtained directly, instead, the estimated $(\hat{\boldsymbol{X}}_i, \hat{Y}_i)$ can be remitted. Now, we propose the empirical process:

$$\mathcal{T}_n(x_1, \dots, x_p) = \frac{1}{\sqrt{n}} \sum_{i=1}^n I\{\hat{X}_{i1} \le x_1, \dots, \hat{X}_{ip} \le x_p\} \hat{\varepsilon}_i, \tag{3.2}$$

where $\hat{\varepsilon}_i = \hat{Y}_i - f(\hat{\boldsymbol{X}}_i, \hat{\boldsymbol{\beta}})$, $i = 1, \dots, n$ and $I\{\cdot\}$ is the indicator function. Based on (3.2), the Cramér-von Mises (CvM) statistic is proposed by

$$\mathcal{Z}_n^2 = \frac{1}{n} \sum_{i=1}^n \mathcal{T}_n^2(\hat{X}_{i1}, \dots, \hat{X}_{ip}). \tag{3.3}$$

This test procedure (3.3) is easy to implement. The null hypothesis $\mathcal{H}_0$ is rejected for large values of $\mathcal{Z}_n^2$. To define its p-values, we propose a wild bootstrap technique to mimic the null distribution of $\mathcal{Z}_n^2$.

Step 1: Compute the test statistic $\mathcal{Z}_n$ from (3.3).

Step 2: Generate B times i.i.d. variables e_{ib}, $i = 1, \dots, n$, $b = 1, \dots, B$ with a two-point distribution which respectively takes values $\frac{1\mp\sqrt{5}}{2}$ with probability $\frac{5\pm\sqrt{5}}{10}$. For each b, compute the arguments $\{\hat{\varepsilon}_1 e_{1b}, \hat{\varepsilon}_2 e_{2b}, \dots, \hat{\varepsilon}_n e_{nb}\}$ and

$$\hat{Y}_i^b = f(\hat{\boldsymbol{X}}_i, \hat{\boldsymbol{\beta}}) + \hat{\varepsilon}_i e_{ib}, \qquad i = 1, \dots, n.$$

Step 3: For each b, we re-calculate the bootstrap nonlinear least squares estimator $\hat{\boldsymbol{\beta}}^b$ based on $(\hat{Y}_i^b, \hat{\boldsymbol{X}}_i)$, and the bootstrap fitted value $f(\hat{\boldsymbol{X}}_i, \hat{\boldsymbol{\beta}}^b)$ and residuals $\hat{\varepsilon}_i^b = \hat{Y}_i^b - f(\hat{\boldsymbol{X}}_i, \hat{\boldsymbol{\beta}}^b)$. Then, we define the bootstrap test statistic:

$$\mathcal{T}_{nb}(x_1, \dots, x_p) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} I\{\hat{X}_{i1} \le x_1, \dots, \hat{X}_{ip} \le x_p\} \hat{\varepsilon}_i^b,$$

$$\mathcal{Z}_{nb}^2 \stackrel{\text{def}}{=} \frac{1}{n} \sum_{i=1}^{n} \mathcal{T}_{nb}^2(\hat{X}_{i1}, \dots, \hat{X}_{ip}).$$

Step 4: We calculate the $1-\kappa$ quantile of the bootstrap test statistic $\mathcal{Z}_{nb}^2$ as the κ-level critical value.

We use the wild bootstrap method to define p-values. There are a number of literature on the wild bootstrap procedure, for example, Wu (1986), Stute, González Manteiga and Presedo Quindimil (1998a), Xia et al. (2004), Escanciano (2006). Especially for the parametric model check, the wild bootstrap method performs well in general, which is shown in Stute, González Manteiga and Presedo Quindimil (1998a). It is noted that the wild bootstrap samples e_{ib} should satisfy $E^*(e_{ib}) = 0$, $\text{Var}^*(e_{ib}) = 1$ and $|e_{ib}| \le c_1 \infty$ for some positive constant c_1. One can easily show that $E^*(\hat{\varepsilon}_i e_{ib}) = 0$ and $\text{Var}^*(\hat{\varepsilon}_i e_{ib}) = \hat{\varepsilon}_i^2$, where the $E^*(\cdot)$ and $\text{Var}^*(\cdot)$ denote the expectation and variance carrying the bootstrap samples e_{ib}'s. The two-point distribution $\frac{1\mp\sqrt{5}}{2}$ for e_{ib} attaching masses $\frac{5\pm\sqrt{5}}{10}$ is suggested by Mammen (1993). Note that $E^*(\xi_{ib}^3) = 1$, and this extra third moment condition can generally improve the performance of wild bootstrap. See more details in Stute, González Manteiga and Presedo Quindimil (1998b), Zhang, Li and Feng (2015), Zhang et al. (2014).

4 Simulation studies

In this section, we present simulation results to evaluate the performance of the proposed estimation and testing procedures. In Example 1, we consider the estimation under the additive partial linear distortion measurement error settings. In Example 2, we conduct a simulation to evaluate model checking problem. Cubic B-splines were used to approximate the additive distortion functions as described in Section 2. The number of knots in the approximation was selected by BIC criteria (Zhang and Liang (2011, Section 4)). The range of J_n is selected from a neighborhood of $\lfloor n^{\frac{1}{5.5}} \rfloor$($\lfloor a \rfloor$ stands for the smallest integer not less than number a). The optimal knot number, N_n, is the minimizer of the following BIC values:

$$N_n = \arg \min_{N'_n \in [\lfloor n^{\frac{1}{5.5}} \rfloor, 4\lfloor n^{\frac{1}{5.5}} \rfloor]} \{-\mathrm{Re}(N'_n) + q_n \log n\},$$

where $\mathrm{Re}(N'_n)$ is the residuals obtained from (2.8), q_n is the total number of parameter in the mean regression model after the spline approximations. It is noted that the nonlinear function (4.1) in the Example 1 in this section and nonlinear function (5.1) in the following Section 5 is a special case of $Y = \boldsymbol{\lambda}_1 + X_1(\boldsymbol{\lambda}_2 + \boldsymbol{\lambda}_3 X_2)^{\boldsymbol{\lambda}_4} + \varepsilon$. For example, in the nonlinear function (4.1) of the Example 1 in this section, we let $\boldsymbol{\lambda}_2 = \boldsymbol{\lambda}_3 = 1$ and set $\boldsymbol{\lambda}_1 = \boldsymbol{\beta}_1$ and $\boldsymbol{\lambda}_4 = \boldsymbol{\beta}_2$; in the nonlinear function (5.1) in Section 5, we set $X_1 \equiv 1$, $X_2 = \mathrm{GA}$ and let $(\boldsymbol{\lambda}_1, \boldsymbol{\lambda}_2, \boldsymbol{\lambda}_3, \boldsymbol{\lambda}_4) = (\boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \boldsymbol{\beta}_3, \boldsymbol{\beta}_4)$. The reason of using model $Y = \boldsymbol{\lambda}_1 + X_1(\boldsymbol{\lambda}_2 + \boldsymbol{\lambda}_3 X_2)^{\boldsymbol{\lambda}_4} + \varepsilon$ is based on our experience for the simulation studies and real data analysis. The nonlinear function (5.1) in Section 5 is also be used in Zhang et al. (2014).

Example 1. We generate 500 data sets from the models:

$$Y = \boldsymbol{\beta}_1 + X_1(1 + X_2)^{\boldsymbol{\beta}_2} + \varepsilon, \tag{4.1}$$

$$Y = \boldsymbol{\beta}_0 + \boldsymbol{\beta}_1 X_1 + \boldsymbol{\beta}_2 X_2 + \boldsymbol{\beta}_3 X_3 + \varepsilon. \tag{4.2}$$

- For nonlinear model (4.1), we set $(\boldsymbol{\beta}_1, \boldsymbol{\beta}_2) = (1, -0.5)$, the model error ε follows $N(0, 0.5^2)$ and is independent with $(X_1, X_2)^\tau$. The predictors $(X_1, X_2)^\tau$ follow $N_2(\mu_X, \Sigma)$ with $\mu_X = (2, 10)^\tau$ and $\Sigma = (\sigma_{ij})_{1 \le i,j \le 2}, \sigma_{ij} = 0.5^{|i-j|}$. For the distorting functions ϕ, ψ_r's, we use (4.3), (4.4)–(4.5) for Y, X_1 and X_2, respectively. The sample size is chosen as $n = 200, 400, 600$.
- For linear model (4.2), we set $(\boldsymbol{\beta}_0, \boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \boldsymbol{\beta}_3) = (1, 0.1, 1, -0.1)$, $X_1 \sim N(2, 1.2^2)$, $X_2 \sim N(0.5, 0.5^2)$ and $X_3 \sim N(1, 1)$. For the distorting functions ϕ, ψ_r's, we use (4.3), (4.4)–(4.6) for Y, X_1, X_2 and X_3. The sample size is chosen as $n = 300, 400, 600$.

