Statistica Sinica Preprint No: SS-2024-0378	
Title	Semiparametric Efficient Estimation of Quantile Regression
Manuscript ID	SS-2024-0378
URL	http://www.stat.sinica.edu.tw/statistica/
DOI	10.5705/ss.202024.0378
Complete List of Authors	Zhanfeng Wang, Kani Chen, Yuanyuan Lin and Zhiliang Ying
Corresponding Authors	Kani Chen
E-mails	makchen@ust.hk
Notice: Accepted author version.	

Statistica Sinica

Statistica Sinica: Newly accepted Paper

Semiparametric Efficient Estimation of Quantile Regression

Zhanfeng Wang, Kani Chen, Yuanyuan Lin, Zhiliang Ying

School of Management, University of Science and Technology of China, Hefei, China

Department of Mathematics, The Hong Kong University of Science and Technology, Hong Kong, China

Department of Statistics and Data Science, The Chinese University of Hong Kong, Hong Kong, China

Department of Statistics, Columbia University, New York, NY 10027, United States

Abstract: Linear quantile regression model assumes quantiles of a response at certain levels are linearly related with covariates. If the model is assumed for one single quantile level, the semiparametric efficient estimation involves estimation of the conditional density of an error given covariates, which could be prohibitively difficult because of the curse of dimensionality. However, if the model is assumed for all quantile levels, estimation of conditional density becomes estimation of the derivative of regression coefficient functions, which is naturally available from initial estimators such as the Koenker-Bassett estimator. This paper derives the semiparametric efficient scores and the corresponding efficiency bounds for the regression coefficients. Although there is no closed form expression of the estimator or estimating function, we propose a computationally feasible procedure leading to semiparametrically efficient estimation. Simulation studies show that the proposed method could lead to substantial efficiency gain over the standard methods.

Statistica Sinica: Newly accepted Paper

Key words and phrases: Quantile regression; Semiparametric efficient score; Least favorable submodel; One-step estimation.

1. Introduction

Let Y be the response variable and $X = (1, \tilde{X}^\top)^\top$, where $\tilde{X}$ is a $(p-1)$-vector of covariates. Denote the conditional quantile of Y given X by $Q(\tau|X) = \inf\{t : P(Y \le t|X) \ge \tau\}$. This paper is concerned with the semi-parametric quantile regression model (Koenker and Bassett, 1978), which specifies, for all $0 < \tau < 1$,

$$Q(\tau|X) = X^\top \boldsymbol{\beta}(\tau), \tag{1.1}$$

where $\boldsymbol{\beta}(\cdot) = (\beta_1(\cdot), \beta_2(\cdot), \cdots, \beta_p(\cdot))^\top$ is the p-dimensional quantile regression function defined on $(0, 1)$. Throughout the paper, $\boldsymbol{\beta}(\cdot)$ is assumed to be twice continuously differentiable.

Beginning with the pioneering work of Koenker and Bassett (1978), there is a vast literature on quantile regression under various framework; see, e.g., Portnoy and Koenker (1989, 1997), Zheng and Portnoy (1998), Yu and Jones (1998). Furthermore, the quantile crossing and testing problems were addressed in He (1997), Koenker and Xiao (2002), He and Zhu (2003) and Bondell, Reich and Wang (2010). For survival data, estimation of the quantile function can be found in Koenker and Geling (2001), Portnoy

Statistica Sinica: Newly accepted Paper

(2003), Peng and Huang (2008), Peng and Fine (2009) and Wang and Wang (2009). The composite quantile regression that combines the commonality shared across quantiles was studied in Zou and Yuan (2008), Wang and Wang (2009), Kai, Li and Zou (2011), Wang, Li and He (2012) and Wang and Li (2013). Findings in Bayesian inference for quantile regression were reported in Yang and He (2011), Kim and Yang (2011) and Feng, Chen and He (2015). Bayesian nonparametric quantile regression were studied by Müller and Quintana (2004), Dunson and Taylor (2005) and Chung and Dunson (2009), Reich, Fuentes and Dunson(2011) and Qu and Yoon (2015). Recently, quantile regression in high dimensional settings attracted considerable interests; see Kato (2011), Wang, Wu and Li (2012), He, Wang and Hong (2013), Jiang, Wang and Bondell (2013) and Zheng, Peng and He (2015, 2017). A comprehensive review of quantile regression models can be found in Koenker (2005).

Asymptotic efficiency is of fundamental importance for semiparametric models; see Newey (1990), Bickel et al. (1993) and Tsiatis (2007) and references therein. The classical Koenker-Bassett estimator is easy to compute and enjoys robustness and many other advantages. However, it is not semiparametric efficient. For model (1.1) assumed at one single quantile level τ, the efficient estimation involves estimating the conditional density of an

Statistica Sinica: Newly accepted Paper

error term given the covariates. This estimation could be prohibitively difficult because of the curse of dimensionality, especially when the covariates are of high dimension. We believe this is the main reason that blocks the pursuit of efficient estimation. However, for quantile process (1.1) for all $\tau \in (0, 1)$, the conditional density is simply the reciprocal of the derivative of Q. This key observation makes efficient estimation computationally feasible as the curse of dimensionality is avoided. In this paper, we first derive the semiparametric efficient scores and the efficiency lower bounds for model (1.1). We then specify an estimation procedure leading to the semiparametric efficient estimation. This procedure has several advantages: (a) it could be regarded as an optimal way to pool information across multiple/other quantiles for efficiency gain; (b) it is computationally feasible and easy to implement, partly because the initial estimator is easily available; and (c) due to the nature of model (1.1), the conditional density estimation is straightforward by plugging in an initial estimator. We show that the resulting estimator can achieve the corresponding semiparametric efficiency lower bound under regularity conditions. Numerical studies confirm that the proposed estimator leads to better accuaracy compared with the Koenker-Bassett estimator for quantile regression.

The rest of the paper is organized as follows. Section 2 presents the

Statistica Sinica: Newly accepted Paper

semiparametric efficient scores for model (1.1) and a procedure of efficient estimation, along with several insightful discussions. Extensive simulation studies with supportive evidence are reported in Section 3. All technical derivations and proofs are in Supplementary.

2. Main results

Let $(X_i, Y_i),\ i = 1, 2, ..., n$, be independent and identically distributed (iid) copies of (X, Y) from model (1.1).

2.1 Semiparametric efficient score and efficiency bound

Let $F(\cdot|X)$ and $f(\cdot|X)$ be the conditional distribution and density functions of Y given X, respectively. Let $\boldsymbol{\beta}_0(\tau) = (\beta_{10}(\tau), \cdots, \beta_{p0}(\tau))^\top$ be the true value of $\boldsymbol{\beta}(\tau) = (\beta_1(\tau), \cdots, \beta_p(\tau))^\top$. Let $\eta(t|X)$ and $\eta_0(t|X)$ be the inverse functions of $X^\top\boldsymbol{\beta}(\tau) = t$ and $X^\top\boldsymbol{\beta}_0(\tau) = t$, respectively, for given X. Since $X^\top\boldsymbol{\beta}(\tau)$ and $X^\top\boldsymbol{\beta}_0(\tau)$ are monotonic functions in τ according to the definition of the quantile process model (1.1), the functions $\eta(t|X)$ and $\eta_0(t|X)$ exist for any given X. By the definition of quantile, $F(Q(\tau|X)|X) = \tau$ with $\boldsymbol{\beta}(\tau) = \boldsymbol{\beta}_0(\tau)$ and, under model (1.1), $F(X^\top\boldsymbol{\beta}_0(\tau)|X) = \tau$.

