NEW EFFICIENT ESTIMATION AND VARIABLE SELECTION METHODS FOR SEMIPARAMETRIC VARYING-COEFFICIENT PARTIALLY LINEAR MODELS

BY BO KAI[1], RUNZE LI[2] AND HUI ZOU[3]

College of Charleston, Pennsylvania State University and University of Minnesota

The complexity of semiparametric models poses new challenges to statistical inference and model selection that frequently arise from real applications. In this work, we propose new estimation and variable selection procedures for the semiparametric varying-coefficient partially linear model. We first study quantile regression estimates for the nonparametric varying-coefficient functions and the parametric regression coefficients. To achieve nice efficiency properties, we further develop a semiparametric composite quantile regression procedure. We establish the asymptotic normality of proposed estimators for both the parametric and nonparametric parts and show that the estimators achieve the best convergence rate. Moreover, we show that the proposed method is much more efficient than the least-squares-based method for many non-normal errors and that it only loses a small amount of efficiency for normal errors. In addition, it is shown that the loss in efficiency is at most 11.1% for estimating varying coefficient functions and is no greater than 13.6% for estimating parametric components. To achieve sparsity with high-dimensional covariates, we propose adaptive penalization methods for variable selection in the semiparametric varying-coefficient partially linear model and prove that the methods possess the oracle property. Extensive Monte Carlo simulation studies are conducted to examine the finite-sample performance of the proposed procedures. Finally, we apply the new methods to analyze the plasma beta-carotene level data.

1. Introduction. Semiparametric regression modeling has recently become popular in the statistics literature because it keeps the flexibility of nonparametric models while maintaining the explanatory power of parametric models. The partially linear model, the most commonly used semiparametric regression model, has received a lot of attention in the literature; see Härdle, Liang and Gao [9], Yatchew [32] and references therein for theory and applications of partially linear

models. Various extensions of the partially linear model have been proposed in the literature; see Ruppert, Wand and Carroll [26] for applications and theoretical developments of semiparametric regression models. The semiparametric varying-coefficient partially linear model, as an important extension of the partially linear model, is becoming popular in the literature. Let Y be a response variable and $\{U, \mathbf{X}, \mathbf{Z}\}$ its covariates. The semiparametric varying-coefficient partially linear model is defined to be

$$(1.1) \qquad Y = \alpha_0(U) + \mathbf{X}^T \boldsymbol{\alpha}(U) + \mathbf{Z}^T \boldsymbol{\beta} + \varepsilon,$$

where $\alpha_0(U)$ is a baseline function, $\boldsymbol{\alpha}(U) = \{\alpha_1(U), \ldots, \alpha_{d_1}(U)\}^T$ consists of d_1 unknown varying coefficient functions, $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_{d_2})^T$ is a d_2-dimensional coefficient vector and ε is random error. In this paper, we will focus on univariate U only, although the proposed procedure is directly applicable for multivariate $\mathbf{U}$. Zhang, Lee and Song [33] proposed an estimation procedure for the model (1.1), based on local polynomial regression techniques. Xia, Zhang and Tong [31] proposed a semilocal estimation procedure to further reduce the bias of the estimator for $\boldsymbol{\beta}$ suggested in Zhang, Lee and Song [33]. Fan and Huang [5] proposed a profile least-squares estimator for model (1.1) and developed statistical inference procedures. As an extension of Fan and Huang [5], a profile likelihood estimation procedure was developed in Lam and Fan [18], under the generalized linear model framework with a diverging number of covariates.

Existing estimation procedures for model (1.1) were built on either least-squares- or likelihood-based methods. Thus, the existing procedures are expected to be sensitive to outliers and their efficiency may be significantly improved for many commonly used non-normal errors. In this paper, we propose new estimation procedures for model (1.1). This paper contains three major developments: (a) semiparametric quantile regression; (b) semiparametric composite quantile regression; (c) adaptive penalization methods for achieving sparsity in semiparametric composite quantile regression.

Quantile regression is often considered as an alternative to least-squares in the literature. For a complete review on quantile regression, see Koenker [17]. Quantile-regression-based inference procedures have been considered in the literature; see, for example, Cai and Xu [2], He and Shi [10], He, Zhu and Fung [11], Lee [19], among others. In Section 2, we propose a new semiparametric quantile regression procedure for model (1.1). We investigate the sampling properties of the proposed method and their asymptotic normality. When applying semiparametric quantile regression to model (1.1), we observe that all quantile regression estimators can estimate $\boldsymbol{\alpha}(u)$ and $\boldsymbol{\beta}$ with the optimal rate of convergence. This fact motivates us to combine the information across multiple quantile estimates to obtain improved estimates of $\boldsymbol{\alpha}(u)$ and $\boldsymbol{\beta}$. Such an idea has been studied for the parametric regression model in Zou and Yuan [35] and it leads to the composite quantile regression (CQR) estimator that is shown to enjoy nice asymptotic

efficiency properties compared with the classical least-squares estimator. In Section 3, we propose the semiparametric composite quantile regression (semi-CQR) estimators for estimating both nonparametric and parametric parts in model (1.1). We show that the semi-CQR estimators achieve the best convergence rates. We also prove the asymptotic normality of the semi-CQR estimators. The asymptotic theory shows that, compared with the semiparametric least-squares estimators, the semi-CQR estimators can have substantial efficiency gain for many non-normal errors and only lose a small amount of efficiency for normal errors. Moreover, the relative efficiency is at least 88.9% for estimating varying-coefficient functions and is at least 86.4% for estimating parametric components.

In practice, there are often many covariates in the parametric part of model (1.1). With high-dimensional covariates, sparse modeling is often considered superior, owing to enhanced model predictability and interpretability [7]. Variable selection for model (1.1) is challenging because it involves both nonparametric and parametric parts. Traditional variable selection methods, such as stepwise regression or best subset variable selection, do not work effectively for the semiparametric model because they need to choose smoothing parameters for each submodel and cannot cope with high-dimensionality. In Section 4, we develop an effective variable selection procedure to select significant parametric components in model (1.1). We demonstrate that the proposed procedure possesses the oracle property, in the sense of Fan and Li [6].

In Section 5, we conduct simulation studies to examine the finite-sample performance of the proposed procedures. The proposed methods are illustrated with the plasma beta-carotene level data. Regularity conditions and technical proofs are given in Section 6.

2. Semiparametric quantile regression. In this section, we develop the semiparametric quantile regression method and theory. Let $\rho_\tau(r) = \tau r - rI$ $(r < 0)$ be the check loss function at $\tau \in (0, 1)$. Quantile regression is often used to estimate the conditional quantile functions of Y,

$$Q_\tau(u, \mathbf{x}, \mathbf{z}) = \operatorname*{argmin}_a E\{\rho_\tau(Y - a)|(U, \mathbf{X}, \mathbf{Z}) = (u, \mathbf{x}, \mathbf{z})\}.$$

The semiparametric varying-coefficient partially linear model assumes that the conditional quantile function is expressed as $Q_\tau(u, \mathbf{x}, \mathbf{z}) = \alpha_{0,\tau}(u) + \mathbf{x}^T \boldsymbol{\alpha}_\tau(u) + \mathbf{z}^T \boldsymbol{\beta}_\tau$.

Suppose that $\{U_i, \mathbf{X}_i, \mathbf{Z}_i, Y_i\}$, $i = 1, \ldots, n$, is an independent and identically distributed sample from the model

$$(2.1) \qquad Y = \alpha_{0,\tau}(U) + \mathbf{X}^T \boldsymbol{\alpha}_\tau(U) + \mathbf{Z}^T \boldsymbol{\beta}_\tau + \varepsilon_\tau,$$

where ε_τ is random error with conditional τth quantile being zero. We obtain quantile regression estimates of $\alpha_{0,\tau}(\cdot)$, $\boldsymbol{\alpha}_\tau(\cdot)$ and $\boldsymbol{\beta}_\tau$ by minimizing the quantile loss

function

$$(2.2) \qquad \sum_{i=1}^{n} \rho_\tau \{Y_i - \alpha_0(U_i) - \mathbf{X}_i^T \boldsymbol{\alpha}(U_i) - \mathbf{Z}_i^T \boldsymbol{\beta}\}.$$

Because (2.2) involves both nonparametric and parametric components, and because they can be estimated by different rates of convergence, we propose a three-stage estimation procedure. In the first stage, we employ local linear regression techniques to derive an initial estimates of $\alpha_{0,\tau}(\cdot)$, $\boldsymbol{\alpha}_\tau(\cdot)$ and $\boldsymbol{\beta}_\tau$. Then, in the second and third stages, we further improve the estimation efficiency of the initial estimates for $\boldsymbol{\beta}_\tau$ and $(\alpha_{0,\tau}(\cdot), \boldsymbol{\alpha}_\tau(\cdot))$, respectively.

For U in the neighborhood of u, we use a local linear approximation

$$\alpha_j(U) \approx \alpha_j(u) + \alpha_j'(u)(U - u) \triangleq a_j + b_j(U - u)$$

for $j = 0, \ldots, d_1$. Let $\{\tilde{a}_{0,\tau}, \tilde{b}_{0,\tau}, \tilde{\mathbf{a}}_\tau, \tilde{\mathbf{b}}_\tau, \tilde{\boldsymbol{\beta}}_\tau\}$ be the minimizer of the local weighted quantile loss function

$$\sum_{i=1}^{n} \rho_\tau \{Y_i - a_0 - b_0(U_i - u) - \mathbf{X}_i^T\{\mathbf{a} + \mathbf{b}(U_i - u)\} - \mathbf{Z}_i^T \boldsymbol{\beta}\} K_h(U_i - u),$$

where $\mathbf{a} = (a_1, \ldots, a_{d_1})^T$, $\mathbf{b} = (b_1, \ldots, b_{d_1})^T$, $K(\cdot)$ is a given kernel function and $K_h(\cdot) = K(\cdot/h)/h$ with a bandwidth h. Then,

$$\tilde{\alpha}_{0,\tau}(u) = \tilde{a}_0, \qquad \tilde{\boldsymbol{\alpha}}_\tau(u) = \tilde{\mathbf{a}}_\tau.$$

We take $\{\tilde{\alpha}_{0,\tau}(u), \tilde{\boldsymbol{\alpha}}_\tau(u), \tilde{\boldsymbol{\beta}}_\tau\}$ as the initial estimates.

We now provide theoretical justifications for the initial estimates. First, we give some notation. Let $f_\tau(\cdot|u, \mathbf{x}, \mathbf{z})$ and $F_\tau(\cdot|u, \mathbf{x}, \mathbf{z})$ be the density function and cumulative distribution function of the error conditional on $(U, \mathbf{X}, \mathbf{Z}) = (u, \mathbf{x}, \mathbf{z})$, respectively. Denote by $f_U(\cdot)$ the marginal density function of the covariate U. The kernel $K(\cdot)$ is chosen as a symmetric density function and we let

$$\mu_j = \int u^j K(u)\, du \quad \text{and} \quad v_j = \int u^j K^2(u)\, du, \qquad j = 0, 1, 2, \ldots.$$

We then have the following result.

THEOREM 2.1. *Under the regularity conditions given in Section 7, if $h \to 0$ and $nh \to \infty$ as $n \to \infty$, then*

$$(2.3) \qquad \begin{aligned} &\sqrt{nh}\left[\begin{pmatrix} \tilde{\alpha}_{0,\tau}(u) - \alpha_{0,\tau}(u) \\ \tilde{\boldsymbol{\alpha}}_\tau(u) - \boldsymbol{\alpha}_\tau(u) \\ \tilde{\boldsymbol{\beta}}_\tau - \boldsymbol{\beta}_\tau \end{pmatrix} - \frac{\mu_2 h^2}{2}\begin{pmatrix} \alpha_{0,\tau}''(u) \\ \boldsymbol{\alpha}_\tau''(u) \\ \mathbf{0} \end{pmatrix}\right] \\ &\qquad \xrightarrow{\mathcal{D}} N\left(\mathbf{0}, \frac{v_0 \tau(1-\tau)}{f_U(u)} \mathbf{A}_1^{-1}(u)\mathbf{B}_1(u)\mathbf{A}_1^{-1}(u)\right), \end{aligned}$$

where $\mathbf{A}_1(u) = E[f_\tau(0|U, \mathbf{X}, \mathbf{Z})(1, \mathbf{X}^T, \mathbf{Z}^T)^T(1, \mathbf{X}^T, \mathbf{Z}^T)|U = u]$ *and* $\mathbf{B}_1(u) = E[(1, \mathbf{X}^T, \mathbf{Z}^T)^T(1, \mathbf{X}^T, \mathbf{Z}^T)|U = u]$.

Theorem 2.1 implies that $\tilde{\boldsymbol{\beta}}_\tau$ is a $\sqrt{nh}$-consistent estimator—this is because we only use data in a local neighborhood of u to estimate $\boldsymbol{\beta}_\tau$. Define $Y_{i,\tau}^* = Y_i - \tilde{\alpha}_{0,\tau}(U_i) - \mathbf{X}_i^T \tilde{\boldsymbol{\alpha}}_\tau(U_i)$ and compute an improved estimator of $\boldsymbol{\beta}_\tau$ by

$$(2.4) \qquad \hat{\boldsymbol{\beta}}_\tau = \underset{\boldsymbol{\beta}}{\operatorname{argmin}} \sum_{i=1}^n \rho_\tau(Y_{i,\tau}^* - \mathbf{Z}_i^T \boldsymbol{\beta}).$$

We call it the *semi-QR estimator* of $\boldsymbol{\beta}_\tau$. The next theorem shows the asymptotic properties of $\hat{\boldsymbol{\beta}}_\tau$.

