Statistica Sinica Preprint No: SS-2025-0456	
Title	Locally Efficient Estimators for Partially Linear Single Index Quantile Regression
Manuscript ID	SS-2025-0456
URL	http://www.stat.sinica.edu.tw/statistica/
DOI	10.5705/ss.202025.0456
Complete List of Authors	Ruimiao Luo, Seungchul Baek and Yanyuan Ma
Corresponding Authors	Yanyuan Ma
E-mails	yzm63@psu.edu
Notice: Accepted author version.	

LOCALLY EFFICIENT ESTIMATORS FOR PARTIALLY LINEAR SINGLE INDEX QUANTILE REGRESSION

Ruimiao Luo, Seungchul Baek and Yanyuan Ma

Shanxi University, University of Maryland, Baltimore County and Penn State University

Abstract: We study the partially linear single index model in the quantile regression context. We propose a class of locally efficient estimators which rely on a working regression error distribution model. We show the robustness of our estimator with respect to the misspecified error model. We further clarify the fundamental difference between the quantile regression with the corresponding partially linear single index mean regression problem. Asymptotic properties of the estimators are established rigorously. We further study the efficiency issue in this model. Numerical examples are implemented to illustrate the finite sample properties of our estimators. A real data application has been done to show our estimator's good performance.

Key words and phrases: Locally efficient, Quantile regression, Semiparametric efficiency.

1. Introduction

Quantile regression is widely known as an alternative to mean regression. Oftentimes, quantile regression is even preferred due to its flexibility to adjust to different quantile levels and its robustness to data contamination. Many mean regression models have their quantile correspondence, such as linear models, partially linear models, single index models, and many survival analysis models. Careful studies have been devoted to these quantile models and statistical properties are well studied.

To our surprise, partially linear single index model, a widely used model in the mean regression context, has received only limited attention in the quantile regression context (Lv, Zhang, Zhao and Liu, 2015; Zhang, Lian and Yu, 2017). For example, the efficiency bound and locally efficient estimation procedures in the partially linear single index quantile regression model have not been developed. This motivates us to conduct a thorough study on this model and to fill the gap in the literature. During the process of studying the partially linear single index model, we start to appreciate the difficulties involved in this model, even though on the surface, it appears very similar to partially linear quantile regression model and partially linear single index mean regression model. Indeed, compared to the former model, the unknown index coefficients inside an unknown function brought much additional complexity both in the estimation procedure and the theoretical treatment. Compared to the latter model, the shift from mean regression to quantile regression created many mathematical challenges which are not necessarily acknowledged and handled rigorously in some existing literatures.

We benefit from many existing works in quantile regression models that are related to the partially linear single index model under our study. For example, Newey and Powell (1990) considered the linear quantile regression model, with possibly censored dependent variable, and derived the semiparametric efficiency bound and proposed an efficient estimator. He and Shi (1998) considered B-Spline estimation for nonparametric quantile regression model. Lee (2003) derived the semiparametric efficiency bound and proposed an efficient estimator for the partially linear quantile regression model. Horowitz and Shi (2005) used kernel method and B-Spline to estimate the unknown function for an additive quantile regression model. Wu, Yu and Yu (2010) and Kong and Xia (2012) considered single index quantile regression, and Ma and He (2016) used B-Spline method and profile optimization to estimate the parameters in single index quantile regression model in the high dimensional covariates setting. Sherwood and Wang (2016) used B-Spline method to estimate parameters of partially linear additive quantile regression model in the high dimensional covariates case. Interestingly, estimators in all these works were obtained via minimizing the check function plus a possible penalty, and the efficiency issue of the resulting estimator was not considered. The only works involving efficiency are Newey and Powell (1990) and Lee (2003), where the semiparametric efficiency theory was respectively studied for the linear quantile regression model and for the partially linear quantile regression model. The existence yet scarcity of works on efficiency indicates both the importance and difficulty of such analysis for quantile regression models. As for the partially linear single index quantile regression model itself, Lv, Zhang, Zhao and Liu (2015) proposed a kernel method while Zhang, Lian and Yu (2017) utilized a B-Spline method, and these are the only works we are aware. However, it is unclear how these two methods compare and if any one of them reaches the theoretical optimality. All above considerations motivate us to investigate the semiparametric efficiency theory in the partially linear single index quantile regression model, with the hope of devising the locally efficient estimator and studying the efficiency bound. We indeed achieve this goal in this paper, and the results we obtain here also apply directly to the single index case since the latter can be considered as a special case. This also allows us to employ the locally efficient estimator to analyze a real data and achieve the optimal efficiency.

The rest of the paper is organized as the following. In Section 2, we study the semiparametric efficiency and propose a novel procedure that leads to the locally efficient estimator. This procedure does not base on minimizing the check function, which is a popular procedure in quantile regression. The theoretical properties of the resulting estimator and the required conditions are presented in Section 3. The rigorous mathematical proofs are provided in Supplementary Materials. The proof involves empirical process, kernel treatment in nonparametrics and special handling of nondifferentiable functions caused by quantile regression, and is a major contribution in itself. The methodology and theory are supported by numerical examples in Section 4. We analyze a real data example in Section 5 and conclude the paper with a discussion in Section 6.

2. Semiparametric Efficiency Bound and Locally Efficient Estimators

Consider the partially linear single index model

$$Q_\tau(Y_i \mid \mathbf{x}_i, \mathbf{z}_i) = \mathbf{x}_i^{\mathrm{T}}\boldsymbol{\gamma} + g(\mathbf{z}_i^{\mathrm{T}}\boldsymbol{\beta}), i = 1, 2, \ldots, N, \tag{2.1}$$

where $\mathbf{x}_i \in R^{d_{\boldsymbol{\gamma}}}$ and $\mathbf{z}_i \in R^{d_{\boldsymbol{\beta}}}$ respectively. Here, $Q_\tau(\cdot)$ stands for the conditional quantile function at level τ, where $0 < \tau < 1$. For identifiability reason, we assume the first element of $\boldsymbol{\beta}$ is 1 and write $\boldsymbol{\beta}_{-1}$ as the last $d_{\boldsymbol{\beta}} - 1$ elements of $\boldsymbol{\beta}$. In practice, it is important that the first element of $\mathbf{z}_i$ does not happen to be an insignificant covariate. For the purpose of avoiding such degenerated case, we can make use of empirical knowledge or preliminary analysis, or we can experiment with putting different element of $\mathbf{z}_i$ as the first one and discard the unstable results. Similar parameterization is used frequently to describe single index or dimension reduction structures (Ma and Zhu , 2013a,b,c, 2014). We aim to study the efficiency bound for the estimation of $\boldsymbol{\gamma}$, $\boldsymbol{\beta}_{-1}$ and devise locally efficient estimators. We also construct the estimation for $g(\cdot)$. We denote the last $d_{\boldsymbol{\beta}} - 1$ components of $\mathbf{z}$ by $\mathbf{z}_{-1}$. Without loss of generality, we assume that $E(\mathbf{z}_i) = \mathbf{0}$. We write the conditional probability density function (pdf) of Y_i given $\mathbf{x}_i, \mathbf{z}_i$, which is a function of $\epsilon_i \equiv Y_i - \mathbf{x}_i^{\mathrm{T}}\boldsymbol{\gamma} - g(\mathbf{z}_i^{\mathrm{T}}\boldsymbol{\beta}), \mathbf{x}_i$ and $\mathbf{z}_i$, as $f_\epsilon(\epsilon, \mathbf{x}, \mathbf{z})$. Note that (2.1) implies that $f_\epsilon(\epsilon, \mathbf{x}, \mathbf{z})$ satisfies $Q_\tau(\epsilon \mid \mathbf{x}, \mathbf{z}) = 0$, in addition to being a valid conditional pdf.

To obtain the semiparametric efficiency bound in estimating the parameters $\boldsymbol{\beta}_{-1}$ and $\boldsymbol{\gamma}$, the key step is to derive the efficient score functions, the inverse of the variance of which is then the efficiency bound. We obtain the efficient score functions through projecting the score functions with respect to $\boldsymbol{\beta}_{-1}$ and $\boldsymbol{\gamma}$ onto the nuisance tangent space corresponding to the three nonparametric functions involved in model (2.1), respectively the marginal pdf of the covariates $\mathbf{x}, \mathbf{z}$, $f_{\mathbf{x},\mathbf{z}}(\mathbf{x}, \mathbf{z})$, conditional pdf $f_\epsilon(\epsilon, \mathbf{x}, \mathbf{z})$ and the unknown function $g(\cdot)$. Below we directly state the result regarding the efficient score functions in Proposition 1, while we consolidate the derivation details in the proof in Section S1 of the Supplementary Materials.

Proposition 1. *Assume $f_\epsilon(\epsilon, \mathbf{x}, \mathbf{z})$ is differentiable with respect to ϵ. The semiparametric efficient score of model (2.1) is*

$$\mathbf{S}_{\text{eff}}(y, \mathbf{x}, \mathbf{z}) = -\frac{f_\epsilon(0, \mathbf{x}, \mathbf{z})}{\tau(1-\tau)}\{I(\epsilon < 0) - \tau\} \begin{pmatrix} \mathbf{x} - \dfrac{E\{f_\epsilon^2(0, \mathbf{x}, \mathbf{z})\mathbf{x}|\mathbf{z}^{\mathrm{T}}\boldsymbol{\beta}\}}{E\{f_\epsilon^2(0, \mathbf{x}, \mathbf{z})|\mathbf{z}^{\mathrm{T}}\boldsymbol{\beta}\}} \\ g'(\mathbf{z}^{\mathrm{T}}\boldsymbol{\beta})\left[\mathbf{z}_{-1} - \dfrac{E\{f_\epsilon^2(0, \mathbf{x}, \mathbf{z})\mathbf{z}_{-1}|\mathbf{z}^{\mathrm{T}}\boldsymbol{\beta}\}}{E\{f_\epsilon^2(0, \mathbf{x}, \mathbf{z})|\mathbf{z}^{\mathrm{T}}\boldsymbol{\beta}\}}\right] \end{pmatrix}.$$

Ideally, one would use the efficient score to construct estimating equations for the estimation of $\boldsymbol{\beta}_{-1}$ and $\boldsymbol{\gamma}$. However, this would require to compute the conditional expectations $E(\cdot \mid \mathbf{z}^{\mathrm{T}}\boldsymbol{\beta})$, to estimate the unknown function $g(\mathbf{z}^{\mathrm{T}}\boldsymbol{\beta})$ and $g'(\mathbf{z}^{\mathrm{T}}\boldsymbol{\beta})$, and to estimate the pdf $f_\epsilon(\epsilon, \mathbf{x}, \mathbf{z})$ evaluated at $\epsilon = 0$. Because $\mathbf{z}^{\mathrm{T}}\boldsymbol{\beta}$ is a one-dimensional quantity, hence the conditional expectations and the estimations of g and g' are feasible. However, $f_\epsilon(\epsilon, \mathbf{x}, \mathbf{z})$ is very difficult to assess since it is the regression error pdf conditional on a possibly high dimensional covariate $(\mathbf{x}^{\mathrm{T}}, \mathbf{z}^{\mathrm{T}})^{\mathrm{T}}$, in addition to that it is also subject to the zero τth quantile constraint. Interestingly, we find that even in the quantile regression context, the efficient score possesses a certain robustness property, in that it retains the mean zero property even if we use a misspecified functional form for $f_\epsilon(\epsilon, \mathbf{x}, \mathbf{z})$, as long as we obey the conditional quantile requirement. Specifically, we assume that we use a working model $f_\epsilon^*(\epsilon, \mathbf{x}, \mathbf{z})$ for $f_\epsilon(\epsilon, \mathbf{x}, \mathbf{z})$, and compute the efficient score accordingly. Let the resulting efficient score functions be $\mathbf{S}_{\text{eff}}^*(\epsilon, \mathbf{x}, \mathbf{z})$. Then it is easy to verify that $E\{\mathbf{S}_{\text{eff}}^*(\epsilon, \mathbf{x}, \mathbf{z})\} = E[E\{\mathbf{S}_{\text{eff}}^*(\epsilon, \mathbf{x}, \mathbf{z}) \mid \mathbf{x}, \mathbf{z}\}] = \mathbf{0}$. Thus, instead of constructing an efficient estimator for $\boldsymbol{\beta}_{-1}, \boldsymbol{\gamma}$, we aim at constructing a class of locally efficient estimators which rely on a working quantile regression error pdf evaluated at 0, i.e., $f_\epsilon^*(0, \mathbf{x}, \mathbf{z})$. If the working model is correct, i.e., if $f_\epsilon^*(0, \mathbf{x}, \mathbf{z}) = f_\epsilon(0, \mathbf{x}, \mathbf{z})$, then the estimator is efficient. If the working model is misspecified, i.e., $f_\epsilon^*(0, \mathbf{x}, \mathbf{z}) \neq f_\epsilon(0, \mathbf{x}, \mathbf{z})$, then the estimator is still consistent.

