

Generalized varying coefficient partially linear measurement errors models

Jun Zhang¹ · Zhenghui Feng² · Peirong Xu³ ·
Hua Liang⁴

Received: 25 September 2014 / Revised: 5 April 2015 / Published online: 29 July 2015
© The Institute of Statistical Mathematics, Tokyo 2015

Abstract We study generalized varying coefficient partially linear models when some linear covariates are error prone, but their ancillary variables are available. We first calibrate the error-prone covariates, then develop a quasi-likelihood-based estimation procedure. To select significant variables in the parametric part, we develop a penalized quasi-likelihood variable selection procedure, and the resulting penalized estimators are shown to be asymptotically normal and have the oracle property. Moreover, to select significant variables in the nonparametric component, we investigate asymptotic behavior of the semiparametric generalized likelihood ratio test. The limiting null distribution is shown to follow a Chi-square distribution, and a new Wilks phenomenon

Electronic supplementary material The online version of this article (doi:[10.1007/s10463-015-0532-y](https://doi.org/10.1007/s10463-015-0532-y)) contains supplementary material, which is available to authorized users.

✉ Jun Zhang
zhangjunstat@gmail.com

Zhenghui Feng
zhfengwise@gmail.com

Peirong Xu
xupeirong@seu.edu.cn

Hua Liang
hliang@gwu.edu

- ¹ Shen Zhen-Hong Kong Joint Research Centre for Applied Statistical Sciences, School of Mathematics and Statistics, Institute of Statistical Sciences, Shenzhen University, Shenzhen 518060, China
- ² School of Economics, and the Wang Yanan Institute for Studies in Economics, Xiamen University, Xiamen 361005, China
- ³ Department of Mathematics, Southeast University, Nanjing 211189, China
- ⁴ Department of Statistics, George Washington University, Washington, DC 20052, USA

is unveiled in the context of error-prone semiparametric modeling. Simulation studies and a real data analysis are conducted to evaluate the performance of the proposed methods.

Keywords Ancillary variables · Errors-in-variable · Error prone · LASSO · Measurement errors · Quasi-likelihood · Penalized quasi-likelihood · SCAD · Varying coefficient models

1 Introduction

Generalized varying coefficient partially linear models (GVCPLM) (Li and Liang 2008) are powerful extensions of generalized partially linear models (GPLM). These models offer additional flexibility compared to GPLM when modeling data with discrete response variable, because they further relax model assumptions imposed on GPLM and allow interactions between covariates and certain unknown functions depending on other covariates, while keep some linear components there. GVCPLM are also useful generalizations of varying coefficient models (Hastie and Tibshirani 1993), which have been applied to parsimoniously describe data structure and uncover scientific feature, and have been studied in the context of quasi-likelihood principle. As well known in the literature, several useful semiparametric models can be classified as special cases of GVCPLM in one way or another to name a few such as GPLM (Hunsberger 1994; Hunsberger et al. 2002; Lin and Carroll 2001; Severini and Staniswalis 1994); partially linear models (Härdle et al. 2000; Robinson 1988; Speckman 1988); semivarying-coefficient models (Fan and Huang 2005; Xia et al. 2004; Zhang et al. 2002) and varying coefficient models (Cai et al. 2000; Hastie and Tibshirani 1993).

Li and Liang (2008) studied variable selection for GVCPLM using the SCAD (Fan and Li 2001) to identify parametric components and generalized likelihood ratio test (Fan et al. 2001) to select nonparametric components. Wang and Xia (2009) proposed a shrinkage method for selecting nonparametric components in varying coefficient models. Wang et al. (2011) developed an estimation procedure and variable selection procedure for generalized additive partial linear models (PLM) with an incorporation of polynomial spline smoothing to estimate nonparametric functions and penalized SCAD quasi-likelihood-based estimators to select linear covariates. Li et al. (2011) considered variable selection on varying coefficient partially linear models when both the number of parametric and nonparametric components diverge at appropriate rates. Wei et al. (2011) further considered variable selection and estimation in “large p , small n ” setting using the group Lasso idea (Yuan and Lin 2006).

Measurement errors are often encountered in biomedical research. Simply ignoring the errors can cause bias in estimation and lead to a loss of power for accurately detecting the relationship among variables. Regression calibration and simulation extrapolation (SIMEX, Cook and Stefanski 1994) are two widely useful methods for eliminating or reducing bias caused by measurement errors. But the corresponding estimators are consistent only in special cases such as linear or loglinear regression, and approximately consistent in general cases. There are possible alternative methods to remedy consistency concerns by deriving unbiased score functions in the pres-

ence of measurement error, for example, the conditional score method (Stefanski and Carroll 1987) and corrected score method (Stefanski 1989), which are essentially M-estimation methods, and the usual numerical methods and asymptotic theory for M-estimators are applicable. But just like other methods, these two methods have their own limitations. In particular, conditional scores can be derived under specific assumptions on the model for response given covariates and the error model for the surrogates, and some conditional score methods may require integration, while corrected scores also impose sufficient assumptions on the error model to ensure unbiased estimation of the true-data estimating function. See Carroll et al. (2006) for more detailed discussions on a variety of estimation and inference methods for nonlinear measurement errors models. Ma and Carroll (2006), Ma and Tsiatis (2006) and Tsiatis and Ma (2004) investigated estimation efficiency for semiparametric models with measurement errors. Hall and Ma (2007) studied semiparametric estimators of functional measurement error models. Yi et al. (2012) considered marginal analysis of longitudinal data when errors-in-variables and missing response appear simultaneously.

Efforts have been made to address various scientific questions using semiparametric models in the presence of measurement errors. For example, Sinha et al. (2010) proposed a semiparametric Bayesian method for handling measurement errors commonly appeared in nutritional epidemiological studies. Carroll and Wang (2008) studied effects of measurement errors on microarray data analysis, and noticed that a direct application of the simulation extrapolation method leads to inconsistent estimators. The authors proposed a permutation SIMEX method which leads to consistent estimators in theory. In environmental research, environmental factors are generally measured with error. Lobach et al. (2008, 2010) developed a genotype-based approach for association analysis of case-control studies of gene-environment interactions using pseudo-likelihood principle to reduce bias caused by measurement errors.

Recently, variable selection in semiparametric regressions with measurement errors has been considered. Liang and Li (2009) developed two variable selection procedures, penalized least squares and penalized quantile regression, for PLM with measurement errors. Ma and Li (2010) proposed a penalized estimating equation with SCAD penalty for a class of parametric measurement error models and semiparametric measurement error models. As observed in Liang and Li (2009), if measurement errors are ignored, some variable selection procedures may falsely choose variables and result in a final biased model.

In this article, we study estimation and variable selection for GVCPLM when the covariates are error prone. We consider three problems: first, calibrating the error-prone covariates using ancillary information and applying nonparametric regression techniques; second, developing quasi-likelihood profile estimating procedures and justifying that the corresponding estimators of parameters of interest are *asymptotically normal*; third, proposing a penalized quasi-likelihood procedure for selecting significant parameters and a generalized likelihood ratio test for selecting nonzero nonparametric functions. Zhou and Liang (2009) once studied the case where the link function is identity one, and gave a variety of examples to illustrate the flexibility of the model. The authors developed a profile-based estimation procedure to estimate unknown parameters of interest.

It is remarkable that extension of the profile estimation procedure of Zhou and Liang (2009) to GVCPLM is by no means trivial. For the case of identity link function, the profile least-square technique can be used and a closed form of estimators is available. Nevertheless, for GVCPLM with measurement errors, only quasi-likelihood-based objective function is available. Whether the resulting estimators still have nice properties such as asymptotic normality is theoretically difficult to address. In GVCPLM, Li and Liang (2008) proposed SCAD-type procedure for parametric component selection and theoretically showed its oracle properties under certain assumptions. Whether such a procedure can be developed under a measurement error framework is not clear and has not been investigated in the literature. No measurement errors, Fan et al. (2001) proposed a generalized quasi-likelihood ratio test (GLRT) to investigate whether the coefficient functions in GVCPLM are constant or not. In this paper, we investigate Wilks phenomenon in the context of error-prone semiparametric setting. We further propose a bootstrap procedure to estimate null distribution of GLRT. To the best of our knowledge, this Wilks phenomenon under error prone is new and the findings contribute to the literature on semiparametric modeling.

The remainder of the paper is organized as follows. In Sect. 2, we propose the quasi-likelihood procedure for estimation of parametric components, then we develop a penalized quasi-likelihood for variable selection. Sampling properties of the proposed procedures are investigated. In Sect. 3, estimation procedure and variable selection procedure are proposed for the nonparametric component. The null distribution of the GLRT is also established. Simulation results and a real data analysis are presented in Sect. 4. Regularity conditions and technical proofs are given in the Appendix.

2 Estimation and variable selection for parametric components

Let $X = (X_1, \dots, X_p)^T \in \mathbb{R}^p$, $\xi = (\xi_1, \dots, \xi_d)^T \in \mathbb{R}^d$, $W = (W_1, \dots, W_r)^T \in \mathbb{R}^r$, $U \in \mathbb{R}$ be the covariates and Y be the response variable. The GVCPLM are of form:

$$g\{\mu(U, \xi, W, X)\} = \beta^T \xi + \theta^T W + \alpha(U)^T X, \quad (1)$$

where $g(\cdot)$ is a known link function, β and θ are vectors of unknown regression coefficients and $\alpha(\cdot)$ is a vector of *unknown* smooth nonparametric functions of U . The response Y is related to covariates (U, ξ, W, X) through an unknown mean function $\mu(U, \xi, W, X) = E(Y|U, \xi, W, X)$ and the conditional variance determined by a known positive function $T(\cdot)$, i.e., $\text{Var}(Y|U, \xi, W, X) = \sigma^2 T\{\mu(U, \xi, W, X)\}$. The components ξ are unobserved directly, but auxiliary variables (η, V) are available to remit ξ . η is related to V via

$$\eta = \xi(V) + e, \quad (2)$$

where e is a measurement error and independent of (X, W, V, U, Y) , and has a positive finite covariance matrix $\Sigma_e = E(ee^T)$. We term (1) and (2) generalized varying coefficient partially linear measurement error models (GVCPLMeM).

Let $\{(Y_i, U_i, \eta_i, V_i, W_i, X_i)\}_{i=1}^n$ be an *i.i.d.* sample from (Y, U, η, V, W, X) . When the covariates ξ are measured with error, we first calibrate ξ using ancillary observed sample $\{(\eta_i, V_i)\}_{i=1}^n$.

2.1 Covariate calibration

We introduce the calibration estimation procedure for ξ in this section. For notational simplicity, we assume V is univariate throughout this paper. Let η_{ik} be the k th entry of vector η_i for $i = 1, \dots, n$. To estimate $\xi_k(v)$, the k th component of $\xi(v)$, we employ the local linear smoothing technique (Fan and Gijbels 1996). That is, to minimize

$$\sum_{i=1}^n \{\eta_{ik} - c_{0k} - c_{1k}(V_i - v)\}^2 L_{b_k}(V_i - v) \tag{3}$$

with respect to c_{0k}, c_{1k} , where $L_b(\cdot) = L(\cdot/b)/b$ with $L(\cdot)$ be a kernel function, $b = b_k$ ($k = 1, \dots, p$) is a bandwidth. Let $\hat{c}_{0k}, \hat{c}_{1k}$ be the minimizers of (3). Write

$$\hat{\xi}_k(v) = \hat{c}_{0k} = \frac{D_{20,k}(v)D_{01,k}(v) - D_{10,k}(v)D_{11,k}(v)}{D_{00,k}(v)D_{20,k}(v) - D_{10,k}^2(v)}, \tag{4}$$

where $D_{s_1s_2,k}(v) = \sum_{i=1}^n L_{b_k}(V_i - v)(V_i - v)^{s_1} \eta_{ik}^{s_2}$ for $s_1 = 0, 1, 2, s_2 = 0, 1, k = 1, \dots, d$.