THEOREM 2.2. *Let* $\boldsymbol{\xi}_\tau(u, \mathbf{x}, \mathbf{z}) = E[f_\tau(0|U, \mathbf{X}, \mathbf{Z})\mathbf{Z}(1, \mathbf{X}^T, \mathbf{0})|U = u] \times \mathbf{A}_1^{-1}(u)(1, \mathbf{x}^T, \mathbf{z}^T)^T$. *Under the regularity conditions given in Section* 7, *if* $nh^4 \to 0$ *and* $nh^2/\log(1/h) \to \infty$ *as* $n \to \infty$, *then the asymptotic distribution of* $\hat{\boldsymbol{\beta}}_\tau$ *is given by*

$$(2.5) \qquad \sqrt{n}(\hat{\boldsymbol{\beta}}_\tau - \boldsymbol{\beta}_\tau) \xrightarrow{\mathcal{D}} N(0, \mathbf{S}_\tau^{-1} \boldsymbol{\Xi}_\tau \mathbf{S}_\tau^{-1}),$$

where $\mathbf{S}_\tau = E[f_\tau(0|U, \mathbf{X}, \mathbf{Z})\mathbf{Z}\mathbf{Z}^T]$ *and* $\boldsymbol{\Xi}_\tau = \tau(1-\tau)E[\{\mathbf{Z} - \boldsymbol{\xi}_\tau(U, \mathbf{X}, \mathbf{Z})\}\{\mathbf{Z} - \boldsymbol{\xi}_\tau(U, \mathbf{X}, \mathbf{Z})\}^T]$.

The optimal bandwidth in Theorem 2.1 is $h \sim n^{-1/5}$. This bandwidth does not satisfy the condition in Theorem 2.2. Hence, in order to obtain the root-n consistency and asymptotic normality for $\hat{\boldsymbol{\beta}}_\tau$, undersmoothing for $\tilde{\alpha}_{0,\tau}(u)$ and $\tilde{\boldsymbol{\alpha}}_\tau(u)$ is necessary. This is a common requirement in semiparametric models; see Carroll et al. [3] for a detailed discussion.

After obtaining the root-n consistent estimator $\hat{\boldsymbol{\beta}}_\tau$, we can further improve the efficiency of $\tilde{\alpha}_{0,\tau}(u)$ and $\tilde{\boldsymbol{\alpha}}_\tau(u)$. To this end, let $\{\hat{a}_{0,\tau}, \hat{b}_{0,\tau}, \hat{\mathbf{a}}_\tau, \hat{\mathbf{b}}_\tau\}$ be the minimizer of

$$\sum_{i=1}^n \rho_\tau\{Y_i - \mathbf{Z}_i^T \hat{\boldsymbol{\beta}}_\tau - a_0 - b_0(U_i - u) - \mathbf{X}_i^T\{\mathbf{a} + \mathbf{b}(U_i - u)\}\}K_h(U_i - u).$$

We define

$$(2.6) \qquad \hat{\alpha}_{0,\tau}(u) = \hat{a}_{0,\tau}, \qquad \hat{\boldsymbol{\alpha}}_\tau(u) = \hat{\mathbf{a}}_\tau.$$

THEOREM 2.3. *Under the regularity conditions given in Section* 7, *if* $h \to 0$ *and* $nh \to \infty$ *as* $n \to \infty$, *then*

$$(2.7) \quad \begin{aligned} &\sqrt{nh}\left[\begin{pmatrix} \hat{\alpha}_{0,\tau}(u) - \alpha_{0,\tau}(u) \\ \hat{\boldsymbol{\alpha}}_\tau(u) - \boldsymbol{\alpha}_\tau(u) \end{pmatrix} - \frac{\mu_2 h^2}{2}\begin{pmatrix} \alpha_{0,\tau}''(u) \\ \boldsymbol{\alpha}_\tau''(u) \end{pmatrix}\right] \\ &\xrightarrow{\mathcal{D}} N\left(\mathbf{0}, \frac{\nu_0 \tau(1-\tau)}{f_U(u)}\mathbf{A}_2^{-1}(u)\mathbf{B}_2(u)\mathbf{A}_2^{-1}(u)\right), \end{aligned}$$

where $\mathbf{A}_2(u) = E[f_\tau(0|U, \mathbf{X}, \mathbf{Z})(1, \mathbf{X}^T)^T(1, \mathbf{X}^T)|U = u]$ *and* $\mathbf{B}_2(u) = E[(1, \mathbf{X}^T)^T(1, \mathbf{X}^T)|U = u]$.

Theorem 2.3 shows that $\hat{\alpha}_{0,\tau}(u)$ and $\hat{\boldsymbol{\alpha}}_\tau(u)$ have the same conditional asymptotic biases as $\tilde{\alpha}_{0,\tau}(u)$ and $\tilde{\boldsymbol{\alpha}}_\tau(u)$, while they have smaller conditional asymptotic variances than $\tilde{\alpha}_{0,\tau}(u)$ and $\tilde{\boldsymbol{\alpha}}_\tau(u)$, respectively. Hence, they are asymptotically more efficient than $\tilde{\alpha}_{0,\tau}(u)$ and $\tilde{\boldsymbol{\alpha}}_\tau(u)$.

3. Semiparametric composite quantile regression. The analysis of semiparametric quantile regression in Section 2 provides a solid foundation for developing the semiparametric composite quantile regression (CQR) estimates. We consider the connection between the quantile regression model (2.1) and model (1.1) in the situations where the random error ε is independent of $(U, \mathbf{X}, \mathbf{Z})$. Let us assume that $Y = \alpha_0(U) + \mathbf{X}^T \boldsymbol{\alpha}(U) + \mathbf{Z}^T \boldsymbol{\beta} + \varepsilon$, where ε follows a distribution F with mean zero. In such situations, $Q_\tau(u, \mathbf{x}, \mathbf{z}) = \alpha_0(u) + c_\tau + \mathbf{x}^T \boldsymbol{\alpha}(u) + \mathbf{z}^T \boldsymbol{\beta}$, where $c_\tau = F^{-1}(\tau)$. Thus, all quantile regression estimates [$\hat{\boldsymbol{\alpha}}_\tau(u)$ and $\hat{\boldsymbol{\beta}}_\tau$ for all τ] estimate the same target quantities [$\boldsymbol{\alpha}(u)$ and $\boldsymbol{\beta}$] with the optimal rate of convergence. Therefore, we can consider combining the information across multiple quantile estimates to obtain improved estimates of $\boldsymbol{\alpha}(u)$ and $\boldsymbol{\beta}$. Such an idea has been studied for the parametric regression model, in Zou and Yuan [35], and it leads to the CQR estimator that is shown to enjoy nice asymptotic efficiency properties compared with the classical least-squares estimator. Kai, Li and Zou [13] proposed the local polynomial CQR estimator for estimating the nonparametric regression function and its derivative. It is shown that the local CQR method can significantly improve the estimation efficiency of the local least-squares estimator for commonly used non-normal error distributions. Inspired by these nice results, we study semiparametric CQR estimates for model (1.1).

Suppose $\{U_i, \mathbf{X}_i, \mathbf{Z}_i, Y_i, i = 1, \ldots, n\}$ is an independent and identically distributed sample from model (1.1) and ε has mean zero. For a given q, let $\tau_k = k/(q+1)$ for $k = 1, 2, \ldots, q$. The CQR procedure estimates $\alpha_0(\cdot)$, $\boldsymbol{\alpha}(\cdot)$ and $\boldsymbol{\beta}$ by minimizing the CQR loss function,

$$\sum_{k=1}^{q} \sum_{i=1}^{n} \rho_{\tau_k} \{Y_i - \alpha_{0k}(U_i) - \mathbf{x}_i^T \boldsymbol{\alpha}(U_i) - \mathbf{z}_i^T \boldsymbol{\beta}\}.$$

To this end, we adapt the three-stage estimation procedure from Section 2.

First, we derive good initial semi-CQR estimates. Let $\{\tilde{\mathbf{a}}_0, \tilde{b}_0, \tilde{\mathbf{a}}, \tilde{\mathbf{b}}, \tilde{\boldsymbol{\beta}}\}$ be the minimizer of the local CQR loss function

$$\sum_{k=1}^{q} \sum_{i=1}^{n} \rho_{\tau_k} \{Y_i - a_{0k} - b_0(U_i - u) - \mathbf{X}_i^T \{\mathbf{a} + \mathbf{b}(U_i - u)\} - \mathbf{Z}_i^T \boldsymbol{\beta}\} K_h(U_i - u),$$

where $\mathbf{a}_0 = (a_{01}, \ldots, a_{0q})^T$, $\mathbf{a} = (a_1, \ldots, a_{d_1})^T$ and $\mathbf{b} = (b_1, \ldots, b_{d_1})^T$. Initial estimates of $\alpha_0(u)$ and $\boldsymbol{\alpha}(u)$ are then given by

$$(3.1) \qquad \tilde{\alpha}_0(u) = \frac{1}{q} \sum_{k=1}^{q} \tilde{a}_{0k}, \qquad \tilde{\boldsymbol{\alpha}}(u) = \tilde{\mathbf{a}}.$$

To investigate asymptotic behaviors of $\tilde{\alpha}_0(u)$, $\tilde{\boldsymbol{\alpha}}(u)$ and $\tilde{\boldsymbol{\beta}}$, let us begin with some new notation. Denote by $f(\cdot)$ and $F(\cdot)$ the density function and cumulative distribution function of the error, respectively. Let $c_k = F^{-1}(\tau_k)$ and C be a $q \times q$ diagonal matrix with $C_{jj} = f(c_j)$. Write $\mathbf{c} = C\mathbf{1}$, $c = \mathbf{1}^T C\mathbf{1}$ and

$$\mathbf{D}_1(u) = E\left[\left.\begin{pmatrix} C & \mathbf{c}\mathbf{X}^T & \mathbf{c}\mathbf{Z}^T \\ \mathbf{X}\mathbf{c}^T & c\mathbf{X}\mathbf{X}^T & c\mathbf{X}\mathbf{Z}^T \\ \mathbf{Z}\mathbf{c}^T & c\mathbf{Z}\mathbf{X}^T & c\mathbf{Z}\mathbf{Z}^T \end{pmatrix}\right| U = u\right].$$

Let $\tau_{kk'} = \tau_k \wedge \tau_{k'} - \tau_k \tau_{k'}$ and let T be a $q \times q$ matrix with the (k, k') element being $\tau_{kk'}$. Write $\mathbf{t} = T\mathbf{1}$, $t = \mathbf{1}^T T\mathbf{1}$ and

$$\boldsymbol{\Sigma}_1(u) = E\left[\left.\begin{pmatrix} T & \mathbf{t}\mathbf{X}^T & \mathbf{t}\mathbf{Z}^T \\ \mathbf{X}\mathbf{t}^T & t\mathbf{X}\mathbf{X}^T & t\mathbf{X}\mathbf{Z}^T \\ \mathbf{Z}\mathbf{t}^T & t\mathbf{Z}\mathbf{X}^T & t\mathbf{Z}\mathbf{Z}^T \end{pmatrix}\right| U = u\right].$$

The following theorem describes the asymptotic sampling distribution of $\{\tilde{\mathbf{a}}_0, \tilde{b}_0, \tilde{\mathbf{a}}, \tilde{\mathbf{b}}, \tilde{\boldsymbol{\beta}}\}$.

THEOREM 3.1. *Under the regularity conditions given in Section 7, if $h \to 0$ and $nh \to \infty$ as $n \to \infty$, then*

$$\sqrt{nh}\left[\begin{pmatrix} \tilde{\mathbf{a}}_0 - \boldsymbol{\alpha}_0(u) \\ \tilde{\mathbf{a}} - \boldsymbol{\alpha}(u) \\ \tilde{\boldsymbol{\beta}} - \boldsymbol{\beta}_0 \end{pmatrix} - \frac{\mu_2 h^2}{2}\begin{pmatrix} \boldsymbol{\alpha}_0''(u) \\ \boldsymbol{\alpha}''(u) \\ \mathbf{0} \end{pmatrix}\right]$$

$$\xrightarrow{\mathcal{D}} N\left(\mathbf{0}, \frac{\nu_0}{f_U(u)}\mathbf{D}_1^{-1}(u)\boldsymbol{\Sigma}_1(u)\mathbf{D}_1^{-1}(u)\right),$$

where $\boldsymbol{\alpha}_0(u) = (\alpha_0(u) + c_1, \ldots, \alpha_0(u) + c_q)^T$ and $\boldsymbol{\beta}_0$ is the true value of $\boldsymbol{\beta}$.