The additive partial linear distorting functions are designed as

$$\phi(\boldsymbol{U}) = 1 + 0.2\cos(2\pi U_1) + 0.2(U_2 - 0.5), \tag{4.3}$$

$$\psi_1(\boldsymbol{U}) = 1 + 0.2\big(\exp(-6.5U_1) - 0.1536\big) + 0.25(U_2 - 0.5), \tag{4.4}$$

$$\psi_2(\boldsymbol{U}) = 1 + 0.2\big(\exp(-3.5U_1) - 0.2771\big) - 0.30(U_2 - 0.5), \tag{4.5}$$

$$\psi_3(\boldsymbol{U}) = 1 + 0.1\big(\exp(-9.5U_1) - 0.1052\big) + 0.05(U_2 - 0.5). \tag{4.6}$$

Distorting functions (4.3), (4.4)–(4.5) for Y, X_1 and X_2 are used for nonlinear model (4.1), and (4.3), (4.4)–(4.6) for Y, X_1, X_2 and X_3 are used for linear model

Table 1 *The estimated mean and associated standard error for nonlinear model (4.1)*

		α_0	ζ_0	α_1	ζ_1	α_2	ζ_2
$n = 200$	Bias	0.0899	0.0001	0.0489	0.0003	−0.0059	0.0012
	SD	0.1686	0.0835	0.2173	0.1148	0.0442	0.0238
$n = 400$	Bias	0.0862	0.0021	0.0338	0.0037	0.0085	0.0004
	SD	0.1196	0.0637	0.1655	0.0823	0.0342	0.0170
$n = 600$	Bias	0.0875	−0.0014	0.0446	−0.0022	−0.0074	−0.0007
	SD	0.0984	0.0507	0.1330	0.0706	0.0281	0.0139

(4.2). Confounding variables U_1 and U_2 are independently uniformly distributed on [0, 1].

Simulation results are reported in Tables 1–4. In Table 1 and Table 3, we report the simulation results for estimators of $\boldsymbol{\zeta}_0$, $\boldsymbol{\zeta}_1$, $\boldsymbol{\zeta}_2$ and $\boldsymbol{\zeta}_3$ involved in additive partial linear distorting functions (4.3)–(4.6). When sample size n increases, the performance of those estimators of $\boldsymbol{\zeta}_0$, $\boldsymbol{\zeta}_1$, $\boldsymbol{\zeta}_2$ and $\boldsymbol{\zeta}_3$ gets better. From Table 2 and Table 4, we can see that the estimated values of $\boldsymbol{\beta}$ in the partial linear additive distorting setting are also close to the true values in both nonlinear model (4.1) and linear model (4.2). For the coverage probabilities and confidence intervals, we also compare the EL approach and normal approximation (NA) approach by directly estimating the asymptotic covariance matrices. Both in nonlinear model (4.1) and linear model (4.2), the confidence intervals based on EL are uniformly better than NA, and the average lengths of confidence intervals based on EL have smaller values than NA. For nonlinear model (4.1), the coverage probabilities are close to nominal level for both EL and NA. For linear model (4.2), the EL has a better coverage probabilities than NA. We also conduct one simulation run to construct the confidence regions of $(\boldsymbol{\beta}_1, \boldsymbol{\beta}_2)$ for nonlinear model (4.1) based on EL approach, and delineate them in Figure 1 for illustration.

Example 2. In this example, we generated 500 samples of size $n = 300$ of data from models (4.7) and (4.8):

$$Y = \boldsymbol{\beta}_1 + X_1(1 + X_2)^{\boldsymbol{\beta}_2} + C_o X_1 X_2 + \varepsilon, \tag{4.7}$$

$$Y = \boldsymbol{\beta}_0 + \boldsymbol{\beta}_1 X_1 + \boldsymbol{\beta}_2 X_2 + \boldsymbol{\beta}_3 X_3 + C_o\left(X_1^2 + X_2^2 + X_3^2\right) + \varepsilon. \tag{4.8}$$

The null model holds if and only if $C_o = 0$, in other words, we aim to test whether the data are from model (4.1) if $C_o = 0$ for model (4.7), and test whether the data are from model (4.2) if $C_o = 0$ for model (4.8). In each case $B = 500$ bootstrap samples were generated and corresponding values of $\mathcal{Z}_n^2$ and $\mathcal{Z}_{nb}^2$ were calculated. The model settings for $\boldsymbol{X}$, $\boldsymbol{U}$, ε and ϕ, ψ's are the same as Example 1. The critical levels were $\kappa = 0.01, 0.025, 0.05, 0.10$. In Table 5, Figure 2, we present the

Table 2 *The estimated mean and associated standard error for nonlinear model* (4.1)

		Bias	SD	EL				NA			
				Lower	Upper	AL	Coverage	Lower	Upper	AL	Coverage
$n = 200$	$\boldsymbol{\beta}_1$	0.0113	0.0805	0.9504	1.0494	0.0990	93.6%	0.9421	1.0581	0.1160	93.8%
	$\boldsymbol{\beta}_2$	−0.0103	0.0509	−0.5165	−0.4835	0.0330		−0.5192	−0.4806	0.0386	
$n = 400$	$\boldsymbol{\beta}_1$	0.0055	0.0582	0.9641	1.0353	0.0712	95.4%	0.9593	1.0414	0.0821	94.4%
	$\boldsymbol{\beta}_2$	−0.0055	0.0358	−0.5119	−0.4882	0.0237		−0.5135	−0.4862	0.0273	
$n = 600$	$\boldsymbol{\beta}_1$	0.0043	0.0437	0.9701	1.0277	0.0576	94.8%	0.9655	1.0330	0.0675	95.4%
	$\boldsymbol{\beta}_2$	−0.0049	0.0280	−0.5100	−0.4906	0.0194		−0.5115	−0.4890	0.0225	

Table 3 *The estimated mean and associated standard error for linear model* (4.2)

		α_0	ζ_0	α_1	ζ_1	α_2	ζ_2	α_3	ζ_3
$n = 300$	Bias	0.0790	−0.0088	0.0308	0.0064	−0.0154	−0.0109	0.0523	0.0204
	SD	0.2718	0.1407	0.2254	0.1201	0.3737	0.1966	0.3465	0.1888
$n = 400$	Bias	0.0895	−0.0079	0.0407	−0.0013	−0.0195	0.0083	0.0693	−0.0020
	SD	0.2420	0.1325	0.1998	0.1105	0.3212	0.1771	0.3155	0.1757
$n = 600$	Bias	0.0922	−0.0049	0.0412	−0.0035	−0.0216	−0.0018	0.0707	−0.0046
	SD	0.1842	0.0989	0.1555	0.0817	0.2518	0.1502	0.2466	0.1449

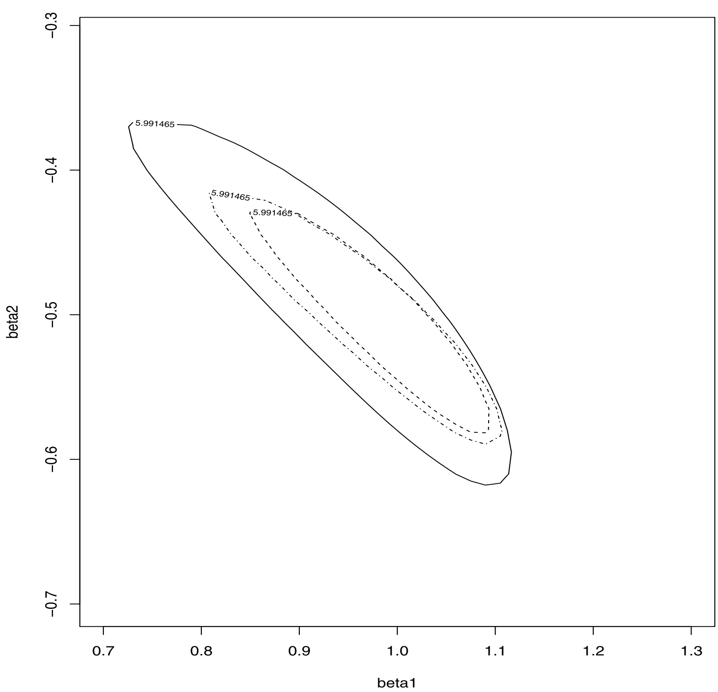

Figure 1 *Empirical likelihood confidence regions for nonlinear model* (4.1) *in the case of partial linear additive distorting setting.* $n = 200$ (*solid lines*), $n = 400$ (*dashed lines*) *and* $n = 600$ (*dotted lines*).

simulation results for the power calculations. We can see that when $C_o = 0$, all empirical levels obtained by these two distorting settings are close to the four nominal levels, which indicates that the bootstrap method gives proper Type I errors. As C_o increases, the power functions increases rapidly. This indicates that the proposed bootstrap test under the distorting measurement error setting works well.