We now list the assumptions needed in the following proposition and theorems. The following are the regularity conditions for our asymptotic results.

- (C1) $q_2(x, y) < 0$ for $x \in \mathbb{R}$ and y in the range of the response variable.
- (C2) The functions $T''(\cdot)$ and $g'''(\cdot)$ are continuous.
- (C3) The random variable U has bounded support \mathcal{U} . The elements of the function $\alpha''(u)$ are continuous in $u \in \mathcal{U}$.
- (C4) The density functions $f_U(u), f_V(v)$ of U, V are Lipschitz continuous and bounded away from 0 and infinite on their supports, respectively. Moreover, the joint density function $f_{U,V}(u, v)$ of (U, V) is continuous on the support $\mathcal{U} \times \mathcal{V}$.
- (C5) Let $Z = \beta^T \xi + \theta^T W + \alpha(U)^T X$. Then, $sE[q_l^s(Z, Y)N^{\otimes 2} | U = u], E[q_l^s(Z, Y)N^{\otimes 2} | V = v]$ and $E[q_l^s(Z, Y)N^{\otimes 2} | U = u, V = v]$ for $l = 1, 2, s = 1, 2$ are Lipschitz continuous and twice differentiable on $u \in \mathcal{U}$ and $v \in \mathcal{V}$. Moreover, $E\{q_2^2(Z, Y)\} < \infty, E\{q_1^{2+\delta}(Z, Y)\} < \infty$ for some $\delta > 2$ and $E[\rho_2(Z)N^{\otimes 2} | U = u]$ is nonsingular for each $u \in \mathcal{U}$.
- (C6) The kernel functions $K(\cdot), L(\cdot)$ are univariate bounded, continuous and symmetric density functions satisfying that $\int t^2 K(t)dt \neq 0, \int t^2 L(t)dt \neq 0$, and $\int |t|^j K(t)dt < \infty, \int |t|^j L(t)dt < \infty$. for $j = 1, 2, 3, 4$. Moreover, the second derivatives of $K(\cdot)$ and $L(\cdot)$ are bounded on \mathbb{R} .
- (C7) The bandwidths h and b satisfy:
 - (i) $b = b_k, k = 1, \dots, d, b_k \asymp c_b h_o$ for some constant $c_b > 0; h \asymp c_h h_o$ for some constant $c_h > 0$.
 - (ii) $h_o \rightarrow 0$ as $n \rightarrow \infty, nh_o^2/(\log h_o^{-1})^4 \rightarrow \infty, nh_o^4 \rightarrow 0$.
- (C8) For all $\lambda_{1j}, \lambda_{2s}, j = 1, \dots, d, s = 1, \dots, r, \lambda_{1j} \rightarrow 0, \sqrt{n}\lambda_{1j} \rightarrow \infty, \lambda_{2s} \rightarrow 0, \sqrt{n}\lambda_{2s} \rightarrow \infty$, and $\liminf_{n \rightarrow \infty} \liminf_{u \rightarrow 0^+} p'_{\lambda_{1j}}(u)/\lambda_{1j} > 0, \liminf_{n \rightarrow \infty} \liminf_{u \rightarrow 0^+} p'_{\lambda_{2s}}(u)/\lambda_{2s} > 0$.

Condition (C1) is imposed to ensure the local likelihood concave and guarantees the solution unique. Conditions (C2) and (C3) are usual smooth conditions (Li and Liang 2008). Condition (C4) is a technique condition commonly imposed for conventional nonparametric regression analysis. Condition (C5) is needed for Taylor expansion and ensures asymptotic variance finite. Condition (C6) is commonly imposed for nonparametric kernel smoothing. Condition (C7) is generally required for bandwidths h and b_k in semiparametric setting. Condition (C8) is a technique condition involved in the SCAD variable selection procedure (Fan and Li 2001; Liang and Li 2009).

Proposition 1 *Under the conditions (C4), (C6) and (C7), we have*

$$\begin{aligned} & \hat{\xi}_k(v) - \xi_k(v) \\ &= \frac{\mu_{L2}}{2} b_k^2 \xi_k^{(2)}(v) + \frac{1}{nf_V(v)} \sum_{i=1}^n L_{b_k}(V_i - v) e_{ki} + o\left(b_k^2 + \log b_k^{-1} / \sqrt{nb_k}\right) \end{aligned} \tag{5}$$

holds uniformly on $v \in \mathcal{V}$, where $\mu_{Lj} = \int u^j L(u) du$, $\xi_k^{(2)}(v)$ is the second derivatives of $\xi_k(v)$, e_{ki} is the k th component of e_i , $i = 1, \dots, n$.

The proof of (5) can be completed in a way similar to Zhou and Liang (2009).

2.2 Quasi-likelihood-based estimation

After we calibrate ξ , we model the ‘‘synthesis’’ data $\{Y_i, U_i, \hat{\xi}_i, W_i, X_i; 1 \leq i \leq n\}$ using the local likelihood principle (Fan and Gijbels 1996) to estimate $\beta, \theta, \alpha(\cdot)$ based on the model:

$$g\{\mu(U, \hat{\xi}, W, X)\} = \beta^T \hat{\xi} + \theta^T W + \alpha(U)^T X. \tag{6}$$

Specifically, let h be the bandwidth, $K(\cdot)$ be the kernel function satisfying the condition (C6), and $K_h(\cdot) = h^{-1}K(\cdot/h)$. For each u in a neighborhood of U , we approximate $\alpha_j(U)$ by $\alpha_{0j}(u) + \alpha'_{0j}(u)(U - u)$, $j = 1, \dots, p$. Let $\alpha(u) = (\alpha_{01}(u), \dots, \alpha_{0p}(u))^T$, $b(u) = (\alpha'_{01}(u), \dots, \alpha'_{0p}(u))^T$. The estimators of $\beta, \theta, \alpha_j(u)$'s and $\alpha'_j(u)$'s are obtained by maximizing the following local quasi-likelihood function with respect to $\alpha(u), b(u), \beta, \theta$,

$$\begin{aligned} & \mathcal{L}_{loc}(a(u), b(u), \beta, \theta) \\ &= \sum_{i=1}^n \mathcal{Q} \left[g^{-1}(\beta^T \hat{\xi}_i + \theta^T W_i + \alpha(u)^T X_i + b(u)^T X_i(U_i - u)), Y_i \right] K_h(U_i - u), \end{aligned} \tag{7}$$

where $\mathcal{Q}(x, y)$ is the quasi-likelihood function and is defined as $\mathcal{Q}(x, y) = \int_y^x \frac{y-u}{T(u)} du$. Denote the local quasi-likelihood estimators from (7) by $\hat{\alpha}_*(u), \hat{b}_*(u), \hat{\beta}_*, \hat{\theta}_*$. As

demonstrated in Lemma A.2 in Appendix that these estimators are all \sqrt{nh} -consistent (or $\sqrt{nh_o}$ -consistent, under Condition (C7)).

We now update estimates of β and θ using all data, through by considering a global quasi-likelihood procedure for improving efficiency. Define

$$\mathcal{L}_{gol}(\beta, \theta) = \sum_{i=1}^n \mathcal{Q} \left[g^{-1}(\beta^T \hat{\xi}_i + \theta^T W_i + \hat{\alpha}_*(U_i)^T X_i), Y_i \right], \tag{8}$$

where $\hat{\alpha}_*(u)$ is obtained from (7). As a result, we have global quasi-likelihood estimators $\hat{\beta}$ and $\hat{\theta}$ by maximizing $\mathcal{L}_{gol}(\beta, \theta)$. The corresponding estimators have the same merit as one-step backfitting algorithm estimates. One may also consider a full iterative backfitting algorithm or a profile likelihood approach to obtain estimators of β, θ .

In the following, we introduce some notations for presenting the properties of the estimators. Denote $A^{\otimes 2} = AA^T$ for any matrix or vector A. Let $q_\ell(x, y) = \frac{\partial^\ell}{\partial x^\ell} \mathcal{Q}\{g^{-1}(x), y\}$ for $\ell = 1, 2$. Then $q_1(x, y) = \{y - g^{-1}(x)\}\rho_1(x)$, $q_2(x, y) = \{y - g^{-1}(x)\}\rho_1'(x) - \rho_2(x)$ with $\rho_\ell(x) = \left\{ \frac{dg^{-1}(x)}{dx} \right\}^\ell / [\sigma^2 T\{g^{-1}(x)\}]$. Let $Z = \beta^T \xi + \theta^T W + \alpha(U)^T X$, $Q = (\xi^T, W^T)^T$, $N = (\xi^T, W^T, X^T)^T$, $\Sigma = E[\rho_2(Z)Q^{\otimes 2}]$. Denote by $\kappa_k(u)$ the k th element of $E[\rho_2(Z)N^{\otimes 2}|U = u]^{-1}N$, $\iota_k(u, v)$ is the k th element of $E[\rho_2(Z)N^{\otimes 2}|U = u]^{-1}E[\rho_2(Z)N|U = u, V = v]$. Moreover,

$$\begin{aligned} \Gamma(u) &= \left\{ Q - \sum_{k=1}^p \kappa_k(u) E[\rho_2(Z)QX_k|U = u] \right\} q_1(Z, Y), \\ x_k(v) &= E \left[QX_k \rho_2(Z) \iota_k(U, v) \frac{f_{U,V}(U, v)}{f_U(U) f_V(v)} \right], \\ \Lambda(v) &= \left\{ \sum_{k=1}^p x_k(v) - E[\rho_2(Z)Q|V = v] \right\} e^T \beta. \end{aligned}$$

We have the following asymptotic results.

Theorem 1 Under Conditions (C1)–(C7) given in the Appendix, we have

$$\begin{aligned} &\sqrt{n}((\hat{\beta} - \beta)^T, (\hat{\theta} - \theta)^T)^T \\ &\xrightarrow{\mathcal{L}} N_q \left(0, \Sigma^{-1} E[\Gamma(U)^{\otimes 2}] \Sigma^{-1} + \Sigma^{-1} E[\Lambda(V)^{\otimes 2}] \Sigma^{-1} \right). \end{aligned}$$

Remark 1 To ensure Theorem 1 holds, undersmoothing is necessary. This strategy concurs with that adapted in modeling GPLM (Severini and Staniswalis 1994). In the asymptotic variance, the first term $\Sigma^{-1} E[\Gamma(U)^{\otimes 2}] \Sigma^{-1}$ is similar to that obtained by Li and Liang (2008) for partially linear models, while the second term $\Sigma^{-1} E[\Lambda(V)^{\otimes 2}] \Sigma^{-1}$ is owing to the impact of measurement error and a bias correction in virtue of the ancillary variable V .