With the initial estimates in hand, we are now ready to derive a $\sqrt{n}$-consistent estimator of $\boldsymbol{\beta}$ by

$$(3.2) \qquad \hat{\boldsymbol{\beta}} = \underset{\boldsymbol{\beta}}{\arg\min}\sum_{k=1}^{q}\sum_{i=1}^{n}\rho_{\tau_k}\{Y_i - \tilde{a}_{0k}(U_i) - \mathbf{X}_i^T\tilde{\mathbf{a}}(U_i) - \mathbf{Z}_i^T\boldsymbol{\beta}\},$$

which is called the *semi-CQR estimator* of $\boldsymbol{\beta}$.

THEOREM 3.2. *Under the regularity conditions given in Section 7, if $nh^4 \to 0$ and $nh^2/\log(1/h) \to \infty$ as $n \to \infty$, then the asymptotic distribution of $\hat{\boldsymbol{\beta}}$ is given by*

$$(3.3) \qquad \sqrt{n}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) \xrightarrow{\mathcal{D}} N\left(0, \frac{1}{c^2}\mathbf{S}^{-1}\boldsymbol{\Delta}\mathbf{S}^{-1}\right),$$

where $\mathbf{S} = E(\mathbf{Z}\mathbf{Z}^T)$ and $\boldsymbol{\Delta} = \sum_{k=1}^{q}\sum_{k'=1}^{q}\tau_{kk'}E[\{\mathbf{Z} - \boldsymbol{\delta}_k(U, \mathbf{X}, \mathbf{Z})\}\{\mathbf{Z} - \boldsymbol{\delta}_{k'}(U, \mathbf{X}, \mathbf{Z})\}^T]$, with $\boldsymbol{\delta}_k(u, \mathbf{x}, \mathbf{z})$ being the kth column of the $d_2 \times q$ matrix

$$\boldsymbol{\delta}(u, \mathbf{x}, \mathbf{z}) = E[\mathbf{Z}(\mathbf{c}^T, c\mathbf{X}^T, \mathbf{0})|U = u]\mathbf{D}_1^{-1}(u)(I_q, \mathbf{1}^T\mathbf{x}, \mathbf{1}^T\mathbf{z})^T.$$

Finally, $\hat{\boldsymbol{\beta}}$ can also be used to further refine the estimates for the nonparametric part. Let $\{\hat{\mathbf{a}}_0, \hat{b}_0, \hat{\mathbf{a}}, \hat{\mathbf{b}}\}$ be the minimizer of

$$\sum_{k=1}^{q} \sum_{i=1}^{n} \rho_{\tau_k}[Y_i - \mathbf{Z}_i^T \hat{\boldsymbol{\beta}} - a_{0k} - b_0(U_i - u) - \mathbf{X}_i^T \{\mathbf{a} + \mathbf{b}(U_i - u)\}] K_h(U_i - u),$$

where $\mathbf{a}_0 = (a_{01}, \ldots, a_{0q})^T$. We then define the *semi-CQR estimators* for $\alpha_0(u)$ and $\boldsymbol{\alpha}(u)$ as

$$(3.4) \qquad \hat{\alpha}_0(u) = \frac{1}{q} \sum_{k=1}^{q} \hat{a}_{0k}, \qquad \hat{\boldsymbol{\alpha}}(u) = \hat{\mathbf{a}}.$$

We now study the asymptotic properties of $\hat{\alpha}_0(u)$ and $\hat{\boldsymbol{\alpha}}(u)$. Let

$$\mathbf{D}_2(u) = E\left[\begin{pmatrix} C & \mathbf{c}\mathbf{X}^T \\ \mathbf{X}\mathbf{c}^T & c\mathbf{X}\mathbf{X}^T \end{pmatrix} \middle| U = u\right],$$

$$\boldsymbol{\Sigma}_2(u) = E\left[\begin{pmatrix} T & \mathbf{t}\mathbf{X}^T \\ \mathbf{X}\mathbf{t}^T & t\mathbf{X}\mathbf{X}^T \end{pmatrix} \middle| U = u\right].$$

THEOREM 3.3.   *Under the regularity conditions given in Section 7, if $h \to 0$ and $nh \to \infty$ as $n \to \infty$, the asymptotic distributions of $\hat{\alpha}_0(u)$ and $\hat{\boldsymbol{\alpha}}(u)$ are given by*

$$\sqrt{nh}\left(\hat{\alpha}_0(u) - \alpha_0(u) - \frac{1}{q}\sum_{k=1}^{q} c_k - \frac{\mu_2 h^2}{2}\alpha_0''(u)\right)$$

$$\xrightarrow{\mathcal{D}} N\left(0, \frac{v_0}{f_U(u)}\frac{1}{q^2}\mathbf{1}^T[\mathbf{D}_2^{-1}(u)\boldsymbol{\Sigma}_2(u)\mathbf{D}_2^{-1}(u)]_{11}\mathbf{1}\right)$$

and

$$\sqrt{nh}\left(\hat{\boldsymbol{\alpha}}(u) - \boldsymbol{\alpha}(u) - \frac{\mu_2 h^2}{2}\boldsymbol{\alpha}''(u)\right) \xrightarrow{\mathcal{D}} N\left(0, \frac{v_0}{f_U(u)}[\mathbf{D}_2^{-1}(u)\boldsymbol{\Sigma}_2(u)\mathbf{D}_2^{-1}(u)]_{22}\right),$$

where $[\cdot]_{11}$ denotes the upper-left $q \times q$ submatrix and $[\cdot]_{22}$ denotes the lower-right $d_1 \times d_1$ submatrix.

REMARK 1. $\boldsymbol{\alpha}(u)$ and $\boldsymbol{\beta}$ represent the contributions of covariates. They are the central quantities of interest in semiparametric inference. Li and Liang [21] studied the least-squares-based semiparametric estimation, which we will refer to as "semi-LS" in this work. The major advantage of semi-CQR over the classical semi-LS is that semi-CQR has competitive asymptotic efficiency. Furthermore, semi-CQR is also more stable and robust. Intuitively speaking, these advantages come from the fact that semi-CQR utilizes information shared across multiple quantile functions, whereas semi-LS only uses the information contained in the mean function.

To elaborate on Remark 1, we discuss the relative efficiency of semi-CQR relative to semi-LS. Note that $E(Y|U) = \alpha_0(U) + E(\mathbf{X}|U)^T \boldsymbol{\alpha}(U) + E(\mathbf{Z}|U)^T \boldsymbol{\beta}$. It then follows that $Y = E(Y|U) + \{\mathbf{X} - E(\mathbf{X}|U)\}^T \boldsymbol{\alpha}(U) + \{\mathbf{Z} - E(\mathbf{Z}|U)\}^T \boldsymbol{\beta} + \epsilon$. Without loss of generality, let us consider the situation in which $E(\mathbf{X}|U) = 0$ and $E(\mathbf{Z}|U) = 0$. Then, all $\mathbf{D}_1(u), \mathbf{D}_2(u), \boldsymbol{\Sigma}_1(u)$ and $\boldsymbol{\Sigma}_2(u)$ become block diagonal matrices. Thus, from Theorem 3.3, we have

$$\sqrt{nh}\left(\hat{\alpha}_0(u) - \alpha_0(u) - \frac{1}{q}\sum_{k=1}^{q} c_k - \frac{\mu_2 h^2}{2}\alpha_0''(u)\right) \xrightarrow{\mathcal{D}} N\left(0, R_1(q)\frac{v_0}{f_U(u)}\right)$$

and

$$\sqrt{nh}\left(\hat{\boldsymbol{\alpha}}(u) - \boldsymbol{\alpha}(u) - \frac{\mu_2 h^2}{2}\boldsymbol{\alpha}''(u)\right) \xrightarrow{\mathcal{D}} N\left(0, R_2(q)\frac{v_0}{f_U(u)}E^{-1}(\mathbf{XX}^T|U=u)\right),$$

where

$$R_1(q) = \frac{1}{q^2}\mathbf{1}^T C^{-1} T C^{-1} \mathbf{1} = \frac{1}{q^2}\sum_{k=1}^{q}\sum_{k'=1}^{q}\frac{\tau_{kk'}}{f(c_k)f(c_{k'})}$$

and

$$R_2(q) = \frac{t}{c^2} = \frac{\sum_{k=1}^{q}\sum_{k'=1}^{q}\tau_{kk'}}{\{\sum_{k=1}^{q}f(c_k)\}^2}.$$

Note that

$$\boldsymbol{\delta}(u, \mathbf{x}, \mathbf{z}) = E(\mathbf{ZX}^T, \mathbf{0}|U=u)E\left[\begin{pmatrix}\mathbf{XX}^T & \mathbf{XZ}^T \\ \mathbf{ZX}^T & \mathbf{ZZ}^T\end{pmatrix}\middle| U=u\right]^{-1}(\mathbf{1}^T\mathbf{x}, \mathbf{1}^T\mathbf{z})^T$$

with all columns of $\boldsymbol{\delta}(u, \mathbf{x}, \mathbf{z})$ the same. Thus, $\boldsymbol{\Delta} = t\boldsymbol{\Delta}_0$ with $\boldsymbol{\Delta}_0 = E[\{\mathbf{Z} - \boldsymbol{\delta}_1(U, \mathbf{X}, \mathbf{Z})\}\{\mathbf{Z} - \boldsymbol{\delta}_1(U, \mathbf{X}, \mathbf{Z})\}^T]$. It is easy to show that $E\{\boldsymbol{\delta}_1(U, \mathbf{X}, \mathbf{Z})\mathbf{Z}^T\} = 0$ and we then have

$$\boldsymbol{\Delta}_0 = E[E(\mathbf{ZZ}^T|U)$$
$$\times \{E(\mathbf{ZZ}^T|U) - E(\mathbf{ZX}^T|U)E(\mathbf{XX}^T|U)^{-1}E(\mathbf{XZ}^T|U)\}^{-1}E(\mathbf{ZZ}^T|U)].$$

Therefore,

(3.5) $$\sqrt{n}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) \xrightarrow{\mathcal{D}} N(0, R_2(q)\mathbf{S}^{-1}\boldsymbol{\Delta}_0\mathbf{S}^{-1}).$$

If we replace $R_2(q)$ with 1 in equations (23) and (24), we end up with the asymptotic normal distributions of the semi-LS estimators, as studied in Li and Liang [21]. Thus, $R_2(q)$ determines the asymptotic relative efficiency (ARE) of semi-CQR relative to semi-LS. By direct calculations, we see that the ARE for estimating $\boldsymbol{\alpha}(u)$ is $R_2(q)^{-4/5}$ and the ARE for estimating $\boldsymbol{\beta}$ is $R_2(q)^{-1}$. It is interesting to see that the same factor, $R_2(q)$, also appears in the asymptotic efficiency analysis of parametric CQR [35] and nonparametric local CQR smoothing [13].

The basic message is that, with a relatively large q ($q \geq 9$), $R_2(q)$ is very close to 1 for the normal errors, but can be much smaller than 1, meaning a huge gain in efficiency, for the commonly seen non-normal errors. It is also shown in [13] that $\lim_{q \to \infty} R_2(q)^{-1} \geq 0.864$ and hence $\lim_{q \to \infty} R_2(q)^{-4/5} \geq 0.8896$, which implies that when a large q is used, the ARE is at least 88.9% for estimating varying-coefficient functions and at least 86.4% for estimating parametric components.

REMARK 2.   The baseline function estimator $\hat{\alpha}_0(u)$ converges to $\alpha_0(u)$ plus the average of uniform quantiles of the error distribution. Therefore, the bias term is zero when the error distribution is symmetric. Even for nonasymmetric distributions, the additional bias term converge to the mean of the error, which is zero for a large value of q. Nevertheless, its asymptotic variance differs from that of the semi-LS estimator by a factor of $R_1(q)$. The study in Kai, Li and Zou [13] shows that $R_1(q)$ approaches 1 as q becomes large and $R_1(q)$ could be much smaller than 1 with a smaller q ($q \leq 9$) for commonly used non-normal distributions.

REMARK 3.   The factors $R_1(q)$ and $R_2(q)$ only depend on the error distribution. We have observed from our simulation study that, as a function of q, the maximum of $R_2(q)$ is often closely approximated by R_2 ($q = 9$). Hence, if we only care about the inference of $\boldsymbol{\alpha}(u)$ and $\boldsymbol{\beta}$, then $q = 9$ seems to be a good default value. On the other hand, R_1 ($q = 5$) is often close to the maximum of $R_1(q)$ based on our numerical study and hence $q = 5$ is a good default value for estimating the baseline function. If prediction accuracy is the primary interest, then we should use a proper q to maximize the total contributions from $R_1(q)$ and $R_2(q)$. Practically speaking, one can choose a q from the interval [5, 9] by some popular tuning methods such as K-fold cross-validation. However, we do not expect these CQR models to have significant differences in terms of model fitting and prediction because, in many cases, $R_1(q)$ and $R_2(q)$ vary little in the interval [5, 9].

4. Variable selection.   Variable selection is a crucial step in high-dimensional modeling. Various powerful penalization methods have been developed for variable selection in parametric models; see Fan and Li [7] for a good review. In the literature, there are only a few papers on variable selection in semiparametric regression models. Li and Liang [21] proposed the nonconcave penalized quasi-likelihood method for variable selection in semiparametric varying-coefficient models. In this section, we study the penalized semiparametric CQR estimator.