Table 4 *The estimated mean and associated standard error for linear model* (4.2)

		Bias	SD	EL				NA			
				Lower	Upper	AL	Coverage	Lower	Upper	AL	Coverage
$n = 300$	$\boldsymbol{\beta}_0$	0.0312	0.1416	0.9688	1.0314	0.0626	93.0%	0.9684	1.0329	0.0645	91.8%
	$\boldsymbol{\beta}_1$	−0.0046	0.0522	0.0875	0.1126	0.0251		0.0874	0.1131	0.0257	
	$\boldsymbol{\beta}_2$	−0.0399	0.1461	0.9377	1.0628	0.1251		0.9368	1.0657	0.1289	
	$\boldsymbol{\beta}_3$	0.0007	0.0569	−0.1374	−0.0623	0.0750		−0.1379	−0.0606	0.0773	
$n = 400$	$\boldsymbol{\beta}_0$	0.0385	0.1247	0.9720	1.0277	0.0557	94.2%	0.9701	1.0289	0.0588	93.0%
	$\boldsymbol{\beta}_1$	−0.0021	0.0440	0.0888	0.1112	0.0224		0.0884	0.1116	0.0232	
	$\boldsymbol{\beta}_2$	−0.0200	0.1069	0.9441	1.0553	0.1112		0.9419	1.0578	0.1159	
	$\boldsymbol{\beta}_3$	0.0021	0.0506	−0.1335	−0.0669	0.0666		−0.1349	−0.0653	0.0696	
$n = 600$	$\boldsymbol{\beta}_0$	0.0056	0.1041	0.9770	1.0221	0.0451	94.4%	0.9750	1.0235	0.0485	94.2%
	$\boldsymbol{\beta}_1$	−0.0006	0.0358	0.0908	0.1088	0.0180		0.0900	0.1095	0.0195	
	$\boldsymbol{\beta}_2$	−0.0170	0.0874	0.9540	1.0443	0.0903		0.9499	1.0473	0.0974	
	$\boldsymbol{\beta}_3$	0.0037	0.0395	−0.1276	−0.0734	0.0542		−0.1300	−0.0716	0.0584	

Table 5 *The estimated mean and associated standard error for nonlinear model* (4.7) *and linear model* (4.8)

		$\kappa = 0.01$	$\kappa = 0.025$	$\kappa = 0.05$	$\kappa = 0.10$
Model (4.7)	$C_o = 0.00$	0.010	0.030	0.057	0.110
	$C_o = 0.05$	0.200	0.307	0.421	0.536
	$C_o = 0.10$	0.636	0.742	0.824	0.894
	$C_o = 0.15$	0.816	0.900	0.950	0.978
	$C_o = 0.20$	0.956	0.982	0.992	1.000
	$C_o = 0.25$	0.966	0.986	0.996	1.000
	$C_o = 0.30$	0.988	0.998	1.000	1.000
	$C_o = 0.35$	0.992	1.000	1.000	1.000
Model (4.8)	$C_o = 0.00$	0.012	0.028	0.042	0.102
	$C_o = 0.05$	0.028	0.054	0.124	0.228
	$C_o = 0.10$	0.166	0.274	0.374	0.516
	$C_o = 0.15$	0.380	0.564	0.714	0.840
	$C_o = 0.20$	0.736	0.860	0.934	0.974
	$C_o = 0.25$	0.898	0.960	0.982	0.996
	$C_o = 0.30$	0.974	0.992	0.998	0.999
	$C_o = 0.35$	0.984	0.994	1.000	1.000

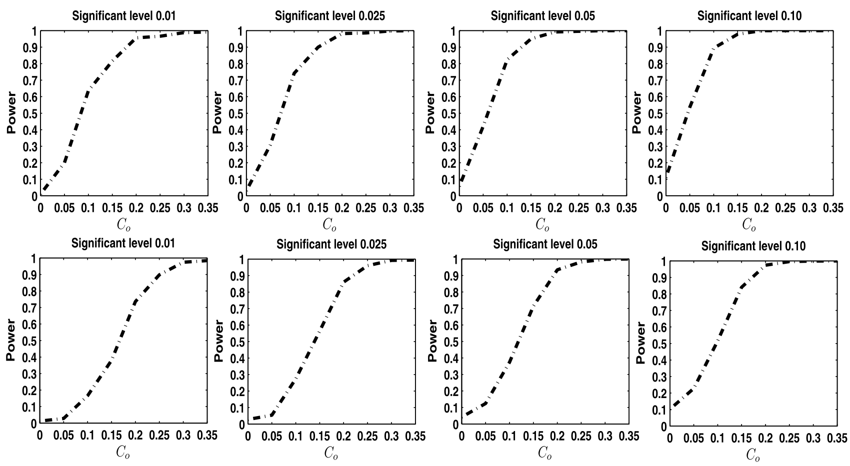

Figure 2 *Power plots* (*above four panels*) *for nonlinear model* (4.7), *and power plots* (*below four panels*) *for linear model* (4.8).

5 Real data analysis

In this section, we re-analyze a dataset on low birth infants weight. These infants with very low (<1600 grams) birth weight (BW) from 1981–1987 were collected at Duke University Medical Center by Dr. Michael O'Shea (Bowman Gray Medical Center). Our interest is the relationship between the response "birth weight"

(BW) and the predictor "gestational age" (GA). The covariates "lowest pH in first four days of life" (Lowest.pH) and "the number of platelet count" (Platelet.count) may affect BW and GA. Therefore, we consider Lowest.pH and Platelet.count as confounding variables in this data analysis.

We first explore that partial linear additive distortion model is more proper for this data set, compared with Zhang et al. (2014). The test procedure proposed by Chen, Liang and Wang (2011) is adopted. For the distorting function $\phi(\cdot)$ of response BW, the covariate Lowest.pH is linear and the covariate Platelet.count is nonlinear, and the estimated parameter for Lowest. pH is $\boldsymbol{\zeta}_0 = 0.5617$. While for the distorting function $\psi(\cdot)$ of predictor GA, the covariate Lowest.pH is nonlinear and the covariate Platelet.count is linear, and the estimated parameter for Platelet.count is $\boldsymbol{\zeta}_1 = -3.4186 \times 10^{-5}$. These test results also coincides with the plots for $\phi(\cdot)$ and $\psi(\cdot)$ presented in Zhang et al. (2014).

We then obtain synthesis data $\hat{X}_i$ and $\hat{Y}_i$ using (2.13) and substitute them in (2.14) to obtain estimated values of BW and GA. These intermediate estimated values are displayed in Figure 3, in which we depict the local linear smoothing curve (thin solid line) and the 95% pointwise confidence band. A linear regres-

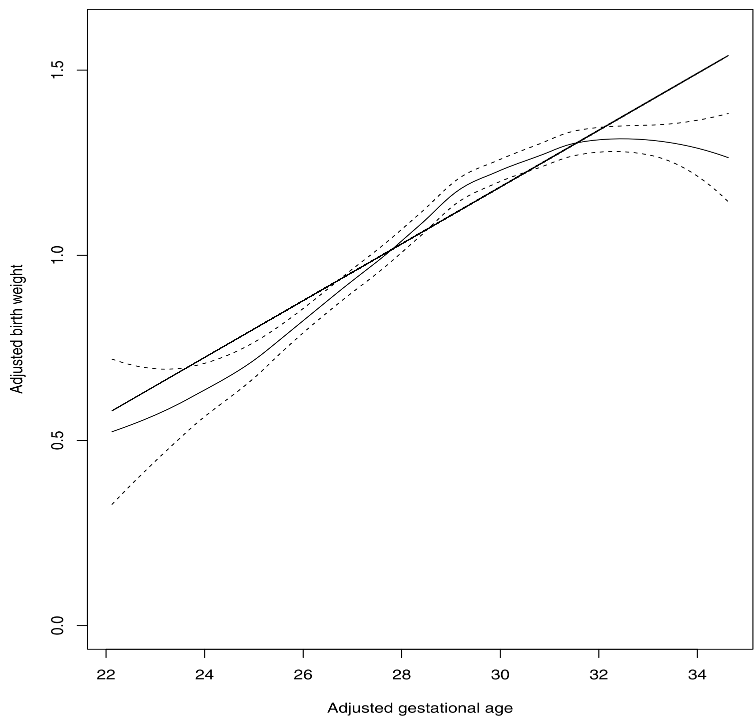

Figure 3 *The local linear estimators (thin solid line) of adjusted birth weight (unit kg) against adjusted gestational age (unit week), along with the 95% pointwise confidence intervals (dotted lines) and a linear fitting (straight line).*

sion is also fitted for this dataset (straight line in Figure 3). Figure 3 shows that the straight line is not encapsulated in the 95% pointwise confidence band. Moreover, we also use the Cramér-von Mises (CvM) statistics proposed in Section 3 to test whether a linear model is appropriate for this dataset. The number of bootstrap replications is 1000. The 99%, 97.5%, 95%, 90% quantiles of $\mathcal{Z}_{nb}^2$ is 0.0085, 0.0061, 0.0052 and 0.0041, and the test statistic for the linear model $\mathcal{Z}_n^2$ is 0.0335, which indicates that the linear model is not appropriate for this data set. As a consequence, a nonlinear model is a proper choice for this dataset. In what follows, we used the following nonlinear model:

$$\mathrm{BW} = \boldsymbol{\beta}_1 + (\boldsymbol{\beta}_2 + \boldsymbol{\beta}_3 \mathrm{GA})^{\boldsymbol{\beta}_4} + \varepsilon. \tag{5.1}$$