For model (1.1), we consider semiparametric efficient estimation of regression coefficient function $\boldsymbol{\beta}(\tau)$ at a fixed $\tau = \tau^* \in (0,\ 1)$. To construct a

Statistica Sinica: Newly accepted Paper

semiparametric efficient score for the j-th component of $\boldsymbol{\beta}(\tau)$, $j = 1, \cdots, p$, a parametric submodel of model (1.1) is considered,

$$Q(\tau, \theta|X) = X^\top \boldsymbol{\beta}(\tau; \theta), \tag{2.2}$$

where $\boldsymbol{\beta}(\tau; \theta) = \boldsymbol{\beta}_0(\tau) + \theta \boldsymbol{d}(\tau)$, the parameter θ is in a neighborhood of 0, $\boldsymbol{d}(\tau) = (d_1(\tau), ..., d_p(\tau))^\top$ is a continuously differentiable function of τ on $(0, 1)$, $d_j(\tau^*) = 1$, and $\boldsymbol{d}(0) = \boldsymbol{d}(1) = 0$.

Semiparametric efficient scores and efficiency bounds for $\boldsymbol{\beta}(\tau^*)$ are constructed via the parametric submodel (2.2) for model (1.1). The main results are presented in the following theorem with proof given in Supplementary I.

Theorem 1. *Under model (1.1) with the parametric submodel (2.2), we show that the semiparametric efficient score of $\boldsymbol{\beta}(\tau^*)$ for $\tau^* \in (0,\ 1)$, is $\boldsymbol{S}(Y, X; \tau^*) = (S_1(Y, X; \tau^*), ..., S_p(Y, X; \tau^*))^\top$, where*

$$S_j(Y, X; \tau^*) = -\frac{\partial}{\partial \tau} \Big\{ f(X^\top \boldsymbol{\beta}(\tau)|X) X^\top \boldsymbol{d}^*(\tau) \Big\} \Big|_{\tau = \eta(Y|X)}, \tag{2.3}$$

$\boldsymbol{d}^* \equiv \arg\min_{\boldsymbol{d}} \{\mathcal{I}(\boldsymbol{d}) : d_j(\tau^*) = 1, \boldsymbol{d}(0) = \boldsymbol{d}(1) = 0\}$, *and*

$$\mathcal{I}(\boldsymbol{d}) \equiv E\left(\left[\frac{\partial}{\partial \tau} \Big\{ f(X^\top \boldsymbol{\beta}(\tau)|X) X^\top \boldsymbol{d}(\tau) \Big\} \Big|_{\tau = \eta(Y|X)} \right]^2 \right). \tag{2.4}$$

Moreover, the semiparametric efficiency lower bound for the estimation of *the j-th component of $\boldsymbol{\beta}(\tau^*)$ is*

$$\sigma_j^2(\tau^*) = \frac{1}{\mathcal{I}(\boldsymbol{d}^*)}, \qquad j = 1, 2, \ldots, p. \tag{2.5}$$

Theorem 1 provides a theoretical foundation for efficient estimation. Under certain regularity conditions, $\sigma_j^2(\tau^*)$ is a lower bound for the asymptotic variance of the estimates of $\boldsymbol{\beta}_j(\tau^*)$. For the special case when $\beta_2(\tau), \ldots, \beta_p(\tau)$ are the same for all τ, i.e., $(\beta_2(\tau), \ldots, \beta_p(\tau))^T = \boldsymbol{b}$, a $(p-1)$-vector, model (1.1) is equivalent to the classical linear regression model

$$Y = \tilde{X}^\top \boldsymbol{b} + \epsilon$$

where ϵ is independent of $\tilde{X}$, and $\boldsymbol{\beta}(\tau) = (Q_\epsilon(\tau), \boldsymbol{b}^\top)^\top$, where $Q_\epsilon(\tau)$ is the τ-quantile of ϵ. Under such a model assumption, Portnoy and Koenker (1989) provided a construction of semiparametrically efficient (adaptive) estimator of the slope parameter vector $\boldsymbol{b}$ based on regression quantiles. On the other hand, instead of all quantiles, suppose only a specific quantile level τ^* is specified for the linear relationship to hold, i.e.,

$$Q(\tau^*|X) = X^\top \boldsymbol{\beta}_{\tau^*}, \tag{2.6}$$

a special case of which is the median regression ($\tau^* = 1/2$). The semiparametric efficiency of $\boldsymbol{\beta}_{\tau^*}$ in this case was dealt with by Newey and Powell (1990).

Statistica Sinica: Newly accepted Paper

Remark 1. Existence of $\boldsymbol{d}^*$ is shown in Supplementary I. Let $\dot{g}(\tau) = \partial g(\tau)/\partial\tau$ for a function $g(\cdot)$. Under a special case of model (1.1) with only an intercept term or one-dimensional covariate $\tilde{X}$ without intercept, that is $X = 1$ or $\tilde{X}$, according to Theorem 1, it can be shown that (details can be found in Supplementary I)

$$\frac{d^*(\tau)}{\dot{\beta}_0(\tau)} = \begin{cases} \frac{\tau}{\dot{\beta}_0(\tau^*)\tau^*}, & \tau \leq \tau^*, \\ \frac{1-\tau}{\dot{\beta}_0(\tau^*)(1-\tau^*)}, & \tau > \tau^*, \end{cases} \tag{2.7}$$

and the lower bound is

$$\sigma_1^2(\tau^*) = \frac{1}{\mathcal{I}(d^*)} = \dot{\beta}_0(\tau^*)^2\tau^*(1-\tau^*) = \begin{cases} \frac{\tau^*(1-\tau^*)}{\{f(\beta_0(\tau^*)|X)\}^2}, & X = 1, \\ \frac{\tau^*(1-\tau^*)}{E\{X^2(f(X\beta_0(\tau^*)|X))^2\}}, & X = \tilde{X}. \end{cases} \tag{2.8}$$

This fact indicates that when $p = 1$, the semiparametric efficient score and the efficiency lower bound of model (1.1) are consistent with those of the Koenker-Bassett estimator for quantile regression.