Bandwidth selection The proposed procedure involves the bandwidth b_k and h , to be selected. As indicated in Zhou and Liang (2009), the undersmoothing is necessary when we estimate ξ . So, the optimal bandwidth for b_k has to be violated. The consequence of undersmoothing ξ is to keep the bias small and preclude the optimal bandwidth for b_k . As suggested by Carroll et al. (1997), an ad hoc but reasonable choice is $O(n^{-1/5}) \times n^{-2/15} = O(n^{-1/3})$. The suitable bandwidth b_k is $b_k = C_1 n^{-1/3}$, where C_1 is a positive constant. One can use a plug-in rule to estimate the constant C_1 , i.e., $b_k = \hat{\sigma}_V n^{-1/3}$. Another selection for b_k can be chosen as $b_k = n^{-2/15} \hat{b}_{k*}$, where $\hat{b}_{k*} = \arg \min_{b_*} CV_k(b_*)$, $CV_k(b_*) = n^{-1} \sum_{i=1}^n \left\{ \eta_{ik} - \hat{\xi}_{k,b_*}^{(-i)}(V_i) \right\}^2$, where $\hat{\xi}_{k,b_*}^{(-i)}(V_i)$ is computed analogous to (3) from the data with the i th observation η_i, V_i deleted and bandwidth b_* . To select h , we first define the “leave-one-sample out” method $h_1 = \arg \min_{h_*} \sum_{i=1}^n \mathcal{Q} \left[g^{-1}(\hat{\beta}_{-i}^T \hat{\xi}_i + \hat{\theta}_{-i}^T W_i + \hat{\alpha}_{-i,h_*}(u)^T X_i), Y_i \right]$, where $\hat{\beta}_{-i}, \hat{\theta}_{-i}$ are obtained from (8), and $\hat{\alpha}_{-i,h_*}(U_i)$ is obtained from (7) with the fixed bandwidth h_* and the leave-one-out sample $\{Y_j, \hat{\xi}_j, W_j, X_j, U_j\}_{1 \leq j \neq i \leq n}$.

2.3 Penalized quasi-likelihood-based variable selection

In this section, we consider the variable selection problem. We define the penalized quasi-likelihood as

$$\mathcal{L}_P(\beta, \theta) = \mathcal{L}_{gol}(\beta, \theta) - n \sum_{j=1}^d p_{\lambda_{1j}}(|\beta_j|) - n \sum_{s=1}^r p_{\lambda_{2s}}(|\theta_s|), \tag{9}$$

where $p_{\lambda_{1j}}(\cdot), p_{\lambda_{2s}}(\cdot)$ are penalty functions, and λ_{1j} and λ_{2s} are tuning parameters. We distinctively choose tuning parameters λ_1 's, λ_2 's for identifying nonzero elements of β and θ . If we are only interested in selecting W -variable, then we set $p_{\lambda_{1j}}(\cdot) = 0, j = 1, \dots, p$. Similarly, we can commit only on ξ -variable.

We first briefly discuss the choice of penalty functions. There have been many penalty functions in the variables selection literature. For example, L_0 -penalty, $p_{\lambda_{1j}}(|\beta_j|) = 0.5\lambda_{1j}^2 I\{|\beta_j| \neq 0\}$, where $I\{\cdot\}$ is an indicator function. Specially, if we further let $\lambda_{1j} = \sigma\sqrt{2/n}, \sigma\sqrt{\log(n)/n}$ and $\sigma\sqrt{\log(d)/n}$, those penalty functions correspond to the popular variable selection criteria such as AIC (Akaike 1973), BIC (Schwarz 1978) and RIC (Foster and George 1994). We adopt SCAD penalty (Fan and Li 2001), whose first derivative is

$$p'_\lambda(\gamma) = \lambda \left\{ I(\gamma \leq \lambda) + \frac{(a\lambda - \gamma)_+}{(a - 1)\lambda} I(\gamma > \lambda) \right\},$$

where $(s)_+ = sI(s > 0)$ is the hinge loss function and $a = 3.7$.

We next study the asymptotic properties of the resulting penalized quasi-likelihood estimates. Without loss of generality, assume the first d_1 components of β are nonzeros, the first r_1 components of θ are nonzeros. I.e., $\beta_s \neq 0, s = 1, \dots, d_1, \theta_l \neq 0, l = 1, \dots, r_1$ and $\beta_k \equiv 0, k = d_1 + 1, \dots, d, \theta_t \equiv 0, t = r_1 + 1, \dots, r$.

For notational simplicity, denote $\mathcal{R}_{n,\lambda_1,\lambda_2} = (\mathcal{R}_{n,\lambda_1}^T, \mathcal{R}_{n,\lambda_2}^T)^T$ with

$$\begin{aligned} \mathcal{R}_{n,\lambda_1} &= \{p'_{\lambda_{1j}}(|\beta_j|)\text{sign}(\beta_j), \dots, p'_{\lambda_{1d_1}}(|\beta_{d_1}|)\text{sign}(\beta_{d_1})\}^T, \\ \mathcal{R}_{n,\lambda_2} &= \{p'_{\lambda_{21}}(|\theta_1|)\text{sign}(\theta_1), \dots, p'_{\lambda_{2r_1}}(|\theta_{r_1}|)\text{sign}(\theta_{r_1})\}^T, \end{aligned}$$

and we further define

$$\begin{aligned} a_n^* &= \max_{1 \leq j \leq d} \{p'_{\lambda_{1j}}(|\beta_j|), \beta_j \neq 0\}, \quad b_n^* = \max_{1 \leq s \leq r} \{p'_{\lambda_{2s}}(|\theta_s|), \theta_s \neq 0\}, \\ a_n^{**} &= \max_{1 \leq j \leq d} \{p''_{\lambda_{1j}}(|\beta_j|), \beta_j \neq 0\}, \quad b_n^{**} = \max_{1 \leq s \leq r} \{p''_{\lambda_{2s}}(|\theta_s|), \theta_s \neq 0\}, \\ \Sigma_{n,\lambda_1,\lambda_2} &= \text{diag}\{p''_{\lambda_{11}}(|\beta_1|), \dots, p''_{\lambda_{1d_1}}(|\beta_{d_1}|), p''_{\lambda_{21}}(|\theta_1|), \dots, p''_{\lambda_{2r_1}}(|\theta_{r_1}|)\}. \end{aligned}$$

Denote the resulting penalized estimators from (9) by $\hat{\beta}_{\lambda_1}, \hat{\theta}_{\lambda_2}$. We have the following asymptotic results.

Theorem 2 *Under Conditions (C1)–(C8) given in the Appendix, moreover, suppose $a_n^* = O(n^{-1/2}), b_n^* = O(n^{-1/2}), a_n^{**} \rightarrow 0, b_n^{**} \rightarrow 0$, then there exist local maximizers $\hat{\beta}_{\lambda_1}, \hat{\theta}_{\lambda_2}$ of (9) such that their rates of convergence are $\hat{\beta}_{\lambda_1} = \beta + O_P(n^{-1/2})$ and $\hat{\theta}_{\lambda_2} = \theta + O_P(n^{-1/2})$.*

We further introduce notations for presenting the oracle properties of the resulting penalized likelihood estimates. Without loss of generality, denote $\beta = (\beta_{(1)}^T, \beta_{(2)}^T)^T$, $\theta = (\theta_{(1)}^T, \theta_{(2)}^T)^T$, where $\beta_{(1)}$ and $\theta_{(1)}$ are d_1 and r_1 nonzero components of β and θ , respectively, and $\beta_{(2)}$ and $\theta_{(2)}$ are two $(d - d_1)$ - and $(r - r_1) \times 1$ -zero vectors. Accordingly, $\xi_{(1)}$ and $W_{(1)}$ are the first d_1 covariates of ξ , and the first r_1 covariates of W . Let $Z_{(1)} = \beta_{(1)}^T \xi_{(1)} + \theta_{(1)}^T W_{(1)} + \alpha(U)^T X$, $Q_{(1)} = (\xi_{(1)}^T, W_{(1)}^T)^T$, $N_{(1)} = (\xi_{(1)}^T, W_{(1)}^T, X^T)^T$, and $e_{(1)}$ be the first d_1 covariates of the error e . Moreover, the definitions of $\Sigma_{(1)}, \Gamma_{(1)}(u)$ and $\Lambda_{(1)}(v)$ and the terms involved in these definitions are accordingly to $\Sigma, \Gamma(u), \Lambda(v)$ by substituting β, Z, Q, N, e with $\beta_{(1)}, Z_{(1)}, Q_{(1)}, N_{(1)}, e_{(1)}$, respectively.

Theorem 3 *Under Conditions (C1)–(C8), the penalized estimators $\hat{\beta}_{\lambda_1} = (\hat{\beta}_{\lambda_1(1)}^T, \hat{\beta}_{\lambda_1(2)}^T)^T$ and $\hat{\theta}_{\lambda_2} = (\hat{\theta}_{\lambda_2(1)}^T, \hat{\theta}_{\lambda_2(2)}^T)^T$ satisfy: (a) with probability tending to one, $\hat{\beta}_{\lambda_1(2)} = \mathbf{0}, \hat{\theta}_{\lambda_2(2)} = \mathbf{0}$; and (b) $\hat{\beta}_{\lambda_1(1)}$ and $\hat{\theta}_{\lambda_2(1)}$ are asymptotically normal, i.e.,*

$$\begin{aligned} &\sqrt{n}(\Sigma_{(1)} + \Sigma_{n,\lambda_1,\lambda_2}) \left\{ ((\hat{\beta}_{\lambda_1(1)} - \beta_{(1)})^T, (\hat{\theta}_{\lambda_2(1)} - \theta_{(1)})^T)^T \right. \\ &\quad \left. + (\Sigma_{(1)} + \Sigma_{n,\lambda_1,\lambda_2})^{-1} \mathcal{R}_{n,\lambda_1,\lambda_2} \right\} \\ &\xrightarrow{\mathcal{L}} N_{d_1+r_1} \left(\mathbf{0}, \Sigma_{(1)}^{-1} E \left[\Gamma_{(1)}(U)^{\otimes 2} \right] \Sigma_{(1)}^{-1} + \Sigma_{(1)}^{-1} E \left[\Lambda_{(1)}(V)^{\otimes 2} \right] \Sigma_{(1)}^{-1} \right). \end{aligned}$$

Remark 2 Theorem 3 indicates that the proposed variable selection procedure processes the oracle property with proper choices of tuning parameters λ_{1j} 's, λ_{2s} 's. If we further demand that $\sqrt{n}\mathcal{R}_{n,\lambda_1,\lambda_2} \rightarrow 0$, and $\Sigma_{n,\lambda_1,\lambda_2} \rightarrow 0$, the asymptotic variance simplifies to summand of $\Sigma_{(1)}^{-1} E \left[\Gamma_{(1)}(U)^{\otimes 2} \right] \Sigma_{(1)}^{-1}$ and $\Sigma_{(1)}^{-1} E \left[\Lambda_{(1)}(V)^{\otimes 2} \right] \Sigma_{(1)}^{-1}$.