CvM statistics aims to testing the goodness of fit for the given model (5.1), and AIC or BIC aim to estimate the quality of each model (linear or nonlinear model), relative to each of the other models. Then, we also calculate the AIC and BIC criteria for the linear model and the nonlinear model (5.1). The AIC and BIC for the linear model are obtained as $\mathrm{AIC_L} = -2055.089$ and $\mathrm{BIC_L} = -2046.342$, and The AIC and BIC for the nonlinear model (5.1) are obtained as $\mathrm{AIC_{NL}} = -2725.2737$ and $\mathrm{BIC_{NL}} = -2707.7804$. Both the $\mathrm{AIC_{NL}}$ and $\mathrm{BIC_{NL}}$ indicate that the nonlinear model (5.1) is a better choice than the linear model. The estimates are $(\hat{\boldsymbol{\beta}}_1, \hat{\boldsymbol{\beta}}_2, \hat{\boldsymbol{\beta}}_3, \hat{\boldsymbol{\beta}}_4) = (-0.2439, -12.3003, 0.5248, 0.2976)$. The corresponding empirical likelihood intervals are as $(-0.2588, -0.2315)$, $(-12.3947, -12.2196)$, $(0.5215, 0.5277)$ and $(0.2871, 0.3062)$, respectively. The marginal EL-based confidence intervals are calculated using (5.1) by treating the estimated values of the remaining parameters as the true values. We also apply the Cramér-von Mises (CvM) statistic to check whether model (5.1) is appropriate or not. Based on 1000 bootstrap samples, the 99%, 97.5%, 95%, 90% quantiles of $\mathcal{Z}_{nb}^2$ is 0.7946, 0.4944, 0.3250 and 0.1644, and the test statistics $\mathcal{Z}_n^2$ for the nonlinear model (5.1) is 0.0195, which indicates that the nonlinear model (5.1) is appropriate for this dataset. In Figure 4, the fitted nonlinear curve along with 95% pointwise confidence intervals is displayed. It also shows a nonlinear pattern between BW and GA. We also fitted model (5.1) for the original data. The square-root of the sum square residual error based on the naive method is 226, while the square-root of the sum of the residual square error based on the adjusted method is 168. Therefore, the confounding variables do have a substantial impact on the improvement of model fitting in this dataset.

Lastly, we compare our partial linear distortion estimation procedure with the single-index distortion estimation procedure proposed by Zhang, Zhu and Liang (2012). After we modeled the distorting functions $\phi(\boldsymbol{U}) = \phi(\theta_1 \mathrm{Lowest.pH} + \theta_2 \mathrm{Platelet.count})$ and $\psi(\boldsymbol{U}) = \psi(\theta_1 \mathrm{Lowest.pH} + \theta_2 \mathrm{Platelet.count})$, the resulting estimator of (θ_1, θ_2) is obtained as $(0.9998, 0.0191)$. For model (5.1), the esti-

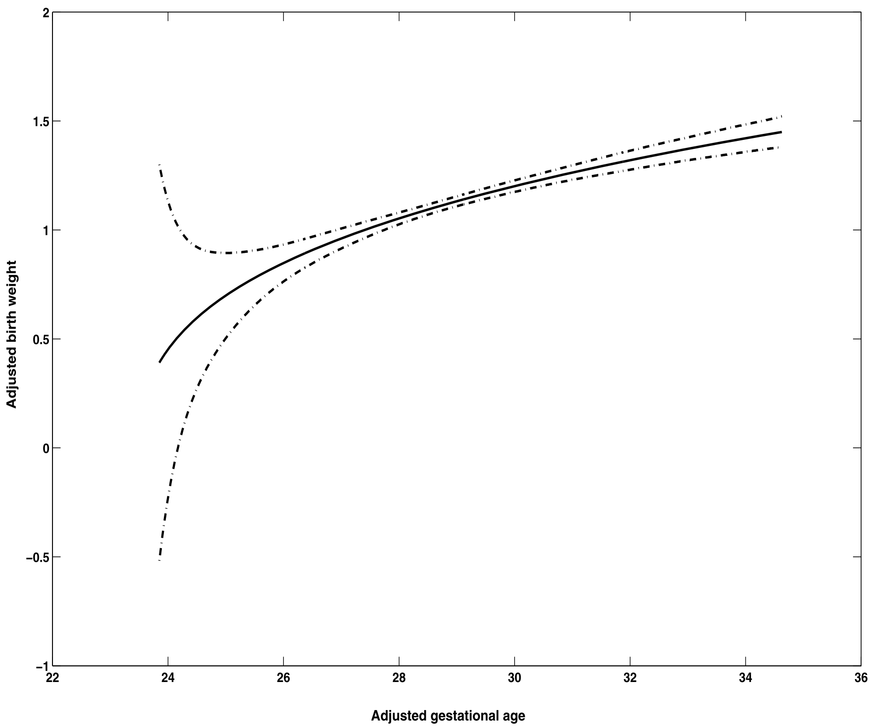

Figure 4 *The estimated curve of adjusted birth weigh (unit kg) against adjusted gestational age (unit week) and the associated* 95% *pointwise confidence intervals (dotted lines).*

mates under the single-index distortion setting are obtained as $(\hat{\boldsymbol{\beta}}_1, \hat{\boldsymbol{\beta}}_2, \hat{\boldsymbol{\beta}}_3, \hat{\boldsymbol{\beta}}_4) = (-0.4164, -13.3984, 0.5999, 0.3096)$. The corresponding empirical likelihood intervals are $(-0.4206, -0.3942)$, $(-13.4212, -13.2150)$, $(0.5988, 0.6061)$ and $(0.3076, 0.3201)$, respectively. The square-root of the sum of the residual square error based on the single-index distortion estimation procedure is 169, a slightly larger than that obtained from additive partial linear distortion estimation procedure. Compare with the average lengths between additive partial linear distortion and single-index distortion, the EL-based confidence intervals of $\boldsymbol{\beta}_2$, $\boldsymbol{\beta}_3$ based on additive partial linear distortion are smaller than those obtained by single-index distortion, while confidence intervals for $\boldsymbol{\beta}_1$, $\boldsymbol{\beta}_4$ are wider than those obtained by single-index distortion. However, all the four confidence intervals exclude 0 whether we consider partial linear additive distortion or single-index distortion. Based on above analysis, we can see that this additive partial linear distortion modeling is also a strong competitor to vie with the single-index distortion in this data analysis.

Appendix

A.1 Conditions

Let $\|\cdot\|_2$ be the Euclidean norm and $\|f\|_\infty = \sup_t |f(t)|$ be the supremum norm of a function f on [0, 1]. Let ν be a positive integer and $\iota \in (0, 1]$ such that $\upsilon = \nu + \iota > 2$. Let $\mathcal{F}$ be the collection of functions f on [0, 1] whose νth derivative, $f^{(\nu)}$ exists and satisfies the Lipschitz condition of order ν:

$$\left|f^{(\nu)}(t_1) - f^{(\nu)}(t_2)\right| \le C_o|t_1 - t_2|^\iota, \qquad 0 \le t_1, t_2 \le 1,$$

where C_o is a positive constant. We now list the conditions needed in the proofs of the theorems and corollaries.

(C1) Distorting functions $\phi_s \in \mathcal{F}$, $\psi_{rs} \in \mathcal{F}$, $r = 1, \ldots, p$, $s = d_1 + 1, \ldots, d$.
(C2) The absolute values of ϕ and ψ_r's are greater than a positive constant on the support of $\boldsymbol{U}$.
(C3) $E\varepsilon = 0$ and $E\varepsilon^4 < \infty$, and the covariance matrix of $\boldsymbol{X}$ is positive definite and finite.
(C4) For $l_1, l_2, l_3, l_4 = 0, 1, 2$, $l_1 + l_2 + l_3 + l_4 \le 3$, $1 \le s_1, s_2 \le p$, $1 \le t_1, t_2 \le q$ and $\boldsymbol{\beta}' \in \Theta_{\boldsymbol{\beta}}$, the partial derivatives

$$\frac{\partial^{l_1+l_2+l_3+l_4} f(\boldsymbol{x}, \boldsymbol{\beta}')}{\partial^{l_1}\beta'_{t_1}\partial^{l_2}\beta'_{t_2}\partial^{l_3}x_{s_1}\partial^{l_4}x_{s_2}}$$

exist, and

$$\left|\frac{\partial^{l_1+l_2+l_3+l_4} f(\boldsymbol{x}, \boldsymbol{\beta}')}{\partial^{l_1}\beta'_{t_1}\partial^{l_2}\beta'_{t_2}\partial^{l_3}x_{s_1}\partial^{l_4}x_{s_2}}\right| \le C, \qquad \text{when } l_3 + l_4 \ge 1,$$

for some positive constant C and

$$\mathbf{E}\left\{\sup_{\boldsymbol{\beta}'}\left|\frac{\partial^{l_1+l_2+l_3+l_4} f(\boldsymbol{x}, \boldsymbol{\beta}')}{\partial^{l_1}\beta'_{t_1}\partial^{l_2}\beta'_{t_2}\partial^{l_3}x_{s_1}\partial^{l_4}x_{s_2}}\bigg|_{\boldsymbol{x}=\boldsymbol{X}}\right|\right\} < \infty,$$

when $1 \le l_1 + l_2 \le 2$, and $l_3 + l_4 = 0$.