Remark 1. *The robustness of the efficient score to the misspecification of $f_\epsilon(0, \mathbf{x}, \mathbf{z})$ is similar to the robustness of the efficient score to the regression error variance in the partially linear single index mean regression model (Ma and Zhu , 2013c). However, in the quantile regression context, the robustness to the function of the single index, $g(\mathbf{z}^{\mathrm{T}}\boldsymbol{\beta})$, is lost even though the mean regression has such robustness as well. This is caused by two key differences between quantile regression and mean regression. First, conditional mean has a linearity property, in that the mean of the sum is the same as the sum of the means, which facilitates that $E\{\epsilon + g(\mathbf{z}^{\mathrm{T}}\boldsymbol{\beta}) - g^*(\mathbf{z}^{\mathrm{T}}\boldsymbol{\beta}) \mid \mathbf{x}, \mathbf{z}\} = g(\mathbf{z}^{\mathrm{T}}\boldsymbol{\beta}) - g^*(\mathbf{z}^{\mathrm{T}}\boldsymbol{\beta})$, which is a function of $\mathbf{z}^{\mathrm{T}}\boldsymbol{\beta}$ only. However, quantile does not have this property, in that $E[I\{\epsilon + g(\mathbf{z}^{\mathrm{T}}\boldsymbol{\beta}) - g^*(\mathbf{z}^{\mathrm{T}}\boldsymbol{\beta}) < 0\} - \tau \mid$*

$\mathbf{x}, \mathbf{z}] \neq E\{I(\epsilon < 0) - \tau \mid \mathbf{x}, \mathbf{z}\} + g(\mathbf{z}^{\mathrm{T}}\boldsymbol{\beta}) - g^*(\mathbf{z}^{\mathrm{T}}\boldsymbol{\beta})$. *Second, the form of the efficient score in the mean model ensures that replacing* ϵ *with an arbitrary function of* $\mathbf{z}^{\mathrm{T}}\boldsymbol{\beta}$ *will still result in zero mean of the entire efficient score function, while this property does not hold in the quantile regression setting, in that if we replace* $I(\epsilon < 0) - \tau$ *with a function of* $\mathbf{z}^{\mathrm{T}}\boldsymbol{\beta}$*, the resulting efficient score no longer has zero mean. These two factors combined together contributes to the nonrobustness of the quantile model to* $g(\mathbf{z}^{\mathrm{T}}\boldsymbol{\beta})$*. In fact, under a misspecified* $g^*(\mathbf{z}^{\mathrm{T}}\boldsymbol{\beta})$, $E[I\{\epsilon + g(\mathbf{z}^{\mathrm{T}}\boldsymbol{\beta}) - g^*(\mathbf{z}^{\mathrm{T}}\boldsymbol{\beta}) < 0\} - \tau \mid \mathbf{x}, \mathbf{z}] = F_\epsilon\{g^*(\mathbf{z}^{\mathrm{T}}\boldsymbol{\beta}) - g(\mathbf{z}^{\mathrm{T}}\boldsymbol{\beta}), \mathbf{x}, \mathbf{z}\} - \tau$*. If we happen to have* $F_\epsilon\{g^*(\mathbf{z}^{\mathrm{T}}\boldsymbol{\beta}) - g(\mathbf{z}^{\mathrm{T}}\boldsymbol{\beta}), \mathbf{x}, \mathbf{z}\} - \tau = f_\epsilon^*(0, \mathbf{x}, \mathbf{z}) h_0(\mathbf{z}^{\mathrm{T}}\boldsymbol{\beta})$ *for some function* $h_0(\mathbf{z}^{\mathrm{T}}\boldsymbol{\beta})$*, then the robustness to* $g(\mathbf{z}^{\mathrm{T}}\boldsymbol{\beta})$ *could be retained. But it is certainly unrealistic to expect such property to hold in general.*

Taking into account the locally efficient score function form, as well as the fact that consistent estimators have already been developed, we use the check function based estimator as an initial estimator of $\boldsymbol{\gamma}$, $\boldsymbol{\beta}_{-1}$, and propose a one step estimator for $\boldsymbol{\gamma}$, $\boldsymbol{\beta}_{-1}$ given as follows. For technical reasons we split the data into two groups. Let $n = \lceil N/2 \rceil$, which means n is the largest integer less than or equal to $N/2$, and let $m = N - n$. Define the index sets $I_1 = \{1, 2, \cdots, n\}$ and $I_2 = \{n+1, \cdots, N\}$. Write the initial estimators of $\boldsymbol{\beta}_{-1}$ and $\boldsymbol{\gamma}$ as $\widetilde{\boldsymbol{\beta}}_{-1}$ and $\widetilde{\boldsymbol{\gamma}}$, which are based on the data from the group I_1. For example, we can extend the method proposed in Ma and He (2016) by incorporating the linear term into the criterion function, then divide the resulting estimator of $\boldsymbol{\beta}$ by its first element to accommodate the different parameterizations. Following the similar proof procedure, we can obtain that $\widetilde{\boldsymbol{\beta}}_{-1} - \boldsymbol{\beta}_{-1} = O_p(n^{-1/2})$ and $\widetilde{\boldsymbol{\gamma}} - \boldsymbol{\gamma} = O_p(n^{-1/2})$. Let $\widetilde{\boldsymbol{\beta}} \equiv (1, \widetilde{\boldsymbol{\beta}}_{-1}^{\mathrm{T}})^{\mathrm{T}}$. Let ϕ satisfy

$$\sum_{j \in I_1} K_{h_1}(\mathbf{z}_j^{\mathrm{T}}\widetilde{\boldsymbol{\beta}} - \mathbf{z}_i^{\mathrm{T}}\widetilde{\boldsymbol{\beta}})\{I(Y_j - \mathbf{x}_j^{\mathrm{T}}\widetilde{\boldsymbol{\gamma}} - \phi < 0) - \tau\} = 0, \tag{2.2}$$

where K is a kernel function, h_1 is a bandwidth, $K_{h_1}(\cdot) = K(\cdot/h_1)/h_1$. We then set $\widetilde{g}(\mathbf{z}_i^{\mathrm{T}}\widetilde{\boldsymbol{\beta}}, \widetilde{\boldsymbol{\gamma}}) \equiv \phi$ to be an estimator of $g(\mathbf{z}_i^{\mathrm{T}}\boldsymbol{\beta})$ based on data in I_1, evaluated at observation i, where $i \in I_2$. Further, we let the one-step estimator of $g'(\mathbf{z}_i^{\mathrm{T}}\boldsymbol{\beta})$ for $i \in I_2$ be

$$\widetilde{g}'(\mathbf{z}_i^{\mathrm{T}}\widetilde{\boldsymbol{\beta}}, \widetilde{\boldsymbol{\gamma}}) = \frac{\sum_{j\in I_1} h_1^{-2} K'(\frac{\mathbf{z}_j^{\mathrm{T}}\widetilde{\boldsymbol{\beta}} - \mathbf{z}_i^{\mathrm{T}}\widetilde{\boldsymbol{\beta}}}{h_1})[I\{Y_j - \mathbf{x}_j^{\mathrm{T}}\widetilde{\boldsymbol{\gamma}} - \widetilde{g}(\mathbf{z}_i^{\mathrm{T}}\widetilde{\boldsymbol{\beta}}, \widetilde{\boldsymbol{\gamma}}) < 0\} - \tau]}{\sum_{j\in I_1} K_{h_1}(\mathbf{z}_j^{\mathrm{T}}\widetilde{\boldsymbol{\beta}} - \mathbf{z}_i^{\mathrm{T}}\widetilde{\boldsymbol{\beta}})\mathcal{K}_{h_3}\{Y_j - \mathbf{x}_j^{\mathrm{T}}\widetilde{\boldsymbol{\gamma}} - \widetilde{g}(\mathbf{z}_i^{\mathrm{T}}\widetilde{\boldsymbol{\beta}}, \widetilde{\boldsymbol{\gamma}})\}}, \tag{2.3}$$

where $K'(\cdot)$ is the first derivative of $K(\cdot)$, $\mathcal{K}$ is a kernel function possibly different from K, h_3 is a bandwidth, and $\mathcal{K}_{h_3}(\cdot) = \mathcal{K}(\cdot/h_3)/h_3$. Note that in estimating $g(\cdot)$ and $g'(\cdot)$, we do not aim to minimize the variability of the estimation. This is reflected in the form of (2.2) and (2.3), where no working model $f_\epsilon^*(\epsilon, \mathbf{x}, \mathbf{z})$ is used. This is because the estimation variability of the locally efficient score based estimator for $\boldsymbol{\beta}_{-1}$ and $\boldsymbol{\gamma}$ will be shown to be unaffected by the variability of the $g(\cdot)$ and $g'(\cdot)$ estimation. We now insert these estimated quantities into the efficient score function and construct the estimators of $\boldsymbol{\gamma}$ and $\boldsymbol{\beta}_{-1}$ as

$$\begin{aligned}\begin{pmatrix}\widehat{\boldsymbol{\gamma}}\\ \widehat{\boldsymbol{\beta}}_{-1}\end{pmatrix} &= \begin{pmatrix}\widetilde{\boldsymbol{\gamma}}\\ \widetilde{\boldsymbol{\beta}}_{-1}\end{pmatrix} + \widehat{V}_0^{-1}\frac{1}{m}\sum_{i\in I_2}\frac{f_\epsilon^*(0, \mathbf{x}_i, \mathbf{z}_i)}{\tau(1-\tau)}\psi_i\{\widetilde{\boldsymbol{\gamma}}, \widetilde{g}(\mathbf{z}_i^{\mathrm{T}}\widetilde{\boldsymbol{\beta}}, \widetilde{\boldsymbol{\gamma}})\}\\ &\quad\times\left\{\begin{matrix}\widehat{\mathbf{x}}_i^c\\ \widetilde{g}'(\mathbf{z}_i^{\mathrm{T}}\widetilde{\boldsymbol{\beta}}, \widetilde{\boldsymbol{\gamma}})\widehat{\mathbf{z}}_{-1,i}^c\end{matrix}\right\},\end{aligned} \tag{2.4}$$

where for $i \in I_2$, $\psi_i\{\widetilde{\boldsymbol{\gamma}}, \widetilde{g}(\mathbf{z}_i^{\mathrm{T}}\widetilde{\boldsymbol{\beta}}, \widetilde{\boldsymbol{\gamma}})\} \equiv \tau - I\{Y_i - \mathbf{x}_i^{\mathrm{T}}\widetilde{\boldsymbol{\gamma}} - \widetilde{g}(\mathbf{z}_i^{\mathrm{T}}\widetilde{\boldsymbol{\beta}}, \widetilde{\boldsymbol{\gamma}}) < 0\}$,

$$\begin{aligned}\widehat{\mathbf{x}}_i^c &\equiv \mathbf{x}_i - \frac{\widehat{E}\{f_\epsilon^{*2}(0, \mathbf{x}_i, \mathbf{z}_i)\mathbf{x}_i|\mathbf{z}_i^{\mathrm{T}}\widetilde{\boldsymbol{\beta}}\}}{\widehat{E}\{f_\epsilon^{*2}(0, \mathbf{x}_i, \mathbf{z}_i)|\mathbf{z}_i^{\mathrm{T}}\widetilde{\boldsymbol{\beta}}\}},\\ \widehat{\mathbf{z}}_{-1,i}^c &\equiv \mathbf{z}_{-1,i} - \frac{\widehat{E}\{f_\epsilon^{*2}(0, \mathbf{x}_i, \mathbf{z}_i)\mathbf{z}_{-1,i}|\mathbf{z}_i^{\mathrm{T}}\widetilde{\boldsymbol{\beta}}\}}{\widehat{E}\{f_\epsilon^{*2}(0, \mathbf{x}_i, \mathbf{z}_i)|\mathbf{z}_i^{\mathrm{T}}\widetilde{\boldsymbol{\beta}}\}},\end{aligned}$$

$$\widehat{V}_0 \equiv \frac{1}{m}\sum_{i\in I_2}\frac{f_\epsilon^{*2}(0,\mathbf{x}_i,\mathbf{z}_i)}{\tau(1-\tau)}\begin{pmatrix}\widehat{\mathbf{x}}_i^c\\ \widetilde{g}'(\mathbf{z}_i^{\mathrm{T}}\widetilde{\boldsymbol\beta},\widetilde{\boldsymbol\gamma})\widehat{\mathbf{z}}_{-1,i}^c\end{pmatrix}\begin{pmatrix}\mathbf{x}_i^{\mathrm{T}} & \widetilde{g}'(\mathbf{z}_i^{\mathrm{T}}\widetilde{\boldsymbol\beta},\widetilde{\boldsymbol\gamma})\mathbf{z}_{-1,i}^{\mathrm{T}}\end{pmatrix}.$$

Here, $\widehat{E}$ denotes the standard kernel based nonparametric estimators, i.e.,

$$\widehat{E}\{f_\epsilon^{*2}(0,\mathbf{x}_i,\mathbf{z}_i)\mathbf{x}_i|\mathbf{z}_i^{\mathrm{T}}\widetilde{\boldsymbol\beta}\} \equiv \sum_{k\in I_2}\mathcal{K}_{h_2}(\mathbf{z}_k^{\mathrm{T}}\widetilde{\boldsymbol\beta}-\mathbf{z}_i^{\mathrm{T}}\widetilde{\boldsymbol\beta})f_\epsilon^{*2}(0,\mathbf{x}_k,\mathbf{z}_k)\mathbf{x}_k/\sum_{k\in I_2}\mathcal{K}_{h_2}(\mathbf{z}_k^{\mathrm{T}}\widetilde{\boldsymbol\beta}-\mathbf{z}_i^{\mathrm{T}}\widetilde{\boldsymbol\beta}),$$
$$\widehat{E}\{f_\epsilon^{*2}(0,\mathbf{x}_i,\mathbf{z}_i)\mathbf{z}_{-1,i}|\mathbf{z}_i^{\mathrm{T}}\widetilde{\boldsymbol\beta}\} \equiv \sum_{k\in I_2}\mathcal{K}_{h_2}(\mathbf{z}_k^{\mathrm{T}}\widetilde{\boldsymbol\beta}-\mathbf{z}_i^{\mathrm{T}}\widetilde{\boldsymbol\beta})f_\epsilon^{*2}(0,\mathbf{x}_k,\mathbf{z}_k)\mathbf{z}_{-1,k}/\sum_{k\in I_2}\mathcal{K}_{h_2}(\mathbf{z}_k^{\mathrm{T}}\widetilde{\boldsymbol\beta}-\mathbf{z}_i^{\mathrm{T}}\widetilde{\boldsymbol\beta}),$$
$$\widehat{E}\{f_\epsilon^{*2}(0,\mathbf{x}_i,\mathbf{z}_i)|\mathbf{z}_i^{\mathrm{T}}\widetilde{\boldsymbol\beta}\} \equiv \sum_{k\in I_2}\mathcal{K}_{h_2}(\mathbf{z}_k^{\mathrm{T}}\widetilde{\boldsymbol\beta}-\mathbf{z}_i^{\mathrm{T}}\widetilde{\boldsymbol\beta})f_\epsilon^{*2}(0,\mathbf{x}_k,\mathbf{z}_k)/\sum_{k\in I_2}\mathcal{K}_{h_2}(\mathbf{z}_k^{\mathrm{T}}\widetilde{\boldsymbol\beta}-\mathbf{z}_i^{\mathrm{T}}\widetilde{\boldsymbol\beta}),$$

where h_2 is a bandwidth, $\mathcal{K}_{h_2}(\cdot)=\mathcal{K}(\cdot/h_2)/h_2$.