2.2 Semiparametric efficient estimation

For a fixed quantile level $\tau^* \in (0,1)$, we are interested in the estimation of $\boldsymbol{\beta}(\tau^*)$. A set of quantile grid points $0 < \tau_1 < \tau_2 < \cdots < \tau_L < 1$ are chosen with $\tau_k = \tau^*$ for some $1 \leq k \leq L$. For simplicity, the estimation of $\boldsymbol{\beta}(\tau^*)$ is presented in the form of the estimation of $\boldsymbol{\beta}(\tau_k)$. Instead of model (1.1), we consider a quantile regression model

$$Q(\tau_l|X) = X^\top \boldsymbol{\beta}(\tau_l), \qquad \text{for all } l = 1, 2, \ldots, L. \tag{2.9}$$

By the definition of quantile, $X^\top \boldsymbol{\beta}(\tau_l)$ is τ_l-quantile of Y given X. Under the boundedness condition of X in Assumption A_1 in Supplementary II, for any x in the support of X, model (2.9) implies $x^\top \boldsymbol{\beta}(\tau_1) < x^\top \boldsymbol{\beta}(\tau_2) < \cdots < x^\top \boldsymbol{\beta}(\tau_L)$.

It can be derived in a similar fashion that the semiparametric efficient score for $\boldsymbol{\beta}(\tau^*)$ under model (2.9) is

$$\begin{aligned} &\boldsymbol{S}^*(Y, X; \tau^*) = (S_1^*(Y, X; \tau^*), ..., S_p^*(Y, X; \tau^*))^\top \\ = &\sum_{l=1}^{L+1} \frac{f(X^\top \boldsymbol{\beta}(\tau_{l-1})|X) X^\top \boldsymbol{D}_{l-1}^* - f(X^\top \boldsymbol{\beta}(\tau_l)|X) X^\top \boldsymbol{D}_l^*}{\tau_l - \tau_{l-1}} \times \\ &\Big[I\{X^\top \boldsymbol{\beta}(\tau_{l-1}) < Y < X^\top \boldsymbol{\beta}(\tau_l)\} - (\tau_l - \tau_{l-1}) \Big], \end{aligned} \tag{2.10}$$

and the efficiency lower bound for the estimation of the j-th component of $\boldsymbol{\beta}(\tau^*)$ is

$$\sigma_j^{*2}(\tau^*) = \frac{1}{\boldsymbol{u}_{kj}^\top \boldsymbol{U} \boldsymbol{u}_{kj}}, \qquad j = 1, 2, \ldots, p, \tag{2.11}$$

where $X^\top \boldsymbol{\beta}(0) = -\infty$ and $X^\top \boldsymbol{\beta}(1) = +\infty$ for all X, $\tau_0 = 0$, $\tau_{L+1} = 1$, $\boldsymbol{D}_l^*$, $l = 0, 1, \ldots, L+1$, $\boldsymbol{u}_{kj}$, $j = 1, 2, \ldots, p$, and $\boldsymbol{U}$ are defined in Lemma 1 of Supplementary II.

Statistica Sinica: Newly accepted Paper

The semiparametric efficient score (2.10) can be regarded as an optimal way to combine information across the quantile levels $\tau_1, ..., \tau_L$. To estimate the density function of Y given X in (2.10), we use the expression

$$f(X^\top \boldsymbol{\beta}(\tau_l)|X) = \frac{1}{X^\top \dot{\boldsymbol{\beta}}(\tau_l)}, \qquad l = 1, 2, \ldots, L. \tag{2.12}$$

Thus, $f(X^\top \boldsymbol{\beta}(\tau_l)|X)$ can be estimated by $1/X^\top \hat{\dot{\boldsymbol{\beta}}}(\tau_l)$ for $l = 1, 2, \ldots, L$, where $\hat{\dot{\boldsymbol{\beta}}}(\tau_l) = \{\hat{\boldsymbol{\beta}}^c(\tau_l + h) - \hat{\boldsymbol{\beta}}^c(\tau_l - h)\}/(2h)$, $\hat{\boldsymbol{\beta}}^c(\tau)$ is the Koenker-Bassett estimate of $\boldsymbol{\beta}(\tau)$ and h is the bandwidth. Next, we define our proposed one-step estimator of $\boldsymbol{\beta}(\tau^*)$, denoted by $\hat{\boldsymbol{\beta}}(\tau^*)$, as

$$\hat{\beta}_j(\tau^*) = \hat{\beta}_j^c(\tau^*) + \hat{\sigma}_j^{*2}(\tau^*) \frac{\sum_{i=1}^n \hat{S}_j^*(Y_i, X_i; \tau^*)}{n}, \quad j = 1, 2, \ldots, p, \tag{2.13}$$

where $\hat{S}_j^*(Y, X; \tau^*)$ is the j-th component of the estimated score $\hat{\boldsymbol{S}}^*(Y, X; \tau^*)$ and $\hat{\sigma}_j^{*2}(\tau^*)$ is the estimated efficiency lower bound by plugging $\hat{\dot{\boldsymbol{\beta}}}(\tau_l)$ and $\hat{\boldsymbol{\beta}}^c(\tau_l)$, $l = 1, \ldots, L$, into (2.10) and $\sigma_j^{*2}(\tau^*)$, respectively.

Model (1.1) can be viewed as a submodel of model (2.9). When τ_l in model (2.9) becomes dense, model (2.9) approaches model (1.1). Conceptually, when this is the case, the efficient estimation in model (2.9) would also approximate that of model (1.1). This is presented in the following theorem. Set $\tau_{\max} = \max\{|\tau_l - \tau_{l-1}|, l = 1, 2, \ldots, L + 1\}$.

Before presenting the properties of $\hat{\beta}_j(\tau^*)$, we need the following conditions:

Statistica Sinica: Newly accepted Paper

Assumption (A_1). Each component of the covariate X is bounded by M with probability 1 for some constant M.

Assumption (A_2). The function $\dot{\boldsymbol{\beta}}(\tau)$ and $\ddot{\boldsymbol{\beta}}(\tau)$ are continuous, and there exists $0 < \epsilon < 1$ such that for all $\tau \in (\epsilon, 1-\epsilon)$, $\dot{\boldsymbol{\beta}}(\tau)$ and $\ddot{\boldsymbol{\beta}}(\tau)$ are bounded away from 0. In addition, $\tau_l \in (\epsilon, 1-\epsilon)$, $l = 1, ..., L$.

Assumption (A_3). The bandwidth h for the derivative estimation satisfies $h = o(n^{-\delta})$ with $1/8 < \delta < 1/4$.

Theorem 2. *Assume model (1.1) with the parametric submodel (2.2), and conditions* $(A_1) - (A_3)$ *hold. Suppose that* $L \to \infty$ *and* $\tau_{\max} \to 0$. *Then, for* $j = 1, ..., p$, $|S_j^*(Y, X; \tau^*) - S_j(Y, X; \tau^*)| \to 0$ *with probability 1 and* $|\sigma_j^{*2}(\tau^*) - \sigma_j^2(\tau^*)| \to 0$. *Moreover, if* $\tau_{\max} = O(1/\log(n))$ *and* $\log(n)/c < L < c\log(n)$ *for some positive constant* $c > 1$, *for* $j = 1, ..., p$,

$$\sqrt{n}\left\{\hat{\beta}_j(\tau^*) - \beta_{0j}(\tau^*)\right\} \to N(0,\ \sigma_j^2(\tau^*)), \tag{2.14}$$

in distribution as $n \to \infty$.