(C5) $\mathbf{E}[f(\boldsymbol{X}, \boldsymbol{\beta}') - f(\boldsymbol{X}, \boldsymbol{\beta})]^2$ admits one unique minimum at $\boldsymbol{\beta}' = \boldsymbol{\beta}$.
(C6) The number of interior knots J_n satisfies: $n^{1/(2\upsilon)} \ll J_n \ll n^{1/3}$.
(C7) Let $\boldsymbol{Z} = (1, \boldsymbol{U}_{d_1}\boldsymbol{B}_{d-d_1}(\boldsymbol{U}_{d-d_1})^\tau)^\tau$, $E\boldsymbol{Z}^{\otimes 2}$ exists and is nonsingular. The largest and smallest eigenvalues of $E\boldsymbol{Z}^{\otimes 2}$ are bounded above and below by a finite positive constant.
(C8) $\mathbf{E}Y$ and $\mathbf{E}(X_r)$, $r = 1, \ldots, p$ are bounded away from 0.

A.2 Proofs of the main results

In the following, we present the proofs of Theorem 1 and Theorem 2. The proofs of Theorem 3 and Theorem 4 are similar to the Corollary 1 and Theorem 2 in Zhang et al. (2014), so we omit the details.

Proof of Theorem 1. The proof of Theorem 1 is similar to the proof of Theorem 3 in Liu, Wang and Liang (2011). We first consider the distorting functions for $\phi(\boldsymbol{U})$ and define a class of functions

$$\mathcal{M}_n = \{m(\boldsymbol{U}, \alpha_0, \boldsymbol{\zeta}_0) = \alpha_0 + \boldsymbol{\zeta}_0^\tau(\boldsymbol{U}_{d_1} - \bar{\boldsymbol{U}}_{d_1}) + g(\boldsymbol{U}_{d-d_i}), g(\cdot) \in \mathcal{Q}_n\}. \quad \text{(A.1)}$$

We define $\hat{m}_i = \hat{\alpha}_0 + \hat{\boldsymbol{\zeta}}_0^\tau(\boldsymbol{U}_{i,d_1} - \bar{\boldsymbol{U}}_{d_1}) + \hat{\boldsymbol{\gamma}}_{0,d_1}^\tau \boldsymbol{B}_{d-d_1}(\boldsymbol{U}_{i,d-d_1})$, $\hat{m} = \hat{\alpha}_0 + \hat{\boldsymbol{\zeta}}_0^\tau(\boldsymbol{U}_{d_1} - \bar{\boldsymbol{U}}_{d_1}) + \hat{\boldsymbol{\gamma}}_{0,d_1}^\tau \boldsymbol{B}_{d-d_1}(\boldsymbol{U}_{d-d_1})$ and $\hat{m}_{\mathbf{v}} = \hat{m} + (1, (\boldsymbol{U}_{d_1}^\tau - \breve{\boldsymbol{\Upsilon}}^\tau(\boldsymbol{U}_{d-d_1}))^\tau \mathbf{v}$ for any $\mathbf{v} \in \mathbb{R}^{d_1+1}$, where $\breve{\boldsymbol{\Upsilon}}(\boldsymbol{u}_{d-d_1}) = \sum_{s=d_1+1}^d \breve{\Upsilon}_s(u_s)$ such that $\breve{\Upsilon}_s(u_s) \in \mathcal{Q}_n$ and $\|\breve{\Upsilon}_s - \Upsilon_s\|_\infty = O(h^v)$. Moreover, note that $\hat{m}_{\mathbf{v}}$ minimizes $\ell(m) = \sum_{i=1}^n \{\tilde{Y}_i/\bar{\tilde{Y}} - m(\boldsymbol{U}_i, \alpha_0, \boldsymbol{\zeta}_0)\}^2$ for all $m \in \mathcal{M}_n$ when $\mathbf{v} = 0$. As such, $\frac{\partial}{\partial \mathbf{v}} \ell(\hat{m}_{\mathbf{v}})|_{\mathbf{v}=0} = 0$, and we obtain that

$$\begin{aligned} 0 &= -\sum_{i=1}^n (\tilde{Y}_i/\bar{\tilde{Y}} - \hat{m}_i)(1, (\boldsymbol{U}_{d_1}^\tau - \breve{\boldsymbol{\Upsilon}}^\tau(\boldsymbol{U}_{d-d_1}))) \\ &= -\sum_{i=1}^n (\tilde{Y}_i/\bar{\tilde{Y}} - \hat{m}_i)(1, \breve{\boldsymbol{U}}_{i,d_1}^\tau)^\tau + O_P(nh^v). \end{aligned} \quad \text{(A.2)}$$

Note that

$$\begin{aligned} \tilde{Y}_i/\bar{\tilde{Y}} - \hat{m}_i &= \{\tilde{Y}_i\alpha_0 - \bar{\tilde{Y}}\hat{\alpha}_0\}/\bar{\tilde{Y}} + \{Y_i\boldsymbol{\zeta}_0^\tau(\boldsymbol{U}_{i,d_1} - E\boldsymbol{U}_{d_1}) - \bar{\tilde{Y}}\hat{\boldsymbol{\zeta}}_0^\tau(\boldsymbol{U}_{i,d_1} - \bar{\boldsymbol{U}}_{d_1})\}/\bar{\tilde{Y}} \\ &\quad + \left\{Y_i \sum_{s=d_1+1}^d \phi_s(U_{is}) - \bar{\tilde{Y}}\hat{\boldsymbol{\gamma}}_{0,d-d_1}^\tau \boldsymbol{B}_{d-d_1}(\boldsymbol{U}_{i,d-d_1})\right\}/\bar{\tilde{Y}} \\ &\stackrel{\text{def}}{=} V_{n1i} + V_{n2i} + V_{n3i}, \end{aligned} \quad \text{(A.3)}$$

together with (A.2) and (A.3), and $\alpha_0 = 1$ and $\bar{\tilde{Y}} = E\tilde{Y} + O_P(n^{-1/2}) = EY + O_P(n^{-1/2})$, we have that

$$\begin{aligned} \sum_{i=1}^n V_{n1i}(1, \breve{\boldsymbol{U}}_{i,d_1}^\tau)^\tau &= \sum_{i=1}^n (1, \breve{\boldsymbol{U}}_{i,d_1}^\tau)^\tau (Y_i - EY)/EY + \sum_{i=1}^n (1, \breve{\boldsymbol{U}}_{i,d_1}^\tau)^\tau (\alpha_0 - \hat{\alpha}_0) \\ &\quad + \sum_{i=1}^n (1, \breve{\boldsymbol{U}}_{i,d_1}^\tau)^\tau (EY - \tilde{Y})/EY + o_P(n^{1/2}), \end{aligned} \quad \text{(A.4)}$$

$$\sum_{i=1}^n V_{n2i}(1, \breve{\boldsymbol{U}}_{i,d_1}^\tau)^\tau = \sum_{i=1}^n (1, \breve{\boldsymbol{U}}_{i,d_1}^\tau)^\tau (Y_i - EY)(\boldsymbol{U}_{i,d_1} - E\boldsymbol{U}_{d_1})^\tau \boldsymbol{\zeta}_0/EY$$

$$
\begin{aligned}
&+\sum_{i=1}^{n}(1,\check{\boldsymbol{U}}_{i,d_1}^{\tau})^{\tau}(\boldsymbol{U}_{i,d_1}-E\boldsymbol{U}_{d_1})^{\tau}(\boldsymbol{\zeta}_0-\hat{\boldsymbol{\zeta}}_0)\\
&+\sum_{i=1}^{n}(1,\check{\boldsymbol{U}}_{i,d_1}^{\tau})^{\tau}(\boldsymbol{U}_{i,d_1}-E\boldsymbol{U}_{d_1})^{\tau}\boldsymbol{\zeta}_0(EY-\bar{\tilde{Y}})/EY \qquad \text{(A.5)}\\
&+\sum_{i=1}^{n}(1,\check{\boldsymbol{U}}_{i,d_1}^{\tau})^{\tau}(\bar{\boldsymbol{U}}_{d_1}-E\boldsymbol{U}_{d_1})^{\tau}\boldsymbol{\zeta}_0+o_P(n^{1/2}).
\end{aligned}
$$

By the results of de Boor (2001, p. 149), we can find a $\check{\boldsymbol{\gamma}}_{0,d-d_1}$ such that $\|\sum_{s=d_1+1}^{d}\phi_s(u_s)-\check{\boldsymbol{\gamma}}_{0,d-d_1}^{\tau}\boldsymbol{B}_{d-d_1}(\boldsymbol{u}_{d-d_1})\|_\infty=O(h^v)$. Thus,

$$
\sum_{i=1}^{n}(1,\check{\boldsymbol{U}}_{i,d_1}^{\tau})^{\tau}\Bigg(\sum_{s=d_1+1}^{d}\phi_s(U_{is})-\check{\boldsymbol{\gamma}}_{0,d-d_1}^{\tau}\boldsymbol{B}_{d-d_1}(\boldsymbol{U}_{i,d-d_1})\Bigg)=O_P(nh^v).
$$