Remark 2. *It is easy to see that once the locally efficient score function can be approximated based on the initial estimator and the nonparametric regressions, the above one-step estimator is motivated by the result of a one-step Newton-Raphson update. However, the typical one-step Newton-Raphson update requires the derivative of the estimating equation, which will involve taking derivative of the locally efficient score function. Because the efficient score function itself is discontinuous, this will involve further complication even though numerically it is feasible via a numerical difference operation. Thus, we modified the one-step Newton-Raphson step and used a working derivative matrix* $\widehat{V}_0$*. This modification leads to several special features. First, our one step estimator does not share the typical one-step Newton-Raphson estimator property, in that the first order asymptotic estimator property of our one-step estimator will be different from a multiple step procedure. Second, if the working model* $f_\epsilon^*(\epsilon,\mathbf{x},\mathbf{z})$ *is correctly specified, then our one-step estimator shares the standard one-step estimator property and will be shown to result in the semiparametrically efficient estimator.*

Finally, when the working model $f^*_\epsilon(\epsilon, \mathbf{x}, \mathbf{z})$ is misspecified, there is no definitive relation between our one-step estimator and the standard Newton-Raphson estimator in terms of estimation variability, hence we opt for our one-step estimator due to its computational simplicity and numerical stability.

Of course, we can repeat the above process while switching the I_1, I_2 data groups, and obtain another set of estimators for $\boldsymbol{\gamma}$ and $\boldsymbol{\beta}_{-1}$, written as $\widehat{\boldsymbol{\gamma}}'$ and $\widehat{\boldsymbol{\beta}}'_{-1}$. We then define our final one-step estimators of $\boldsymbol{\gamma}$ and $\boldsymbol{\beta}_{-1}$ as

$$\overline{\boldsymbol{\gamma}} = \frac{1}{2}(\widehat{\boldsymbol{\gamma}} + \widehat{\boldsymbol{\gamma}}'), \text{ and } \overline{\boldsymbol{\beta}}_{-1} = \frac{1}{2}(\widehat{\boldsymbol{\beta}}_{-1} + \widehat{\boldsymbol{\beta}}'_{-1}).$$

Write $\overline{\boldsymbol{\beta}} \equiv (1, (\overline{\boldsymbol{\beta}}_{-1})^{\mathrm{T}})^{\mathrm{T}}$. Let ϕ satisfy

$$\sum_{j=1}^{N} K_{h_1}(\mathbf{z}_j^{\mathrm{T}}\overline{\boldsymbol{\beta}} - \mathbf{z}_i\overline{\boldsymbol{\beta}})\{I(Y_j - \mathbf{x}_j^{\mathrm{T}}\overline{\boldsymbol{\gamma}} - \phi < 0) - \tau\} = 0. \tag{2.5}$$

We then set $\overline{g}(\mathbf{z}_i^{\mathrm{T}}\overline{\boldsymbol{\beta}}, \overline{\boldsymbol{\gamma}}) \equiv \phi$ to be our final one-step estimator of $g(\mathbf{z}_i^{\mathrm{T}}\boldsymbol{\beta})$, $i = 1, 2, \cdots, N$. Further, we set our final one step estimator of $g'(\mathbf{z}_i^{\mathrm{T}}\boldsymbol{\beta})$ to be

$$\overline{g'}(\mathbf{z}_i^{\mathrm{T}}\overline{\boldsymbol{\beta}}, \overline{\boldsymbol{\gamma}}) = \frac{\sum_{j=1}^{N} h_1^{-2} K'(\frac{\mathbf{z}_j^{\mathrm{T}}\overline{\boldsymbol{\beta}} - \mathbf{z}_i^{\mathrm{T}}\overline{\boldsymbol{\beta}}}{h_1})[I\{Y_j - \mathbf{x}_j^{\mathrm{T}}\overline{\boldsymbol{\gamma}} - \overline{g}(\mathbf{z}_i^{\mathrm{T}}\overline{\boldsymbol{\beta}}, \overline{\boldsymbol{\gamma}}) < 0\} - \tau]}{\sum_{j=1}^{N} K_{h_1}(\mathbf{z}_j^{\mathrm{T}}\overline{\boldsymbol{\beta}} - \mathbf{z}_i^{\mathrm{T}}\overline{\boldsymbol{\beta}})\mathcal{K}_{h_3}\{Y_j - \mathbf{x}_j^{\mathrm{T}}\overline{\boldsymbol{\gamma}} - \overline{g}(\mathbf{z}_i^{\mathrm{T}}\overline{\boldsymbol{\beta}}, \overline{\boldsymbol{\gamma}})\}}, \quad \text{for } i = 1, 2, \cdots, N. \tag{2.6}$$

<u>Remark</u> 3. *The data splitting above is performed to ensure that the initial estimator and the estimation of the nuisance components are based on the data that is independent of the data that facilitates the subsequent one-step update. Such data splitting procedure is commonly used to ease the theoretical development (Newey and Powell, 1990) and can sometimes be* *avoided at the expense of more complex and careful analysis, thus in practice, it is often not necessarily followed. In our context, however, we do not see a clear way to push through the theoretical development without performing data splitting. An interesting observation is that data splitting does not cause an efficiency loss, in that if we had used the truth for all the nuisance components, and used the entire data set to construct one-step estimators for* $\boldsymbol{\beta}_{-1}$ *and* $\boldsymbol{\gamma}$*, we would have obtained the same efficiency.*

To clarify the estimation procedure, we provide the algorithm below. **Algorithm:**

1. Data: $(Y_i, \mathbf{x}_i, \mathbf{z}_i)$, $i = 1, 2, \cdots, N$.

2. Splitting: Let $I_1 = \{1, 2, \cdots, n\}$, $I_2 = \{n+1, \cdots, N\}$, $n = \lceil N/2 \rceil$ and $m = N - n$.

3a. Initial estimator: Obtain $\widetilde{\boldsymbol{\beta}}_{-1}$ and $\widetilde{\boldsymbol{\gamma}}$ based on the data in group I_1.

3b. Estimate $g(\cdot)$: Obtain $\widetilde{g}(\mathbf{z}^{\mathrm{T}}\widetilde{\boldsymbol{\beta}}, \widetilde{\boldsymbol{\gamma}})$ at $\mathbf{z} = \mathbf{z}_i : i \in I_2$ by solving (2.2).

3c. Estimate $g'(\cdot)$: Obtain $\widetilde{g}'(\mathbf{z}^{\mathrm{T}}\widetilde{\boldsymbol{\beta}}, \widetilde{\boldsymbol{\gamma}})$ at $\mathbf{z} = \mathbf{z}_i : i \in I_2$ using (2.3).

3d. One-step estimator: Obtain $\widehat{\boldsymbol{\gamma}}$ and $\widehat{\boldsymbol{\beta}}_{-1}$ by (2.4).

4. Data swap: Switch the roles of I_1 and I_2 and repeat the last four steps to get $\widehat{\boldsymbol{\gamma}}'$ and $\widehat{\boldsymbol{\beta}}'_{-1}$.

5. Final one-step estimator: Obtain the final $\overline{\boldsymbol{\gamma}} \equiv (\widehat{\boldsymbol{\gamma}} + \widehat{\boldsymbol{\gamma}}')/2$ and $\overline{\boldsymbol{\beta}}_{-1} \equiv (\widehat{\boldsymbol{\beta}}_{-1} + \widehat{\boldsymbol{\beta}}'_{-1})/2$.

6. Final estimators of $g(\cdot)$ and $g'(\cdot)$: Obtain $\overline{g}(\mathbf{z}^{\mathrm{T}}\overline{\boldsymbol{\beta}}, \overline{\boldsymbol{\gamma}})$ and $\overline{g}'(\mathbf{z}^{\mathrm{T}}\overline{\boldsymbol{\beta}}, \overline{\boldsymbol{\gamma}})$ at any $\mathbf{z} = \mathbf{z}_i, i \in I_1 \cup I_2$ by (2.5) and (2.6).

Note that the estimation in steps 3b and 3c are based on data I_1, while the step in 3d combines the preparations in steps 3a, 3b, 3c, which are based on I_1, and the data in I_2.

3. Theoretical Results

Before we present the theoretical results of the locally efficient estimator, we first list some regularity conditions.

C1 (*covariates*) The covariates $(\mathbf{x}^{\mathrm{T}}, \mathbf{z}^{\mathrm{T}})^{\mathrm{T}}$ belong to a finite region.

C2 (*error pdf*) There are constants $0 < c_f < C_f < \infty$ such that $c_f \leqslant f_\epsilon^*(\epsilon, \mathbf{x}, \mathbf{z}) \leqslant C_f$, $c_f \leqslant f_\epsilon(\epsilon, \mathbf{x}, \mathbf{z}) \leqslant C_f$, $c_f \leqslant f_\epsilon'(\epsilon, \mathbf{x}, \mathbf{z}) \leqslant C_f$, $c_f \leqslant f_{\mathbf{z}^{\mathrm{T}}\boldsymbol{\beta}}(\mathbf{z}^{\mathrm{T}}\boldsymbol{\beta}) \leqslant C_f$ and $c_f \leqslant f_{\epsilon|\mathbf{z}^{\mathrm{T}}\boldsymbol{\beta}^*}(\epsilon, \mathbf{z}^{\mathrm{T}}\boldsymbol{\beta}^*) \leqslant C_f$, for all ϵ in a neighborhood of 0, for any possibly random $\boldsymbol{\beta}^*$ satisfying $\|\boldsymbol{\beta}^* - \boldsymbol{\beta}\| = O_p(n^{-1/2})$, and for all $\mathbf{x}, \mathbf{z}$.

C3 (*error cdf*) Let $F_\epsilon(\epsilon, \mathbf{x}, \mathbf{z})$ be the conditional distribution function of ϵ given $\mathbf{x}, \mathbf{z}$. There exists a finite constant L, so that for any ϵ_1 and ϵ_2, $|F_\epsilon(\epsilon_1, \mathbf{x}, \mathbf{z}) - F_\epsilon(\epsilon_2, \mathbf{x}, \mathbf{z})| \leqslant L|\epsilon_1 - \epsilon_2|$, for all $\mathbf{x}, \mathbf{z}$.

C4 (*bandwidth*) The bandwidths satisfy $h_1 = n^{-\alpha_1}$, $1/8 < \alpha_1 < 1/5$, $h_2 = n^{-\alpha_2}$, $0 < \alpha_2 < 1$, $h_3 = n^{-\alpha_3}$, $0 < \alpha_3 < 1$.

C5 (*kernel function*) The kernel functions $K(\cdot)$ and $\mathcal{K}(\cdot)$ are continuously differentiable, symmetric, bounded, and have compact support $[-1, 1]$. The function $K(\cdot)$ also satisfies $\int K(u)u^2 \mathrm{d}u = 0$.

C6 (*density function of* $\mathbf{z}$) The density function $f_{\mathbf{z}}(\mathbf{z})$ of $\mathbf{z}$ is bounded and has compact support.