Let $p_{\lambda_n}(\cdot)$ be a pre-specified penalty function with regularization parameter λ_n. We consider the penalized CQR loss

$$(4.1) \quad \sum_{k=1}^{q} \sum_{i=1}^{n} \rho_{\tau_k}\{Y_i - \tilde{\alpha}_{0k}(U_i) - \mathbf{X}_i^T \tilde{\boldsymbol{\alpha}}(U_i) - \mathbf{Z}_i^T \boldsymbol{\beta}\} + nq \sum_{j=1}^{d_2} p_{\lambda_n}(|\beta_j|).$$

By minimizing the above objective function with a proper penalty parameter λ_n, we can get a sparse estimator of $\boldsymbol{\beta}$ and hence conduct variable selection.

Fan and Li [6] suggested using a concave penalty function since it is able to produce an oracular estimator, that is, the penalized estimator performs as well as if the subset model were known in advance. However, optimizing (4.1) with a concave penalty function is very challenging because the objective function is nonconvex and both loss and penalty parts are nondifferentiable. Various numerical algorithms have been proposed to address the computational difficulties. Fan and Li [6] suggested using local quadratic approximation (LQA) to substitute for the penalty function and then optimizing using the Newton–Raphson algorithm. Hunter and Li [12] further proposed a perturbed version of LQA to alleviate one drawback of LQA. Recently, Zou and Li [34] proposed a new unified algorithm based on local linear approximation (LLA) and further suggested using the one-step LLA estimator because the one-step LLA automatically adopts a sparse representation and is as efficient as the fully iterated LLA estimator. Thus, the one-step LLA estimator is computationally and statistically efficient.

We proposed to follow the one-step sparse estimate scheme in Zou and Li [34] to derive a one-step sparse semi-CQR estimator, as follows. First, we compute the unpenalized semi-CQR estimate $\hat{\boldsymbol{\beta}}^{(0)}$, as described in Section 3. We then define

$$G_n(\boldsymbol{\beta}) = \sum_{k=1}^{q} \sum_{i=1}^{n} \rho_{\tau_k}\{Y_i - \tilde{\alpha}_{0k}(U_i) - \mathbf{X}_i^T \tilde{\boldsymbol{\alpha}}(U_i) - \mathbf{Z}_i^T \boldsymbol{\beta}\} + nq \sum_{j=1}^{d_2} p'_{\lambda_n}(|\beta_j^{(0)}|)|\beta_j|.$$

We define $\hat{\boldsymbol{\beta}}^{\mathrm{OSE}} = \mathrm{argmin}_{\boldsymbol{\beta}} G_n(\boldsymbol{\beta})$ and call this the *one-step sparse semi-CQR* estimator. Indeed, this is a weighted L_1 regularization procedure.

We now show that the one-step sparse semi-CQR estimator $\hat{\boldsymbol{\beta}}^{\mathrm{OSE}}$ enjoys the oracle property. This property holds for a wide class of concave penalties. To establish the idea, we focus on the SCAD penalty from Fan and Li [6], which is perhaps the most popular concave penalty in the literature. Let $\boldsymbol{\beta}_0 = (\boldsymbol{\beta}_{10}^T, \boldsymbol{\beta}_{20}^T)^T$ denote the true value of $\boldsymbol{\beta}$, where $\boldsymbol{\beta}_{10}$ is a s-vector. Without loss of generality, we assume that $\boldsymbol{\beta}_{20} = \mathbf{0}$ and that $\boldsymbol{\beta}_{10}$ contains all nonzero components of $\boldsymbol{\beta}_0$. Furthermore, let $\mathbf{Z}_1$ be the first s elements of $\mathbf{Z}$ and define

$$\boldsymbol{\lambda}(u, \mathbf{x}, \mathbf{z}) = E[\mathbf{Z}_1(\mathbf{c}^T, c\mathbf{X}^T, \mathbf{0})|U = u]\mathbf{D}_1^{-1}(u)(I_q, \mathbf{1}^T\mathbf{x}, \mathbf{1}^T\mathbf{z})^T.$$

THEOREM 4.1 (Oracle property).   *Let* $p_\lambda(\cdot)$ *be the SCAD penalty. Assume that the regularity conditions* (B1)–(B6) *given in the Appendix hold. If* $\sqrt{n}\lambda_n \to \infty$, $\lambda_n \to 0$, $nh^4 \to 0$ *and* $nh^2/\log(1/h) \to \infty$ *as* $n \to \infty$, *then the one-step semi-CQR estimator* $\hat{\boldsymbol{\beta}}^{\mathrm{OSE}}$ *must satisfy*:

  (a) *sparsity, that is,* $\hat{\boldsymbol{\beta}}_2^{\mathrm{OSE}} = 0$, *with probability tending to one*;
  (b) *asymptotic normality, that is,*

$$\text{(4.2)} \qquad \sqrt{n}(\hat{\boldsymbol{\beta}}_1^{\mathrm{OSE}} - \boldsymbol{\beta}_{10}) \xrightarrow{\mathcal{D}} N\left(0, \frac{1}{c^2}\mathbf{S}_1^{-1}\boldsymbol{\Lambda}\mathbf{S}_1^{-1}\right),$$

where $\mathbf{S}_1 = E(\mathbf{Z}_1\mathbf{Z}_1^T)$ *and* $\boldsymbol{\Lambda} = \sum_{k=1}^{q}\sum_{k'=1}^{q} \tau_{kk'} E[\{\mathbf{Z}_1 - \boldsymbol{\lambda}_k(U, \mathbf{X}, \mathbf{Z})\}\{\mathbf{Z}_1 - \boldsymbol{\lambda}_{k'}(U, \mathbf{X}, \mathbf{Z})\}^T]$ *with* $\boldsymbol{\lambda}_k(u, \mathbf{x}, \mathbf{z})$ *being the kth column of the matrix* $\boldsymbol{\lambda}(u, \mathbf{x}, \mathbf{z})$.

Theorem 4.1 shows the asymptotic magnitude of the optimal λ_n. For a given data set with finite sample, it is practically important to have a data-driven method to select a good λ_n. Various techniques have been proposed in previous studies, such as the generalized cross-validation selector [6] and the BIC selector [27]. In this work, we use a BIC-like criterion to select the penalization parameter. The BIC criterion is defined as

$$\text{BIC}(\lambda) = \log\left(\sum_{k=1}^{q}\sum_{i=1}^{n}\rho_{\tau_k}\{Y_i - \hat{\alpha}_{0k}(U_i) - \mathbf{X}_i^T\hat{\boldsymbol{\alpha}}(U_i) - \mathbf{Z}_i^T\hat{\boldsymbol{\beta}}^{\text{OSE}}(\lambda)\}\right)$$
$$+ \frac{\log(n)}{n}df_\lambda,$$

where df_λ is the number of nonzero coefficients in the parametric part of the fitted model. We let $\hat{\lambda}_{\text{BIC}} = \arg\min \text{BIC}(\lambda)$. The performance of $\hat{\lambda}_{\text{BIC}}$ will be examined in our simulation studies in the next section.

REMARK 4.   Variable selection in linear quantile regression has been considered in several papers; see Li and Zhu [22] and Wu and Liu [30]. The developed method for sparse semiparametric CQR can be easily adopted for variable selection in semiparametric quantile regression. Consider the penalized check loss

$$(4.3) \qquad \sum_{i=1}^{n}\rho_\tau\{Y_i - \tilde{\alpha}_{0,\tau}(U_i) - \mathbf{X}_i^T\tilde{\boldsymbol{\alpha}}_\tau(U_i) - \mathbf{Z}_i^T\boldsymbol{\beta}\} + n\sum_{j=1}^{d_2}p_{\lambda_n}(|\beta_j|).$$

For its one-step version, we use

$$(4.4) \quad \sum_{i=1}^{n}\rho_\tau\{Y_i - \tilde{\alpha}_{0,\tau}(U_i) - \mathbf{X}_i^T\tilde{\boldsymbol{\alpha}}_\tau(U_i) - \mathbf{Z}_i^T\boldsymbol{\beta}\} + n\sum_{j=1}^{d_2}p'_{\lambda_n}(|\beta_j^{(0)}|)|\beta_j|,$$

where $\hat{\boldsymbol{\beta}}^{(0)}$ denotes the unpenalized semiparametric quantile regression estimator defined in Section 2. We can also prove the oracle property of the one-step sparse semiparametric quantile regression estimator by following the lines of proof for Theorem 4.1. For reasons of brevity, we omit the details here.

5. Numerical studies.   In this section, we conduct simulation studies to assess the finite-sample performance of the proposed procedures and illustrate the proposed methodology on a real-world data set in a health study. In all examples, we fix the kernel function to be the Epanechnikov kernel, that is, $K(u) = \frac{3}{4}(1 - u^2)_+$, and we use the SCAD penalty function for variable selection. Note that all proposed estimators, including semi-QR, semi-CQR and one-step sparse semi-CQR, can be formulated as linear programming (LP) problems. In our study, we solved these estimators by using LP tools.

EXAMPLE 1. In this example, we generate 400 random samples, each consisting of $n = 200$ observations, from the model

$$Y = \alpha_1(U)X_1 + \alpha_2(U)X_2 + \beta_1 Z_1 + \beta_2 Z_2 + \beta_3 Z_3 + \varepsilon,$$

where $\alpha_1(U) = \sin(6\pi U)$, $\alpha_2(U) = \sin(2\pi U)$, $\beta_1 = 2$, $\beta_2 = 1$ and $\beta_3 = 0.5$. The covariate U is from the uniform distribution on $[0, 1]$. The covariates X_1, X_2, Z_1, Z_2 are jointly normally distributed with mean 0, variance 1 and correlation 2/3. The covariate Z_3 is Bernoulli with $\Pr(Z_3 = 1) = 0.4$. Furthermore, U and $(X_1, X_2, Z_1, Z_2, Z_3)$ are independent. In our simulation, we considered the following error distributions: $N(0, 1)$, logistic, standard Cauchy, t-distribution with 3 degrees of freedom, mixture of normals $0.9N(0, 1) + 0.1N(0, 10^2)$ and log-normal distribution. Because the error is independent of the covariates, the least-squares (LS), quantile regression (QR) and composite quantile regression (CQR) procedures provide estimates for the same quantity and hence are directly comparable.

Performance of $\hat{\beta}_\tau$ and $\hat{\beta}$. To examine the performance of the proposed procedures with a wide range of bandwidths, three bandwidths for LS were considered, $h = 0.085, 0.128, 0.192$, which correspond to the undersmoothing, appropriate smoothing and oversmoothing, respectively. By straightforward calculation, as in Kai, Li and Zou [13], we can produce two simple formulas for the asymptotic optimal bandwidths for QR and CQR: $h_{\text{CQR}} = h_{\text{LS}} \cdot R_2(q)^{1/5}$ and $h_{\text{QR},\tau} = h_{\text{LS}} \cdot \{\tau(1-\tau)/f[F^{-1}(\tau)]\}^{1/5}$, where h_{LS} is the asymptotic optimal bandwidth for LS. We considered only the case of normal error. The bias and standard deviation based on 400 simulations are reported in Table 1. First, we see that the estimators are not very sensitive to the choice of bandwidth. As for the estimation accuracy, all three estimators have comparable bias and the differences are shown in standard deviation. The LS estimates have the smallest standard deviation, as expected. The CQR estimates are slightly worse than the LS estimates.

In the second study, we fixed $h = 0.128$ and compared the efficiency of QR and CQR relative to LS. Reported in Table 2 are RMSEs, the ratios of the MSEs of the QR and CQR estimators to the LS estimator for different error distributions. Several observations can be made from Table 2. When the error follows the normal distribution, the RMSEs of CQR are slightly less than 1. For all other non-normal distributions in the table, the RMSE can be much greater than 1, indicating a huge gain in efficiency. These findings agree with the asymptotic theory. For QR estimators, their performance varies and depends heavily on the level of quantile and the error distribution. Overall, CQR outperforms both QR and LS.