Choice of λ_j 's. We adopt a data-driven GCV procedure proposed by Li and Liang (2008) to select the tuning parameters λ_1 's, λ_2 's in a $d + r$ -dimensional space. Let $\lambda_{1j} = \lambda * \text{Se}(\hat{\beta}_j)$, $\lambda_{2i} = \lambda * \text{Se}(\hat{\theta}_i)$, where $\text{Se}(\hat{\beta}_j)$ and $\text{Se}(\hat{\theta}_i)$ are the estimated standard error of $\hat{\beta}_j, \hat{\theta}_i$. Thus, the minimization over λ_1 's, λ_2 's is simplified to an one-dimensional minimization through λ . We first introduce the estimation procedure for the standard errors, which can be obtained from the estimated covariance matrix $\widehat{\text{Cov}}(\hat{\gamma})$, where $\hat{\gamma} = (\hat{\beta}^T, \hat{\theta}^T)^T$ is obtained from (8). Write $\ell'(\gamma) = \frac{\mathcal{L}_{\text{gol}}(\beta, \theta)}{\partial \gamma} \ell''(\gamma) = \frac{\mathcal{L}_{\text{gol}}(\beta, \theta)}{\partial \gamma \partial \gamma^T}$, $\gamma = (\beta^T, \theta^T)^T$ and

$$\Sigma_{n, \lambda_1, \lambda_2}^* = \text{diag} \left(\frac{p'_{\lambda_{11}}(|\beta_{11}|)}{|\beta_{11}|}, \dots, \frac{p'_{\lambda_{1d}}(|\beta_d|)}{|\beta_d|}, \frac{p'_{\lambda_{21}}(|\theta_1|)}{|\theta_1|}, \dots, \frac{p'_{\lambda_{2r}}(|\theta_r|)}{|\theta_r|} \right). \tag{10}$$

A sandwich formula for the covariance matrix of the estimates $\hat{\gamma} = (\hat{\beta}^T, \hat{\theta}^T)^T$ is given by

$$\widehat{\text{Cov}}(\hat{\gamma}) = \{ \ell''(\hat{\gamma}) - n \Sigma_{n, \lambda_1, \lambda_2}^* \}^{-1} \widehat{\text{Cov}}(\ell'(\hat{\gamma})) \{ \ell''(\hat{\gamma}) - n \Sigma_{n, \lambda_1, \lambda_2}^* \}^{-1}.$$

Write $e(\lambda) = \text{tr} \left\{ \left\{ \ell''(\hat{\gamma}) - n \Sigma_{n, \lambda, \lambda}^* \right\}^{-1} \ell''(\hat{\gamma}) \right\}$, where $\Sigma_{n, \lambda_1, \lambda_2}^*$ is obtained from (10)

by substituting $\lambda_{1j}, \lambda_{2i}$ with $\lambda * \text{Se}(\hat{\beta}_j), \lambda * \text{Se}(\hat{\theta}_i)$ respectively. The GCV statistic is defined by

$$\text{GCV}(\lambda) = \frac{\sum_{i=1}^n \mathcal{D} \left\{ Y_i, g^{-1}(\hat{\alpha}^T(U_i)X_i + \hat{\xi}_i^T \hat{\beta}(\lambda) + W_i^T \hat{\theta}(\lambda)) \right\}}{n\{1 - e(\lambda)/n\}^2},$$

where $\mathcal{D}\{Y, \mu\}$ denotes the deviance of Y corresponding to the model fitting with λ . The minimizer of $\text{GCV}(\lambda)$ with respect to λ can be obtained by a grid search.

3 Statistical inference for nonparametric components

In this section, we consider a refined estimator of $\alpha(u)$ and propose a generalized likelihood ratio test to select significant components of X .

3.1 Refined estimator of nonparametric component

After we obtain the final estimators of β and θ from Sect. 2.2, the estimator of $\alpha(u)$ can be refined by maximizing the following local likelihood function:

$$\begin{aligned} &\mathcal{L}_{loc}^*(a(u), b(u)) \\ &= \sum_{i=1}^n \mathcal{Q} \left[g^{-1}(\hat{\beta}^T \hat{\xi}_i + \hat{\theta}^T W_i + \alpha(u)^T X_i + b(u)^T X_i(U_i - u)), Y_i \right] K_h(U_i - u) \end{aligned} \tag{11}$$

with respect to $\alpha(u)$ and $b(u)$. Let $\hat{\alpha}(u)$ be the maximizer of (11). We have the following asymptotic result.

Theorem 4 *Under Conditions (C1)–(C7), we have*

$$\begin{aligned} & \sqrt{nh} \left(\hat{\alpha}(u) - \alpha(u) - \frac{h^2 \mu_{K_2}}{2} \alpha''(u) \right. \\ & \quad \left. + \frac{b^2 \mu_{L_2}}{2} \Sigma_X(u)^{-1} E \left[\rho_2(Z) \xi^{(2)}(V)^T \beta X \mid U = u \right] \right) \\ & \xrightarrow{\mathcal{L}} N \left(\mathbf{0}, \frac{v_{K_0}}{f_U(u)} \Sigma_X(u)^{-1} \right), \end{aligned}$$

where $\Sigma_X(u) = E \left[\rho_2(Z) X^{\otimes 2} \mid U = u \right]$, and $\mu_{K_2} = \int t^2 K(t) dt$, $\mu_{L_2} = \int t^2 L(t) dt$, $v_{K_0} = \int K^2(t) dt$.

Remark 3 The second term in the asymptotic bias of $\hat{\alpha}(u)$ is owing to calibrating the error-prone covariates. In fact, we can eliminate two bias terms $O(h^2)$ and $O(b^2)$ if we adapt the undersmoothing strategy in order for $\hat{\beta}, \hat{\theta}$ being root- n consistent. As such, the bias of $\hat{\alpha}(u)$ tends to zero and the rates of $\hat{\alpha}(u)$ are $(nh_o)^{1/2}$.

3.2 Variable selection for nonparametric component

It is of interest to select nonzero component of $\alpha(u)$ to increase model prediction. In this section, we adopt the GLRT proposed by Fan et al. (2001) to detect significant components of X , achieved by using the backward elimination procedure. In each step, we test $H_0 : \alpha_{j_1}(u) = \dots = \alpha_{j_k}(u) = 0$ versus $H_1 : \text{not all } \alpha_{j_l}(u) \neq 0$. For ease of presentation, we consider the following hypothesis:

$$H_0 : \alpha_1(u) = \dots = \alpha_p(u) = 0 \quad \text{versus} \quad H_1 : \text{not all } \alpha_j(u) \neq 0. \quad (12)$$

Let $\hat{\alpha}(u)$, $\hat{\beta}, \hat{\theta}$ be the estimators obtained from (8) and (11) under the alternative hypothesis, and $\bar{\beta}$ and $\bar{\theta}$ be the estimators of β, θ under the null hypothesis. Write

$$\mathcal{H}_1 = \sum_{i=1}^n \mathcal{Q} \left\{ g^{-1} \left(\hat{\alpha}(U_i)^T X_i + \hat{\theta}^T \hat{\xi}_i + \hat{\theta}^T W_i \right), Y_i \right\}$$

and

$$\mathcal{H}_0 = \sum_{i=1}^n \mathcal{Q} \left\{ g^{-1} \left(\bar{\beta}^T \hat{\xi}_i + \bar{\theta}^T W_i \right), Y_i \right\}.$$

Following Fan et al. (2001) and Li and Liang (2008), we define the GLRT statistic

$$\mathcal{T}_{\text{GLR}} = \mathcal{H}_1 - \mathcal{H}_0.$$

Define $v_{L_0} = \int L^2(t)dt, v_{K_0} = \int K^2(t)dt, \sigma_K^2 = 2p \left\{ \int [2K(t) - K * K(t)]^2 dt \right\}^2$ $|\mathcal{U}|$ with $|\mathcal{U}|$ being the length of the support of U . $\sigma_L^2 = 2 \left\{ \int [L * L(t)]^2 dt \right\}^2$ $E \left\{ \frac{[E[\rho_2(Z)|V]]^2}{f_V(V)} \right\} (\beta^T \Sigma_e \beta)^2$. $K * K(t), L * L(t)$ are the convolutions of $K(t), L(t)$, respectively. c_b and c_h are two positive constants satisfying Condition (C7). We have the following theorem.

Theorem 5 Under Conditions (C1)–(C7), $r_{LK}(\mathcal{T}_{GLR} - \chi_{df_n}^2) \xrightarrow{\mathcal{L}} 0$ under the null hypothesis \mathcal{H}_0 , here

$$r_{LK} = \frac{8c_b^{-1} v_{L_0} \beta^T \Sigma_e \beta E \left\{ \frac{E[\rho_2(Z)|V]}{f_V(V)} \right\} + 8c_h^{-1} p |\mathcal{U}| [K(0) - 0.5v_{K_0}]}{c_b^{-1} \sigma_L^2 + c_h^{-1} \sigma_K^2}, \text{ and}$$

$$df_n = \frac{v_{L_0} \beta^T \Sigma_e \beta}{b} E \left[\frac{E[\rho_2(Z)|V]}{f_V(V)} \right] + \frac{p|\mathcal{U}|}{h} [K(0) - 0.5v_{K_0}].$$

Remark 4 Theorem 5 claims that the Wilks type of phenomenon holds for GVC-PLMeM. The first part of df_n gains insight into the effect of measurement error and ancillary variable. As indicated in Li and Liang (2008), this generalized likelihood ratio theory can be justified using empirical procedure, such as Monte Carlo simulation or a bootstrap procedure, since the degrees of freedom df_n tend to infinity as sample size n increases. It is worth mentioning that the main order of the degree of freedom $r_{LK}df_n$ cannot be obtained similarly to those in Fan et al. (2001), because Σ_e, β and $E \left[\frac{E[\rho_2(Z)|V]}{f_V(V)} \right]$ are usually unknown in practice and needed to be estimated from the data, and their estimators may not perform well when sample sizes are small or moderate. Moreover, those constants c_b, c_h involved in Condition (C7) for the bandwidth h, b are also unknown. If the covariate ξ can be observed without measurement errors, i.e., $\Sigma_e = 0$, the c_b, c_h are vanished in r_{LK} and df_n , and the method of calibration formulas for degree of freedoms proposed by Zhang (2004) can be directly applied. For these reasons and for practical purposes, one can follow the conditional bootstrap procedure suggested by Zhou and Liang (2009) and Cai et al. (2000) to estimate null distribution of \mathcal{T}_{GLR} .

Remark 5 Under the Conditions (C1)–(C7), we can have the following asymptotic

$$\sqrt{n}((\bar{\beta} - \beta)^T, (\bar{\theta} - \theta)^T)^T \xrightarrow{\mathcal{L}} N_q(0, \bar{\Sigma}^{-1} \bar{\Gamma} \bar{\Sigma}^{-1} + \bar{\Sigma}^{-1} E[\bar{\Lambda}(V)^{\otimes 2}] \bar{\Sigma}^{-1}). \tag{13}$$

where $\bar{\Sigma} = E[\rho_2(Z_*)Q^{\otimes 2}], \bar{\Gamma} = E[q_1^2(Z_*, Y)Q^{\otimes 2}], \bar{\Lambda}(v) = E[\rho_2(Z_*)Q|V = v]$ $e^T \beta$ and $Z_* = \beta^T \xi + \theta^T W$. The asymptotic relative efficiency (ARE) of $\bar{\beta}, \bar{\theta}$ with respect to $\hat{\beta}, \hat{\theta}$ obtained in (8) is

$$\text{ARE} \left((\bar{\beta}, \bar{\theta}), (\hat{\beta}, \hat{\theta}) \right) = \frac{\|\Sigma\|_D^{2/q} \|\bar{\Gamma} + E[\bar{\Lambda}(V)^{\otimes 2}]\|_D^{1/q}}{\|\bar{\Sigma}\|_D^{2/q} \|E[\Gamma(U)^{\otimes 2}] + E[\Lambda(V)^{\otimes 2}]\|_D^{1/q}},$$

where $\|\cdot\|_D$ denotes the determinants of the covariance matrices.

4 Numerical studies

In this section, we conduct simulation studies to assess the performance of the proposed method. We then apply our method to analyze a real data set from a diabetes study. We used the Epanechnikov kernel function $L(t) = K(t) = 0.75(1 - t^2)_+$ in all numerical studies below. Note Condition (C7) means that the optimal bandwidth cannot be used because undersmoothing is necessary. As such, we used the rule of thumb (Silverman 1986). The smoothing parameter b was chosen as $\hat{\sigma}_V n^{-1/3}$, where $\hat{\sigma}_V$ is the sample deviation of V . This choice of b naturally meets Condition (C7). As pointed out in Remark 1 of Sect. 3.2, undersmoothing for h is an usual requirement for fitting generalized semiparametric models.

In our simulation studies, we generated 500 data sets consisting of $n = 500$ and $n = 1000$ observations from the semiparametric coefficient logistic regression model:

$$\logit\{p(U, \xi, W, X)\} = \xi^T \beta + \theta^T W + \alpha(U)^T X \tag{14}$$

with covariates, nonparametric functions and parameters being explicitly specified below.