C7 (*Lipschitz condition*) For any possibly random $\boldsymbol{\beta}^*$ satisfying $\|\boldsymbol{\beta}^* - \boldsymbol{\beta}\| = O_p(n^{-1/2})$, any $\mathbf{x}$ and any $\mathbf{z}$, $|f_{\mathbf{x},\mathbf{z}|\mathbf{z}^{\mathrm{T}}\boldsymbol{\beta}^*}(\mathbf{x}, \mathbf{z}, \mathbf{z}^{\mathrm{T}}\boldsymbol{\beta}^*) - f_{\mathbf{x},\mathbf{z}|\mathbf{z}^{\mathrm{T}}\boldsymbol{\beta}}(\mathbf{x}, \mathbf{z}, \mathbf{z}^{\mathrm{T}}\boldsymbol{\beta})| \leqslant C\|\boldsymbol{\beta}^* - \boldsymbol{\beta}\|$, $|f_{\mathbf{z}^{\mathrm{T}}\boldsymbol{\beta}^*}(\mathbf{z}^{\mathrm{T}}\boldsymbol{\beta}^*) - f_{\mathbf{z}^{\mathrm{T}}\boldsymbol{\beta}}(\mathbf{z}^{\mathrm{T}}\boldsymbol{\beta})| \leqslant C\|\boldsymbol{\beta}^* - \boldsymbol{\beta}\|$, $|f^{(1)}_{\mathbf{x},\mathbf{z}|\mathbf{z}^{\mathrm{T}}\boldsymbol{\beta}^*}(\mathbf{x}, \mathbf{z}, \mathbf{z}^{\mathrm{T}}\boldsymbol{\beta}^*) - f^{(1)}_{\mathbf{x},\mathbf{z}|\mathbf{z}^{\mathrm{T}}\boldsymbol{\beta}}(\mathbf{x}, \mathbf{z}, \mathbf{z}^{\mathrm{T}}\boldsymbol{\beta})| \leqslant C\|\boldsymbol{\beta}^* - \boldsymbol{\beta}\|$,

$|f^{(1)}_{\mathbf{z}^{\mathrm{T}}\boldsymbol{\beta}^*}(\mathbf{z}^{\mathrm{T}}\boldsymbol{\beta}^*) - f^{(1)}_{\mathbf{z}^{\mathrm{T}}\boldsymbol{\beta}}(\mathbf{z}^{\mathrm{T}}\boldsymbol{\beta})| \leqslant C\|\boldsymbol{\beta}^* - \boldsymbol{\beta}\|$, $|f^{(2)}_{\mathbf{x},\mathbf{z}|\mathbf{z}^{\mathrm{T}}\boldsymbol{\beta}^*}(\mathbf{x},\mathbf{z},\mathbf{z}^{\mathrm{T}}\boldsymbol{\beta}^*) - f^{(2)}_{\mathbf{x},\mathbf{z}|\mathbf{z}^{\mathrm{T}}\boldsymbol{\beta}}(\mathbf{x},\mathbf{z},\mathbf{z}^{\mathrm{T}}\boldsymbol{\beta})| \leqslant C\|\boldsymbol{\beta}^* - \boldsymbol{\beta}\|$, and $|f^{(2)}_{\mathbf{z}^{\mathrm{T}}\boldsymbol{\beta}^*}(\mathbf{z}^{\mathrm{T}}\boldsymbol{\beta}^*) - f^{(2)}_{\mathbf{z}^{\mathrm{T}}\boldsymbol{\beta}}(\mathbf{z}^{\mathrm{T}}\boldsymbol{\beta})| \leqslant C\|\boldsymbol{\beta}^* - \boldsymbol{\beta}\|$, where $f_{\mathbf{x},\mathbf{z}|\mathbf{z}^{\mathrm{T}}\boldsymbol{\beta}^*}(\mathbf{x},\mathbf{z},\mathbf{z}^{\mathrm{T}}\boldsymbol{\beta}^*)$ is the conditional density function of $\mathbf{x},\mathbf{z}$ given $\mathbf{z}^{\mathrm{T}}\boldsymbol{\beta}^*$, and $f_{\mathbf{z}^{\mathrm{T}}\boldsymbol{\beta}^*}(\mathbf{z}^{\mathrm{T}}\boldsymbol{\beta}^*)$ is the density function of $\mathbf{z}^{\mathrm{T}}\boldsymbol{\beta}^*$. $f_{\mathbf{x},\mathbf{z}|\mathbf{z}^{\mathrm{T}}\boldsymbol{\beta}^*}(\mathbf{x},\mathbf{z},\mathbf{z}^{\mathrm{T}}\boldsymbol{\beta}^*)$ has compact support.

C8 (*boundedness*) For any possibly random $\boldsymbol{\beta}^*$ satisfying $\|\boldsymbol{\beta}^* - \boldsymbol{\beta}\| = O_p(n^{-1/2})$, $f_{\mathbf{x},\mathbf{z}|\mathbf{z}^{\mathrm{T}}\boldsymbol{\beta}^*}(\mathbf{x},\mathbf{z},\mathbf{z}^{\mathrm{T}}\boldsymbol{\beta}^*)$, $f_{\mathbf{z}^{\mathrm{T}}\boldsymbol{\beta}^*}(\mathbf{z}^{\mathrm{T}}\boldsymbol{\beta}^*)$, $f^{(1)}_{\mathbf{x},\mathbf{z}|\mathbf{z}^{\mathrm{T}}\boldsymbol{\beta}^*}(\mathbf{x},\mathbf{z},\mathbf{z}^{\mathrm{T}}\boldsymbol{\beta}^*)$, $f^{(1)}_{\mathbf{z}^{\mathrm{T}}\boldsymbol{\beta}^*}(\mathbf{z}^{\mathrm{T}}\boldsymbol{\beta}^*)$, $f^{(2)}_{\mathbf{x},\mathbf{z}|\mathbf{z}^{\mathrm{T}}\boldsymbol{\beta}^*}(\mathbf{x},\mathbf{z},\mathbf{z}^{\mathrm{T}}\boldsymbol{\beta}^*)$, $f^{(2)}_{\mathbf{z}^{\mathrm{T}}\boldsymbol{\beta}^*}(\mathbf{z}^{\mathrm{T}}\boldsymbol{\beta}^*)$ are bounded.

C9 (*function g*) $g(\cdot)$ is twice continuously differentiable with finite derivatives.

C10 (*differentiability*) Let $r(t) \equiv E_{\mathbf{x},\mathbf{z}|\mathbf{z}^{\mathrm{T}}\boldsymbol{\beta}}\{f^*_\epsilon(0,\mathbf{x}_i,\mathbf{z}_i)f_\epsilon(0,\mathbf{x}_i,\mathbf{z}_i)P_s(\mathbf{x}_i,\mathbf{z}_i)| \mathbf{z}_i^{\mathrm{T}}\boldsymbol{\beta} = t\}\{g(t) - g(\mathbf{z}_j^{\mathrm{T}}\boldsymbol{\beta})\}/f_{\epsilon|\mathbf{z}^{\mathrm{T}}\boldsymbol{\beta}}(0,t)$, where $P_s(\mathbf{x}_i,\mathbf{z}_i)$ is defined in Lemma 8 in Supplementary Materials. Assume that $r(t)$ is four times continuously differentiable with finite derivatives for $i \in I_2$ and $j \in I_1$, or $i \in I_1$ and $j \in I_2$.

These conditions mainly concern the boundedness and smoothness of related functions and are quite standard. The only conditions worth mentioning are Condition C4 which concerns the bandwidth involved in the nonparametric estimation procedures and Condition C5 which requires $K(\cdot)$ to be a kernel with order higher than two. Even though all the nonparametric regression involved in the analysis are conditional on the single index, which is one dimensional, Condition C4 still excludes the optimal bandwidth $O(n^{-1/5})$. Instead, a larger bandwidth is required, which would lead to oversmoothing in combination with the regular second order kernel function and inflated bias. The bandwidth requirement in Condition C4 directly leads to the higher order kernel requirement in Condition C5. This is a special feature of the quantile regression, a direct consequence caused by the unsmoothness of the indicator function. Similar bandwidth requirement is also needed in Newey and Powell (1990).

For notational brevity in stating the theoretical results, define

$$\widetilde{F}(\phi, \boldsymbol{\gamma}, \boldsymbol{\beta}) \equiv \frac{\sum_{j\in I_1} K_{h_1}(\mathbf{z}_j^{\mathrm{T}}\boldsymbol{\beta} - \mathbf{z}_i^{\mathrm{T}}\boldsymbol{\beta}) I(Y_j - \mathbf{x}_j^{\mathrm{T}}\boldsymbol{\gamma} - \phi < 0)}{\sum_{j\in I_1} K_{h_1}(\mathbf{z}_j^{\mathrm{T}}\boldsymbol{\beta} - \mathbf{z}_i^{\mathrm{T}}\boldsymbol{\beta})}.$$

Theorem 1. *Under the Regularity Conditions C1-C8,* $\widetilde{g}(\mathbf{z}^{\mathrm{T}}\widetilde{\boldsymbol{\beta}}, \widetilde{\gamma}) - g(\mathbf{z}^{\mathrm{T}}\boldsymbol{\beta}) = O_p\{k(n, h_1)\}$ *uniformly for all* $\mathbf{z}$*, where* $k(n, h_1) = h_1^4 + (nh_1)^{-1/2}$ *.*

Theorem 1 establishes the uniform convergence rate result for the nonparametric estimator of the single index function $g(\cdot)$. This paves the way to further deriving the asymptotic expansion of the estimator.

Theorem 2. *Under the Regularity Conditions C1-C8, for any* $i \in I_2$*, we have*

$$\begin{aligned}
& \widetilde{g}(\mathbf{z}_i^{\mathrm{T}}\widetilde{\boldsymbol{\beta}}, \widetilde{\gamma}) - g(\mathbf{z}_i^{\mathrm{T}}\widetilde{\boldsymbol{\beta}}) \\
= \;& \{f_{\epsilon|\mathbf{z}^{\mathrm{T}}\boldsymbol{\beta}}(0, \mathbf{z}_i^{\mathrm{T}}\boldsymbol{\beta})\}^{-1}[\tau - \widetilde{F}\{g(\mathbf{z}_i^{\mathrm{T}}\boldsymbol{\beta}), \boldsymbol{\gamma}, \boldsymbol{\beta}\}] \\
& - \{f_{\epsilon|\mathbf{z}^{\mathrm{T}}\boldsymbol{\beta}}(0, \mathbf{z}_i^{\mathrm{T}}\boldsymbol{\beta})\}^{-1} E\{f_\epsilon(0, \mathbf{x}_i, \mathbf{z}_i)\mathbf{x}_i^{\mathrm{T}} | \mathbf{z}_i^{\mathrm{T}}\boldsymbol{\beta}\}(\widetilde{\gamma} - \boldsymbol{\gamma}) \\
& - \{f_{\epsilon|\mathbf{z}^{\mathrm{T}}\boldsymbol{\beta}}(0, \mathbf{z}_i^{\mathrm{T}}\boldsymbol{\beta})\}^{-1} g'(\mathbf{z}_i^{\mathrm{T}}\boldsymbol{\beta}) E\{f_\epsilon(0, \mathbf{x}_i, \mathbf{z}_i)\mathbf{z}_{-1,i}^{\mathrm{T}} | \mathbf{z}_i^{\mathrm{T}}\boldsymbol{\beta}\}(\widetilde{\boldsymbol{\beta}}_{-1} - \boldsymbol{\beta}_{-1}) + o_p(n^{-1/2}),
\end{aligned}$$

where $f_{\epsilon|\mathbf{z}^{\mathrm{T}}\boldsymbol{\beta}}(\epsilon, \mathbf{z}^{\mathrm{T}}\boldsymbol{\beta})$ *is the conditional probability density function of* ϵ *given* $\mathbf{z}^{\mathrm{T}}\boldsymbol{\beta}$*.*

Theorem 2 provides the asymptotic expansion of the nonparametric estimation of $g(\mathbf{z}^{\mathrm{T}}\boldsymbol{\beta})$ at the index formed at the initial estimator of $\boldsymbol{\beta}$ and each observation in I_2. This expansion is later used in facilitating the asymptotic analysis of the locally efficient estimator $\widehat{\boldsymbol{\beta}}_{-1}$ and $\widehat{\gamma}$.

Theorem 3. *Under the Regularity Conditions, C1-C10, we have*

$$\sqrt{N}[\{\overline{\boldsymbol{\gamma}}^{\mathrm{T}},(\overline{\boldsymbol{\beta}}_{-1})^{\mathrm{T}}\}^{\mathrm{T}}-\{\boldsymbol{\gamma}^{\mathrm{T}},(\boldsymbol{\beta}_{-1})^{\mathrm{T}}\}^{\mathrm{T}}]\rightarrow N(0,\boldsymbol{\Sigma})$$

in distribution as $N\rightarrow\infty$*, where* $\boldsymbol{\Sigma}$ *is defined in (S6.42) of the Supplementary Materials. Further, when* $f_\epsilon^*(\epsilon,\mathbf{x},\mathbf{z})=f_\epsilon(\epsilon,\mathbf{x},\mathbf{z})$*,* $\boldsymbol{\Sigma}=[E\{\mathbf{S}_{\mathrm{eff}}(Y,\mathbf{x},\mathbf{z})^{\otimes 2}\}]^{-1}$*, hence the estimator reaches the optimal efficiency. Here and throughout the text, we use* $\mathbf{a}^{\otimes 2}$ *to denote* $\mathbf{a}\mathbf{a}^{\mathrm{T}}$ *for any vector or matrix* $\mathbf{a}$*.*

Theorem 3 is the main theorem of this work. It establishes the asymptotic convergence of the locally efficient estimators of the parameters of our central interest $\boldsymbol{\beta}_{-1}$ and $\boldsymbol{\gamma}$. The form of the asymptotic variance is unfortunately very complex, which we defer to Supplementary Materials. Nevertheless, several observations are worth mentioning. First, the result of Theorem 3 shows that when we perform data splitting, switching and then averaging the results, we obtain an estimator that is as efficient as we would have expected to get intuitively if we had used the entire data set in each estimation step. This is evidenced by the fact that the estimator is efficient when the working model f_ϵ^* is correct. In other words, there is no cost associated with the data splitting procedure. Intuitively, this is because each part of the data still contributes to the estimation of all the unknown components in the model, and the data splitting is conducted in a way only to facilitate the technical analysis. This is very much similar to the data splitting treatment when a general machine learning method is engaged in semiparametric models. Second, when the working model is correctly specified, the need of performing nonparametric estimation of $g(\cdot)$, $g'(\cdot)$ and the various nonparametric expectations conditional on $\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x}$ does not incur any efficiency loss. However, this is no longer the case when the working model is misspecified. This is drastically different from the mean regression case. Indeed, in Ma and Zhu (2013c), it has been shown that nonparametric estimations are not associated with efficiency loss, regardless of whether the regression error model is misspecified or not. A closer look into this reveals that this difference between the partially linear single index quantile and mean regressions lies in the difference of the $\Lambda^{\perp}$ spaces, which is much smaller in the quantile regression context than in the mean regression context. In fact, in the quantile regression context, as soon as the error distribution is misspecified, the resulting locally efficient score is no longer a member of $\Lambda^{\perp}$, hence its orthogonality with the nonparametric estimation is lost. This is not the case for the mean regression, in that the efficient score function computed under a misspecified error distribution model continues to belong to $\Lambda^{\perp}$, hence continues to be orthogonal with the nuisance parameter estimations. This shows the fundamental difference between the quantile model and the mean model, and also reflects the inherent difficulties involved in the partially linear single index quantile model, in that both the semiparametric analysis and the theoretical development is much more involved.