Performance of $\hat{\alpha}_\tau$ and $\hat{\alpha}$. We now compare the LS, QR and CQR estimates for α by using the ratio of average squared errors (RASE). We first compute

$$\text{ASE} = \left\{ \frac{1}{n_{\text{grid}}} \sum_{m=1}^{d_1} \sum_{k=1}^{n_{\text{grid}}} \{\hat{a}_m(u_k) - a_m(u_k)\}^2 \right\},$$

TABLE 1
Summary of the bias and standard deviation over 400 *simulations*

		Bias(SD)		
h	Method	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$
0.085	LSE	−0.012 (0.128)	0.008 (0.121)	−0.009 (0.171)
	CQR_9	−0.009 (0.131)	0.009 (0.125)	−0.007 (0.172)
	$QR_{0.25}$	−0.017 (0.163)	0.009 (0.161)	−0.151 (0.237)
	$QR_{0.50}$	−0.012 (0.155)	0.011 (0.151)	−0.007 (0.198)
	$QR_{0.75}$	−0.007 (0.165)	0.005 (0.158)	0.122 (0.216)
0.128	LSE	−0.009 (0.121)	0.005 (0.117)	−0.008 (0.164)
	CQR_9	−0.010 (0.127)	0.008 (0.121)	−0.005 (0.163)
	$QR_{0.25}$	−0.010 (0.159)	0.003 (0.152)	−0.082 (0.227)
	$QR_{0.50}$	−0.008 (0.154)	0.011 (0.147)	−0.004 (0.193)
	$QR_{0.75}$	−0.012 (0.163)	0.003 (0.161)	0.071 (0.207)
0.192	LSE	−0.007 (0.128)	0.001 (0.123)	−0.008 (0.169)
	CQR_9	−0.009 (0.131)	0.005 (0.127)	−0.005 (0.169)
	$QR_{0.25}$	−0.006 (0.169)	−0.004 (0.169)	−0.061 (0.230)
	$QR_{0.50}$	−0.005 (0.153)	0.006 (0.152)	−0.007 (0.191)
	$QR_{0.75}$	−0.012 (0.170)	0.007 (0.171)	0.049 (0.225)

where $\{u_k : k = 1, \ldots, n_{\text{grid}}\}$ is a set of grid points uniformly placed on $[0, 1]$ with $n_{\text{grid}} = 200$. RASE is then defined to be

$$(5.1) \qquad \qquad \text{RASE}(\hat{g}) = \frac{\text{ASE}(\hat{g}_{\text{LS}})}{\text{ASE}(\hat{g})}$$

for an estimator $\hat{g}$, where $\hat{g}_{\text{LS}}$ is the least-squares-based estimator.

The sample mean and standard deviation of the RASEs over 400 simulations are presented in Table 3, where the values in the parentheses are the standard deviations. The findings are quite similar to those in Table 2. We see that CQR performs almost as well as LS when the error is normally distributed. Also, its RASEs are much larger than 1 for other non-normal error distributions. The efficiency gain can be substantial. Note that for the Cauchy distribution, RASEs of QR and CQR are huge—this is because LS fails when the error variance is infinite.

EXAMPLE 2.   The goal is to compare the proposed one-step sparse semi-CQR estimator with the one-step sparse semi-LS estimator. In this example, 400 random samples, each consisting of $n = 200$ observations, were generated from the varying-coefficient partially linear model

$$Y = \alpha_1(U)X_1 + \alpha_2(U)X_2 + \boldsymbol{\beta}^T \mathbf{Z} + \varepsilon,$$

where $\boldsymbol{\beta} = (3, 1.5, 0, 0, 2, 0, 0, 0)^T$ and the covariate vector $(X_1, X_2, \mathbf{Z}^T)^T$ is normally distributed with mean 0, variance 1 and correlation $0.5^{|i-j|}$ ($i, j =$

TABLE 2
Summary of the ratio of MSE over 400 *simulations*

	RMSE		
Method	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$
Standard normal			
CQR$_9$	0.920	0.932	1.011
QR$_{0.25}$	0.585	0.594	0.460
QR$_{0.50}$	0.621	0.631	0.724
QR$_{0.75}$	0.554	0.528	0.561
Logistic			
CQR$_9$	1.044	1.083	1.016
QR$_{0.25}$	0.651	0.664	0.502
QR$_{0.50}$	0.826	0.871	0.799
QR$_{0.75}$	0.661	0.732	0.527
Standard Cauchy			
CQR$_9$	15,246	106,710	52,544
QR$_{0.25}$	8894	56,704	24,359
QR$_{0.50}$	19,556	137,109	66,560
QR$_{0.75}$	8223	62,282	26,210
t-distribution with $df = 3$			
CQR$_9$	1.554	1.546	1.683
QR$_{0.25}$	1.000	0.948	0.819
QR$_{0.50}$	1.354	1.333	1.451
QR$_{0.75}$	0.935	1.059	0.859
$0.9N(0, 1) + 0.1N(0, 10^2)$			
CQR$_9$	5.752	4.860	5.152
QR$_{0.25}$	3.239	3.096	2.300
QR$_{0.50}$	5.430	4.730	4.994
QR$_{0.75}$	3.790	2.952	2.515
Log-normal			
CQR$_9$	3.079	3.369	3.732
QR$_{0.25}$	5.198	5.361	3.006
QR$_{0.50}$	2.787	2.829	3.139
QR$_{0.75}$	0.819	0.868	0.823

TABLE 3
Summary of the RASE over 400 *simulations*

	Normal	Logistic	Cauchy	t_3	Mixture	Log-normal
CQR$_9$	0.968 (0.104)	1.040 (0.134)	12,872 (176719)	1.428 (1.299)	3.292 (1.405)	2.455 (1.498)
QR$_{0.25}$	0.666 (0.160)	0.720 (0.203)	7621 (110692)	0.958 (0.647)	2.029 (1.003)	3.490 (3.224)
QR$_{0.50}$	0.771 (0.184)	0.881 (0.206)	13,720 (187298)	1.274 (1.166)	3.155 (1.323)	2.155 (1.674)
QR$_{0.75}$	0.681 (0.191)	0.713 (0.201)	5781 (87909)	0.896 (0.325)	1.953 (0.905)	0.824 (0.679)

$1, \dots, 10$). Other model settings are exactly the same as those in Example 1. We use the generalized mean square error (GMSE), as defined in [21],

$$\text{(5.2)} \qquad \text{GMSE}(\hat{\boldsymbol{\beta}}) = (\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})^T E(\mathbf{Z}\mathbf{Z}^T)(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}),$$

to assess the performance of variable selection procedures for the parametric component. For each procedure, we calculate the relative GMSE (RGMSE), which is defined to be the ratio of the GMSE of a selected final model to that of the unpenalized least-squares estimate under the full model.

The results over 400 simulations are summarized in Table 4, where the column "RGMSE" reports both the median and MAD of 400 RGMSEs. Both columns "C" and "IC" are measures of model complexity. Column "C" shows the average number of zero coefficients correctly estimated to be zero and column "IC" presents the average number of nonzero coefficients incorrectly estimated to be zero. In the column labeled "U-fit" (short for "under-fit"), we present the proportion of trials excluding any nonzero coefficients in 400 replications. Likewise, we report the probability of trials selecting the exact subset model and the probability of trials including all three significant variables and some noise variables in the columns "C-fit" ("correct-fit") and "O-fit" ("over-fit"), respectively.

TABLE 4
One-step estimates for variable selection in semiparametric models

Method	RGMSE Median (MAD)	No. of zeros		Proportion of fits		
		C	IC	U-fit	C-fit	O-fit
Standard normal						
One-step LS	0.335 (0.194)	4.825	0.000	0.000	0.867	0.133
One-step CQR	0.288 (0.213)	4.990	0.000	0.000	0.990	0.010
Logistic						
One-step LS	0.352 (0.197)	4.805	0.000	0.000	0.870	0.130
One-step CQR	0.289 (0.206)	4.975	0.000	0.000	0.975	0.025
Standard Cauchy						
One-step LS	0.956 (0.249)	2.920	0.795	0.595	0.108	0.297
One-step CQR	0.005 (0.021)	5.000	0.295	0.210	0.790	0.000
t-distribution with $df = 3$						
One-step LS	0.346 (0.179)	4.803	0.000	0.000	0.860	0.140
One-step CQR	0.183 (0.177)	4.987	0.000	0.000	0.988	0.013
$0.9N(0, 1) + 0.1N(0, 10^2)$						
One-step LS	0.331 (0.190)	4.848	0.000	0.000	0.883	0.117
One-step CQR	0.060 (0.083)	4.997	0.000	0.000	0.998	0.003
Log-normal						
One-step LS	0.303 (0.182)	4.845	0.000	0.000	0.887	0.113
One-step CQR	0.111 (0.118)	4.990	0.000	0.000	0.990	0.010

From Table 4, we see that both variable selection procedures dramatically re-
duce model errors, which clearly show the virtue of variable selection. Second,
the one-step CQR performs better than the one-step LS in terms of all of the
criteria: RGMSE, number of zeros and proportion of fits, and for all of the er-
ror distributions in Table 4. It is also interesting to see that in the normal error
case, the one-step CQR seems to perform no worse than the one-step LS (or even
slightly better). We performed the Mann–Whitney test to compare their RGM-
SEs and the corresponding p-value is 0.0495. This observation appears to be
contradictory to the asymptotic theory. However, this "contradiction" can be ex-
plained by observing that the one-step CQR has better variable selection perfor-
mance. Note that the one-step CQR has significantly higher probability of correct-
select than the one-step LS, which also tends to overselect. Thus, the one-step
LS needs to estimate a larger model than the truth, compared to the one-step
CQR.

EXAMPLE 3.   As an illustration, we apply the proposed procedures to ana-
lyze the plasma beta-carotene level data set collected by a cross-sectional study
[24]. This data set consists of 273 samples. Of interest are the relationships be-
tween the plasma beta-carotene level and the following covariates: age, smoking
status, quetelet index (BMI), vitamin use, number of calories, grams of fat, grams
of fiber, number of alcoholic drinks, cholesterol and dietary beta-carotene. The
complete description of the data can be found in the StatLib database via the link
lib.stat.cmu.edu/datasets/Plasma_Retinol.

We fit the data by using a partially linear model with U being "*dietary beta-
carotene*." The covariates "*smoking status*" and "*vitamin use*" are categorical and
are thus replaced with dummy variables. All of the other covariates are standard-
ized. We applied the one-step sparse CQR and LS estimators to fit the partially lin-
ear regression model. Five-fold cross-validation was used to select the bandwidths
for LS and CQR. We used the first 200 observations as a training data set to fit the
model and to select significant variables, then used the remaining 73 observations
to evaluate the predictive ability of the selected model.

The prediction performance is measured by the median absolute prediction error
(MAPE), which is the median of $\{|y_i - \hat{y}_i|, i = 1, 2, \ldots, 73\}$. To see the effect
of q on the CQR estimate, we tried $q = 5, 7, 9$. We found that the selected Z-
variables are the same for these three values of q and their MAPEs are 58.52,
58.11, 62.43, respectively. Thus, the effect of q is minor. The resulting model
with $q = 7$ is given in Table 5 and the estimated intercept function is depicted in
Figure 1. From Table 5, it can be seen that the CQR model is much sparser than
the LS model. Only two covariates, "*fiber consumption per day*" and "*fairly often
use of vitamin*" are included in the parametric part of the CQR model. Meanwhile,
the CQR model has much better prediction performance than the LS model, whose
MAPE is 111.28.

TABLE 5
Selected parametric components for plasma beta-carotene level data

	$\hat{\beta}_{LS}^{OSE}$	$\hat{\beta}_{CQR}^{OSE}$
Age	0	0
Quetelet index	0	0
Calories	-100.47	0
Fat	52.60	0
Fiber	87.51	29.89
Alcohol	44.61	0
Cholesterol	0	0
Smoking status (never)	51.71	0
Smoking status (former)	72.48	0
Vitamin use (yes, fairly often)	130.39	30.21
Vitamin use (yes, not often)	0	0
MAPE	111.28	58.11

6. Discussion.   We discuss some directions in which this work could be further extended. We have focused on using uniform weights in composite quantile regression. In theory, we can use nonuniform weights, which may provide an even more efficient estimator when a reliable estimate of the error distribution is available. Koenker [16] discussed the theoretically optimal weights. Bradic, Fan and Wang [1] suggested a data-driven weighted CQR for parametric linear regression, in which the weights mimic the optimal weights. The idea in Bradic, Fan and
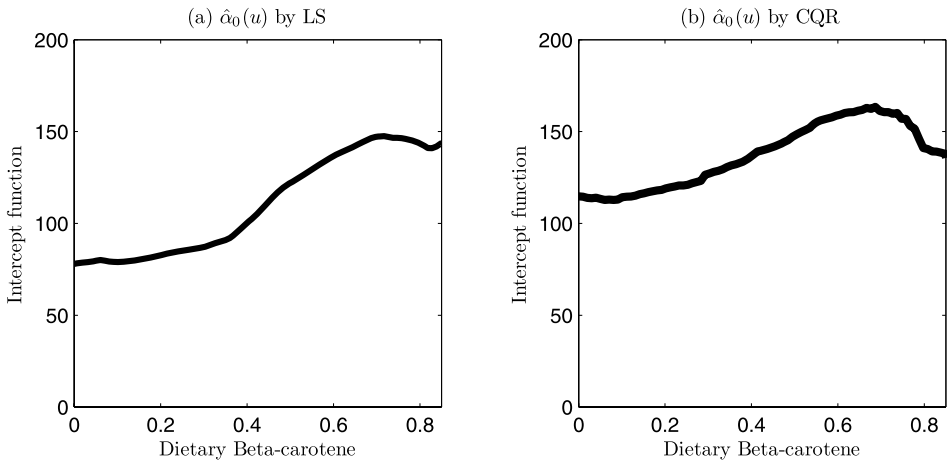

FIG. 1.   *Plot of estimated intercept function of dietary beta-carotene*: (a) *the estimated intercept function by LS method*; (b) *the estimated intercept function by CQR method with q* $= 7$.

Wang [1] can be easily extended to the semi-CQR estimator, which will be investigated in detail in a future paper.