Similar to the analysis of $L_{n1}^{[3]}$ in the proof of Lemma A.3 in Zhang et al. (2014), we further have $(\check{\boldsymbol{\gamma}}_{0,d-d_1}-\hat{\boldsymbol{\gamma}}_{0,d-d_1})^{\tau}\sum_{i=1}^{n}\boldsymbol{B}_{d-d_1}(\boldsymbol{U}_{i,d-d_1})=o_P(n^{1/2})$. Thus,

$$
\begin{aligned}
&\sum_{i=1}^{n}V_{n3i}(1,\check{\boldsymbol{U}}_{i,d_1}^{\tau})^{\tau}\\
&\quad=\sum_{i=1}^{n}(1,\check{\boldsymbol{U}}_{i,d_1}^{\tau})^{\tau}(Y_i-EY)\sum_{s=d_1+1}^{d}\phi_s(U_{is})/EY\\
&\qquad+\sum_{i=1}^{n}(1,\check{\boldsymbol{U}}_{i,d_1}^{\tau})^{\tau}(EY-\bar{\tilde{Y}})\sum_{s=d_1+1}^{d}\phi_s(U_{is})/EY\\
&\qquad+\sum_{i=1}^{n}(1,\check{\boldsymbol{U}}_{i,d_1}^{\tau})^{\tau}\Bigg(\sum_{s=d_1+1}^{d}\phi_s(U_{is})-\check{\boldsymbol{\gamma}}_{0,d-d_1}^{\tau}\boldsymbol{B}_{d-d_1}(\boldsymbol{U}_{i,d-d_1})\Bigg) \qquad \text{(A.6)}\\
&\qquad+(\check{\boldsymbol{\gamma}}_{0,d-d_1}-\hat{\boldsymbol{\gamma}}_{0,d-d_1})^{\tau}\sum_{i=1}^{n}(1,\check{\boldsymbol{U}}_{i,d_1}^{\tau})^{\tau}\boldsymbol{B}_{d-d_1}(\boldsymbol{U}_{i,d-d_1})\\
&\qquad+o_P(n^{1/2}).
\end{aligned}
$$

In fact, the expression (A.6) can be further expressed as

$$
\begin{aligned}
\sum_{i=1}^{n}V_{n3i}(1,\check{\boldsymbol{U}}_{i,d_1}^{\tau})^{\tau}&=\sum_{i=1}^{n}(1,\check{\boldsymbol{U}}_{i,d_1}^{\tau})^{\tau}(Y_i-EY)\sum_{s=d_1+1}^{d}\phi_s(U_{is})/EY\\
&\quad+\sum_{i=1}^{n}(1,\check{\boldsymbol{U}}_{i,d_1}^{\tau})^{\tau}(EY-\bar{\tilde{Y}})\sum_{s=d_1+1}^{d}\phi_s(U_{is})/EY \qquad \text{(A.7)}\\
&\quad+o_P(n^{1/2})+O_P(nh^v).
\end{aligned}
$$

Define $\hat{\boldsymbol{\vartheta}}_0 = (\hat{\alpha}_0, \hat{\boldsymbol{\xi}}_0^\tau)^\tau$, $\boldsymbol{\vartheta}_0 = (\alpha_0, \boldsymbol{\zeta}_0^\tau)^\tau$ and define $\mathbf{M}_n = n^{-1}\sum_{i=1}^n (1, \check{\boldsymbol{U}}_{i,d_1}^\tau)^\tau (1, \boldsymbol{U}_{i,d_1}^\tau - E\boldsymbol{U}_{d_1}^\tau)$. Together with (A.2)–(A.7), we have that

$$\begin{aligned}
&\mathbf{M}_n\sqrt{n}(\hat{\boldsymbol{\vartheta}}_0 - \boldsymbol{\vartheta}_0) \\
&\quad = \frac{1}{\sqrt{n}}\sum_{i=1}^n (1, \check{\boldsymbol{U}}_{i,d_1}^\tau)^\tau \left\{ \frac{Y_i - EY}{EY}\phi(\boldsymbol{U}_i) - \frac{\bar{\tilde{Y}} - EY}{EY}\phi(\boldsymbol{U}_i) \right. \qquad \text{(A.8)} \\
&\qquad \left. + \boldsymbol{\zeta}_0^\tau(\bar{\boldsymbol{U}}_{d_1} - E\boldsymbol{U}_{d_1}) \right\} + o_P(1) + O_P\big(n^{1/2}h^v\big).
\end{aligned}$$

By condition (C4), we know that $nh^{2v} \to 0$. Together with $\bar{\boldsymbol{U}}_{d_1} - E\boldsymbol{U}_{d_1} = O_P(n^{-1/2})$, $E\tilde{Y} = EY$, $\bar{\tilde{Y}} - EY = O_P(n^{-1/2})$ and (A.8), the condition that $\boldsymbol{U}$ is independent of Y, and $\mathbf{M}_n = E[(1, \check{\boldsymbol{U}}_{d_1}^\tau)^\tau]^{\otimes 2} + o_P(1)$, the asymptotic distribution of $\sqrt{n}(\hat{\boldsymbol{\vartheta}}_0 - \boldsymbol{\vartheta}_0)$ can be directly obtained from (A.8). Moreover, the asymptotic distributions of $\sqrt{n}(\hat{\boldsymbol{\vartheta}}_r - \boldsymbol{\vartheta}_r)$, $r = 1, \dots, p$ can also be obtained similarly. □

Proof of Theorem 3. Directly using the fact that

$$\begin{aligned}
&\hat{\phi}(\boldsymbol{U}_i) - \phi(\boldsymbol{U}_i) \\
&\quad = (\hat{\boldsymbol{\vartheta}}_0 - \boldsymbol{\vartheta}_0)^\tau\big(1, (\boldsymbol{U}_{i,d_1} - E\boldsymbol{U}_{d_1})^\tau\big)^\tau - \hat{\boldsymbol{\xi}}_0^\tau(E\boldsymbol{U}_{d_1} - \bar{\boldsymbol{U}}_{d_1}) \qquad \text{(A.9)} \\
&\qquad + \left\{ \hat{\boldsymbol{\gamma}}_{0,d-d_1}^\tau \boldsymbol{B}_{d-d_1}(\boldsymbol{U}_{i,d-d_1}) - \sum_{s=d_1+1}^d \phi_s(U_{is}) \right\}, \\
&\hat{\psi}_r(\boldsymbol{U}_i) - \psi_r(\boldsymbol{U}_i) \\
&\quad = (\hat{\boldsymbol{\vartheta}}_r - \boldsymbol{\vartheta}_r)^\tau\big(1, (\boldsymbol{U}_{i,d_1} - E\boldsymbol{U}_{d_1})^\tau\big)^\tau - \hat{\boldsymbol{\xi}}_r^\tau(E\boldsymbol{U}_{d_1} - \bar{\boldsymbol{U}}_{d_1}) \qquad \text{(A.10)} \\
&\qquad + \left\{ \hat{\boldsymbol{\gamma}}_{r,d-d_1}^\tau \boldsymbol{B}_{d-d_1}(\boldsymbol{U}_{i,d-d_1}) - \sum_{s=d_1+1}^d \psi_{rs}(U_{is}) \right\}.
\end{aligned}$$

Using (A.9) and (A.10), similar to the analysis of L_{n1} and L_{n2} in Lemma A.3 in Zhang et al. (2014), as $nh^{2v} \to 0$, we have that

$$\begin{aligned}
&n^{-1}\sum_{i=1}^n (\hat{Y}_i - Y_i)T(\boldsymbol{X}_i) \\
&\quad = -n^{-1}\sum_{i=1}^n \frac{Y_i T(\boldsymbol{X}_i)}{\phi(\boldsymbol{U}_i)}\big(1, (\boldsymbol{U}_{i,d_1} - E\boldsymbol{U}_{d_1})^\tau\big)(\hat{\boldsymbol{\vartheta}}_0 - \boldsymbol{\vartheta}_0) \qquad \text{(A.11)} \\
&\qquad - E\left\{ \frac{YT(\boldsymbol{X})}{\phi(\boldsymbol{U})} \right\}(\bar{\boldsymbol{U}}_{d_1} - E\boldsymbol{U}_{d_1})^\tau \boldsymbol{\zeta}_0 + o_P\big(n^{-1/2}\big),
\end{aligned}$$

and

$$
\begin{aligned}
n^{-1}\sum_{i=1}^{n}(\hat{X}_{ri}-X_{ri})T(\boldsymbol{X}_i)
&= -n^{-1}\sum_{i=1}^{n}\frac{X_{ri}T(\boldsymbol{X}_i)}{\psi_r(\boldsymbol{U}_i)}\big(1,(\boldsymbol{U}_{i,d_1}-E\boldsymbol{U}_{d_1})^{\tau}\big)(\hat{\boldsymbol{\vartheta}}_r-\boldsymbol{\vartheta}_r) \\
&\quad -E\left\{\frac{X_rT(\boldsymbol{X})}{\psi_r(\boldsymbol{U})}\right\}(\bar{\boldsymbol{U}}_{d_1}-E\boldsymbol{U}_{d_1})^{\tau}\boldsymbol{\zeta}_r+o_P\big(n^{-1/2}\big).
\end{aligned} \tag{A.12}
$$