The asymptotic property of $\overline{g}(\mathbf{z}^{\mathrm{T}}\overline{\boldsymbol{\beta}}, \overline{\boldsymbol{\gamma}})$ can be obtained by following the similar process for $\widetilde{g}(\mathbf{z}^{\mathrm{T}}\widetilde{\boldsymbol{\beta}}, \widetilde{\boldsymbol{\gamma}})$, so we omit it.

4. Numerical Examples

In this section we carry out simulations to evaluate the finite sample performance of our proposed estimator. Throughout all simulations, the kernel used to obtain all nonparametric regressions is the quartic kernel,

$$K(u) = \frac{15}{16}(1-u^2)^2, \ |u| \leqslant 1.$$

We use bandwidths satisfying Condition C4. More specifically $h_1 = n^{-(1/8+1/5)/2}$, and $h_3 = n^{-1/10}$ are used throughout numerical studies. For the correct error model, $h_2 = n^{-1/2}$. For the two misspecified error cases, $h_2 = n^{-1/3}$. The estimators by Zhang, Lian and Yu (2017) are used as the initial estimators of β_{-1} and γ for our estimation. Although we generated the data from distributions with unbounded support, in the realization, we find that Condition C1 is practically satisfied.

In Simulation 1, we generate data from a sine-bump model

$$Y_i = \mathbf{x}_i^{\mathrm{T}}\boldsymbol{\gamma} + 5\sin\left(\frac{\mathbf{z}_i^{\mathrm{T}}\boldsymbol{\beta} - a}{b-a}\pi\right) + \sigma\epsilon_i. \tag{4.7}$$

Here $\epsilon_i \equiv \sigma(\mathbf{x}_i, \mathbf{z}_i)\{\widetilde{\epsilon}_i - F^{-1}(\tau)\}$, where $\widetilde{\epsilon}_i \sim F(\cdot)$ and independent of the $\mathbf{x}_i$ and $\mathbf{z}_i$. Note that this construction of ϵ_i ensures that the model assumption (2.1) is always satisfied for any τ. We first generate $\widetilde{\mathbf{x}}_i \sim N_5(\mathbf{0}, \boldsymbol{\Sigma})$ and the (i,j)-th element of the 5×5 matrix $\boldsymbol{\Sigma}$ is $0.3^{|i-j|}$ for $i,j = 1,\ldots,5$, where $\widetilde{\mathbf{x}}_i = (\widetilde{\mathbf{x}}_{i1}, \cdots, \widetilde{\mathbf{x}}_{i5})$. Let $\mathbf{x}_i = (\widetilde{\mathbf{x}}_{i1}, \widetilde{\mathbf{x}}_{i5})^{\mathrm{T}}$ and $\mathbf{z}_i = (\widetilde{\mathbf{x}}_{i2}, \widetilde{\mathbf{x}}_{i3}, \widetilde{\mathbf{x}}_{i4})^{\mathrm{T}}$. Here, $\boldsymbol{\gamma} = (2,1)^{\mathrm{T}}/\sqrt{3}$, $\boldsymbol{\beta} = (1,-1,0.5)^{\mathrm{T}}$, $a = 0.4$, $b = 1.5$ and $\widetilde{\epsilon}_i \sim N(0, 0.3^2)$.

We consider three different quantile levels $\tau = 0.4, 0.5, 0.6$, and two different settings for $\sigma(\mathbf{x}_i, \mathbf{z}_i)$.

- Setting 1: $\sigma(\mathbf{x}_i, \mathbf{z}_i) = 1$.

- Setting 2: $\sigma(\mathbf{x}_i, \mathbf{z}_i) = \sqrt{|\widetilde{\mathbf{x}}_{i3}|}$.

Note that in Setting 2, ϵ_i has heteroscedastic distribution.

In addition, we investigate the performance of our estimators under two misspecified error models. Specifically, we specify the error to be $\epsilon_i^* = \widetilde{\epsilon}_i^* - F^{*-1}(\tau)$, where

- Model 1: $\widetilde{\epsilon}_i^* \sim N(0, 0.1^2)$.
- Model 2: $\widetilde{\epsilon}_i^* \sim t(3)$.

Throughout, we consider sample sizes $n = 400$ and 1000, and repeat the Monte Carlo experiment 500 times.

The estimation results are given in Tables 1 and 2. for settings 1 and 2 respectively, where we provide the biases and the empirical standard deviations. In addition, the comparison estimation results based on Zhang, Lian and Yu (2017) are provided in the tables as well. As seen in Tables 1, our estimators based on both the correct error model and the misspecified error models (Model 1 and Model 2) have better performance than that by Zhang, Lian and Yu (2017) when the error is independent of the covariates (Setting 1). When the error is dependent of the covariates (Setting 2), our estimators still have better performance, as is shown in Tables 2. This indicates the robustness property of our proposed estimators. The performance of the nonparametric function estimates, with mean and confidence bands, are presented in Figures 1- 6. One can see that the estimated curves are very close to the true $g(\cdot)$ function across all τ values.

Table 1: Simulation 1, Setting 1. Bias (“bias”) and empirical standard deviation (“std”) for γ and β_{-1} are based on 500 iterations. The true error $\epsilon_i = \sigma(\mathbf{x}_i, \mathbf{z}_i)\{\widetilde{\epsilon}_i - F^{-1}(\tau)\}$, where $\widetilde{\epsilon}_i \sim F(\cdot)$ and $\sigma(\mathbf{x}_i, \mathbf{z}_i) = 1$. Estimation results are provided by our proposed method (“PR”) consisting of the one based on the correct model $\widetilde{\epsilon}_i \sim N(0, 0.3^2)$ (“correct”) and those based on $\widetilde{\epsilon}_i^* \sim N(0, 0.1^2)$ and $\widetilde{\epsilon}_i^* \sim t_{(3)}$, and by Zhang, Lian and Yu (2017) (“ZH”). The true parameter is $\boldsymbol{\gamma} = (2/\sqrt{3}, 1/\sqrt{3})^{\mathrm{T}}$ and $\boldsymbol{\beta} = (1, -1, 0.5)^{\mathrm{T}}$.

$n = 400$									
		PR						ZH	
		correct		Model 1		Model 2			
τ		γ_1	γ_2	γ_1	γ_2	γ_1	γ_2	γ_1	γ_2
0.4	bias	0.0009	-0.0002	0.0018	0.0001	-0.0025	-0.0013	0.0018	0.0035
	std	0.0721	0.0641	0.0921	0.0840	0.0579	0.0621	0.1322	0.1293
0.5	bias	0.0000	0.0019	0.0002	0.0031	-0.0005	-0.0027	-0.0010	-0.0027
	std	0.0679	0.0708	0.0919	0.0965	0.0677	0.0739	0.1281	0.1358
0.6	bias	0.0006	0.0023	0.0021	0.0026	-0.0051	0.0012	0.0020	0.0033
	std	0.0625	0.0674	0.0822	0.0885	0.0644	0.0644	0.1203	0.1344
τ		β_2	β_3	β_2	β_3	β_2	β_3	β_2	β_3
0.4	bias	0.0011	-0.0005	0.0012	-0.0005	0.0005	-0.0007	0.0021	-0.0004
	std	0.0181	0.0173	0.0221	0.0211	0.0098	0.0090	0.0322	0.0284
0.5	bias	0.0001	0.0002	-0.0003	0.0004	0.0009	-0.0009	0.0006	-0.0012
	std	0.0162	0.0153	0.0207	0.0196	0.0128	0.0115	0.0320	0.0288
0.6	bias	0.0014	0.0004	0.0014	0.0006	0.0012	-0.0006	0.0010	-0.0001
	std	0.0167	0.0146	0.0205	0.0180	0.0112	0.0100	0.0269	0.0264
$n = 1000$									
τ		γ_1	γ_2	γ_1	γ_2	γ_1	γ_2	γ_1	γ_2
0.4	bias	0.0020	0.0016	0.0023	0.0020	0.0008	0.0001	0.0035	0.0018
	std	0.0538	0.0505	0.0769	0.0722	0.0598	0.0550	0.0991	0.0979
0.5	bias	0.0026	0.0026	0.0050	0.0037	-0.0068	-0.0014	0.0019	0.0028
	std	0.0500	0.0468	0.0782	0.0731	0.0840	0.0810	0.1129	0.1042
0.6	bias	-0.0001	0.0054	-0.0007	0.0079	0.0021	-0.0041	0.0043	0.0128
	std	0.0461	0.0435	0.0683	0.0642	0.0618	0.0586	0.0900	0.0891
τ		β_2	β_3	β_2	β_3	β_2	β_3	β_2	β_3
0.4	bias	-0.0003	0.0006	-0.0006	0.0009	0.0010	-0.0009	-0.0011	0.0014
	std	0.0194	0.0164	0.0245	0.0206	0.0082	0.0074	0.0299	0.0271
0.5	bias	-0.0004	0.0001	-0.0007	0.0002	0.0010	-0.0011	-0.0009	-0.0001
	std	0.0161	0.0149	0.0215	0.0200	0.0125	0.0115	0.0308	0.0269
0.6	bias	0.0009	0.0000	0.0010	0.0001	0.0004	-0.0001	0.0005	0.0009
	std	0.0136	0.0133	0.0176	0.0172	0.0070	0.0069	0.0260	0.0240

Table 2: Simulation 1, Setting 2. Bias ("bias") and empirical standard deviation ("std") for γ and β_{-1} are based on 500 iterations. The true error $\epsilon_i = \sigma(\mathbf{x}_i, \mathbf{z}_i)\{\widetilde{\epsilon}_i - F^{-1}(\tau)\}$, where $\widetilde{\epsilon}_i \sim F(\cdot)$ and $\sigma(\mathbf{x}_i, \mathbf{z}_i) = \sqrt{|\widetilde{\mathbf{x}}_{i3}|}$. Estimation results are provided by our proposed method ("PR") consisting of the one based on the correct model $\widetilde{\epsilon}_i \sim N(0, 0.3^2)$ ("correct") and those based on $\widetilde{\epsilon}_i^* \sim N(0, 0.1^2)$ and $\widetilde{\epsilon}_i^* \sim t_{(3)}$, and by Zhang, Lian and Yu (2017) ("ZH"). The true parameter is $\boldsymbol{\gamma} = (2/\sqrt{3}, 1/\sqrt{3})^{\mathrm{T}}$ and $\boldsymbol{\beta} = (1, -1, 0.5)^{\mathrm{T}}$.

$n = 400$									
		PR						ZH	
		correct		Model 1		Model 2			
τ		γ_1	γ_2	γ_1	γ_2	γ_1	γ_2	γ_1	γ_2
0.4	bias	0.0049	-0.0002	0.0052	-0.0004	0.0015	-0.0019	0.0023	-0.0015
	std	0.0861	0.0778	0.0919	0.0832	0.0651	0.0635	0.1289	0.1255
0.5	bias	-0.0017	-0.0008	-0.0017	-0.0007	0.0015	-0.0037	0.0025	0.0040
	std	0.0917	0.0881	0.0990	0.0938	0.0762	0.0744	0.1536	0.1303
0.6	bias	-0.0046	0.0035	-0.0049	0.0039	-0.0005	-0.0038	-0.0013	0.0056
	std	0.0803	0.0779	0.0867	0.0834	0.0647	0.0683	0.1218	0.1219
τ		β_2	β_3	β_2	β_3	β_2	β_3	β_2	β_3
0.4	bias	-0.0016	-0.0002	0.0006	-0.0001	0.0018	-0.0010	0.0041	-0.0024
	std	0.0181	0.0185	0.0208	0.0195	0.0094	0.0089	0.0289	0.0294
0.5	bias	-0.0021	0.0006	-0.0002	0.0007	0.0003	0.0000	-0.0020	0.0015
	std	0.0176	0.0193	0.0210	0.0206	0.0123	0.0123	0.0319	0.0333
0.6	bias	-0.0008	-0.0007	0.0001	-0.0008	0.0009	-0.0003	0.0012	0.0014
	std	0.0166	0.0170	0.0193	0.0178	0.0103	0.0102	0.0284	0.0253
$n = 1000$									
τ		γ_1	γ_2	γ_1	γ_2	γ_1	γ_2	γ_1	γ_2
0.4	bias	0.0020	0.0030	0.0029	0.0029	-0.0026	-0.0001	-0.0021	0.0057
	std	0.0678	0.0673	0.0721	0.0718	0.0614	0.0603	0.0961	0.0998
0.5	bias	-0.0013	0.0036	-0.0018	0.0038	0.0011	-0.0040	-0.0003	0.0064
	std	0.0689	0.0706	0.0748	0.0761	0.0874	0.0862	0.0999	0.1073
0.6	bias	0.0051	-0.0015	0.0050	0.0015	-0.0020	0.0041		-0.0030
	std	0.0659	0.0655	0.0700	0.0700	0.0655	0.0682	0.0984	0.0952
τ		β_2	β_3	β_2	β_3	β_2	β_3	β_2	β_3
0.4	bias	-0.0012	0.0008	0.0010	0.0008	0.0010	0.0001	0.0013	-0.0004
	std	0.0191	0.0202	0.0222	0.0210	0.0068	0.0066	0.0292	0.0269
0.5	bias	-0.0012	0.0002	0.0009	0.0003	0.0014	-0.0004	0.0010	-0.0001
	std	0.0189	0.0193	0.0232	0.0204	0.0126	0.0125	0.0295	0.0284
0.6	bias	-0.0004	0.0002	0.0008	0.0002	0.0005	0.0002	0.0011	-0.0001
	std	0.0144	0.0172	0.0173	0.0179	0.0070	0.0081	0.0268	0.0238