Penalized Wilcoxon rank regression has been considered independently in Leng [20] and Wang and Li [29] and found to achieve a similar efficiency property of CQR for variable selection in parametric linear regression. We could also generalize rank regression to handle semiparametric varying-coefficient partially linear models. In a working paper, we show that rank regression is exactly equivalent to CQR using $q = n - 1$ quantiles with uniform weights. This result indicates that CQR is more flexible than rank regression because we can easily use flexible nonuniform weights in CQR to further improve efficiency, as in Bradic, Fan and Wang [1]. Obviously, CQR is also computationally more efficient than rank regression. We note that in parametric linear regression models, rank regression has no efficiency gain over least-squares for estimating the intercept. This result is expected to hold for estimating the baseline function in the semiparametric varying-coefficient partially linear model.

When the number of varying coefficient components is large, it is also desirable to consider selecting a few important components. This problem was studied in Wang and Xia [28], where a LASSO-type penalized local least-squares estimator was proposed. It would be interesting to apply CQR to their method to further improve the estimation efficiency.

7. Proofs. To establish the asymptotic properties of the proposed estimators, the following regularity conditions are imposed:

(C1) the random variable U has bounded support Ω and its density function $f_U(\cdot)$ is positive and has a continuous second derivative;

(C2) the varying coefficients $\alpha_0(\cdot)$ and $\boldsymbol{\alpha}(\cdot)$ have continuous second derivatives in $u \in \Omega$;

(C3) $K(\cdot)$ is a symmetric density function with bounded support and satisfies a Lipschitz condition;

(C4) the random vector $\mathbf{Z}$ has bounded support;

(C5) for the semi-QR procedure,

(i) $F_\tau(0|u, \mathbf{x}, \mathbf{z}) = \tau$ for all $(u, \mathbf{x}, \mathbf{z})$, and $f_\tau(\cdot|u, \mathbf{x}, \mathbf{z})$ is bounded away from zero and has a continuous and uniformly bounded derivative;

(ii) $\mathbf{A}_1(u)$ defined in Theorem 2.1 and $\mathbf{A}_2(u)$ defined in Theorem 2.3 are nonsingular for all $u \in \Omega$;

(C6) for the semi-CQR procedure,

(i) $f(\cdot)$ is bounded away from zero and has a continuous and uniformly bounded derivative;

(ii) $\mathbf{D}_1(u)$ defined in Theorem 3.1 and $\mathbf{D}_2(u)$ defined in Theorem 3.3 are nonsingular for all $u \in \Omega$.

Although the proposed semi-QR and semi-CQR procedures require different regularity conditions, the proofs follow similar strategies. For brevity, we only present the detailed proofs for the semi-CQR procedure. The detailed proofs for the semi-QR procedure was given in the earlier version of this paper. Lemma 7.1 below, which is a direct result of Mack and Silverman [23], will be used repeatedly in our proofs. Throughout the proofs, identities of the form $G(u) = O_p(a_n)$ always stand for $\sup_{u \in \Omega} |G(u)| = O_p(a_n)$.

LEMMA 7.1. *Let $(\mathbf{X}_1, Y_1), \ldots, (\mathbf{X}_n, Y_n)$ be i.i.d. random vectors, where the Y_i's are scalar random variables. Assume, further, that $E|Y|^r < \infty$ and that $\sup_{\mathbf{x}} \int |y|^r f(\mathbf{x}, y)\, dy < \infty$, where f denotes the joint density of $(\mathbf{X}, Y)$. Let K be a bounded positive function with bounded support, satisfying a Lipschitz condition. Then,*

$$\sup_{\mathbf{x} \in D} \left| n^{-1} \sum_{i=1}^{n} \{K_h(\mathbf{X}_i - \mathbf{x}) Y_i - E[K_h(\mathbf{X}_i - \mathbf{x}) Y_i]\} \right| = O_p\left(\frac{\log^{1/2}(1/h)}{\sqrt{nh}} \right),$$

provided that $n^{2\varepsilon - 1} h \to \infty$ for some $\varepsilon < 1 - r^{-1}$.

Let $\eta_{i,k} = I(\varepsilon_i \leq c_k) - \tau_k$ and $\eta_{i,k}^*(u) = I\{\varepsilon_i \leq c_k - r_i(u)\} - \tau_k$, where $r_i(u) = \alpha_0(U_i) - \alpha_0(u) - \alpha_0'(u)(U_i - u) + \mathbf{X}_i^T\{\boldsymbol{\alpha}(U_i) - \boldsymbol{\alpha}(u) - \boldsymbol{\alpha}'(u)(U_i - u)\}$. Furthermore, let $\tilde{\boldsymbol{\theta}}^* = \sqrt{nh}\{\tilde{a}_{01} - \alpha_0(u) - c_1, \ldots, \tilde{a}_{0q} - \alpha_0(u) - c_q, \{\tilde{\mathbf{a}} - \boldsymbol{\alpha}(u)\}^T, \{\tilde{\boldsymbol{\beta}} - \boldsymbol{\beta}_0\}^T, h\{\tilde{b}_0 - \alpha_0'(u)\}, h\{\tilde{\mathbf{b}} - \boldsymbol{\alpha}'(u)\}^T\}^T$ and $\mathbf{X}_{i,k}^*(u) = \{\mathbf{e}_k^T, \mathbf{X}_i^T, \mathbf{Z}_i^T, (U_i - u)/h, \mathbf{X}_i^T(U_i - u)/h\}^T$, where $\mathbf{e}_k$ is a q-vector with 1 at the kth position and 0 elsewhere.

In the proof of Theorem 3.1, we will first show the following asymptotic representation of $\{\tilde{\mathbf{a}}_0, \tilde{b}_0, \tilde{\mathbf{a}}, \tilde{\mathbf{b}}, \tilde{\boldsymbol{\beta}}\}$:

$$(7.1) \qquad \tilde{\boldsymbol{\theta}}^* = -f_U^{-1}(u)\{\mathbf{S}^*(u)\}^{-1} \mathbf{W}_n^*(u) + O_p(h^2 + \log^{1/2}(1/h)/\sqrt{nh}),$$

where $\mathbf{S}^*(u) = \text{diag}\{\mathbf{D}_1(u), c\mu_2 \mathbf{B}_2(u)\}$ and

$$\mathbf{W}_n^*(u) = \frac{1}{\sqrt{nh}} \sum_{k=1}^{q} \sum_{i=1}^{n} K\{(U_i - u)/h\} \eta_{i,k}^*(u) \mathbf{X}_{i,k}^*(u).$$

The asymptotic normality of $\{\tilde{\mathbf{a}}_0, \tilde{b}_0, \tilde{\mathbf{a}}, \tilde{\mathbf{b}}, \tilde{\boldsymbol{\beta}}\}$ then follows by demonstrating the asymptotic normality of $\mathbf{W}_n^*(u)$.

PROOF OF THEOREM 3.1. Recall that $\{\tilde{\mathbf{a}}_0, \tilde{\mathbf{a}}, \tilde{\boldsymbol{\beta}}, \tilde{b}_0, \tilde{\mathbf{b}}\}$ minimizes

$$\sum_{k=1}^{q} \sum_{i=1}^{n} \rho_{\tau_k}[Y_i - a_{0k} - b_0(U_i - u) - \mathbf{X}_i^T\{\mathbf{a} + \mathbf{b}(U_i - u)\} - \mathbf{Z}_i^T\boldsymbol{\beta}]K_h(U_i - u).$$

We write $Y_i - a_{0k} - b_0(U_i - u) - \mathbf{X}_i^T\{\mathbf{a} + \mathbf{b}(U_i - u)\} - \mathbf{Z}_i^T\boldsymbol{\beta} = (\varepsilon_i - c_k) + r_i(u) - \Delta_{i,k}$, where $\Delta_{i,k} = \{\mathbf{X}_{i,k}^*(u)\}^T\boldsymbol{\theta}^*/\sqrt{nh}$. Then, $\tilde{\boldsymbol{\theta}}^*$ is also the minimizer of

$$L_n^*(\boldsymbol{\theta}^*) = \sum_{k=1}^{q}\sum_{i=1}^{n} K_i(u)[\rho_{\tau_k}\{(\varepsilon_i - c_k) + r_i(u) - \Delta_{i,k}\} - \rho_{\tau_k}\{(\varepsilon_i - c_k) + r_i(u)\}],$$

where $K_i(u) = K\{(U_i - u)/h\}$. By applying the identity [14]

$$(7.2) \quad \rho_\tau(x - y) - \rho_\tau(x) = y\{I(x \le 0) - \tau\} + \int_0^y \{I(x \le z) - I(x \le 0)\}\,dz,$$

we have

$$L_n^*(\boldsymbol{\theta}^*) = \sum_{k=1}^{q}\sum_{i=1}^{n} K_i(u)\bigg\{\Delta_{i,k}[I\{\varepsilon_i \le c_k - r_i(u)\} - \tau_k]$$

$$+ \int_0^{\Delta_{i,k}} [I\{\varepsilon_i \le c_k - r_i(u) + z\} - I\{\varepsilon_i \le c_k - r_i(u)\}]\,dz\bigg\}$$

$$= \{\mathbf{W}_n^*(u)\}^T\boldsymbol{\theta}^* + \sum_{k=1}^{q} B_{n,k}^*(\boldsymbol{\theta}^*),$$

where

$$B_{n,k}^*(\boldsymbol{\theta}^*) = \sum_{i=1}^{n} K_i(u)\int_0^{\Delta_{i,k}} [I\{\varepsilon_i \le c_k - r_i(u) + z\} - I\{\varepsilon_i \le c_k - r_i(u)\}]\,dz.$$

Since $B_{n,k}^*(\boldsymbol{\theta}^*)$ is a summation of i.i.d. random variables of the kernel form, it follows, by Lemma 7.1, that

$$B_{n,k}^*(\boldsymbol{\theta}^*) = E[B_{n,k}^*(\boldsymbol{\theta}^*)] + O_p(\log^{1/2}(1/h)/\sqrt{nh}).$$

The conditional expectation of $\sum_{k=1}^{q} B_{n,k}^*(\boldsymbol{\theta}^*)$ can be calculated as

$$\sum_{k=1}^{q} E[B_{n,k}^*(\boldsymbol{\theta}^*)|U, \mathbf{X}, \mathbf{Z}]$$

$$= \sum_{k=1}^{q}\sum_{i=1}^{n} K_i(u)\int_0^{\Delta_{i,k}} \big[F(c_k - r_i(u) + z) - F(c_k - r_i(u))\big]\,dz$$

$$= \frac{1}{2}(\boldsymbol{\theta}^*)^T\left(\frac{1}{nh}\sum_{k=1}^{q}\sum_{i=1}^{n} K_i(u)f(c_k - r_i(u))\{\mathbf{X}_{i,k}^*(u)\}\{\mathbf{X}_{i,k}^*(u)\}^T\right)\boldsymbol{\theta}^*$$

$$+ O_p(\log^{1/2}(1/h)/\sqrt{nh})$$

$$\triangleq \frac{1}{2}(\boldsymbol{\theta}^*)^T\mathbf{S}_n^*(u)\boldsymbol{\theta}^* + O_p(\log^{1/2}(1/h)/\sqrt{nh}).$$

Then,

$$L_n^*(\boldsymbol{\theta}^*) = \{\mathbf{W}_n^*(u)\}^T \boldsymbol{\theta}^* + \sum_{k=1}^{q} E[B_{n,k}^*(\boldsymbol{\theta}^*)] + O_p\big(\log^{1/2}(1/h)/\sqrt{nh}\big)$$

$$= \{\mathbf{W}_n^*(u)\}^T \boldsymbol{\theta}^* + \sum_{k=1}^{q} E\{E[B_{n,k}^*(\boldsymbol{\theta}^*)|U, \mathbf{X}, \mathbf{Z}]\} + O_p\big(\log^{1/2}(1/h)/\sqrt{nh}\big)$$

$$= \{\mathbf{W}_n^*(u)\}^T \boldsymbol{\theta}^* + \frac{1}{2}(\boldsymbol{\theta}^*)^T E[\mathbf{S}_n^*(u)]\boldsymbol{\theta}^* + O_p\big(\log^{1/2}(1/h)/\sqrt{nh}\big).$$

It can be shown that $E[\mathbf{S}_n^*(u)] = f_U(u)\mathbf{S}^*(u) + O(h^2)$. Therefore, we can write $L_n(\boldsymbol{\theta}^*)$ as

$$L_n^*(\boldsymbol{\theta}^*) = \{\mathbf{W}_n^*(u)\}^T \boldsymbol{\theta}^* + \frac{f_U(u)}{2}(\boldsymbol{\theta}^*)^T \mathbf{S}^*(u)\boldsymbol{\theta}^* + O_p\big(h^2 + \log^{1/2}(1/h)/\sqrt{nh}\big).$$

By applying the convexity lemma [25] and the quadratic approximation lemma [4], the minimizer of $L_n^*(\boldsymbol{\theta}^*)$ can be expressed as

$$(7.3) \qquad \tilde{\boldsymbol{\theta}}^* = -f_U^{-1}(u)\{\mathbf{S}^*(u)\}^{-1}\mathbf{W}_n^*(u) + O_p\big(h^2 + \log^{1/2}(1/h)/\sqrt{nh}\big),$$

which holds uniformly for $u \in \Omega$. Meanwhile, for any point $u \in \Omega$, we have

$$(7.4) \qquad\qquad \tilde{\boldsymbol{\theta}}^* = -f_U^{-1}(u)\{\mathbf{S}^*(u)\}^{-1}\mathbf{W}_n^*(u) + o_p(1).$$