Define $f'_{\boldsymbol{\beta}}(\boldsymbol{x},\boldsymbol{\beta})=(f'_1(\boldsymbol{x},\boldsymbol{\beta}),\ldots,f'_q(\boldsymbol{x},\boldsymbol{\beta}))$, $f'_{\boldsymbol{x}}(\boldsymbol{x},\boldsymbol{\beta})=(f'_{x_1}(\boldsymbol{x},\boldsymbol{\beta}),\ldots,f'_{x_p}(\boldsymbol{x},\boldsymbol{\beta}))$, and $f''_{\boldsymbol{\beta}\boldsymbol{\beta}}(\boldsymbol{x},\boldsymbol{\beta})=\partial f'_{\boldsymbol{\beta}}(\boldsymbol{x},\boldsymbol{\beta})/\partial\boldsymbol{\beta}^{\tau}$. Note that

$$
\begin{aligned}
0 &= n^{-1}\sum_{i=1}^{n}\{\hat{Y}_i-f(\hat{\boldsymbol{X}}_i,\hat{\boldsymbol{\beta}})\}f'_{\boldsymbol{\beta}}(\hat{\boldsymbol{X}}_i,\hat{\boldsymbol{\beta}}) \\
&= n^{-1}\sum_{i=1}^{n}\{\hat{Y}_i-f(\hat{\boldsymbol{X}}_i,\boldsymbol{\beta})\}f'_{\boldsymbol{\beta}}(\hat{\boldsymbol{X}}_i,\boldsymbol{\beta}) \\
&\quad +n^{-1}\sum_{i=1}^{n}\{f(\hat{\boldsymbol{X}}_i,\boldsymbol{\beta})-f(\hat{\boldsymbol{X}}_i,\hat{\boldsymbol{\beta}})\}f'_{\boldsymbol{\beta}}(\hat{\boldsymbol{X}}_i,\boldsymbol{\beta}) \\
&\quad +n^{-1}\sum_{i=1}^{n}\{\hat{Y}_i-f(\hat{\boldsymbol{X}}_i,\boldsymbol{\beta})\}\{f'_{\boldsymbol{\beta}}(\hat{\boldsymbol{X}}_i,\hat{\boldsymbol{\beta}})-f'_{\boldsymbol{\beta}}(\hat{\boldsymbol{X}}_i,\boldsymbol{\beta})\} \\
&\quad +n^{-1}\sum_{i=1}^{n}\{f(\hat{\boldsymbol{X}}_i,\boldsymbol{\beta})-f(\hat{\boldsymbol{X}}_i,\hat{\boldsymbol{\beta}})\}\{f'_{\boldsymbol{\beta}}(\hat{\boldsymbol{X}}_i,\hat{\boldsymbol{\beta}})-f'_{\boldsymbol{\beta}}(\hat{\boldsymbol{X}}_i,\boldsymbol{\beta})\} \\
&\overset{\text{def}}{=} Q_{n1}(\boldsymbol{\beta})+Q_{n2}(\boldsymbol{\beta})+Q_{n3}(\boldsymbol{\beta})+Q_{n4}(\boldsymbol{\beta}).
\end{aligned} \tag{A.13}
$$

Similar to the analysis of Lemma A.4 and Theorem 1 in Zhang et al. (2014), we show that

$$
\begin{aligned}
Q_{n1}(\boldsymbol{\beta}) &\overset{\text{def}}{=} n^{-1}\sum_{i=1}^{n}\{\hat{Y}_i-f(\hat{\boldsymbol{X}}_i,\boldsymbol{\beta})\}f'_{\boldsymbol{\beta}}(\hat{\boldsymbol{X}}_i,\boldsymbol{\beta}) \\
&= n^{-1}\sum_{i=1}^{n}\varepsilon_i f'_{\boldsymbol{\beta}}(\boldsymbol{X}_i,\boldsymbol{\beta})+n^{-1}\sum_{i=1}^{n}\{\hat{Y}_i-Y_i\}f'_{\boldsymbol{\beta}}(\boldsymbol{X}_i,\boldsymbol{\beta}) \\
&\quad -n^{-1}\sum_{i=1}^{n}\sum_{l=1}^{p}f'_{\boldsymbol{\beta}}(\boldsymbol{X}_i,\boldsymbol{\beta})f'_{x_l}(\boldsymbol{X}_i,\boldsymbol{\beta})(\hat{X}_{li}-X_{li})+o_P\big(n^{-1/2}\big).
\end{aligned} \tag{A.14}
$$

Applying (A.11) and (A.12) on (A.14), we obtain that

$$
\begin{aligned}
\mathcal{Q}_{n1}(\boldsymbol{\beta}) = n^{-1}\sum_{i=1}^{n}\varepsilon_i f'_{\boldsymbol{\beta}}(\boldsymbol{X}_i,\boldsymbol{\beta}) \\
& - n^{-1}\sum_{i=1}^{n}\frac{Y_i f'_{\boldsymbol{\beta}}(\boldsymbol{X}_i,\boldsymbol{\beta})}{\phi(\boldsymbol{U}_i)}\big(1,(\boldsymbol{U}_{i,d_1}-E\boldsymbol{U}_{d_1})^{\tau}\big)(\hat{\boldsymbol{\vartheta}}_0-\boldsymbol{\vartheta}_0) \\
& + n^{-1}\sum_{l=1}^{p}\sum_{i=1}^{n} f'_{\boldsymbol{\beta}}(\boldsymbol{X}_i,\boldsymbol{\beta}) \\
& \times \frac{X_{li} f'_{x_l}(\boldsymbol{X}_i,\boldsymbol{\beta})}{\psi_l(\boldsymbol{U}_i)}\big(1,(\boldsymbol{U}_{i,d_1}-E\boldsymbol{U}_{d_1})^{\tau}\big)(\hat{\boldsymbol{\vartheta}}_l-\boldsymbol{\vartheta}_l) \qquad \text{(A.15)} \\
& - (\bar{\boldsymbol{U}}_{d_1}-E\boldsymbol{U}_{d_1})^{\tau}\left\{\boldsymbol{\zeta}_0 E\left\{\frac{Y f'_{\boldsymbol{\beta}}(\boldsymbol{X},\boldsymbol{\beta})}{\phi(\boldsymbol{U})}\right\}\right. \\
& \left. - \sum_{l=1}^{p}\boldsymbol{\zeta}_l E\left\{\frac{X_l f'_{\boldsymbol{\beta}}(\boldsymbol{X},\boldsymbol{\beta}) f'_{x_l}(\boldsymbol{X},\boldsymbol{\beta})}{\psi_l(\boldsymbol{U})}\right\}\right\} \\
& + o_P(n^{-1/2}).
\end{aligned}
$$

Appealing to expression (A.13) again, similar to proof of Theorem 1 in Zhang et al. (2014), we show that $\sqrt{n}\mathcal{Q}_{n3}(\boldsymbol{\beta}) = o_P(1)$, $\sqrt{n}\mathcal{Q}_{n4}(\boldsymbol{\beta}) = o_P(1)$, and also $\mathcal{Q}_{n2}(\boldsymbol{\beta}) = \boldsymbol{\Gamma}_{\boldsymbol{\beta}}(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}) + o_P(n^{-1/2})$. Together with expression (A.13), (A.8) and (A.15), we have $\sqrt{n}(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}) = \boldsymbol{\Gamma}_{\boldsymbol{\beta}}^{-1}\sqrt{n}\mathcal{Q}_{n1} + o_P(1)$, the asymptotic distribution of $\sqrt{n}(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta})$ follows from (A.8) and (A.15). We complete the proof. □

Acknowledgments

The authors thank two anonymous referees and the editor for their many constructive suggestions that helped us improve both the presentation and the substance of this article. Zhang Jun's research was supported by the National Natural Science Foundation of China (Grant No. 11401391).

References

Carroll, R. J., Ruppert, D., Stefanski, L. A. and Crainiceanu, C. M. (2006). *Nonlinear Measurement Error Models, a Modern Perspective*, 2nd ed. New York: Chapman and Hall. MR2243417

Chen, R., Liang, H. and Wang, J. (2011). Determination of linear components in additive models. *Journal of Nonparametric Statistics* **23**, 367–383. MR2801299

Cui, X., Guo, W., Lin, L. and Zhu, L. (2009). Covariate-adjusted nonlinear regression. *The Annals of Statistics* **37**, 1839–1870. MR2533473

Cui, X., Härdle, W. and Zhu, L.-X. (2011). The EFM approach for single-index models. *The Annals of Statistics* **39**, 1658–1688. MR2850216

de Boor, C. (2001). *A Practical Guide to Splines*, Revised ed. *Applied Mathematical Sciences* **27**. New York: Springer. MR1900298

Delaigle, A., Hall, P. and Wen-Xin, Z. (2016). Nonparametric covariate-adjusted regression. *The Annals of Statistics*. To appear. Available at arXiv:1601.02739. MR3546448

Escanciano, J. C. (2006). A consistent diagnostic test for regression models using projections. *Econometric Theory* **22**, 1030–1051. MR2328527

Eubank, R. L. and Spiegelman, C. H. (1990). Testing the goodness of fit of a linear model via nonparametric regression techniques. *Journal of the American Statistical Association* **85**, 387–392. MR1141739

Fan, J. and Li, R. (2001). Variable selection via nonconcave penalized likelihood and its oracle properties. *Journal of the American Statistical Association* **96**, 1348–1360. MR1946581

Härdle, W. and Liang, H. (2007). Partially linear models. In *Statistical Methods for Biostatistics and Related Fields*, 87–103. Berlin: Springer. MR2376405