4.1 Simulation studies

Example 1 $\beta = 2, \theta = (3, 1.5, 2)$ or $\beta = 0.2, \theta = (0.3, 0.15, 0.2)$. $X = (1, X)^T$ and $X \sim N(0, 1)$, $\alpha(u) = (\alpha_1(u), \alpha_2(u))^T$, $\alpha_1(u) = \exp(2u - 1)$, $\alpha_2(u) = 2 \sin^2(2\pi u)$. ξ is unobserved and remitted by (η, V) through $\eta = \xi(V) + e$ with $\xi(V) = 3V - \cos(V)$, $V \sim N(0, 0.5^2)$ and is independent of (U, W, X) , e follows $N(0, 0.5^2)$ and is independent of (U, V, W, X) . We consider three cases: (i) W is independent of U , $W \sim N(\mathbf{0}, \Sigma_W)$, $\Sigma_W = (\sigma_{w,ij})$ with $\sigma_{w,ij} = 0.25^{|i-j|}$, $U \sim \text{Unif}[0, 1]$. (ii) (W, U) follows $\text{Unif}[-1, 1]$, and $\text{Var}((W^T, U)^T) = (\sigma_{ij})$ with $\sigma_{ij} = 0.5^{|i-j|}$. (iii) The first component of W is 0 with probability 0.5 and 1 with probability 0.5, the rest components of W are normally distributed with mean 0, and $\text{Var}(W) = (\sigma_{w,i'j'})$ with $\sigma_{w,i'j'} = 0.5^{|i'-j'|}$, $U \sim \text{Unif}[0, 1]$ and is independent of W . In this example, we use the bandwidth $h = 3 \times n^{-1/3}$.

The simulation results for the benchmark estimator (i.e., all covariates are measured exactly), the proposed estimator and the naive estimator (using η directly) are presented in Tables 1 and 2, which reports the mean and associated standard errors of $(\hat{\beta}, \hat{\theta})$. We can see that the estimated values based on the proposed procedure and the benchmark procedure are close to the true value in all three cases. This indicates our proposed method is promising. However, the naive estimator has severe bias and performs worse, especially when the sample size $n = 500$.

Example 2 In this example, we examined the performance of the proposed variable selection procedure by comparing it with the traditional subset selection criteria such as AIC, BIC and RIC from model (14). Let $\beta = (-0.5, 0.5)^T$. X and $\alpha(u)$ are the same as in Example 1. $\xi(V) = (\xi_1(V), \xi_2(V))^T$, $\xi_1(V) = 2 \cos(V)$, $\xi_2(V) = 0.1 \exp(V) + 3 \sin(V)$, and the ancillary variable $\eta = (\eta_1, \eta_2)^T$ with $\eta_1 = \xi_1(V) + e_1$,

Table 1 The simulation results for Example 1

| | | $\beta_1 = 3$ | $\beta_2 = 1.5$ | $\beta_3 = 2$ | $\theta_1 = 2$ |
|-------------------------|------|---------------|-----------------|---------------|----------------|
| <i>Case 1, n = 500</i> | | | | | |
| <i>B</i> | MEAN | 2.9217 | 1.4416 | 1.9215 | 1.9498 |
| | SD | 0.3147 | 0.2421 | 0.2458 | 0.1876 |
| <i>P</i> | MEAN | 2.9211 | 1.4713 | 1.9329 | 1.9531 |
| | SD | 0.2987 | 0.2342 | 0.2433 | 0.1901 |
| <i>N</i> | MEAN | 2.4321 | 1.2186 | 1.6298 | 1.4623 |
| | SD | 0.2567 | 0.2218 | 0.2169 | 0.1631 |
| <i>Case 1, n = 1000</i> | | | | | |
| <i>B</i> | MEAN | 2.9350 | 1.4643 | 1.9410 | 1.9653 |
| | SD | 0.2699 | 0.2040 | 0.2112 | 0.1710 |
| <i>P</i> | MEAN | 2.9370 | 1.4651 | 1.9426 | 1.9662 |
| | SD | 0.2710 | 0.2048 | 0.2119 | 0.1708 |
| <i>N</i> | MEAN | 2.6123 | 1.3067 | 1.7252 | 1.5802 |
| | SD | 0.2354 | 0.1916 | 0.1828 | 0.1370 |
| <i>Case 2, n = 500</i> | | | | | |
| <i>B</i> | MEAN | 2.9417 | 1.4289 | 1.9921 | 1.9431 |
| | SD | 0.3929 | 0.3621 | 0.4371 | 0.1756 |
| <i>P</i> | MEAN | 2.9518 | 1.4311 | 2.0233 | 1.9530 |
| | SD | 0.3876 | 0.3817 | 0.4231 | 0.1821 |
| <i>N</i> | MEAN | 2.4219 | 1.1190 | 1.7098 | 1.4812 |
| | SD | 0.3623 | 0.3176 | 0.3435 | 0.1451 |
| <i>Case 2, n = 1000</i> | | | | | |
| <i>B</i> | MEAN | 2.9601 | 1.4464 | 2.0143 | 1.9683 |
| | SD | 0.3802 | 0.3576 | 0.4162 | 0.1678 |
| <i>P</i> | MEAN | 2.9660 | 1.4484 | 2.0106 | 1.9666 |
| | SD | 0.3813 | 0.3580 | 0.4165 | 0.1669 |
| <i>N</i> | MEAN | 2.6011 | 1.2820 | 1.7766 | 1.5700 |
| | SD | 0.3373 | 0.3223 | 0.3756 | 0.1258 |
| <i>Case 3, n = 500</i> | | | | | |
| <i>B</i> | MEAN | 2.9287 | 1.4632 | 1.9389 | 1.9590 |
| | SD | 0.3715 | 0.3321 | 0.3219 | 0.2741 |
| <i>P</i> | MEAN | 2.9290 | 1.4731 | 1.9466 | 1.9631 |
| | SD | 0.3801 | 0.3424 | 0.3310 | 0.2565 |
| <i>N</i> | MEAN | 2.5438 | 1.2109 | 1.6654 | 1.4219 |
| | SD | 0.4009 | 0.3817 | 0.3426 | 0.2851 |
| <i>Case 3, n = 1000</i> | | | | | |
| <i>B</i> | MEAN | 2.9355 | 1.4749 | 1.9574 | 1.9669 |
| | SD | 0.3610 | 0.3103 | 0.3132 | 0.2359 |
| <i>P</i> | MEAN | 2.9388 | 1.4764 | 1.9596 | 1.9737 |
| | SD | 0.3605 | 0.3104 | 0.3136 | 0.2366 |
| <i>N</i> | MEAN | 2.6017 | 1.3049 | 1.7302 | 1.5598 |
| | SD | 0.4153 | 0.3028 | 0.3140 | 0.2248 |

MEAN simulation mean, SD standard deviation, *B* benchmark estimator, *P* proposed estimator, and *N* naive estimator

Table 2 The simulation results for Example 1

| | | $\beta_1 = 0.3$ | $\beta_2 = 0.15$ | $\beta_3 = 0.2$ | $\theta_1 = 0.2$ |
|-------------------------|------|-----------------|------------------|-----------------|------------------|
| <i>Case 1, n = 500</i> | | | | | |
| <i>B</i> | MEAN | 0.2791 | 0.1423 | 0.1899 | 0.1956 |
| | SD | 0.0709 | 0.0771 | 0.0765 | 0.0623 |
| <i>P</i> | MEAN | 0.2799 | 0.1456 | 0.1921 | 0.1943 |
| | SD | 0.0689 | 0.0689 | 0.0712 | 0.0601 |
| <i>N</i> | MEAN | 0.2813 | 0.1421 | 0.1876 | 0.1753 |
| | SD | 0.0697 | 0.0680 | 0.0699 | 0.0521 |
| <i>Case 1, n = 1000</i> | | | | | |
| <i>B</i> | MEAN | 0.2880 | 0.1478 | 0.1950 | 0.2036 |
| | SD | 0.0626 | 0.0638 | 0.0627 | 0.0513 |
| <i>P</i> | MEAN | 0.2880 | 0.1477 | 0.1950 | 0.2039 |
| | SD | 0.0628 | 0.0638 | 0.0627 | 0.0515 |
| <i>N</i> | MEAN | 0.2877 | 0.1476 | 0.1947 | 0.1827 |
| | SD | 0.0627 | 0.0637 | 0.0628 | 0.0498 |
| <i>Case 2, n = 500</i> | | | | | |
| <i>B</i> | MEAN | 0.2913 | 0.1474 | 0.1921 | 0.1921 |
| | SD | 0.0790 | 0.0753 | 0.0756 | 0.0523 |
| <i>P</i> | MEAN | 0.2901 | 0.1466 | 0.1914 | 0.1910 |
| | SD | 0.0821 | 0.0760 | 0.0766 | 0.0548 |
| <i>N</i> | MEAN | 0.2914 | 0.1481 | 0.1923 | 0.1611 |
| | SD | 0.0807 | 0.0780 | 0.0791 | 0.0541 |
| <i>Case 2, n = 1000</i> | | | | | |
| <i>B</i> | MEAN | 0.2974 | 0.1501 | 0.1984 | 0.1968 |
| | SD | 0.0733 | 0.0672 | 0.0740 | 0.0447 |
| <i>P</i> | MEAN | 0.2975 | 0.1500 | 0.1984 | 0.1969 |
| | SD | 0.0733 | 0.0672 | 0.0742 | 0.0446 |
| <i>N</i> | MEAN | 0.2971 | 0.1511 | 0.1964 | 0.1772 |
| | SD | 0.0730 | 0.0688 | 0.0768 | 0.0433 |
| <i>Case 3, n = 500</i> | | | | | |
| <i>B</i> | MEAN | 0.2899 | 0.1411 | 0.1919 | 0.1921 |
| | SD | 0.0613 | 0.0579 | 0.0568 | 0.0505 |
| <i>P</i> | MEAN | 0.2910 | 0.1451 | 0.1927 | 0.1919 |
| | SD | 0.0641 | 0.0556 | 0.0578 | 0.0512 |
| <i>N</i> | MEAN | 0.2914 | 0.1427 | 0.1923 | 0.1699 |
| | SD | 0.0693 | 0.0580 | 0.0590 | 0.0473 |
| <i>Case 3, n = 1000</i> | | | | | |
| <i>B</i> | MEAN | 0.2936 | 0.1429 | 0.1939 | 0.1968 |
| | SD | 0.0580 | 0.0540 | 0.0534 | 0.0456 |
| <i>P</i> | MEAN | 0.2939 | 0.1428 | 0.1939 | 0.1969 |
| | SD | 0.0580 | 0.0540 | 0.0534 | 0.0457 |
| <i>N</i> | MEAN | 0.2925 | 0.1430 | 0.1937 | 0.1785 |
| | SD | 0.0593 | 0.0543 | 0.0535 | 0.0437 |

MEAN simulation mean, SD standard deviation, *B* benchmark estimator, *P* proposed estimator, and *N* naive estimator