Note that $\mathbf{S}^*(u) = \text{diag}\{\mathbf{D}_1(u), c\mu_2\mathbf{B}_2(u)\}$ is a quasi-diagonal matrix. So,

$$(7.5) \qquad \sqrt{nh}\begin{pmatrix} \tilde{\mathbf{a}}_0 - \boldsymbol{\alpha}_0(u) \\ \tilde{\mathbf{a}} - \boldsymbol{\alpha}(u) \\ \tilde{\boldsymbol{\beta}} - \boldsymbol{\beta}_0 \end{pmatrix} = -f_U^{-1}(u)\mathbf{D}_1^{-1}(u)\mathbf{W}_{n,1}^*(u) + o_p(1),$$

where $\mathbf{W}_{n,1}^*(u) = \frac{1}{\sqrt{nh}}\sum_{k=1}^{q}\sum_{i=1}^{n} K_i(u)\eta_{i,k}^*(u)(\mathbf{e}_k^T, \mathbf{X}_i^T, \mathbf{Z}_i^T)^T$. Let

$$\mathbf{W}_{n,1}^{\#}(u) = \frac{1}{\sqrt{nh}}\sum_{k=1}^{q}\sum_{i=1}^{n} K_i(u)\eta_{i,k}(\mathbf{e}_k^T, \mathbf{X}_i^T, \mathbf{Z}_i^T)^T.$$

Note that

$$\text{Cov}(\eta_{i,k}, \eta_{i,k'}) = \tau_{kk'}, \qquad \text{Cov}(\eta_{i,k}, \eta_{j,k'}) = 0 \qquad \text{if } i \neq j.$$

By some calculations, we have that $E[\mathbf{W}_{n,1}^{\#}(u)] = \mathbf{0}$ and $\text{Var}[\mathbf{W}_{n,1}^{\#}(u)] \to f_U(u)v_0\boldsymbol{\Sigma}_1(u)$. By the Cramér–Wold theorem, the central limit theorem for $\mathbf{W}_{n,1}(u)$ holds. Therefore,

$$\mathbf{W}_{n,1}^{\#}(u) \xrightarrow{\mathcal{D}} N(\mathbf{0}, f_U(u)v_0\boldsymbol{\Sigma}_1(u)).$$

Moreover, we have $\text{Var}[\mathbf{W}_{n,1}^*(u) - \mathbf{W}_{n,1}^{\#}(u)|U, \mathbf{X}, \mathbf{Z}] \leq \frac{q^2}{nh}\sum_{i=1}^{n} K_i^2(u)(\mathbf{e}_k^T, \mathbf{X}_i^T, \mathbf{Z}_i^T)^T(\mathbf{e}_k^T, \mathbf{X}_i^T, \mathbf{Z}_i^T)\max_k\{F(c_k + |r_i|) - F(c_k)\} = o_p(1)$, thus

$$\text{Var}[\mathbf{W}_{n,1}^*(u) - \mathbf{W}_{n,1}^{\#}(u)] = o(1).$$

So, by Slutsky's theorem, conditioning on $\{U, \mathbf{X}, \mathbf{Z}\}$, we have

(7.6) $$\mathbf{W}_{n,1}^*(u) - E[\mathbf{W}_{n,1}^*(u)] \xrightarrow{\mathcal{D}} N(\mathbf{0}, f_U(u)v_0 \boldsymbol{\Sigma}_1(u)).$$

We now calculate the conditional mean of $\mathbf{W}_{n,1}^*(u)$:

$$\frac{1}{\sqrt{nh}} E[\mathbf{W}_{n,1}^*(u)|U, \mathbf{X}, \mathbf{Z}]$$

(7.7)
$$= \frac{1}{nh} \sum_{k=1}^{q} \sum_{i=1}^{n} K_i(u)\{F(c_k - r_i(u)) - F(c_k)\}(\mathbf{e}_k^t, \mathbf{X}_i^T, \mathbf{Z}_i^T)^T$$

$$= -\frac{1}{nh} \sum_{k=1}^{q} \sum_{i=1}^{n} K_i(u)r_i(u)f(c_k)\{1 + o(1)\}(\mathbf{e}_k^t, \mathbf{X}_i^T, \mathbf{Z}_i^T)^T$$

$$= -\frac{\mu_2 h^2}{2} f_U(u)\mathbf{D}_1(u) \begin{pmatrix} \boldsymbol{\alpha}_0''(u) \\ \boldsymbol{\alpha}''(u) \\ \mathbf{0} \end{pmatrix} + o_p(h^2).$$

The proof is completed by combining (7.5), (7.6) and (7.7). $\square$

PROOF OF THEOREM 3.2. Let $\boldsymbol{\theta} = \sqrt{n}(\boldsymbol{\beta} - \boldsymbol{\beta}_0)$. Then,

$$Y_i - \tilde{a}_{0k}(U_i) - \mathbf{X}_i^T \tilde{\mathbf{a}}(U_i) - \mathbf{Z}_i^T \boldsymbol{\beta}$$

$$= \varepsilon_i - c_k - \{\tilde{a}_{0k}(U_i) - \alpha_0(U_i) - c_k\} - \mathbf{X}_i^T\{\tilde{\mathbf{a}}(U_i) - \boldsymbol{\alpha}(U_i)\} - \mathbf{Z}_i^T(\boldsymbol{\beta} - \boldsymbol{\beta}_0)$$

$$= \varepsilon_i - c_k - \tilde{r}_{i,k} - \mathbf{Z}_i^T \boldsymbol{\theta}/\sqrt{n},$$

where $\tilde{r}_{i,k} = \{\tilde{a}_{0k}(U_i) - \alpha_0(U_i) - c_k\} + \mathbf{X}_i^T\{\tilde{\mathbf{a}}(U_i) - \boldsymbol{\alpha}(U_i)\}$. Then,

$$\hat{\boldsymbol{\theta}} = \operatorname{argmin} \sum_{k=1}^{q} \sum_{i=1}^{n} \rho_{\tau_k}(Y_i - \tilde{a}_{0k}(U_i) - \mathbf{X}_i^T \tilde{\mathbf{a}}(U_i) - \mathbf{Z}_i^T \boldsymbol{\beta})$$

is also the minimizer of

$$L_n(\boldsymbol{\theta}) = \sum_{k=1}^{q} \sum_{i=1}^{n} \{\rho_{\tau_k}(\varepsilon_i - c_k - \tilde{r}_{i,k} - \mathbf{Z}_i^T \boldsymbol{\theta}/\sqrt{n}) - \rho_{\tau_k}(\varepsilon_i - c_k - \tilde{r}_{i,k})\}.$$

By applying the identity (7.2), we can rewrite $L_n(\boldsymbol{\theta})$ as follows:

$$L_n(\boldsymbol{\theta}) = \sum_{k=1}^{q} \sum_{i=1}^{n} \left\{ \frac{\mathbf{Z}_i^T \boldsymbol{\theta}}{\sqrt{n}}[I(\varepsilon_i \leq c_k) - \tau_k] \right.$$

$$\left. + \int_{\tilde{r}_{i,k}}^{\tilde{r}_{i,k} + \mathbf{Z}_i^T \boldsymbol{\theta}/\sqrt{n}} [I(\varepsilon_i \leq c_k + z) - I(\varepsilon_i \leq c_k)] \, dz \right\}$$

$$= \left(\frac{1}{\sqrt{n}} \sum_{k=1}^{q} \sum_{i=1}^{n} \eta_{i,k} \mathbf{Z}_i \right)^T \boldsymbol{\theta} + B_n(\boldsymbol{\theta}),$$

where $B_n(\boldsymbol{\theta}) = \sum_{k=1}^{q} \sum_{i=1}^{n} \int_{\tilde{r}_{i,k}}^{\tilde{r}_{i,k} + \mathbf{Z}_i^T \boldsymbol{\theta}/\sqrt{n}} [I(\varepsilon_i \le c_k + z) - I(\varepsilon_i \le c_k)] \, dz$. Let us now calculate the conditional expectation of $B_n(\boldsymbol{\theta})$:

$$E[B_n(\boldsymbol{\theta}) | U, \mathbf{X}, \mathbf{Z}]$$

$$= \sum_{k=1}^{q} \sum_{i=1}^{n} \int_{\tilde{r}_{i,k}}^{\tilde{r}_{i,k} + \mathbf{Z}_i^T \boldsymbol{\theta}/\sqrt{n}} [z f(c_k)\{1 + o(1)\}] \, dz$$

$$= \frac{1}{2} \boldsymbol{\theta}^T \left(\frac{1}{n} \sum_{k=1}^{q} \sum_{i=1}^{n} f(c_k) \mathbf{Z}_i \mathbf{Z}_i^T \right) \boldsymbol{\theta} - \left(\frac{1}{\sqrt{n}} \sum_{k=1}^{q} \sum_{i=1}^{n} f(c_k) \tilde{r}_{i,k} \mathbf{Z}_i \right)^T \boldsymbol{\theta} + o_p(1).$$

Define $R_n(\boldsymbol{\theta}) = B_n(\boldsymbol{\theta}) - E[B_n(\boldsymbol{\theta}) | U, \mathbf{X}, \mathbf{Z}]$. It can be shown that $R_n(\boldsymbol{\theta}) = o_p(1)$. Hence,

$$L_n(\boldsymbol{\theta}) = \left(\frac{1}{\sqrt{n}} \sum_{k=1}^{q} \sum_{i=1}^{n} \eta_{i,k} \mathbf{Z}_i \right)^T \boldsymbol{\theta} + E[B_n(\boldsymbol{\theta}) | U, \mathbf{X}, \mathbf{Z}] + R_n(\boldsymbol{\theta})$$

$$= \frac{1}{2} \boldsymbol{\theta}^T \mathbf{S}_n \boldsymbol{\theta} + \left(\frac{1}{\sqrt{n}} \sum_{k=1}^{q} \sum_{i=1}^{n} \eta_{i,k} \mathbf{Z}_i \right)^T \boldsymbol{\theta} - \left(\frac{1}{\sqrt{n}} \sum_{k=1}^{q} \sum_{i=1}^{n} f(c_k) \tilde{r}_{i,k} \mathbf{Z}_i \right)^T \boldsymbol{\theta}$$

$$+ o_p(1),$$

where $\mathbf{S}_n = \frac{1}{n} \sum_{k=1}^{q} \sum_{i=1}^{n} f(c_k) \mathbf{Z}_i \mathbf{Z}_i^T$. By (7.3), the third term in the previous expression can be expressed as

$$\frac{1}{\sqrt{n}} \sum_{k=1}^{q} \sum_{i=1}^{n} f(c_k) \tilde{r}_{i,k} \mathbf{Z}_i$$

$$= \frac{1}{\sqrt{n}} \sum_{k=1}^{q} \sum_{i=1}^{n} \frac{f(c_k)}{f_U(U_i)} \mathbf{Z}_i \begin{pmatrix} \mathbf{e}_k \\ \mathbf{X}_i \\ \mathbf{0} \end{pmatrix}^T$$

$$\times \mathbf{D}_1^{-1}(U_i) \left(\frac{1}{nh} \sum_{k'=1}^{q} \sum_{i'=1}^{n} \eta_{i',k'}^*(U_i) \begin{pmatrix} \mathbf{e}_{k'} \\ \mathbf{X}_{i'} \\ \mathbf{Z}_{i'} \end{pmatrix} K_{i'}(U_i) \right)$$

$$+ O_p(h^{3/2} + \log^{1/2}(1/h)/\sqrt{nh^2})$$

$$= \frac{1}{\sqrt{n}} \sum_{k'=1}^{q} \sum_{i'=1}^{n} \eta_{i',k'} \boldsymbol{\delta}_{k'}(U_{i'}, \mathbf{X}_{i'}, \mathbf{Z}_{i'}) + O_p(n^{1/2}h^2 + \log^{1/2}(1/h)/\sqrt{nh^2})$$

$$= \frac{1}{\sqrt{n}} \sum_{k=1}^{q} \sum_{i=1}^{n} \eta_{i,k} \boldsymbol{\delta}_k(U_i, \mathbf{X}_i, \mathbf{Z}_i) + o_p(1),$$

where

$$\boldsymbol{\delta}(U_i, \mathbf{X}_i, \mathbf{Z}_i) = E[\mathbf{Z}(\mathbf{c}^T, c\mathbf{X}^T, \mathbf{0}) | U = U_i] \mathbf{D}_1^{-1}(U_i)(I_q, \mathbf{1}^T \mathbf{X}_i, \mathbf{1}^T \mathbf{Z}_i)^T.$$