Härdle, W., Liang, H. and Gao, J. T. (2000). *Partially Linear Models*. Heidelberg: Springer. MR1787637

Härdle, W. and Mammen, E. (1993). Comparing nonparametric versus parametric regression fits. *The Annals of Statistics* **21**, 1926–1947. MR1245774

Hart, J. D. (1997). *Nonparametric Smoothing and Lack-of-Fit Tests*. *Springer Series in Statistics*. New York: Springer. MR1461272

Hastie, T. and Tibshirani, R. (1990). *Generalized Additive Models*. London: Chapman and Hall. MR1082147

Heckman, N. E. (1986). Spline smoothing in partly linear models. *Journal of the Royal Statistical Society, Series B* **48**, 244–248. MR0868002

Huang, J. (1999). Efficient estimation of the partly linear additive Cox model. *The Annals of Statistics* **27**, 1536–1563. MR1742499

Huang, J. Z. (2003). Local asymptotics for polynomial spline regression. *The Annals of Statistics* **31**, 1600–1635. MR2012827

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

Li, F., Lin, L. and Cui, X. (2010). Covariate-adjusted partially linear regression models. *Communications in Statistics. Theory and Methods* **39**, 1054–1074. MR2745361

Li, G., Lin, L. and Zhu, L. (2012). Empirical likelihood for a varying coefficient partially linear model with diverging number of parameters. *Journal of Multivariate Analysis* **105**, 85–111. MR2877505

Li, Q. (2000). Efficient estimation of additive partially linear models. *International Economic Review* **41**, 1073–1092. MR1790072

Li, X., Du, J., Li, G. and Fan, M. (2014). Variable selection for covariate adjusted regression model. *Journal of Systems Science and Complexity* **27**, 1227–1246. MR3284348

Li, Y. and Ruppert, D. (2008). On the asymptotics of penalized splines. *Biometrika* **95**, 415–436. MR2521591

Lian, H. (2012). Empirical likelihood confidence intervals for nonparametric functional data analysis. *Journal of Statistical Planning and Inference* **142**, 1669–1677. MR2903379

Liang, H. and Li, R. (2009). Variable selection for partially linear models with measurement errors. *Journal of the American Statistical Association* **104**, 234–248. MR2504375

Liang, H., Qin, Y., Zhang, X. and Ruppert, D. (2009a). Empirical likelihood-based inferences for generalized partially linear models. *Scandinavian Journal of Statistics. Theory and Applications* **36**, 433–443. MR2549703

Liang, H., Su, H., Thurston, S. W., Meeker, J. D. and Hauser, R. (2009b). Empirical likelihood based inference for additive partial linear measurement error models. *Statistics and its Interface* **2**, 83–90. MR2500770

Liang, H., Thurston, S. W., Ruppert, D., Apanasovich, T. and Hauser, R. (2008). Additive partial linear models with measurement errors. *Biometrika* **95**, 667–678. MR2443182

Liu, X., Wang, L. and Liang, H. (2011). Estimation and variable selection for semiparametric additive partial linear models. *Statistica Sinica* **21**, 1225–1248. MR2827522

Mammen, E. (1993). Bootstrap and wild bootstrap for high-dimensional linear models. *The Annals of Statistics* **21**, 255–285. MR1212176

Nguyen, D. V. and Şentürk, D. (2008). Multicovariate-adjusted regression models. *Journal of Statistical Computation and Simulation* **78**, 813–827. MR2458498

Opsomer, J. D. and Ruppert, D. (1997). Fitting a bivariate additive model by local polynomial regression. *The Annals of Statistics* **25**, 186–211. MR1429922

Owen, A. B. (2001). *Empirical Likelihood*. London: Chapman and Hall/CRC.

Qin, J. and Lawless, J. (1994). Empirical likelihood and general estimating equations. *The Annals of Statistics* **22**, 300–325. MR1272085

Şentürk, D. and Müller, H.-G. (2005). Covariate-adjusted regression. *Biometrika* **92**, 75–89. MR2158611

Şentürk, D. and Müller, H.-G. (2006). Inference for covariate adjusted regression via varying coefficient models. *The Annals of Statistics* **34**, 654–679. MR2281880

Şentürk, D. and Müller, H.-G. (2009). Covariate-adjusted generalized linear models. *Biometrika* **96**, 357–370. MR2507148

Şentürk, D. and Nguyen, D. V. (2009). Partial covariate adjusted regression. *Journal of Statistical Planning and Inference* **139**, 454–468. MR2474018

Speckman, P. E. (1988). Kernel smoothing in partial linear models. *Journal of the Royal Statistical Society, Series B* **50**, 413–436. MR0970977

Stone, C. J. (1985). Additive regression and other nonparametric models. *The Annals of Statistics* **13**, 689–705. MR0790566

Stute, W., González Manteiga, W. and Presedo Quindimil, M. (1998a). Bootstrap approximations in model checks for regression. *Journal of the American Statistical Association* **93**, 141–149. MR1614600

Stute, W., González Manteiga, W. and Presedo Quindimil, M. (1998b). Bootstrap approximations in model checks for regression. *Journal of the American Statistical Association* **93**, 141–149. MR1614600

Stute, W., Thies, S. and Zhu, L.-X. (1998). Model checks for regression: An innovation process approach. *The Annals of Statistics* **26**, 1916–1934. MR1673284

Tang, N.-S. and Zhao, P.-Y. (2013a). Empirical likelihood-based inference in nonlinear regression models with missing responses at random. *Statistics. A Journal of Theoretical and Applied Statistics* **47**, 1141–1159. MR3175740

Tang, N.-S. and Zhao, P.-Y. (2013b). Empirical likelihood semiparametric nonlinear regression analysis for longitudinal data with responses missing at random. *Annals of the Institute of Statistical Mathematics* **65**, 639–665. MR3094950

Wang, L., Liu, X., Liang, H. and Carroll, R. J. (2011). Estimation and variable selection for generalized additive partial linear models. *The Annals of Statistics* **39**, 1827–1851. MR2893854

Wang, X., Li, G. and Lin, L. (2011). Empirical likelihood inference for semi-parametric varying-coefficient partially linear EV models. *Metrika. International Journal for Theoretical and Applied Statistics* **73**, 171–185. MR2769269

Wei, Z. and Zhu, L. (2010). Evaluation of value at risk: An empirical likelihood approach. *Statistica Sinica* **20**, 455–468. MR2640704

Wu, C.-F. J. (1986). Jackknife, bootstrap and other resampling methods in regression analysis. *The Annals of Statistics* **14**, 1261–1350. MR0868303

Xia, Y., Li, W. K., Tong, H. and Zhang, D. (2004). A goodness-of-fit test for single-index models. *Statistica Sinica* **14**, 1–39. MR2036761

Zhang, J., Feng, S., Li, G. and Lian, H. (2011). Empirical likelihood inference for partially linear panel data models with fixed effects. *Economics Letters* **113**, 165–167. MR2894863

Zhang, J., Feng, Z. and Zhou, B. (2014). A revisit to correlation analysis for distortion measurement error data. *Journal of Multivariate Analysis* **124**, 116–129. MR3147315

Zhang, J., Gai, Y. and Wu, P. (2013). Estimation in linear regression models with measurement errors subject to single-indexed distortion. *Computational Statistics & Data Analysis* **59**, 103–120. MR3000045

Zhang, J., Li, G. and Feng, Z. (2015). Checking the adequacy for a distortion errors-in-variables parametric regression model. *Computational Statistics & Data Analysis* **83**, 52–64. MR3281795

Zhang, J., Yu, Y., Zhou, B. and Liang, H. (2014). Nonlinear measurement errors models subject to additive distortion. *Journal of Statistical Planning and Inference* **150**, 49–65. MR3206720

Zhang, J., Yu, Y., Zhu, L. and Liang, H. (2013). Partial linear single index models with distortion measurement errors. *Annals of the Institute of Statistical Mathematics* **65**, 237–267. MR3011622

Zhang, J., Zhou, N., Sun, Z., Li, G. and Wei, Z. (2016). Statistical inference on restricted partial linear regression models with partial distortion measurement errors. *Statistica Neerlandica. Journal of the Netherlands Society for Statistics and Operations Research*. **70**, 304–331. Available at DOI:10.1111/stan.12089. MR3565895

Zhang, J., Zhu, L. and Liang, H. (2012). Nonlinear models with measurement errors subject to single-indexed distortion. *Journal of Multivariate Analysis* **112**, 1–23. MR2957282

Zhang, X. and Liang, H. (2011). Focused information criterion and model averaging for generalized additive partial linear models. *The Annals of Statistics* **39**, 194–200. MR2797843

Zhu, L., Lin, L., Cui, X. and Li, G. (2010). Bias-corrected empirical likelihood in a multi-link semiparametric model. *Journal of Multivariate Analysis* **101**, 850–868. MR2584904

J. Zhang
College of Mathematics and Statistics
Institute of Statistical Sciences
Shen Zhen-Hong Kong Joint Research Centre
for Applied Statistical Sciences
Shenzhen University
Shenzhen 518060
China
E-mail: zhangjunstat@gmail.com

N. Zhou
Q. Chen
T. Chu
College of Mathematics and Statistics
Shenzhen University
Shenzhen 518060
China