The proof of Theorem 2 is given in Supplementary III. The first part of Theorem 2 justifies the approximation of model (2.9) and model (1.1). The second part of Theorem 2 shows the semiparametric efficiency of the proposed estimator, with computation algorithm given below.

Algorithm:

Statistica Sinica: Newly accepted Paper

Step 1. For each $l = 1, \cdots, L$, compute the initial estimator $\hat{\boldsymbol{\beta}}^c(\tau_l)$;

Step 2. For each $l = 1, \cdots, L$, calculate $\hat{\dot{\boldsymbol{\beta}}}(\tau_l)$ and the conditional density function $f(X_i^\top \boldsymbol{\beta}(\tau_l)|X_i)$ with $1/X_i^\top \hat{\dot{\boldsymbol{\beta}}}(\tau_l)$;

Step 3. For each $j = 1, \cdots, p$, compute $\hat{S}_j^*(Y_i, X_i; \tau^*)$ and $\hat{\sigma}_j^{*2}(\tau^*)$ by plugging the initial estimator in step 1 and the estimated density in step 2 into $S_j^*(Y_i, X_i; \tau^*)$ and $\sigma_j^{*2}(\tau^*)$;

Step 4. Obtain $\hat{\beta}_j(\tau^*)$ by (2.13).

Remark 2. For the proposed one-step efficient estimation, we only need to estimate the conditional density function $f(X^\top \boldsymbol{\beta}(\tau_l)|X)$ at quantile levels $\{\tau_l, l = 1, \ldots, L\}$. Hence, we only need to assume the linear quantile regression model is specified in a neighborhood of each τ_l, $l = 1, \ldots, L$, rather than assuming a linear quantile regression model for all $\tau \in (0, 1)$.

The asymptotic normality in Theorem 2 is a pointwise result. It would be interesting to study the global behavior of $\hat{\beta}_j(s)$ as a function of $s \in (\epsilon,\ 1 - \epsilon)$ with $0 < \epsilon < 1$ in assumption A_2 of Supplementary II. For any quantile level s, let $S_j(Y, X; s)$ be the semiparametric efficient score for $\beta_j(s)$ given in (2.3), and let $\sigma_j^2(s)$ be the corresponding semiparametric efficiency bound. The weak convergence of the quantile process $\hat{\beta}_j(s)$ is presented in the following Theorem.

Statistica Sinica: Newly accepted Paper

Theorem 3. *Under model (1.1) with the parametric submodel (2.2), and conditions of Theorem 2, then, for* $j = 1, 2, \ldots, p$ *and* $s \in (\epsilon,\ 1-\epsilon)$, $\sqrt{n}\{\hat{\beta}_j(s) - \beta_{0j}(s)\}$ *converges weakly to a Gaussian process with 0 mean function and kernel function* g, *denoted by* $GP(0,\ g)$, *where* $g(u,v) = \sigma_j^2(u)\sigma_j^2(v)\ E[S_j^0(Y,X;u)S_j^0(Y,X;v)]$ *for* $u, v \in (\epsilon,\ 1-\epsilon)$, *and* $S_j^0(Y,X;\tau)$ *is the semiparametric efficient score* $S_j(Y,X;\tau)$ *by plugging in the true parameter* $\boldsymbol{\beta}_0(\tau)$.

The proof of Theorem 3 is given in Supplementary IV.

2.3 Properties of the semiparametric efficient score

Given the semiparametric efficient score in (2.10), there are several important implications:

Remark 3. When $L = 1$, model (2.9) reduces to the classical quantile regression model (2.6) specified at one single quantile point τ^*. By Lemma 1, the semiparametric efficient score of $\boldsymbol{\beta}_{\tau^*}$ is

$$\boldsymbol{S}^*(Y,X) = f(X^\top \boldsymbol{\beta}_{\tau^*}|X)\frac{1}{(1-\tau^*)\tau^*}\left\{\tau^* - I(Y < X^\top \boldsymbol{\beta}_{\tau^*})\right\}\boldsymbol{D}^{*\top} X, \tag{2.15}$$

where $\boldsymbol{D}^*$ is a constant matrix not depending on X. The semiparametric efficient estimation of $\boldsymbol{\beta}_{\tau^*}$ is addressed in Newey and Powell (1990); see also Portnoy and Koenker (1989) and Zhou and Portnoy (1998).

Statistica Sinica: Newly accepted Paper

Remark 4. In the special case of $p = 1$ (without covariate), the efficient scores for β_{τ_k} with $\tau_k = \tau^*$ is irrelevant to the information at other quantiles $\{\tau_l, l \neq k\}$, as derived and presented in Supplementary V. In other words, for model (2.9) with $p = 1$, the semiparametric efficiency lower bound for the estimation of $\beta_1(\tau^*)$ can be achieved using only the information at τ^*. When there is at least one more covariate in the model, namely $p \geq 2$, the efficient estimator of $\beta_j(\tau^*)$ generally depends on the information at other quantiles. In view of this fact, borrowing information across other quantiles via the efficient score (2.10) should be able to improve the estimation accuracy of $\boldsymbol{\beta}(\tau^*)$ for $p \geq 2$. For illustration, we provide a toy example for model (2.9) with $L = 2$ in Supplementary VI, which confirms that combining information across different quantile levels leads to efficiency gain in estimating $\boldsymbol{\beta}(\tau^*)$.

Remark 5. We provide a data-driven way to determine the selection of L in practice. The main idea is, when one focus on estimation of some one quantile among the L quantiles, such as median i.e. $\tau = 0.5$, we can use the estimation performance of median quantile as a criterion to determine L. We randomly partition the data into two parts, one (i.e. 60% of the whole data) is used to estimate the regression quantiles by the proposed method, and the other one with the remaining data (denoted by D_t) is

Statistica Sinica: Newly accepted Paper

used to evaluate performance of the built model. We compute empirical median of responses in D_t and take the corresponding covariate, denoted by $(y^*, \boldsymbol{X}_*)$. Then, compute the square error $SE = (y^* - \boldsymbol{X}_*^\top \hat{\boldsymbol{\beta}}(0.5))^2$. Repeat this procedure B times and compute the mean of the SEs (MSE). Finally, we select L with the smallest MSE.

3. Simulation Studies

Simulations are conducted to evaluate the performance of our proposed method. In the simulation, for a quantile level τ_k of interest, we consider three methods for the estimation of $\hat{\beta}_j(\tau_k)$: the traditional Koenker-Bassett quantile estimate $\hat{\boldsymbol{\beta}}^c_\tau$, denoted by TQE; the proposed one-step estimate based on the semiparametric efficient score of $\boldsymbol{\beta}(\tau_k)$, referred as EFF; the one-step estimate based on the score function (2.15) ignoring the model information at other quantiles, referred as (SEF). The simulated data is generated from the following quantile regression model with two covariates,

$$Q(\tau|X) = X_1\beta_1(\tau) + X_2\beta_2(\tau), \tag{3.16}$$

where $\beta_1(\tau)$ and $\beta_2(\tau)$ takes each of the following 5 forms:

$M1 : \beta_1(\tau) = 2$ and $\beta_2(\tau) = 1 + \Phi^{-1}(\tau)$;

$M2 : \beta_1(\tau) = 2 + \Phi^{-1}(\tau)$ and $\beta_2(\tau) = 2 + \Phi^{-1}(\tau)$;

$M3 : \beta_1(\tau) = 2$ and $\beta_2(\tau) = 1 + \log\{\tau/(1-\tau)\}$;

Statistica Sinica: Newly accepted Paper

$M4 : \beta_1(\tau) = 2$ and $\beta_2(\tau) = 1 + \tan\{\pi(\tau - 0.5)\}$;

$M5 : \beta_1(\tau) = 1 + \log\{\tau/(1-\tau)\}$ and $\beta_2(\tau) = 2 + \tan\{\pi(\tau - 0.5)\}$.