Therefore,

$$L_n(\boldsymbol{\theta}) = \frac{1}{2}\boldsymbol{\theta}^T \mathbf{S}_n \boldsymbol{\theta} + \left(\frac{1}{\sqrt{n}}\sum_{k=1}^{q}\sum_{i=1}^{n}\eta_{i,k}\{\mathbf{Z}_i - \boldsymbol{\delta}_k(U_i, \mathbf{X}_i, \mathbf{Z}_i)\}\right)^T \boldsymbol{\theta} + o_p(1)$$

$$\triangleq \frac{1}{2}\boldsymbol{\theta}^T \mathbf{S}_n \boldsymbol{\theta} + \mathbf{W}_n^T \boldsymbol{\theta} + o_p(1).$$

It can be shown that $\mathbf{S}_n = E(\mathbf{S}_n) + o_p(1) = c\mathbf{S} + o_p(1)$. Hence,

$$L_n(\boldsymbol{\theta}) = \frac{c}{2}\boldsymbol{\theta}^T \mathbf{S}\boldsymbol{\theta} + \mathbf{W}_n^T \boldsymbol{\theta} + o_p(1).$$

Since the convex function $L_n(\boldsymbol{\theta}) - \mathbf{W}_n^T \boldsymbol{\theta}$ converges in probability to the convex function $\frac{c}{2}\boldsymbol{\theta}^T \mathbf{S}\boldsymbol{\theta}$, it follows, by the convexity lemma [25], that the quadratic approximation to $L_n(\boldsymbol{\theta})$ holds uniformly for $\boldsymbol{\theta}$ in any compact set Θ. Thus, it follows that

(7.8)
$$\hat{\boldsymbol{\theta}} = -\frac{1}{c}\mathbf{S}^{-1}\mathbf{W}_n + o_p(1).$$

By the Cramér–Wold theorem, the central limit theorem for $\mathbf{W}_n$ holds and $\mathrm{Var}(\mathbf{W}_n) \to \boldsymbol{\Delta} = \sum_{k=1}^{q}\sum_{k'=1}^{q}\tau_{kk'}E\{\mathbf{Z} - \boldsymbol{\delta}_k(U, \mathbf{X}, \mathbf{Z})\}\{\mathbf{Z} - \boldsymbol{\delta}_{k'}(U, \mathbf{X}, \mathbf{Z})\}^T$. Therefore, the asymptotic normality of $\hat{\boldsymbol{\beta}}$ is followed by

$$\sqrt{n}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) \xrightarrow{\mathcal{D}} N\left(0, \frac{1}{c^2}\mathbf{S}^{-1}\boldsymbol{\Delta}\mathbf{S}^{-1}\right).$$

This completes the proof. $\square$

PROOF OF THEOREM 3.3.   The asymptotic normality of $\hat{\alpha}_0(u)$ and $\hat{\boldsymbol{\alpha}}(u)$ can be obtained by following the ideas in the proof of Theorem 3.1. $\square$

PROOF OF THEOREM 4.1.   Use the same notation as in the proof of Theorem 3.2. Minimizing

$$\sum_{k=1}^{q}\sum_{i=1}^{n}\rho_{\tau_k}\{Y_i - \tilde{a}_{0k}(U_i) - \mathbf{X}_i^T \tilde{\mathbf{a}}(U_i) - \mathbf{Z}_i^T \boldsymbol{\beta}\} + nq\sum_{j=1}^{d}p'_{\lambda_j}(|\beta_j^{(0)}|)|\beta_j|$$

is equivalent to minimizing

$$G_n(\boldsymbol{\theta}) = \sum_{k=1}^{q}\sum_{i=1}^{n}\{\rho_{\tau_k}(\varepsilon_i - c_k - \tilde{r}_{i,k} - \mathbf{Z}_i^T \boldsymbol{\theta}/\sqrt{n}) - \rho_{\tau_k}(\varepsilon_i - c_k - \tilde{r}_{i,k})\}$$

$$+ nq\sum_{j=1}^{d}p'_{\lambda_j}(|\beta_j^{(0)}|)(|\beta_j| - |\beta_{0j}|)$$

$$= \frac{c}{2}\boldsymbol{\theta}^T \mathbf{S}\boldsymbol{\theta} + \mathbf{W}_n^T \boldsymbol{\theta} + nq\sum_{j=1}^{d}p'_{\lambda_j}(|\beta_j^{(0)}|)(|\beta_j| - |\beta_{0j}|) + o_p(1),$$

where $\boldsymbol{\theta} = \sqrt{n}(\boldsymbol{\beta} - \boldsymbol{\beta}_0)$ and $\tilde{r}_{i,k} = \{\tilde{a}_{0k}(U_i) - \alpha_0(U_i) - c_k\} + \mathbf{X}_i^T\{\tilde{\mathbf{a}}(U_i) - \boldsymbol{\alpha}(U_i)\}$. Similar to the derivation in the proof of Theorem 5 in Zou and Li [34], the third term above can be expressed as

$$(7.9) \qquad nq \sum_{j=1}^{d} p_{\lambda_j}'(|\beta_j^{(0)}|)(|\beta_j| - |\beta_{0j}|) \xrightarrow{\mathcal{P}} \begin{cases} 0, & \text{if } \boldsymbol{\beta}_2 = \boldsymbol{\beta}_{20}, \\ \infty, & \text{otherwise.} \end{cases}$$

Therefore, by the epiconvergence results [8, 15], we have $\hat{\boldsymbol{\beta}}_2^{\text{OSE}} \xrightarrow{\mathcal{P}} 0$ and the asymptotic results for $\hat{\boldsymbol{\beta}}_1^{\text{OSE}}$ holds.

To prove sparsity, we only need to show that $\hat{\boldsymbol{\beta}}_2^{\text{OSE}} = 0$ with probability tending to 1. It suffices to prove that if $\beta_{0j} = 0$, then $P(\hat{\beta}_j^{\text{OSE}} \neq 0) \to 0$. By using the fact that $|\frac{\rho_\tau(t_1) - \rho_\tau(t_2)}{t_1 - t_2}| \leq \max(\tau, 1 - \tau) < 1$, if $\hat{\beta}_j^{\text{OSE}} \neq 0$, then we must have $\sqrt{n} p_{\lambda_j}'(|\beta_j^{(0)}|) < \frac{1}{n}\sum_{i=1}^{n}|Z_{ij}|$. Thus, we have $P(\hat{\beta}_j^{\text{OSE}} \neq 0) \leq P(\sqrt{n} p_{\lambda_j}'(|\beta_j^{(0)}|) < \frac{1}{n}\sum_{i=1}^{n}|Z_{ij}|)$. However, under the assumptions, we have $\sqrt{n} p_{\lambda_j}'(|\beta_j^{(0)}|) \to \infty$. Therefore, $P(\hat{\beta}_j^{\text{OSE}} \neq 0) \to 0$. This completes the proof. $\square$

REFERENCES

[1] BRADIC, J., FAN, J. and WANG, W. (2010). Penalized composite quasi-likelihood for ultrahigh-dimensional variable selection. Available at arXiv:0912.5200v1.

[2] CAI, Z. and XU, X. (2009). Nonparametric quantile estimations for dynamic smooth coefficient models. *J. Amer. Statist. Assoc.* **104** 371–383. MR2504383

[3] CARROLL, R., FAN, J., GIJBELS, I. and WAND, M. (1997). Generalized partially linear single-index models. *J. Amer. Statist. Assoc.* **92** 477–489. MR1467842

[4] FAN, J. and GIJBELS, I. (1996). *Local Polynomial Modelling and Its Applications*. Chapman & Hall, London. MR1383587

[5] FAN, J. and HUANG, T. (2005). Profile likelihood inferences on semiparametric varying-coefficient partially linear models. *Bernoulli* **11** 1031–1057. MR2189080

[6] FAN, J. and LI, R. (2001). Variable selection via nonconcave penalized likelihood and its oracle properties. *J. Amer. Statist. Assoc.* **96** 1348–1361. MR1946581

[7] FAN, J. and LI, R. (2006). Statistical challenges with high dimensionality: Feature selection in knowledge discovery. In *Proceedings of the International Congress of Mathematicians* (M. Sanz-Sole, J. Soria, J. Varona and J. Verdera, eds.) **III** 595–622. Eur. Math. Soc., Zürich. MR2275698

[8] GEYER, C. (1994). On the asymptotics of constrained *M*-estimation. *Ann. Statist.* **22** 1993–2010. MR1329179

[9] HÄRDLE, W., LIANG, H. and GAO, J. (2000). *Partially Linear Models*. Physica Verlag, Heidelberg. MR1787637

[10] HE, X. and SHI, P. (1996). Bivariate tensor-product B-splines in a partly linear model. *J. Multivariate Anal.* **58** 162–181. MR1405586

[11] HE, X., ZHU, Z. and FUNG, W. (2002). Estimation in a semiparametric model for longitudinal data with unspecified dependence structure. *Biometrika* **89** 579–590. MR1929164

[12] HUNTER, D. and LI, R. (2005). Variable selection using MM algorithms. *Ann. Statist.* **33** 1617–1642. MR2166557

[13] KAI, B., LI, R. and ZOU, H. (2010). Local composite quantile regression smoothing: An efficient and safe alternative to local polynomial regression. *J. Roy. Statist. Soc. Ser. B* **72** 49–69.

[14] KNIGHT, K. (1998). Limiting distributions for L_1 regression estimators under general conditions. *Ann. Statist.* **26** 755–770. MR1626024

[15] KNIGHT, K. and FU, W. (2000). Asymptotics for lasso-type estimators. *Ann. Statist.* **28** 1356–1378. MR1805787

[16] KOENKER, R. (1984). A note on L-estimates for linear models. *Statist. Probab. Lett.* **2** 323-325. MR0782652

[17] KOENKER, R. (2005). *Quantile Regression*. Cambridge Univ. Press, Cambridge. MR2268657

[18] LAM, C. and FAN, J. (2008). Profile-kernel likelihood inference with diverging number of parameters. *Ann. Statist.* **36** 2232–2260. MR2458186

[19] LEE, S. (2003). Efficient semiparametric estimation of a partially linear quantile regression model. *Econometric Theory* **19** 1–31. MR1965840

[20] LENG, C. (2010). Variable selection and coefficient estimation via regularized rank regression. *Statist. Sinica* **20** 167–181. MR2640661

[21] LI, R. and LIANG, H. (2008). Variable selection in semiparametric regression modeling. *Ann. Statist.* **36** 261–286. MR2387971

[22] LI, Y. and ZHU, J. (2007). L_1-norm quantile regression. *J. Comput. Graph. Statist.* **17** 163–185. MR2424800

[23] MACK, Y. and SILVERMAN, B. (1982). Weak and strong uniform consistency of kernel regression estimates. *Probab. Theory Related Fields* **61** 405–415. MR0679685

[24] NIERENBERG, D., STUKEL, T., BARON, J., DAIN, B. and GREENBERG, E. (1989). Determinants of plasma levels of beta-carotene and retinol. *American Journal of Epidemiology* **130** 511–521.

[25] POLLARD, D. (1991). Asymptotics for least absolute deviation regression estimators. *Econometric Theory* **7** 186–199. MR1128411

[26] RUPPERT, D., WAND, M. and CARROLL, R. (2003). *Semiparametric Regression*. Cambridge Univ. Press, Cambridge. MR1998720

[27] WANG, H., LI, R. and TSAI, C. (2007). Tuning parameter selectors for the smoothly clipped absolute deviation method. *Biometrika* **94** 553–568. MR2410008

[28] WANG, H. and XIA, Y. (2009). Shrinkage estimation of the varying coefficient model. *J. Amer. Statist. Assoc.* **104** 747–757. MR2541592

[29] WANG, L. and LI, R. (2009). Weighted Wilcoxon-type smoothly clipped absolute deviation method. *Biometrics* **65** 564–571.

[30] WU, Y. and LIU, Y. (2009). Variable selection in quantile regression. *Statist. Sinica* **19** 801–817. MR2514189

[31] XIA, Y., ZHANG, W. and TONG, H. (2004). Efficient estimation for semivarying-coefficient models. *Biometrika* **91** 661–681. MR2090629

[32] YATCHEW, A. (2003). *Semiparametric Regression for the Applied Econometrician*. Cambridge Univ. Press, Cambridge.

[33] ZHANG, W., LEE, S. and SONG, X. (2002). Local polynomial fitting in semivarying coefficient model. *J. Multivariate Anal.* **82** 166–188. MR1918619

[34] ZOU, H. and LI, R. (2008). One-step sparse estimates in nonconcave penalized likelihood models (with discussion). *Ann. Statist.* **36** 1509–1533. MR2435443

[35] ZOU, H. and YUAN, M. (2008). Composite quantile regression and the oracle model selection
      theory. *Ann. Statist.* **36** 1108–1126. MR2418651

B. KAI   R. LI
DEPARTMENT OF MATHEMATICS   DEPARTMENT OF STATISTICS
COLLEGE OF CHARLESTON   PENNSYLVANIA STATE UNIVERSITY
CHARLESTON, SOUTH CAROLINA 29424   UNIVERSITY PARK, PENNSYLVANIA 16802
USA   USA
E-MAIL: KaiB@cofc.edu   E-MAIL: rli@stat.psu.edu

                        H. ZOU
                        SCHOOL OF STATISTICS
                        UNIVERSITY OF MINNESOTA
                        MINNEAPOLIS, MINNESOTA 55455
                        USA
                        E-MAIL: hzou@stat.umn.edu