Table 3 The simulation results for Example 2

| Penalty | B | | | P | | | N | | |
|-------------------------|--------|--------|--------|--------|--------|--------|--------|--------|--------|
| | MedSE | C | I | MedSE | C | I | MedSE | C | I |
| <i>Case 1, n = 500</i> | | | | | | | | | |
| SCAD | 0.0768 | 2.9921 | 0.1100 | 0.0691 | 2.9923 | 0.1123 | 0.4437 | 2.9876 | 1.2313 |
| AIC | 0.1123 | 2.3521 | 0.0231 | 0.0997 | 2.3931 | 0.0210 | 0.4081 | 2.1098 | 0.6697 |
| BIC | 0.0789 | 2.8504 | 0.0450 | 0.0707 | 2.8602 | 0.0421 | 0.4137 | 2.7098 | 0.9980 |
| RIC | 0.0921 | 2.5541 | 0.0390 | 0.0887 | 2.5432 | 0.0387 | 0.3989 | 2.5431 | 0.7213 |
| Oracle | 0.0697 | 3 | 0 | 0.0704 | 3 | 0 | 0.3521 | 3 | 0 |
| <i>Case 1, n = 1000</i> | | | | | | | | | |
| SCAD | 0.0698 | 2.9980 | 0.0900 | 0.0583 | 2.9980 | 0.0980 | 0.3266 | 2.9960 | 0.9540 |
| AIC | 0.0995 | 2.4340 | 0.0080 | 0.0854 | 2.4430 | 0.0060 | 0.3142 | 2.3440 | 0.4480 |
| BIC | 0.0733 | 2.9300 | 0.0220 | 0.0619 | 2.9360 | 0.0180 | 0.3226 | 2.8640 | 0.7520 |
| RIC | 0.0836 | 2.7720 | 0.0160 | 0.0721 | 2.7680 | 0.0160 | 0.3183 | 2.6880 | 0.6300 |
| Oracle | 0.0578 | 3 | 0 | 0.0573 | 3 | 0 | 0.2688 | 3 | 0 |
| <i>Case 2, n = 500</i> | | | | | | | | | |
| SCAD | 0.0756 | 5.8867 | 0.1417 | 0.0776 | 5.8976 | 0.1534 | 0.4351 | 5.4109 | 1.2413 |
| AIC | 0.1320 | 4.6700 | 0.0221 | 0.1215 | 4.5600 | 0.0141 | 0.4213 | 4.4109 | 0.7890 |
| BIC | 0.0734 | 5.7876 | 0.0342 | 0.0765 | 5.6789 | 0.0401 | 0.4670 | 5.2341 | 0.9879 |
| RIC | 0.0756 | 5.6678 | 0.0176 | 0.0876 | 5.4567 | 0.0234 | 0.4567 | 5.0989 | 0.7790 |
| Oracle | 0.0709 | 6 | 0 | 0.0717 | 6 | 0 | 0.4098 | 6 | 0 |
| <i>Case 2, n = 1000</i> | | | | | | | | | |
| SCAD | 0.0684 | 6.0000 | 0.1100 | 0.0578 | 6.0000 | 0.1340 | 0.3215 | 6.0000 | 0.9880 |
| AIC | 0.1120 | 4.9200 | 0.0100 | 0.1000 | 4.9100 | 0.0060 | 0.3192 | 4.8060 | 0.4160 |
| BIC | 0.0709 | 5.9120 | 0.0200 | 0.0582 | 5.9080 | 0.0200 | 0.3102 | 5.8340 | 0.6980 |
| RIC | 0.0783 | 5.7260 | 0.0080 | 0.0682 | 5.7320 | 0.0140 | 0.3102 | 5.6180 | 0.5960 |
| Oracle | 0.0541 | 6 | 0 | 0.0541 | 6 | 0 | 0.2469 | 6 | 0 |

$\eta_1 = \xi_2(V) + e_2$. V is independent of $(e_1, e_2)^T$ and follows $N(0, 1)$. $(e_1, e_2)^T$ follows $N_2((0, 0)^T, \Sigma_e)$ with $\Sigma_e = (\sigma_{e,ij})_{1 \leq i, j \leq 2}$, $\sigma_{e,ij} = (-0.5)^{|i-j|}$. U follows $Unif[0, 1]$. Moreover, $\text{Var}((\xi^T, W)^T) = (\sigma_{o,ij})_{1 \leq i, j \leq q}$ with $\sigma_{o,ij} = 0.5^{|i-j|}$. We considered two cases: $\theta = (1, 0, 0, 1, 0)^T \in \mathbb{R}^5$ and $\theta = (1, 0, 0, 1, 0, 0, 0, 0)^T \in \mathbb{R}^8$.

We examined the following quantities: the median of the squares errors (MedSE) of $\|\hat{\gamma} - \gamma\|_2^2$, the average number (labeled ‘‘C’’) of the three or eight true zero coefficients correctly set to zero, and the average number (labeled ‘‘I’’) of the four true nonzeros incorrectly set to zero. Similar to Example 1, we considered three estimators: the benchmark estimator, the proposed estimator and the naive estimator. The GCV procedure introduced in Sect. 2.3 was used for selecting λ_j ’s. 30 grid points were set to be evenly distributed over the range of λ . The simulation results are reported in Table 3.

We can see that both benchmark estimator and our proposed estimator perform better as the sample size increase to 1000. The values of ‘‘C’’ and ‘‘I’’ are close to the true values 3 in case 1 or 6 in case 2, and 0, respectively. The performance of the SCAD

procedure is similar to that of the oracle procedure and better than the penalized best subset variable selection procedure using AIC and RIC. Moreover, the performance of the SCAD is similar to that of BIC, which costs much more computational time, however. The MedSE of the SCAD and BIC procedures for both benchmark estimator and the proposed estimator are also close to those obtained from the oracle procedure. As anticipated, the naive procedure has a much higher rate of incorrectly setting nonzero coefficients to zero. Especially, the number of SCAD incorrectly setting nonzero coefficients closes to 1 instead of 0 in the two cases. When sample size $n = 1000$, the number of best subset variable selection incorrectly setting nonzero coefficients is at least 0.4 instead of 0. At the same time, the MedSE of the naive estimator is about 0.37 when $n = 500$ and 0.26 when $n = 1000$ even for the oracle setting. This means that ignoring measurement error e increases the chance of identifying more significant components, and causes that one may falsely choose variables and result in an inappropriate model. $\hat{\xi}(v)$ performs well for variable selection.

Example 3 In this example, we examined the performance of the estimation procedure for nonparametric components introduced in Sect. 3.1. $\beta = (1, 1, 1)^T$, $\theta = (-1, 0.5)^T$, $X = (1, X)^T$, where X follows $N(0, 1)$. $\alpha(u) = (\alpha_1(u), \alpha_2(u))^T$, $\alpha_1(u) = 2 \exp(-2u)$, $\alpha_2(u) = 2 \sin^2(\pi u)$. $U, \xi(V), V$ and e are the same as in Example 2. Moreover, $\text{Var}\{(\xi^T, W)^T\} = (\sigma_{o,ij})_{1 \leq i, j \leq q}$ with $\sigma_{o,ij} = 0.5^{|i-j|}$. In this example, we set $h = 0.2$. The performance of the estimator $\hat{\alpha}(u) = (\hat{\alpha}_1(u), \hat{\alpha}_2(u))^T$ was assessed by the square root of average square errors (RASE)

$$\begin{aligned} \text{RASE}_1 &= \left\{ n_0^{-1} \sum_{i=1}^{n_0} \|\hat{\alpha}_1(u_i) - \alpha_1(u_i)\| \right\}^{1/2}, \\ \text{RASE}_2 &= \left\{ n_0^{-1} \sum_{i=1}^{n_0} \|\hat{\alpha}_2(u_i) - \alpha_2(u_i)\| \right\}^{1/2}, \end{aligned}$$

where $\{u_1, \dots, u_{n_0}\}$ are the given grid points, and $n_0 = 200$ is the number of grid points.

We evaluated the estimation procedure (11) for two scenarios: (i) using the estimated $\hat{\gamma} = (\hat{\beta}^T, \hat{\theta}^T)^T$, (ii) using the true value $\gamma = (\beta^T, \theta^T)^T$. We report the simulation mean and standard derivation of RASE_1 and RASE_2 , and the simulation mean and associated stand derivation of $\|\hat{\gamma} - \gamma\|^2$ in Table 4. These results indicate that the performance of both the benchmark estimator and the proposed estimator works well regardless $\hat{\gamma}$ or γ being used. This is not surprising because $\hat{\gamma}$ is root- n consistent with higher convergence rates than nonparametric estimates. As a result, the benchmark estimator and the proposed estimator work satisfactorily under the two scenarios in term of RASE. On the other hand, the naive procedure results in no-ignorable biases in estimation of γ and the biased estimators $\hat{\gamma}$ deteriorate the estimation procedure for $\alpha(\cdot)$ and eventually make $\hat{\alpha}(u)$ larger biases. It is worthy mention that the naive estimator through by using true γ works well since no biases are caused (see the third row under the ‘‘Exact γ ’’ column in Table 4). The estimation for function $\alpha_2(u)$ with the estimated $\hat{\gamma}$ performs as well as if we knew the true value of γ regardless the

Table 4 The simulation results for Example 3

| Penalty | Exact γ | | | | Estimated $\hat{\gamma}$ | | | | | |
|-----------------|----------------|--------|------------|--------|--------------------------|--------|------------|--------|------------|--------|
| | $RASE_1^2$ | SD | $RASE_2^2$ | SD | ME | SD | $RASE_1^2$ | SD | $RASE_2^2$ | SD |
| <i>n</i> = 500 | | | | | | | | | | |
| B | 0.0567 | 0.0357 | 0.0680 | 0.0398 | 0.0713 | 0.0332 | 0.0709 | 0.0520 | 0.0743 | 0.0459 |
| P | 0.0735 | 0.0431 | 0.0754 | 0.0432 | 0.0807 | 0.0413 | 0.0724 | 0.0538 | 0.0776 | 0.0468 |
| N | 0.0776 | 0.0491 | 0.0784 | 0.0451 | 0.8721 | 0.2341 | 0.9124 | 0.2439 | 0.1123 | 0.0676 |
| <i>n</i> = 1000 | | | | | | | | | | |
| B | 0.0359 | 0.0235 | 0.0531 | 0.0333 | 0.0507 | 0.0294 | 0.0527 | 0.0409 | 0.0549 | 0.0330 |
| P | 0.0615 | 0.0387 | 0.0554 | 0.0346 | 0.0710 | 0.0367 | 0.0554 | 0.0438 | 0.0549 | 0.0350 |
| N | 0.0554 | 0.0389 | 0.0648 | 0.0405 | 0.6493 | 0.1285 | 0.7393 | 0.2193 | 0.0885 | 0.0510 |

ME simulation mean of $\|\hat{\gamma} - \gamma\|^2$, SD associated standard deviation, $RASE_2^2$ square root of average square errors

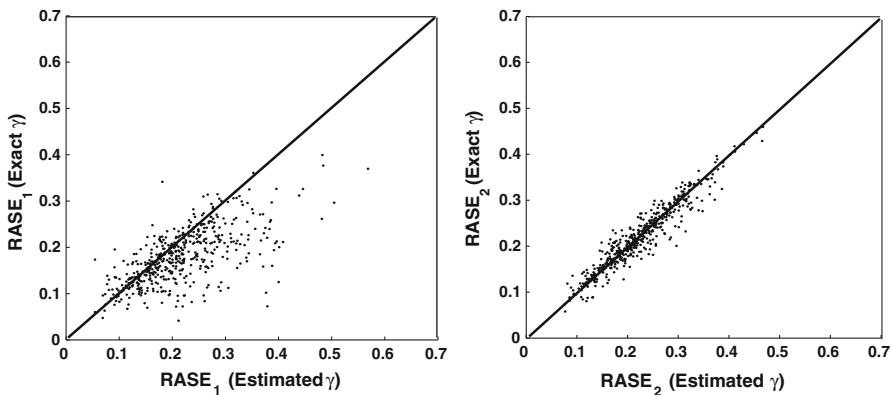


Fig. 1 Simulation results (*n* = 1000) for Example 2 along with the benchmark procedure, i.e., using $\xi(v)$. RASE based on the true γ against RASE based on estimated $\hat{\gamma}$ for $\alpha_1(u)$ (left panel) and $\alpha_2(u)$ (right panel)

proposed estimation method or naive estimation. But the estimation procedure for $\alpha_1(u)$ does not have such a nice property; i.e., the RASE value for $\alpha_1(u)$ increases from 0.0527 for the proposed method to 0.7993 for the naive estimation when the sample size $n = 1000$, whereas the RASE value for $\alpha_2(u)$ keeps the same scale for the proposed method and the naive estimation. This substantial difference is because $\alpha_1(u)$ but $\alpha_2(u)$ includes the biases caused by ignoring measurement errors. These features can further be observed in the left panels and right panels of Fig. 1 for the benchmark procedure when $n = 1000$, i.e., using $\xi(v)$ and Fig. 2 for the proposed procedure, i.e., using $\hat{\xi}(v)$, where we plot the RASE values based on the true γ against the RASE values based on the estimated $\hat{\gamma}$ for $\alpha_1(u)$ the (left panel) and $\alpha_2(u)$ (right panel). It can be seen that the estimation procedure (11) for $\alpha_2(u)$ with the estimated $\hat{\gamma}$ performs as well as if we knew the true value of γ .