The covariate X_1 is constant 1 for $M1$, $M3$ and $M4$, and it follows standard log-normal distribution for $M2$ and $M5$. Another covariate X_2 follows log-normal distribution for all cases. In particular, model (3.16) with cases $M1$ and $M2$ are equivalent to

$$Y = 2 + X_2 + X_2\epsilon,$$

and

$$Y = 2X_1 + 2X_2 + (X_1 + X_2)\epsilon,$$

respectively, where ϵ follows the standard normal distribution. The sample size $n = 1000$ and 2000. All simulations are repeated 1000 times.

We first consider two quantiles at levels 0.5 and 0.7. The simulation results are summarized in Table 1. One can see that the parameter estimates are generally unbiased. For all configurations with $n = 1000$ or 2000, EFF has the smallest standard deviation (SD) compared with TQE and SEF. And SEF have much smaller SD compared to TQE. For example, for case M3 and $n = 1000$, the ratio of the standard deviations of TQE and EFF ranges from 1.343 to 2.214. And the ratio of the standard deviations of SEF and EFF ranges from 1.026 to 1.062. In other words, EFF improves

Statistica Sinica: Newly accepted Paper

efficiency of TQE for at least 80% and it improves efficiency of the SEF for around 5% to 12%, which confirms our theoretical findings.

In addition, we compare the numerical performance of the three methods with higher quantiles. Table 2 reports the estimation results with quantiles 0.5 and 0.9 for the 5 cases, while results with quantiles 0.8 and 0.9 are shown in Table 1 of Supplementary VII. One can see that similar conclusion to that of $\tau = 0.5$ and 0.7 can be drawn. EFF has the smallest standard errors and SEF is more efficient than TQE. As discussed in Remark 5, if a quantile at certain level is of particular interest, it is beneficial to combine the model information across other quantile levels for more efficient and stable estimation, even if the combined quantiles may be distant from the quantile of interest.

In addition, we investigate the performance of the proposed method with more covariates in model (1.1). Consider

$$Q(\tau|X) = X_1\beta_1(\tau) + X_2\beta_2(\tau) + X_3\beta_3(\tau) + X_4\beta_4(\tau), \tag{3.17}$$

where $\beta_1(\tau) = 2$, $\beta_2(\tau) = 1 + \Phi^{-1}(\tau)$, $\beta_3(\tau) = 1 + \log\{\tau/(1-\tau)\}$, $\beta_4(\tau) = 1 + \tan\{\pi(\tau - 0.5)\}$, X_1 is constant 1, and $X_i, i = 2, 3, 4$ are independent standard log-normal variables. The results are summarized in Table 2 of Supplementary VII, from which one can see that EFF and SEF are comparable and have smaller SD than TQE. This is understandable in our view,

Statistica Sinica: Newly accepted Paper

as density estimation involved in our proposed procedure becomes more challenging when the number of covariates increases.

Lastly, sensitivity of the proposed method is also studied when the model is misspecified. The simulated data are generated from model M2 but contaminated with data generated from

$$Q(\tau|X) = \{X_1\beta_1(\tau) + X_2\beta_2(\tau)\}^3, \tag{3.18}$$

where $\beta_1(\tau) = \beta_2(\tau) = 1 + \log\{\tau/(1-\tau)\}$, the covariate X_1 is constant 1, and X_2 follows log-normal distribution. Three cases are considered:

S1: for $\tau < 0.1$ or > 0.9, data are generated from the misspecified model (3.18), otherwise from M2;

S2: for $\tau < 0.2$ or > 0.9, data are generated from the misspecified model (3.18), otherwise from M2;

S3: for $\tau < 0.2$ or > 0.8, data are from the misspecified model (3.18), otherwise from M2.

The results are presented in Table 3. One can observe that the estimation results are similar to those in Table 1 for case S1. When there are more contaminated data as in S2 and S3, the performance of EFF is somewhat discounted, in the sense that EFF has smaller SD than TQE and comparable SD with SEF when $n = 2000$.

Statistica Sinica: Newly accepted Paper

4. Real data

We apply the proposed method to analyze a birth dataset (birth) released annually by the National Center for Health Statistics. The data includes information on nearly all live births from United States. Education of mother of each birth is recorded as 5 classes based on years of education. For illustration, we only consider the births that occurred in the month of June, 1997, and had mothers with smoking cigarettes and education class 2 (7 to 11 years of education). There are 9832 birth children consisting of 4861 female and 4971 male. The goal of this analysis is to study the relationship of the birth weight of child (in grams) and the covariates: the age of mother (Mage), the age of father (Fage) and total number of prenatal care visits (Nprevist). All variables have logarithm transformation before analysis.

As shown in Figure 2 in Supplementary VII, the MSE of median quantile estimation from the combination of quantiles $\tau = 0.5, 0.7$ and 0.9 is the smallest. Hence, they are taken for our method. Table 4 presents results of parameter estimation by 3 methods: TQE, SEF and EFF, for female child; those for male child are given in Table 3 in Supplementary VII. For each τ, TQE gives parameter estimates of the traditional quantile regression, and SEF shows one step parameter estimates via the single quantile score

Statistica Sinica: Newly accepted Paper

(2.15). Results of EFF is obtained by combining these 3 quantiles by quantile score (2.10). In the tables, Est stands for parameter estimate, Esd is variance estimate of Est by 1000 boostrap resampling method and P value is computed by $1-\Phi(|Est/Esd|)$, where $\Phi(\cdot)$ is the cumulative distribution function of standard normal variable. When nominal significance level is 0.05, all these 3 methods detect Nprevist for all quantiles, ages of parents for $\tau = 0.5$, and father age for female data with $\tau = 0.7$. When $\tau = 0.7$, Fage and Mage for male data do not have significantly nonzero coefficients, however, for female data, EFF has significant nonzero variable Mage while TQE and SEF do not detect it.

From the two tables, we observe that Nprevist and ages of parents have positive and negative coefficients, respectively, which suggests that birth weights of children become heavier when their mothers are younger and have more prenatal care visits. In addition, these 3 covariates for median quantile ($\tau = 0.5$) have greater effect to birth weights of children compared to for higher quantile ($\tau = 0.9$). When light birth children are considered, the lower quantile is of more concern. Moreover, for female children, Table 5 shows that the coefficients of Mage are overestimated by EFF, which suggests that effects of ages of parents from EFF become much larger.