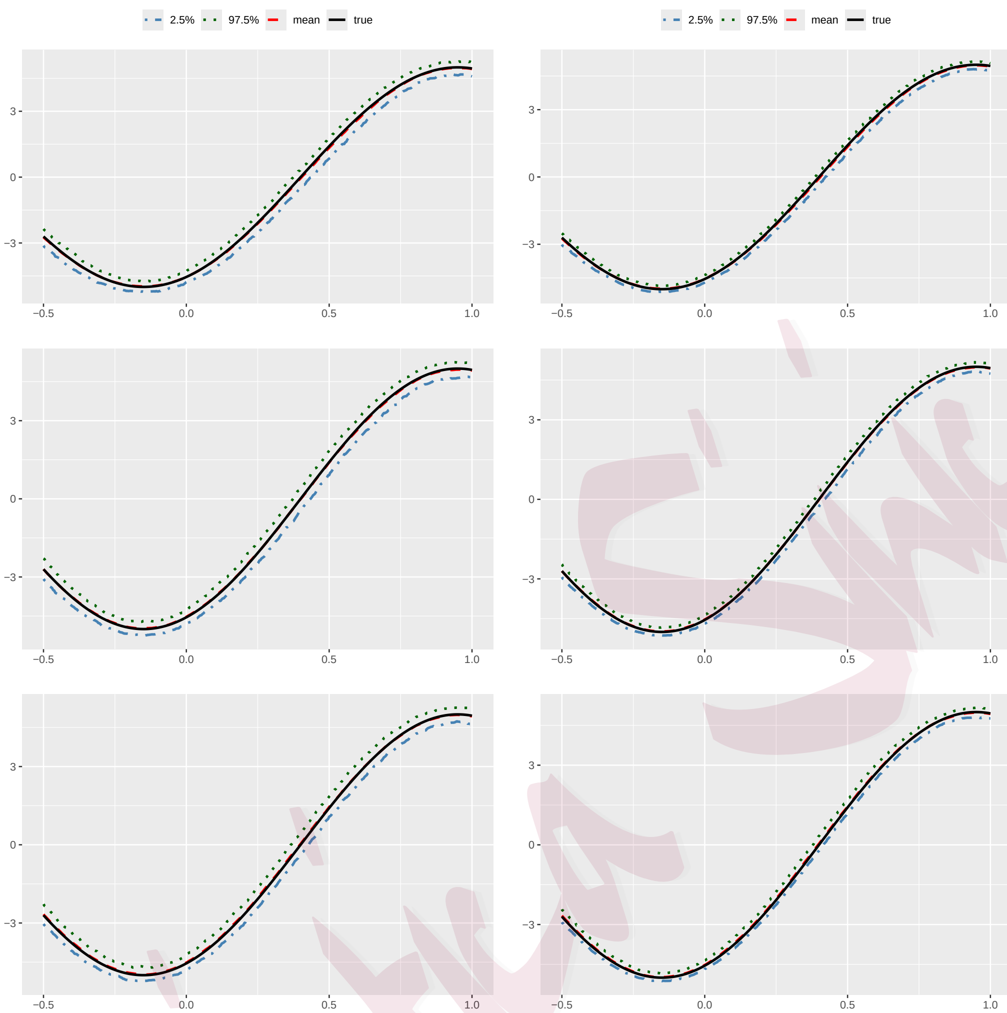

Figure 1: Plots of estimators for $g(\cdot)$ in Simulation 1, Setting 1. The true error $\epsilon_i = \sigma(\mathbf{x}_i, \mathbf{z}_i)\{\widetilde{\epsilon}_i - F^{-1}(\tau)\}$, where $\widetilde{\epsilon}_i = F(\cdot)$ and $\sigma(\mathbf{x}_i, \mathbf{z}_i) = 1$. Estimation is based on our proposed method with the correct model $\widetilde{\epsilon}_i \sim N(0, 0.3^2)$. Top to bottom panels: $\tau = 0.4, 0.5, 0.6$. Left panels: $n = 400$, right panels: $n = 1000$. The black sold lines are the true, the dashed red lines are the mean curve, the blue dot-dashed lines are 2.5% quantile, and the darkgreen dotted lines are 97.5% quantile.

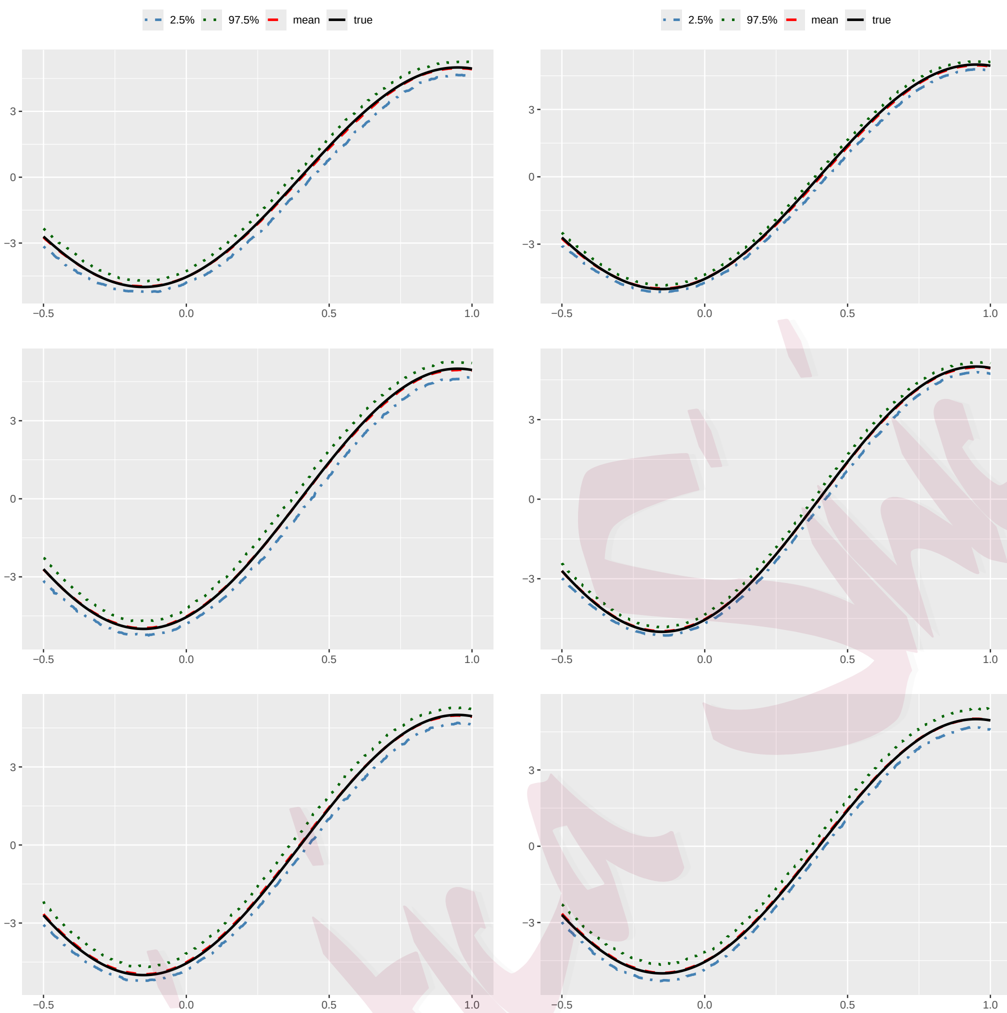

Figure 2: Plots of estimators for $g(\cdot)$ in Simulation 1, Setting 1. The true error $\epsilon_i = \sigma(\mathbf{x}_i, \mathbf{z}_i)\{\widetilde{\epsilon}_i - F^{-1}(\tau)\}$, where $\widetilde{\epsilon}_i = F(\cdot)$ and $\sigma(\mathbf{x}_i, \mathbf{z}_i) = 1$. Estimation is based on our proposed method with the first misspecified error model $\widetilde{\epsilon}_i^* \sim N(0, 0.1^2)$. Top to bottom panels: $\tau = 0.4, 0.5, 0.6$. Left panels: $n = 400$, right panels: $n = 1000$. The black sold lines are the true, the dashed red lines are the mean curve, the blue dot-dashed lines are 2.5% quantile, and the darkgreen dotted lines are 97.5% quantile.

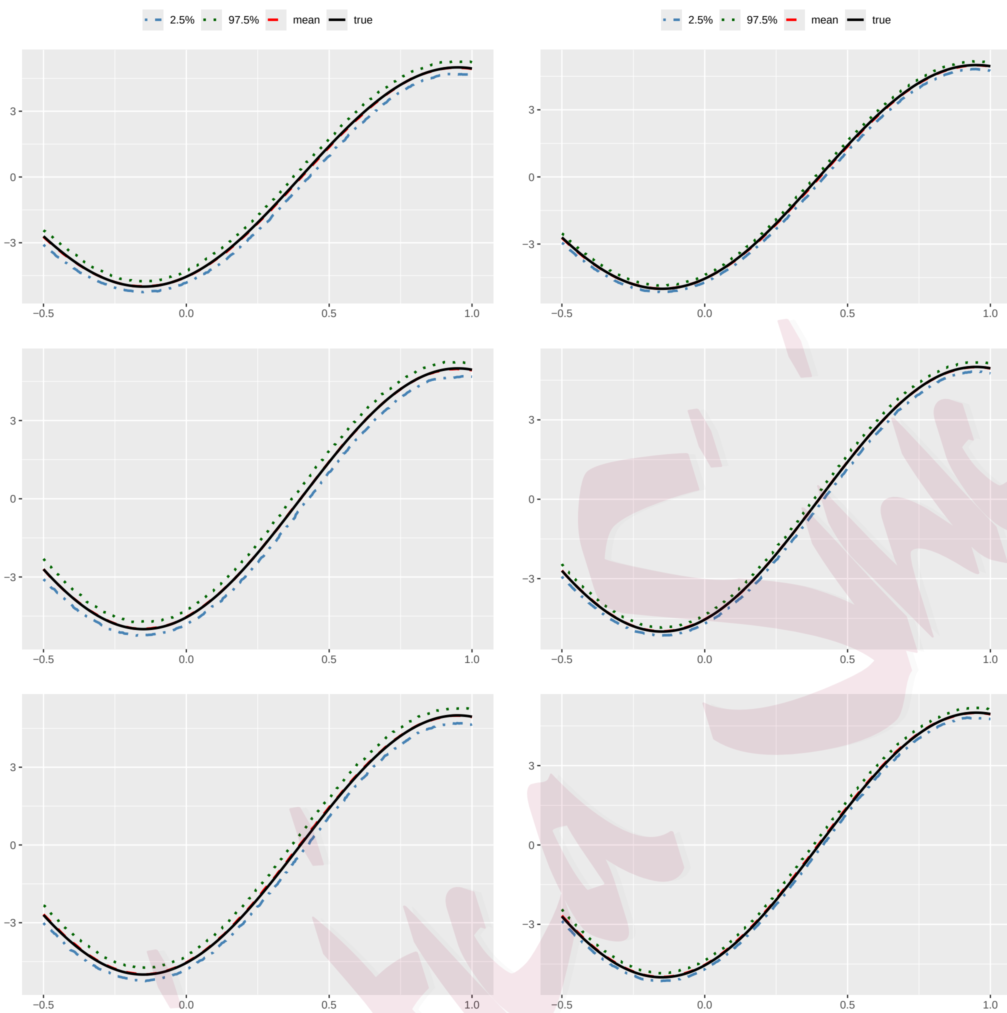

Figure 3: Plots of estimators for $g(\cdot)$ in Simulation 1, Setting 1. The true error $\epsilon_i = \sigma(\mathbf{x}_i, \mathbf{z}_i)\{\widetilde{\epsilon}_i - F^{-1}(\tau)\}$, where $\widetilde{\epsilon}_i = F(\cdot)$ and $\sigma(\mathbf{x}_i, \mathbf{z}_i) = 1$. Estimation is based on our proposed method with the second misspecified error model $\widetilde{\epsilon}_i^* \sim t_{(3)}$. Top to bottom panels: $\tau = 0.4, 0.5, 0.6$. Left panels: $n = 400$, right panels: $n = 1000$. The black sold lines are the true, the dashed red lines are the mean curve, the blue dot-dashed lines are 2.5% quantile, and the darkgreen dotted lines are 97.5% quantile.