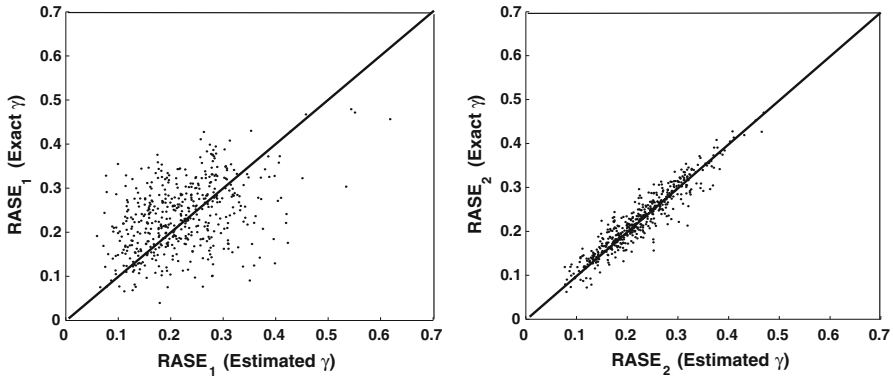


Fig. 2 Simulation results ($n = 1000$) for Example 2 along with the proposed procedure, i.e., using $\hat{\xi}(v)$. RASE based on the true γ against RASE based on estimated $\hat{\gamma}$ for $\alpha_1(u)$ (left panel) and $\alpha_2(u)$ (right panel)

Example 4 In this example, we examine the performance of the test procedures proposed in Sect. 4.2. The simulation setting is the same as in Example 3. Consider the hypothesis

$$H_0 : \alpha_2(u) = 0 \text{ vs } H_1 : \alpha_2(u) \neq 0, \tag{15}$$

where $\alpha_2(u)$ is a sequence of alternative models indexed by C_o of form $\alpha_2(u) = C_o \times u(1 - u)$. We conducted 400 simulations at four different significance levels: 0.01, 0.025, 0.05 and 0.10 for the benchmark procedure and the proposed procedure. 500 conditional bootstrap (Cai et al. 2000) samples were generated in each simulation for power calculation. The simulation results are reported in Table 5 and Fig. 3. We can see that when $C_o = 0$, all empirical levels obtained by these two procedures 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. It is worth noting that the simulation results for the benchmark procedure concur with what Li and Liang (2008) observed, and the proposed estimation procedure performs also well. This indicates that the proposed GLRT under the measurement error setting works well numerically and confirms our theoretical findings.

4.2 An empirical example

We analyzed a data set with 358 complete observations from a diabetes study conducted in central Virginia for African Americans, whose aim was at understanding the relationship between the prevalence of obesity, diabetes, and other cardiovascular risk factors. There are 14 covariates of potential interest: “TC, Total Cholesterol”; “SG, Stabilized Glucose”; “HDL, High-Density Lipoprotein”; “Ratio, Cholesterol/HDL”; “GH, Glycosolated Hemoglobin”; “age”; “gender”; “height”; “weight”; “frame”; “FSBP, First Systolic Blood Pressure”; “FDBP, First Diastolic Blood Pressure”; “waist” and “hip”.

Table 5 The simulation results for Example 4

| Significant level | 0.01 | 0.025 | 0.05 | 0.10 |
|---|-------|-------|-------|-------|
| <i>Benchmark procedure using $\xi(v)$</i> | | | | |
| <i>n = 500</i> | | | | |
| $C_o = 0.00$ | 0.011 | 0.029 | 0.047 | 0.090 |
| $C_o = 0.50$ | 0.017 | 0.023 | 0.065 | 0.142 |
| $C_o = 1.00$ | 0.079 | 0.198 | 0.278 | 0.357 |
| $C_o = 1.50$ | 0.254 | 0.456 | 0.503 | 0.656 |
| $C_o = 2.00$ | 0.687 | 0.774 | 0.876 | 0.891 |
| $C_o = 2.50$ | 0.856 | 0.904 | 0.923 | 0.941 |
| $C_o = 3.00$ | 0.917 | 0.946 | 0.975 | 0.991 |
| <i>n = 1000</i> | | | | |
| $C_o = 0.00$ | 0.010 | 0.027 | 0.043 | 0.092 |
| $C_o = 0.50$ | 0.027 | 0.032 | 0.071 | 0.151 |
| $C_o = 1.00$ | 0.089 | 0.223 | 0.313 | 0.393 |
| $C_o = 1.50$ | 0.295 | 0.482 | 0.598 | 0.714 |
| $C_o = 2.00$ | 0.759 | 0.821 | 0.911 | 0.955 |
| $C_o = 2.50$ | 0.911 | 0.955 | 0.964 | 0.982 |
| $C_o = 3.00$ | 1.000 | 1.000 | 1.000 | 1.000 |
| <i>Proposed procedure using estimated $\hat{\xi}(v)$</i> | | | | |
| <i>n = 500</i> | | | | |
| $C_o = 0.00$ | 0.008 | 0.021 | 0.044 | 0.104 |
| $C_o = 0.50$ | 0.014 | 0.048 | 0.089 | 0.176 |
| $C_o = 1.00$ | 0.165 | 0.212 | 0.309 | 0.423 |
| $C_o = 1.50$ | 0.406 | 0.498 | 0.579 | 0.623 |
| $C_o = 2.00$ | 0.643 | 0.798 | 0.857 | 0.904 |
| $C_o = 2.50$ | 0.889 | 0.913 | 0.937 | 0.957 |
| $C_o = 3.00$ | 0.924 | 0.943 | 0.987 | 0.992 |
| <i>n = 1000</i> | | | | |
| $C_o = 0.00$ | 0.009 | 0.022 | 0.041 | 0.103 |
| $C_o = 0.50$ | 0.018 | 0.054 | 0.107 | 0.188 |
| $C_o = 1.00$ | 0.179 | 0.253 | 0.337 | 0.495 |
| $C_o = 1.50$ | 0.467 | 0.539 | 0.618 | 0.696 |
| $C_o = 2.00$ | 0.701 | 0.836 | 0.911 | 0.936 |
| $C_o = 2.50$ | 0.938 | 0.955 | 0.964 | 0.991 |
| $C_o = 3.00$ | 0.991 | 0.993 | 1.000 | 1.000 |

Usually, GH over 7.0 indicates a positive diagnosis of diabetes. So Y was assigned 1 if $GH > 7.0$ and 0 otherwise. We are interested in the relationship between the probability being diabetes and the collected covariates. Cambien et al. (1987) found that blood pressure is strongly associated with glucose. Han et al. (1995) also found that Ratio is associated with TC and HDL. On the basis of preliminary results, we treat $\eta = (FSBP, FDBP)^T$ and $V = SG$ as ancillary variables to remit unobservable variables $\xi = (\xi_1(V), \xi_2(V))^T$. Take $X = (TC, HDL)^T$ and $U = \text{Ratio}$ to

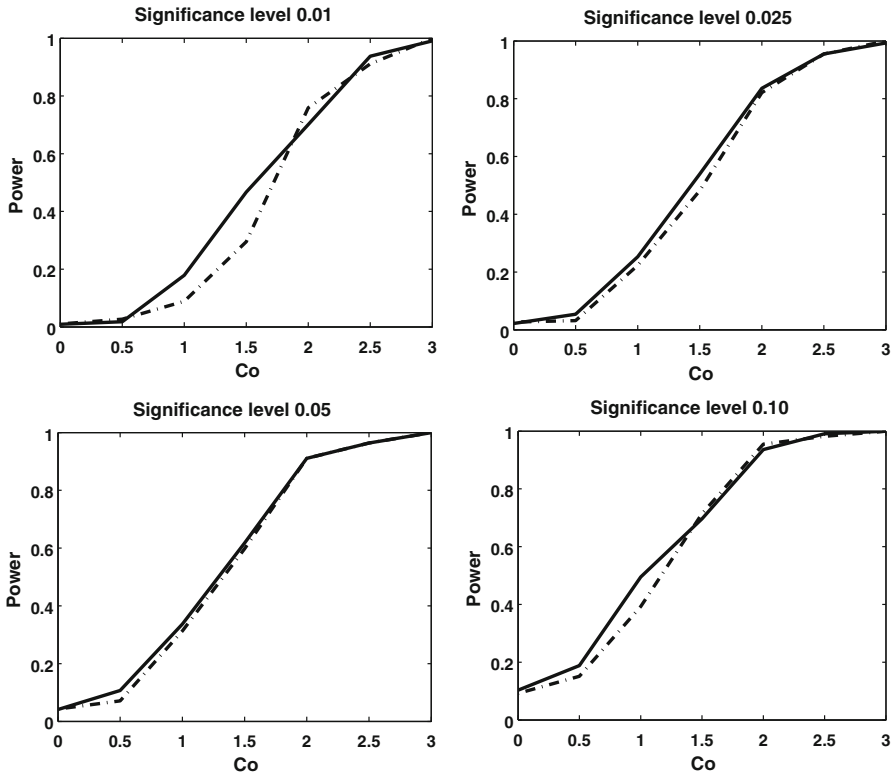


Fig. 3 Simulation results ($n = 1000$) for Example 3—power plot for the bootstrap test proposed in Sect. 4.2. The significance level is 0.01, 0.025, 0.05 and 0.01. The dotted lines represent the power functions for the benchmark procedure by directly using $\xi(v)$. The solid lines represent the proposed procedure using estimated $\hat{\xi}(v)$

possibly investigate the varying coefficient functions $\alpha(\cdot) = (\alpha_1(U), \alpha_2(U))^T$. Other W -variables include age, gender, height, weight, frame, waist, hip. Both gender and frame are discrete variables of 1 and 0 for male and female, and of 1, 2, 3 for small, medium and large frames, respectively. All continuous covariates were standardized.

We used the proposed quasi-likelihood method and the penalized quasi-likelihood with SCAD penalty for estimation and variable selection. To make a comparison, we also considered AIC, BIC and RIC variable selection procedures. The bandwidth $h = 0.5n^{-1/3}$ was used for local regression fitting. The results are reported in Table 6. We can see that the SCAD procedure is in conjunction with BIC and RIC procedures, and all the three methods indicate that only the possibly remitted variables $\xi = (\xi_1(V), \xi_2(V))^T$ are significant, while all W -variable are not significant. Compared with the estimated value of β , the SCAD-based estimates for β are close to those obtained using the quasi-likelihood. AIC selects extra 2 W -variables: waist and hip. Recalling the simulation performance in Sect. 4.1, AIC may suggest an over-fitted model. As such, the model selected through SCAD, BIC and RIC may be more proper.