Statistica Sinica: Newly accepted Paper

5. Concluding remarks

Semiparametric quantile regression is an important and widely studied model in statistics. This paper addresses the fundamental issue of semiparametric efficiency. It derives the semiparametrically efficient score function and the information lower bound, and provides a construction of efficient estimation. These theoretical findings, however, do not cover extensions of the quantile regression for longitudinal data and survival data, which are certainly of interest. In particular, for the censored quantile regression as being considered in Portnoy (2003) and Peng and Huang (2008), it would be of interest to study semiparametrically efficient estimation. The technical challenges appear to be formidable as it has to deal with the issue of censoring and to use the counting process and related martingale tools.

Acknowledgements

The authors thank the Editor, the Associate Editor and two anonymous reviewers for their insightful comments and constructive suggestions that helped improve the paper significantly. Z. Wang's research was supported by the National Science Foundation of China (No. 12371277, No. 12231017). Y. Lin's research was partially supported by the Hong Kong Research Grants Council (No. 14306620 and 14304523), and Direct Grants for Re-

Statistica Sinica: Newly accepted Paper

search, The Chinese University of Hong Kong.

References

[1] Bickel, P. J., Klaassen, C. A., Ritov, Y. and Wellner, J. A. (1993). *Efficient and Adaptive Estimation for Semiparametric Models.* New York: Springer-Verlag.

[2] Bondell, H. D., Reich, B. J. and Wang, H. (2010). Noncrossing quantile regression curve estimation. *Biometrika*, **97**, 825-838.

[3] Chung, Y. and Dunson, D. B. (2009). Nonparametric Bayes conditional distribution modeling with variable selection. *J. Amer. Statist. Assoc.* **104**, 1646-1660.

[4] Dunson, D. B. and Taylor, J. A. (2005). Approximate Bayesian inference for quantiles. *J. Nonparametri. Stat.* **17**, 385-400.

[5] Feng, Y., Chen, Y., and He, X. (2015). Bayesian quantile regression with approximate likelihood. *Bernoulli* **21**, 832-850.

[6] He, X. (1997). Quantile curves without crossing. *The American Statistician* **51**, 186-192.

Statistica Sinica: Newly accepted Paper

[7] He, X., Wang, L., and Hong, H.G. (2013). Quantile-adaptive model-free variable screening for high-dimensional heterogeneous data. *Ann. Statist.* **41**, 342-369.

[8] He, X. and Zhu, L. X. (2003). A lack-of-fit test for quantile regression. *J. Amer. Statist. Assoc.* **98**, 1013-1022.

[9] Jiang, L., Wang, H. J. and Bondell, H. D. (2013). Interquartile shrinkage in regression models. *J. Comp. Graph. Statist.*, **22**, 970-986.

[10] Kai, B., Li, R. and Zou, H. (2011). New efficient estimation and variable selection methods for semiparametric varying-coefficient partially linear models. *Ann. Statist.* **39**, 305-332.

[11] Kato, K. (2011). Group Lasso for high dimensional sparse quantile regression models. arXiv preprint arXiv:1103.1458.

[12] Kim, M. O. and Yang, Y. (2011). Semiparametric approach to a random effects quantile regression model. *J. Amer. Statist. Assoc.* **106**, 1405-1417.

[13] Koenker, R. (2005). *Quantile Regression.* Econometric Society Monographs 38. Cambridge Univ. Press, Cambridge.

Statistica Sinica: Newly accepted Paper

[14] Koenker, R. and Bassett, G. JR. (1978). Regression quantiles. *Econometrica* **46**, 33–50.

[15] Koenker, R. and Geling, O. (2001). Reappraising medfly longevity: A quantile regression survival analysis. *J. Amer. Statist. Assoc.* **96**, 458-468.

[16] Koenker, R. and Xiao, Z. (2002). Inference on the quantile regression process. *Econometrica* **70**, 1583-1612.

[17] Müller, P. and Quintana, F. A. (2004). Nonparametric Bayesian data analysis. *Stat. Sci.* **19**, 95-110.

[18] Newey, W. K. (1990). Semiparametric efficiency bounds. *Journal of applied econometrics* **5**, 99-135.

[19] Newey, W. K. and Powell, J. L. (1990). Efficient estimation of linear and type I censored regression models under conditional quantile restrictions. *Econometric Theory* **6**, 295- 317.

[20] Peng, L. and Fine, J. P. (2009). Competing risks quantile regression. *J. Amer. Statist. Assoc.* **104**, 1440–1453.

[21] Peng, L. and Huang, Y. (2008). Survival analysis with quantile regression models. *J. Amer. Statist. Assoc.* **103**, 637-649.

Statistica Sinica: Newly accepted Paper

[22] Portnoy, S. (2003). Censored regression quantiles. *J. Amer. Statist. Assoc.* **98**, 1001-1012.

[23] Portnoy, S. (2012). Nearly root-n approximation for regression quantile processes. *Ann. Statist* **40**, 1714-1736.

[24] Portnoy, S. and Koenker, R. (1989). Adaptive L-estimation for linear models. *Ann. Statist.* **17**, 362-381.

[25] Portnoy, S. and Koenker, R. (1997). The Gaussian hare and the Laplacian tortoise: computability of squared-error versus absolute-error estimators (with discussion). *Statistical Science* **12**, 279-300.

[26] Qu, Z. and Yoon, J. (2015). Nonparametric estimation and inference on conditional quantile processes. *J. Econometrics* **185**, 1-19.

[27] Reich, B. J., Fuentes, M., and Dunson, D. B. (2011). Bayesian spatial quantile regression. *J. Amer. Statist. Assoc.* **106**, 6-20.

[28] Tsiatis A. (2007). *Semiparametric Theory and Missing Data.* Springer Science & Business Media.

[29] Wang, H. J. and Li, D. (2013). Estimation of Extreme Conditional Quantiles Through Power Transformation. *J. Amer. Statist. Assoc.* **108**, 1062-1074.

Statistica Sinica: Newly accepted Paper

[30] Wang, H. J., Li, D. and He, X. (2012). Estimation of high conditional quantiles for heavy-tailed distributions. *J. Amer. Statist. Assoc.* **107**, 1453-1464.

[31] Wang, H. J. and Wang, L. (2009). Locally weighted censored quantile regression. *J. Amer. Statist. Assoc.* **104**, 1117–1128.

[32] Wang, L., Wu, Y., and Li, R. (2012). Quantile regression for analyzing heterogeneity in ultra-high dimension. *J. Amer. Statist. Assoc.* **107**, 214-222.

[33] Yang, Y. and He, X. (2012). Bayesian empirical likelihood for quantile regression. *Ann. Statist.* **40**, 1102-1131.

[34] Yu, K. and Jones, M. C. (1998). Local linear quantile regression. *J. Amer. Statist. Assoc.* **93**, 228-237.