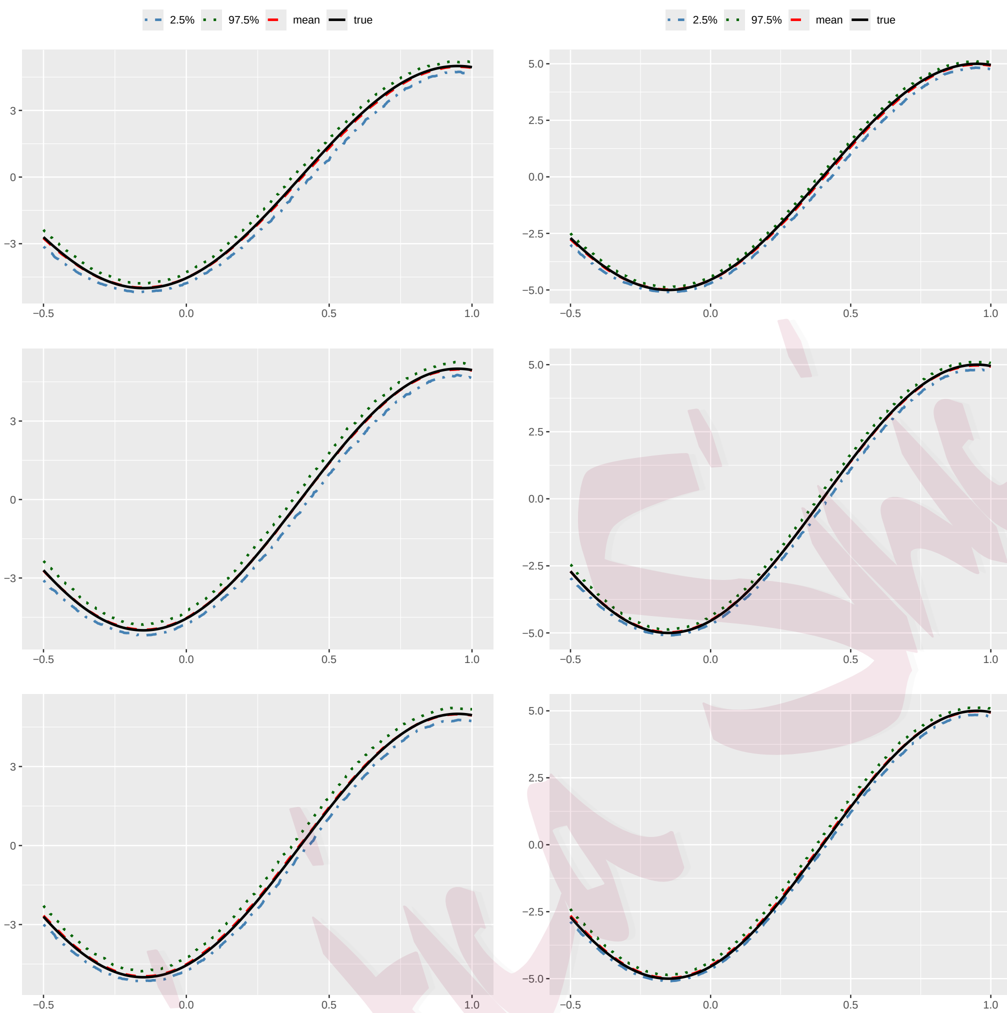

Figure 4: Plots of estimators for $g(\cdot)$ in Simulation 1, Setting 2: The true error $\epsilon_i = \sigma(\mathbf{x}_i, \mathbf{z}_i)\{\widetilde{\epsilon}_i - F^{-1}(\tau)\}$, where $\widetilde{\epsilon}_i = F(\cdot)$ and $\sigma(\mathbf{x}_i, \mathbf{z}_i) = \sqrt{|\widetilde{\mathbf{x}}_{i3}|}$. Estimation is based on our proposed method with the correct model $\widetilde{\epsilon}_i \sim N(0, 0.3^2)$. Top to bottom panels: $\tau = 0.4, 0.5, 0.6$. Left panels: $n = 400$, right panels: $n = 1000$. The black sold lines are the true, the dashed red lines are the mean curve, the blue dot-dashed lines are 2.5% quantile, and the darkgreen dotted lines are 97.5% quantile.

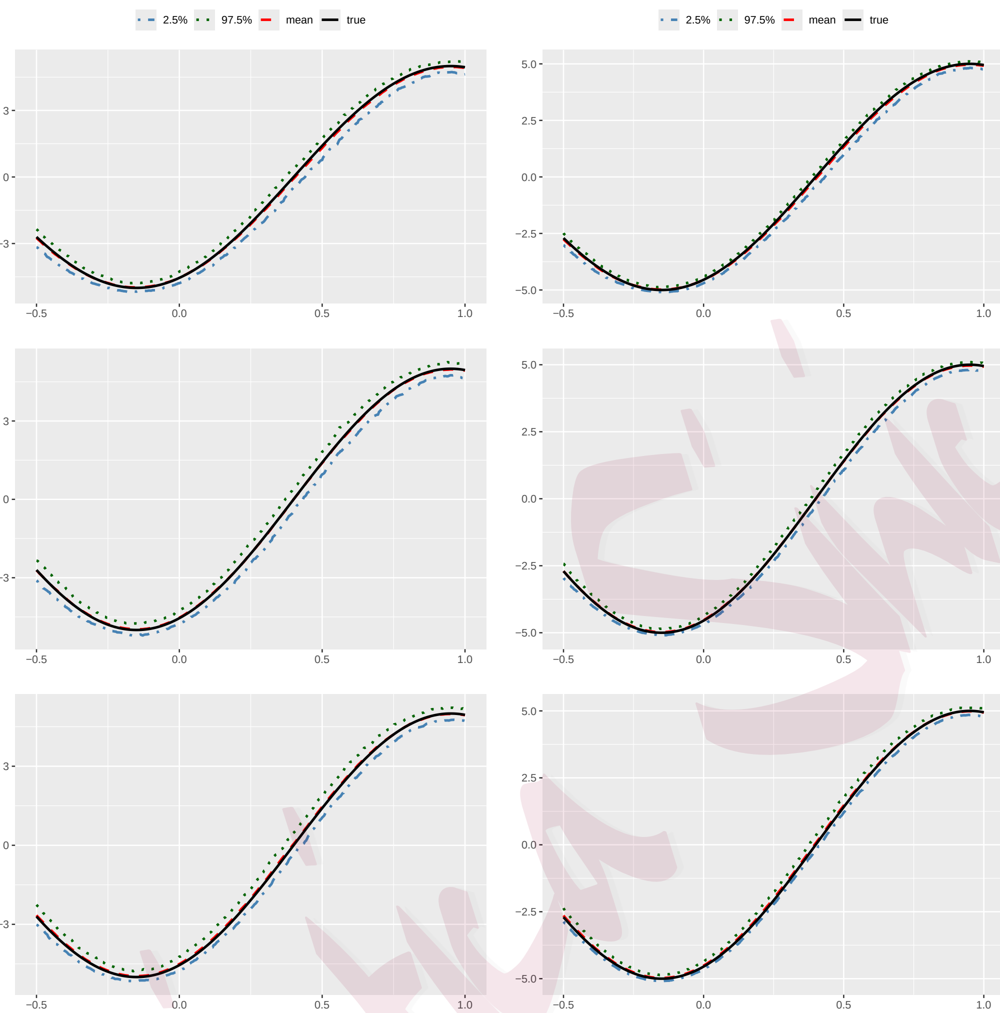

Figure 5: Plots of estimators for $g(\cdot)$ in Simulation 1, Setting 2: The true error $\epsilon_i = \sigma(\mathbf{x}_i, \mathbf{z}_i)\{\widetilde{\epsilon}_i - F^{-1}(\tau)\}$, where $\widetilde{\epsilon}_i = F(\cdot)$ and $\sigma(\mathbf{x}_i, \mathbf{z}_i) = \sqrt{|\widetilde{\mathbf{x}}_{i3}|}$. Estimation is based on our proposed method with the first misspecified error model $\widetilde{\epsilon}_i^* \sim N(0, 0.1^2)$. Top to bottom panels: $\tau = 0.4, 0.5, 0.6$. Left panels: $n = 400$, right panels: $n = 1000$. The black sold lines are the true, the dashed red lines are the mean curve, the blue dot-dashed lines are 2.5% quantile, and the darkgreen dotted lines are 97.5% quantile.

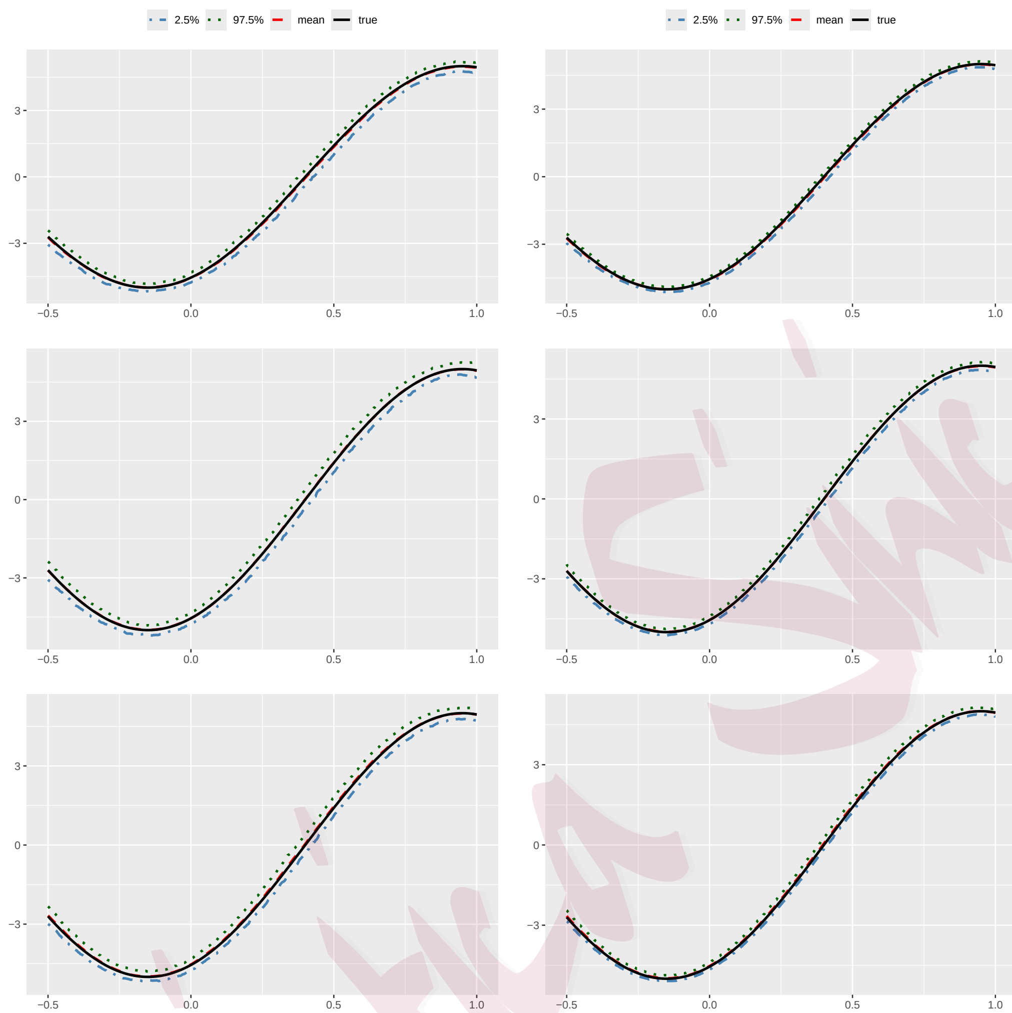

Figure 6: Plots of estimators for $g(\cdot)$ in Simulation 1, Setting 2: The true error $\epsilon_i = \sigma(\mathbf{x}_i, \mathbf{z}_i)\{\widetilde{\epsilon}_i - F^{-1}(\tau)\}$, where $\widetilde{\epsilon}_i \sim F(\cdot)$ and $\sigma(\mathbf{x}_i, \mathbf{z}_i) = \sqrt{|\widetilde{\mathbf{x}}_{i3}|}$. Estimation is based on our proposed method with the second misspecified error model $\widetilde{\epsilon}_i^* \sim t_{(3)}$. Top to bottom panels: $\tau = 0.4, 0.5, 0.6$. Left panels: $n = 400$, right panels: $n = 1000$. The black sold lines are the true, the dashed red lines are the mean curve, the blue dot-dashed lines are 2.5% quantile, and the darkgreen dotted lines are 97.5% quantile.

We further conducted a second simulation. In Simulation 2, we consider the same model as in (4.7) and the same covariates generation. As in Simulation 1, we set $\epsilon_i \equiv \sigma(\mathbf{x}_i, \mathbf{z}_i)\{\widetilde{\epsilon}_i - F^{-1}(\tau)\}$, where $\widetilde{\epsilon}_i \sim F(\cdot)$ and independent of the $\mathbf{x}_i$ and $\mathbf{z}_i$. However, different from Simulation 1, we set $\widetilde{\epsilon}_i \sim SN(0, 1, 3)$, where $SN(0, 1, 3)$ is the Skew-Normal distribution, with 0,1 and 3 being the location, scale and slant parameters. We fix a, b, γ and β at the same values as in Simulation 1.

We still consider three different quantile levels $\tau = 0.4, 0.5, 0.6$, and two different settings for $\sigma(\mathbf{x}_i, \mathbf{z}_i)$.

- Setting 1: $\sigma(\mathbf{x}_i, \mathbf{z}_i) = 1$.
- Setting 2: $\sigma(\mathbf{x}_i, \mathbf{z}_i) = \sqrt{|\widetilde{\mathbf{x}}_{i2}|}$.

In estimation, we considered two misspecified working models. Specifically, we assume the misspecified error $\epsilon_i^* = \widetilde{\epsilon}_i^* - F^{*-1}(\tau)$ where

- Model 1: $\widetilde{\epsilon}_i^* \sim N(0, 1)$,
- Model 2: $\widetilde{\epsilon}_i^* \sim t(3)$.

The estimation results are given in Tables 3 and 4 for settings 1 and 2 respectively, where we provide the biases and the empirical standard deviations. In addition, the comparison estimation results based on Zhang, Lian and Yu (2017) are provided in the tables as well. As seen in Table 3 and 4 , our estimators based on both error models have better performance than that by Zhang, Lian and Yu (2017). The mean and confidence bands of the nonparametric function estimates are presented in Figures 7-9 and Figures S.1-S.3 of the Supplementary Materials, showing close approximation to the true function.

Inspecting the results of simulations 1 and 2, the standard deviations of the estimators under correct specifications are larger than those under misspecifications. We suspect this is caused by the non-satisfactory performance of the initial estimator. Thus, we experimented by setting the initial estimator to be the true value plus a random vector with variance n^{-1}. The results based on such initial estimator for Simulation 2 are given in Table 5 and 6. In this case, the estimators under correct specifications always have better performance and reflect our theoretical findings.