Table 6 Estimation and variable selection results of real data analysis

| Method | Age | Gender | Height | Weight | Frame | Waist | Hip | $\hat{\xi}_1(V)$ | $\hat{\xi}_2(V)$ |
|------------------|--------|---------|--------|---------|---------|---------|---------|------------------|------------------|
| Quasi-likelihood | 0.0365 | -0.2312 | 0.3174 | -0.0863 | -0.2044 | 0.06752 | -0.7062 | 13.7826 | -12.1243 |
| SCAD | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 13.5329 | -10.9978 |
| RAIC | 0 | 0 | 0 | 0 | 0 | 0.6302 | -0.6364 | 13.4154 | -11.3623 |
| RBIC | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 13.6825 | -11.5036 |
| R RIC | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 13.6825 | -11.5036 |

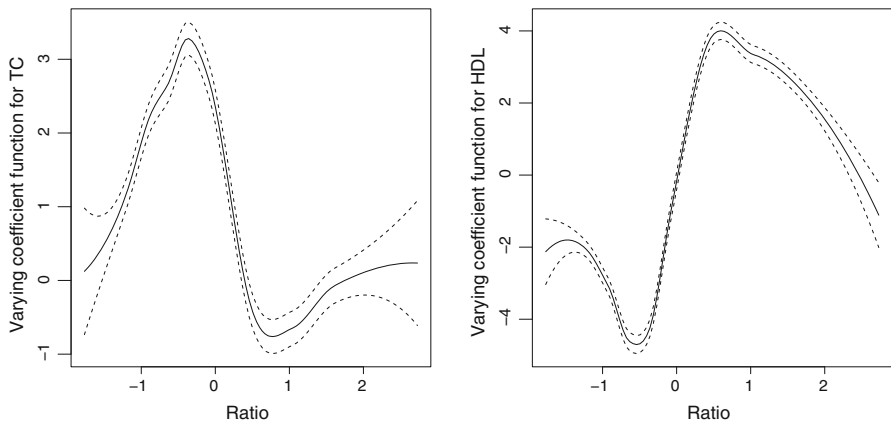


Fig. 4 Results for real data example. The local linear estimators for $TC-\alpha_1(u)$ (the left panel) and $HDL-\alpha_2(u)$ (the right panel) against variable $U = \text{Ratio}$ and the associated 95 % pointwise confidence intervals (dotted lines)

We further considered estimation procedure and variable selection for X -variables. We conducted 500 bootstraps to test $\alpha_1(\cdot) = 0$. The corresponding GLRT-based p value is 0.0711, larger than the 97.5 % quantile of 500 bootstraps, 0.0418, and suggests a rejection of the null hypothesis. In the same way, we tested $\alpha_2(\cdot) = 0$ and got the corresponding p value 0.3027, much larger than the 97.5 % quantile of 500 bootstraps, 0.0355. This also indicates that we should reject the null hypothesis. The estimated curves associated with their 95 % pointwise confidence bands are depicted in Fig. 4, which shows a nonzero and nonlinear pattern. As a result, both $\alpha_1(u)$ and $\alpha_2(u)$ should be included in the final model.

Acknowledgements The authors thank the associate editor, two referees for their constructive suggestions that helped us to improve the early manuscript. Zhang Jun’s research was supported by the National Natural Science Foundation of China (NSFC) Grant No. 11326179 (Tian yuan fund for Mathematics), and NSFC Grant No. 11401391, and the Project of Department of Education of Guangdong Province of China, Grant No. 2014KTSCX112. Feng Zhenghui’s research was supported by the NSFC Grant No. 11301434. Xu Peirong’s research was supported by the Natural Science Foundation of Jiangsu Province, China, Grant No. BK20140617. Liang Hua’s research was partially supported by NSF Grants DMS- 1440121 and DMS-1418042, and by Award Number 11228103, made by National Natural Science Foundation of China.

References

- Akaike, H. (1973). Maximum likelihood identification of Gaussian autoregressive moving average models. *Biometrika*, *60*, 255–265.
- Cai, Z., Fan, J., Li, R. (2000). Efficient estimation and inferences for varying-coefficient models. *Journal of the American Statistical Association*, *95*, 888–902.
- Cambien, F., Warnet, J., Eschwege, E., Jacqueson, A., Richard, J., Rosselin, G. (1987). Body mass, blood pressure, glucose, and lipids. Does plasma insulin explain their relationships? *Arteriosclerosis, Thrombosis, and Vascular Biology*, *7*, 197–202.
- Carroll, R. J., Wang, Y. (2008). Nonparametric variance estimation in the analysis of microarray data: A measurement error approach. *Biometrika*, *95*(2), 437–449.
- Carroll, R. J., Fan, J., Gijbels, I., Wand, M. P. (1997). Generalized partially linear single-index models. *Journal of the American Statistical Association*, *92*(438), 477–489.
- Carroll, R. J., Ruppert, D., Stefanski, L. A., Crainiceanu, C. M. (2006). *Nonlinear measurement error models, a modern perspective* (2nd ed.). New York: Chapman and Hall.
- Cook, J. R., Stefanski, L. A. (1994). Simulation-extrapolation estimation in parametric measurement error models. *Journal of the American Statistical Association*, *89*, 1314–1328.
- Fan, J., Gijbels, I. (1996). *Local polynomial modelling and its applications* (Vol. 66). London: Chapman & Hall.
- Fan, J., Huang, T. (2005). Profile likelihood inferences on semiparametric varying-coefficient partially linear models. *Bernoulli*, *11*, 1031–1057.
- Fan, J., Li, R. (2001). Variable selection via nonconcave penalized likelihood and its oracle properties. *Journal of the American Statistical Association*, *96*, 1348–1360.
- Fan, J., Zhang, C. M., Zhang, J. (2001). Generalized likelihood ratio statistics and Wilks phenomenon. *The Annals of Statistics*, *29*, 153–193.
- Foster, D., George, E. (1994). The risk inflation criterion for multiple regression. *The Annals of Statistics*, *22*, 1947–1975.
- Hall, P., Ma, Y. (2007). Semiparametric estimators of functional measurement error models with unknown error. *Journal of the Royal Statistical Society, Series B Statistical Methodology*, *69*(3), 429–446.
- Han, T. S., van Leer, E. M., Seidell, J. C., Lean, M. E. (1995). Waist circumference action levels in the identification of cardiovascular risk factors: prevalence study in a random sample. *British Medical Journal (BMJ)*, *311*(7017), 1401–1405.
- Härdle, W., Liang, H., Gao, J. (2000). *Partially linear models*. Heidelberg: Physica-Verlag.
- Hastie, T., Tibshirani, R. (1993). Varying-coefficient models (with discussion). *Journal of the Royal Statistical Society, Series B Statistical Methodology*, *55*, 757–796.
- Hastie, T., Tibshirani, R. (1993). Varying-coefficient models (with discussion). *Journal of the Royal Statistical Society, Series B Statistical Methodology*, *55*, 757–796.
- Hunsberger, S. (1994). Semiparametric regression in likelihood-based models. *Journal of the American Statistical Association*, *89*, 1354–1365.
- Hunsberger, S., Albert, P. S., Follmann, D. A., Suh, E. (2002). Parametric and semiparametric approaches to testing for seasonal trend in serial count data. *Biostatistics*, *3*, 289–298.
- Li, G., Xue, L., Lian, H. (2011). Semi-varying coefficient models with a diverging number of components. *Journal of Multivariate Analysis*, *102*(7), 1166–1174.
- Li, R., Liang, H. (2008). Variable selection in semiparametric regression modeling. *The Annals of Statistics*, *36*, 261–286.
- Liang, H., Li, R. (2009). Variable selection for partially linear models with measurement errors. *Journal of the American Statistical Association*, *104*(485), 234–248.
- Lin, X., Carroll, R. J. (2001). Semiparametric regression for clustered data using generalized estimating equations. *Journal of the American Statistical Association*, *96*, 1045–1056.
- Lobach, I., Carroll, R. J., Spinka, C., Gail, M., Chatterjee, N. (2008). Haplotype-based regression analysis and inference of case-control studies with unphased genotypes and measurement errors in environmental exposures. *Biometrics*, *64*, 673–684.
- Lobach, I., Fan, R., Carroll, R. J. (2010). Genotype-based association mapping of complex diseases: gene-environment interactions with multiple genetic markers and measurement error in environmental exposures. *Genetic Epidemiology*, *34*, 792–802.
- Ma, Y., Carroll, R. J. (2006). Locally efficient estimators for semiparametric models with measurement error. *Journal of the American Statistical Association*, *101*(476), 1465–1474.

- Ma, Y., Li, R. (2010). Variable selection in measurement error models. *Bernoulli*, 16(1), 274–300.
- Ma, Y., Tsiatis, A. A. (2006). On closed form semiparametric estimators for measurement error models. *Statistica Sinica*, 16(1), 183–193.
- Robinson, P. M. (1988). Root n -consistent semiparametric regression. *Econometrica*, 56, 931–954.
- Schwarz, G. (1978). Estimating the dimension of a model. *The Annals of Statistics*, 6, 461–464.
- Severini, T. A., Staniswalis, J. G. (1994). Quasi-likelihood estimation in semiparametric models. *Journal of the American Statistical Association*, 89, 501–511.
- Silverman, B. W. (1986). *Density estimation for statistics and data analysis*, Vol. 26 of Monographs on statistics and applied probability. London: Chapman and Hall.
- Sinha, S., Mallick, B. K., Kipnis, V., Carroll, R. J. (2010). Semiparametric Bayesian analysis of nutritional epidemiology data in the presence of measurement error. *Biometrics*, 66(2), 444–454.
- Speckman, P. E. (1988). Kernel smoothing in partial linear models. *Journal of the Royal Statistical Society, Series B Statistical Methodology*, 50, 413–436.
- Stefanski, L. A. (1989). Unbiased estimation of a nonlinear function of a normal mean with application to measurement error models. *Communications in Statistics. Theory and Methods*, 18(12), 4335–4358.
- Stefanski, L. A., Carroll, R. J. (1987). Conditional scores and optimal scores for generalized linear measurement-error models. *Biometrika*, 74(4), 703–716.
- Tsiatis, A. A., Ma, Y. (2004). Locally efficient semiparametric estimators for functional measurement error models. *Biometrika*, 91(4), 835–848.
- Wang, H., Xia, Y. (2009). Shrinkage estimation of the varying coefficient model. *Journal of the American Statistical Association*, 104, 747–757.
- Wang, L., Liu, X., Liang, H., Carroll, R. J. (2011). Estimation and variable selection for generalized additive partial linear models. *The Annals of Statistics*, 39, 1827–1851.
- Wei, F., Huang, J., Li, H. (2011). Variable selection and estimation in high-dimensional varying coefficient models. *Statistica Sinica*, 21(4), 1515–1540.
- Xia, Y., Zhang, W., Tong, H. (2004). Efficient estimation for semivarying-coefficient models. *Biometrika*, 91, 661–681.
- Yi, G. Y., Ma, Y. Y., Carroll, R. J. (2012). A functional generalized method of moments approach for longitudinal studies with missing responses and covariate measurement error. *Biometrika*, 99(1), 151–165.
- Yuan, M., Lin, Y. (2006). Model selection and estimation in regression with grouped variables. *Journal of the Royal Statistical Society, Series B Statistical Methodology*, 68, 49–67.
- Zhang, C. (2003). Calibrating the degrees of freedom for automatic data smoothing and effective curve checking. *Journal of the American Statistical Association*, 98(463), 609–628.
- Zhang, W., Lee, S.-Y., Song, X. (2002). Local polynomial fitting in semivarying coefficient model. *Journal of Multivariate Analysis*, 82, 166–188.
- Zhou, Y., Liang, H. (2009). Statistical inference for semiparametric varying-coefficient partially linear models with error-prone linear covariates. *The Annals of Statistics*, 37, 427–458.