[35] Zheng, Q., Peng, L., and He, X. (2015). Globally adaptive quantile regression with ultra-high dimensional data. *Ann. Statist.* **43**, 2225-2258.

[36] Zhou, K. Q. and Portnoy, S. L. (1998). Statistical inference on heteroscedastic models based on regression quantiles. *J. Nonparametric Stat.***9**, 239-260.

Statistica Sinica: Newly accepted Paper

[37] Zou, H. and Yuan, M. (2008). Composite quantile regression and the oracle model selection theory. *Ann. Statist.* **36**, 1108-1126.

School of Management, University of Science and Technology of China, Hefei, China

E-mail: zfw@ustc.edu.cn

Department of Mathematics, The Hong Kong University of Science and Technology, Hong Kong, China

E-mail: makchen@ust.hk

Department of Statistics, The Chinese University of Hong Kong, Hong Kong, China

E-mail: ylin@sta.cuhk.edu.hk

Department of Statistics, Columbia University, New York, NY 10027, United States

E-mail: zying@stat.columbia.edu

Statistica Sinica: Newly accepted Paper

Table 1: Simulation results for five models with quantiles 0.5 and 0.7.

Model	n		$\tau = 0.5$ $\beta_1(\tau)$	$\tau = 0.5$ $\beta_2(\tau)$	$\tau = 0.7$ $\beta_1(\tau)$	$\tau = 0.7$ $\beta_2(\tau)$
M1		True	2	1	2	1.5244
	1000	TQE	2.0007(0.0512)	0.9974(0.0899)	2.0031(0.0547)	1.5195(0.0961)
		SEF	2.0009(0.0238)	0.9968(0.0547)	2.0050(0.0265)	1.5149(0.0560)
		EFF	2.0015(0.0227)	0.9959(0.0533)	2.0009(0.0247)	1.5200(0.0529)
	2000	TQE	1.9992(0.0365)	1.0010(0.0652)	2.0023(0.0370)	1.5213(0.0653)
		SEF	2.0002(0.0159)	0.9993(0.0361)	2.0034(0.0174)	1.5190(0.0376)
		EFF	2.0002(0.0145)	0.9992(0.0352)	2.0006(0.0150)	1.5224(0.0365)
M2		True	2	2	2.5244	2.5244
	1000	TQE	1.9976(0.1192)	1.9987(0.1155)	2.5240(0.1244)	2.5206(0.1229)
		SEF	1.9989(0.0896)	1.9981(0.0875)	2.5228(0.0891)	2.5209(0.0903)
		EFF	1.9985(0.0881)	1.9982(0.0870)	2.5239(0.0883)	2.5205(0.0881)
	2000	TQE	1.9980(0.0834)	2.0022(0.0844)	2.5230(0.0877)	2.5225(0.0833)
		SEF	1.9990(0.0617)	2.0003(0.0614)	2.5232(0.0631)	2.5236(0.0605)
		EFF	1.9988(0.0608)	2.0002(0.0608)	2.5240(0.0624)	2.5228(0.0602)
M3		True	2	1	2	1.8473
	1000	TQE	2.0011(0.0822)	0.9958(0.1437)	2.0055(0.0907)	1.8397(0.1592)
		SEF	2.0014(0.0381)	0.9949(0.0874)	2.0094(0.0445)	1.8305(0.0929)
		EFF	2.0021(0.0365)	0.9938(0.0852)	2.0019(0.0420)	1.8400(0.0875)
	2000	TQE	1.9987(0.0585)	1.0017(0.1042)	2.0040(0.0615)	1.8424(0.1082)
		SEF	2.0003(0.0256)	0.9990(0.0575)	2.0061(0.0290)	1.8378(0.0622)
		EFF	2.0002(0.0230)	0.9990(0.0561)	2.0012(0.0250)	1.8436(0.0607)
M4		True	2	1	2	1.7265
	1000	TQE	2.0009(0.0669)	0.9966(0.1144)	2.0083(0.0930)	1.7221(0.1621)
		SEF	2.0014(0.0316)	0.9955(0.0699)	2.0166(0.0491)	1.7015(0.0952)
		EFF	2.0023(0.0287)	0.9945(0.0677)	2.0041(0.0480)	1.7172(0.0925)
	2000	TQE	1.9990(0.0469)	1.0013(0.0824)	2.0057(0.0629)	1.7228(0.1097)
		SEF	2.0002(0.0207)	0.9993(0.0461)	2.0103(0.0327)	1.7118(0.0646)
		EFF	2.0005(0.0188)	0.9988(0.0449)	2.0016(0.0289)	1.7227(0.0628)
M5		True	1	2	1.8473	2.7265
	1000	TQE	0.9964(0.1797)	1.9982(0.1555)	1.8467(0.2073)	2.7277(0.2072)
		SEF	0.9979(0.1344)	1.9972(0.1179)	1.8440(0.1488)	2.7214(0.1510)
		EFF	0.9971(0.1315)	1.9984(0.1173)	1.8449(0.1465)	2.7250(0.1474)
	2000	TQE	0.9973(0.1258)	2.003(0.1139)	1.8449(0.1459)	2.7268(0.1396)
		SEF	0.9987(0.0921)	2.0006(0.0831)	1.8448(0.1052)	2.7260(0.1011)
		EFF	0.9982(0.0911)	2.0004(0.0817)	1.8462(0.1039)	2.7264(0.1004)

* Standard deviations are in parentheses.

Statistica Sinica: Newly accepted Paper

Table 2: Simulation results for five models with a higher quantile.

			$\tau = 0.5$		$\tau = 0.9$	
Model	n		$\beta_1(\tau)$	$\beta_2(\tau)$	$\beta_1(\tau)$	$\beta_2(\tau)$
M1		True	2	1	2	2.2816
	1000	TQE	2.0007(0.0512)	0.9974(0.0899)	2.0117(0.0725)	2.2734(0.1296)
		SEF	2.0009(0.0238)	0.9968(0.0547)	2.0158(0.0362)	2.2605(0.0784)
		EFF	2.0014(0.0226)	0.9960(0.0530)	2.0032(0.0377)	2.2757(0.0772)
	2000	TQE	1.9992(0.0365)	1.0010(0.0652)	2.0090(0.0482)	2.2727(0.0877)
		SEF	2.0002(0.0159)	0.9993(0.0361)	2.0088(0.0231)	2.2698(0.0543)
		EFF	2.0004(0.0142)	0.9989(0.0347)	2.0026(0.0207)	2.2777(0.0510)
M2		True	2	2	3.2816	3.2816
	1000	TQE	1.9976(0.1192)	1.9987(0.1155)	3.2885(0.1622)	3.2751(0.1648)
		SEF	1.9989(0.0896)	1.9981(0.0875)	3.2818(0.1227)	3.2764(0.1237)
		EFF	1.9982(0.0879)	1.9984(0.0868)	3.2839(0.1189)	3.2785(0.1200)
	2000	TQE	1.9980(0.0834)	2.0022(0.0844)	3.2861(0.1149)	3.2750(0.1127)
		SEF	1.9990(0.0617)	2.0003(0.0614)	3.2836(0.0851)	3.2784(0.0862)
		EFF	1.9986(0.0607)	2.0004(0.0607)	3.2845(0.0821)	3.2797(0.0839)
M3		True	2	1	2	3.1972
	1000	TQE	2.0011(0.0822)	0.9958(0.1437)	2.0246(0.1410)	3.1834(0.2531)
		SEF	2.0014(0.0381)	0.9949(0.0874)	2.0354(0.0720)	3.1530(0.1533)
		EFF	2.0023(0.0366)	0.9935(0.0848)	2.0107(0.0668)	3.1817(0.1458)
	2000	TQE	1.9987(0.0585)	1.0017(0.1042)	2.0186(0.0938)	3.1807(0.1708)
		SEF	2.0003(0.0256)	0.9990(0.0575)	2.0201(0.0458)	3.1719(0.1061)
		EFF	2.0005(0.0226)	0.9985(0.0555)	2.0074(0.0403)	3.1872(0.0995)
M4		True	2	1	2	4.0777
	1000	TQE	2.0009(0.0669)	0.9966(0.1144)	2.1044(0.4091)	4.0791(0.7658)
		SEF	2.0014(0.0316)	0.9955(0.0699)	2.1874(0.2729)	3.9097(0.4643)
		EFF	2.0023(0.0286)	0.9943(0.0672)	2.0994(0.2469)	3.9866(0.4447)
	2000	TQE	1.999(0.0469)	1.0013(0.0824)	2.0754(0.2667)	4.0444(0.5028)
		SEF	2.0002(0.0207)	0.9993(0.0461)	2.1081(0.1550)	3.9713(0.3199)
		EFF	2.0007(0.0183)	0.9984(0.0444)	2.0579(0.1358)	4.0204(0.3075)
M5		True	1	2	3.1972	5.0777
	1000	TQE	0.9964(0.1797)	1.9982(0.1555)	3.2258(0.4602)	5.1270(0.8171)
		SEF	0.9979(0.1344)	1.9972(0.1179)	3.2108(0.3645)	5.1003(0.6241)
		EFF	0.9968(0.1313)	1.9987(0.1166)	3.2081(0.3424)	5.1070(0.5705)
	2000	TQE	0.9973(0.1258)	2.0030(0.1139)	3.2158(0.3218)	5.0801(0.5341)
		SEF	0.9987(0.0921)	2.0006(0.0831)	3.2074(0.2518)	5.0805(0.4226)
		EFF	0.9981(0.0906)	2.0005(0.0814)	3.2088(0.2400)	5.0860(0.3935)

* Standard deviations are in parentheses.

Statistica Sinica: Newly accepted Paper

Table 3: Simulation results with contaminated data from a misspecified model.

Misspecified			$\tau = 0.5$		$\tau = 0.7$	
case	n		$\beta_1(\tau)$	$\beta_2(\tau)$	$\beta_1(\tau)$	$\beta_2(\tau)$
		True	2	2	2.5244	2.5244
S1	1000	TQE	1.9976(0.1192)	1.9987(0.1155)	2.524(0.1244)	2.5206(0.1229)
		SEF	1.999(0.0896)	1.9978(0.0876)	2.5225(0.0893)	2.5209(0.0904)
		EFF	1.9984(0.0883)	1.9981(0.0873)	2.5234(0.0883)	2.5207(0.0885)
	2000	TQE	1.998(0.0834)	2.0022(0.0844)	2.523(0.0877)	2.5225(0.0833)
		SEF	1.999(0.0617)	2.0004(0.0614)	2.5229(0.0634)	2.5236(0.0606)
		EFF	1.9989(0.0609)	2.0000(0.0607)	2.5238(0.0623)	2.523(0.0602)
S2	1000	TQE	1.9976(0.1192)	1.9987(0.1155)	2.524(0.1244)	2.5206(0.1229)
		SEF	1.9988(0.0897)	1.9978(0.0877)	2.5225(0.0891)	2.5209(0.0902)
		EFF	1.9982(0.0881)	1.9983(0.0872)	2.5236(0.0884)	2.5206(0.0884)
	2000	TQE	1.998(0.0834)	2.0022(0.0844)	2.523(0.0877)	2.5225(0.0833)
		SEF	1.999(0.0617)	2.0003(0.0614)	2.523(0.0633)	2.5235(0.0605)
		EFF	1.9989(0.0609)	2.0000(0.0608)	2.5239(0.0623)	2.5227(0.0602)
S3	1000	TQE	1.9975(0.1191)	1.9986(0.1155)	2.524(0.1246)	2.5205(0.1229)
		SEF	1.998(0.0897)	1.9971(0.0879)	2.5216(0.0906)	2.5192(0.091)
		EFF	1.9976(0.0888)	1.9974(0.0882)	2.5386(0.4368)	2.5201(0.1855)
	2000	TQE	1.9979(0.0834)	2.0022(0.0843)	2.5228(0.0877)	2.5224(0.0834)
		SEF	1.9982(0.0617)	1.9996(0.0614)	2.5216(0.0632)	2.5224(0.0606)
		EFF	1.9982(0.061)	1.9993(0.0609)	2.522(0.0626)	2.5227(0.0719)

* Standard deviations are in parentheses.

Statistica Sinica: Newly accepted Paper

Table 4: Parameter estimate results for birth data with female child.

		Intercept			Mage			Fage			Nprevist		
τ	model	Est	Esd	P value	Est	Esd	P value	Est	Esd	P value	Est	Esd	P value
0.5	TQE	8.1886	0.0457	$< 0.0001^*$	-0.0498	0.0127	< 0.0001	-0.0156	0.0047	0.0004	0.021	0.0052	< 0.0001
	SEF	8.1815	0.0442	< 0.0001	-0.0478	0.0124	0.0001	-0.0153	0.0047	0.0006	0.0209	0.0051	< 0.0001
	EFF	8.1903	0.0493	< 0.0001	-0.0503	0.014	0.0002	-0.0155	0.0048	0.0007	0.0209	0.0054	0.0001
0.7	TQE	8.1628	0.0403	< 0.0001	-0.0175	0.0126	0.0824	-0.0151	0.0039	0.0001	0.0212	0.0043	< 0.0001
	SEF	8.1497	0.0403	< 0.0001	-0.0119	0.0126	0.1735	-0.0155	0.004	< 0.0001	0.0199	0.0043	< 0.0001
	EFF	8.2023	0.0532	< 0.0001	-0.0281	0.0159	0.0385	-0.0155	0.005	0.0009	0.0198	0.0057	0.0003
0.9	TQE	8.2374	0.043	< 0.0001	-0.0185	0.013	0.0774	-0.0054	0.0051	0.1469	0.0203	0.005	< 0.0001
	SEF	8.2345	0.0455	< 0.0001	-0.0189	0.0138	0.0853	-0.0046	0.0051	0.1817	0.0208	0.0053	< 0.0001
	EFF	8.2621	0.0563	< 0.0001	-0.0217	0.016	0.0875	-0.0075	0.0063	0.119	0.0179	0.0066	0.0031

* means that p value is less than 0.0001.