Table 3: Simulation 2, Setting 1. Bias ("bias") and empirical standard deviation ("std") for γ and β_{-1} are based on 500 iterations. The true error $\epsilon_i = \sigma(\mathbf{x}_i, \mathbf{z}_i)\{\widetilde{\epsilon}_i - F^{-1}(\tau)\}$, where $\widetilde{\epsilon}_i \sim F(\cdot)$ and $\sigma(\mathbf{x}_i, \mathbf{z}_i) = 1$. Estimation results are provided by our proposed method ("PR") consisting of the one based on the correct model $\widetilde{\epsilon}_i \sim SN(0, 1, 3)$ ("correct") and those based on $\widetilde{\epsilon}_i^* \sim N(0, 1)$ and $\widetilde{\epsilon}_i^* \sim t_{(3)}$, and by Zhang, Lian and Yu (2017) ("ZH"). The true parameter is $\boldsymbol{\gamma} = (2/\sqrt{3}, 1/\sqrt{3})^{\mathrm{T}}$ and $\boldsymbol{\beta} = (1, -1, 0.5)^{\mathrm{T}}$.

$n = 400$									
		PR						ZH	
		correct		Model 1		Model 2			
τ		γ_1	γ_2	γ_1	γ_2	γ_1	γ_2	γ_1	γ_2
0.4	bias	-0.0012	-0.0015	-0.0005	-0.0012	-0.0003	-0.0011	-0.0075	0.0023
	std	0.0640	0.0624	0.0581	0.0562	0.0626	0.0602	0.1331	0.1296
0.5	bias	0.0006	0.0018	-0.0017	0.0025	-0.0022	0.0026	0.0007	0.0068
	std	0.0646	0.0558	0.0633	0.0548	0.0669	0.0580	0.1467	0.1279
0.6	bias	-0.0004	0.0034	-0.0029	0.0044	-0.0037	0.0048	0.0098	0.0130
	std	0.0629	0.0603	0.0637	0.0584	0.0674	0.0616	0.1185	0.1264
τ		β_2	β_3	β_2	β_3	β_2	β_3	β_2	β_3
0.4	bias	0.0009	-0.0017	0.0008	-0.0013	0.0008	-0.0012	-0.0015	0.0008
	std	0.0146	0.0130	0.0114	0.0103	0.0116	0.0106	0.0310	0.0306
0.5	bias	0.0001	-0.0002	-0.0003	-0.0006	-0.0003	-0.0007	0.0013	-0.0011
	std	0.0145	0.0129	0.0119	0.0107	0.0121	0.0109	0.0328	0.0295
0.6	bias	-0.0003	0.0006	0.0002	0.0004	0.0004	0.0003	-0.0028	0.0023
	std	0.0119	0.0117	0.0104	0.0101	0.0106	0.0102	0.0277	0.0268
$n = 1000$									
τ		γ_1	γ_2	γ_1	γ_2	γ_1	γ_2	γ_1	γ_2
0.4	bias	-0.0018	0.0039	-0.0008	-0.0011	-0.0006	-0.0023	0.0012	0.0092
	std	0.0377	0.0395	0.0382	0.0369	0.0447	0.0438	0.0971	0.1060
0.5	bias	0.0011	-0.0019	-0.0004	-0.0022	-0.0007	-0.0022	0.0115	0.0012
	std	0.0371	0.0374	0.0375	0.0383	0.0427	0.0434	0.1021	0.0985
0.6	bias	0.0025	-0.0011	-0.0006	-0.0016	-0.0017	-0.0017	0.0088	0.0008
	std	0.0388	0.0361	0.0363	0.0357	0.0396	0.0392	0.0995	0.0904
τ		β_2	β_3	β_2	β_3	β_2	β_3	β_2	β_3
0.4	bias	0.0004	0.0010	0.0010	-0.0008	0.0012	-0.0008	0.0013	0.0021
	std	0.0131	0.0117	0.0077	0.0066	0.0077	0.0066	0.0322	0.0284
0.5	bias	0.0006	-0.0009	0.0013	-0.0008	0.0014	-0.0008	-0.0013	-0.0011
	std	0.0111	0.0103	0.0078	0.0077	0.0083	0.0082	0.0296	0.0281
0.6	bias	-0.0002	0.0003	0.0006	-0.0003	0.0009	-0.0005	-0.0019	0.0007
	std	0.0096	0.0095	0.0066	0.0068	0.0064	0.0066	0.0256	0.0232

Table 4: Simulation 2, Setting 2. Bias ("bias") and empirical standard deviation ("std") for γ and β_{-1} are based on 500 iterations. The true error $\epsilon_i = \sigma(\mathbf{x}_i, \mathbf{z}_i)\{\widetilde{\epsilon}_i - F^{-1}(\tau)\}$, where $\widetilde{\epsilon}_i \sim F(\cdot)$ and $\sigma(\mathbf{x}_i, \mathbf{z}_i) = \sqrt{|\widetilde{\mathbf{x}}_{i2}|}$. Estimation results are provided by our proposed method ("PR") consisting of the one based on the correct model $\widetilde{\epsilon}_i \sim SN(0, 1, 3)$ ("correct") and those based on $\widetilde{\epsilon}_i^* \sim N(0, 1)$ and $\widetilde{\epsilon}_i^* \sim t_{(3)}$, and by Zhang, Lian and Yu (2017) ("ZH"). The true parameter is $\boldsymbol{\gamma} = (2/\sqrt{3}, 1/\sqrt{3})^{\mathrm{T}}$ and $\boldsymbol{\beta} = (1, -1, 0.5)^{\mathrm{T}}$.

$n = 400$									
		PR						ZH	
		correct		Model 1		Model 2			
τ		γ_1	γ_2	γ_1	γ_2	γ_1	γ_2	γ_1	γ_2
0.4	bias	0.0032	-0.0038	0.0009	0.0026	0.0004	0.0035	0.0080	-0.0057
	std	0.0766	0.0781	0.0551	0.0567	0.0609	0.0625	0.1226	0.1214
0.5	bias	-0.0036	0.0001	0.0013	0.0034	0.0018	0.0037	-0.0050	0.0106
	std	0.0724	0.0724	0.0555	0.0548	0.0611	0.0606	0.1262	0.1311
0.6	bias	0.0019	-0.0005	-0.0006	0.0013	-0.0009	0.0018	0.0038	-0.0131
	std	0.0708	0.0706	0.0578	0.0548	0.0628	0.0600	0.1195	0.1222
τ		β_2	β_3	β_2	β_3	β_2	β_3	β_2	β_3
0.4	bias	0.0038	-0.0023	-0.0020	0.0010	-0.0018	0.0009	-0.0050	0.0020
	std	0.0170	0.0166	0.0121	0.0106	0.0120	0.0107	0.0339	0.0307
0.5	bias	0.0032	-0.0013	-0.0016	0.0007	-0.0014	0.0005	-0.0025	0.0026
	std	0.0145	0.0152	0.0116	0.0104	0.0119	0.0111	0.0323	0.0288
0.6	bias	0.0004	0.0002	-0.0009	0.0010	-0.0006	0.0008	-0.0049	0.0026
	std	0.0132	0.0147	0.0113	0.0102	0.0115	0.0106	0.0289	0.0248
$n = 1000$									
τ		γ_1	γ_2	γ_1	γ_2	γ_1	γ_2	γ_1	γ_2
0.4	bias	0.0026	0.0008	0.0008	0.0019	0.0006	0.0021	0.0008	0.0016
	std	0.0622	0.0595	0.0405	0.0393	0.0495	0.0483	0.1034	0.0996
0.5	bias	-0.0019	-0.0024	0.0010	-0.0010	0.0014	-0.0010	0.0025	-0.0069
	std	0.0519	0.0590	0.0483	0.0434	0.0562	0.0514	0.1051	0.0999
0.6	bias	0.0030	0.0045	-0.0039	0.0007	-0.0050	0.0000	0.0059	0.0036
	std	0.0521	0.0523	0.0346	0.0365	0.0405	0.0436	0.0905	0.0873
τ		β_2	β_3	β_2	β_3	β_2	β_3	β_2	β_3
0.4	bias	0.0031	-0.0009	-0.0020	0.0008	-0.0017	0.0006	-0.0035	0.0016
	std	0.0152	0.0176	0.0076	0.0080	0.0074	0.0079	0.0316	0.0277
0.5	bias	0.0028	-0.0015	-0.0017	0.0006	-0.0015	0.0005	-0.0042	0.0023
	std	0.0130	0.0168	0.0094	0.0087	0.0099	0.0092	0.0305	0.0287
0.6	bias	0.0008	0.0000	-0.0005	-0.0003	-0.0002	-0.0006	-0.0051	0.0034
	std	0.0101	0.0141	0.0068	0.0067	0.0066	0.0066	0.0229	0.0223

Table 5: Simulation 2, Setting 1. Bias ("bias") and empirical standard deviation ("std") for γ and β are based on 500 iterations. The true error $\epsilon_i = \sigma(\mathbf{x}_i, \mathbf{z}_i)\{\widetilde{\epsilon}_i - F^{-1}(\tau)\}$, where $\widetilde{\epsilon}_i \sim F(\cdot)$ and $\sigma(\mathbf{x}_i, \mathbf{z}_i) = 1$. Estimation results are provided by our proposed method ("PR") consisting of the one based on the correct model $\widetilde{\epsilon}_i \sim SN(0,1,3)$ ("correct") and those based on $\widetilde{\epsilon}_i^* \sim N(0,1)$ and $\widetilde{\epsilon}_i^* \sim t_{(3)}$. The true parameter is $\boldsymbol{\gamma} = (2/\sqrt{3}, 1/\sqrt{3})^{\mathrm{T}}$ and $\boldsymbol{\beta} = (1, -1, 0.5)^{\mathrm{T}}$.

$n = 400$							
		PR					
		correct		Model 1		Model 2	
τ		γ_1	γ_2	γ_1	γ_2	γ_1	γ_2
0.4	bias	-0.0004	-0.0018	-0.0007	-0.0030	-0.0008	-0.0033
	std	0.0337	0.0331	0.0572	0.0562	0.0632	0.0621
0.5	bias	0.0023	-0.0011	0.0035	-0.0018	0.0038	-0.0019
	std	0.0386	0.0390	0.0603	0.0608	0.0654	0.0660
0.6	bias	0.0006	-0.0007	0.0008	-0.0010	0.0009	-0.0012
	std	0.0419	0.0422	0.0604	0.0609	0.0668	0.0673
τ		β_2	β_3	β_2	β_3	β_2	β_3
0.4	bias	-0.0001	0.0001	-0.0003	0.0002	-0.0003	0.0002
	std	0.0047	0.0047	0.0080	0.0079	0.0089	0.0088
0.5	bias	0.0004	0.0000	0.0006	0.0000	0.0006	0.0000
	std	0.0062	00052	0.0097	0.0082	0.0105	0.0089
0.6	bias	0.0004	0.0001	0.0006	0.0001	0.0007	0.0001
	std	0.0064	0.0066	0.0092	0.0080	0.0102	0.0089
$n = 1000$							
τ		γ_1	γ_2	γ_1	γ_2	γ_1	γ_2
0.4	bias	-0.0003	-0.0004	-0.0006	-0.0006	-0.0007	-0.0007
	std	0.0219	0.0213	0.0372	0.0362	0.0411	0.0400
0.5	bias	0.0007	-0.0017	0.0010	-0.0026	0.0011	-0.0028
	std	0.0251	0.0247	0.0392	0.0386	0.0425	0.0418
0.6	bias	-0.0010	-0.0006	-0.0015	-0.0009	-0.0016	-0.0010
	std	0.0257	0.0239	0.0370	0.0346	0.0410	0.0382
τ		β_2	β_3	β_2	β_3	β_2	β_3
0.4	bias	0.0000	-0.0000	0.0001	-0.0001	-0.0001	0.0001
	std	0.0027	0.0024	0.0045	0.0041	0.0051	0.0045
0.5	bias	0.0001	-0.0002	0.0001	-0.0004	0.0001	-0.0004
	std	0.0035	0.0029	0.0055	0.0046	0.0059	0.0050
0.6	bias	-0.0003	0.0003	-0.0004	0.0004	-0.0004	0.0005
	std	0.0033	0.0030	0.0048	0.0044	0.0053	0.0048

Table 6: Simulation 2, Setting 2. Bias ("bias") and empirical standard deviation ("std") for γ and β are based on 500 iterations. The true error $\epsilon_i = \sigma(\mathbf{x}_i, \mathbf{z}_i)\{\widetilde{\epsilon}_i - F^{-1}(\tau)\}$, where $\widetilde{\epsilon}_i \sim F(\cdot)$ and $\sigma(\mathbf{x}_i, \mathbf{z}_i) = \sqrt{|\widetilde{\mathbf{x}}_{i2}|}$. Estimation results are provided by our proposed method ("PR") consisting of the one based on the correct model $\widetilde{\epsilon}_i \sim SN(0,1,3)$ ("correct") and those based on $\widetilde{\epsilon}_i^* \sim N(0,1)$ and $\widetilde{\epsilon}_i^* \sim t_{(3)}$. The true parameter is $\boldsymbol{\gamma} = (2/\sqrt{3}, 1/\sqrt{3})^{\mathrm{T}}$ and $\boldsymbol{\beta} = (1, -1, 0.5)^{\mathrm{T}}$.

$n = 400$							
		PR					
		correct		Model 1		Model 2	
τ		γ_1	γ_2	γ_1	γ_2	γ_1	γ_2
0.4	bias	-0.0019	-0.0019	-0.0055	-0.0005	-0.0061	-0.0006
	std	0.0231	0.0225	0.0539	0.0544	0.0596	0.0602
0.5	bias	0.0002	0.0006	0.0010	0.0026	0.0010	0.0029
	std	0.0250	0.0259	0.0569	0.0590	0.0617	0.0640
0.6	bias	0.0012	-0.0013	-0.0007	-0.0005	-0.0008	-0.0005
	std	0.0268	0.0270	0.0569	0.0516	0.0629	0.0570
τ		β_2	β_3	β_2	β_3	β_2	β_3
0.4	bias	-0.0006	0.0003	-0.0027	0.0017	-0.0030	0.0019
	std	0.0037	0.0032	0.0076	0.0069	0.0085	0.0077
0.5	bias	-0.0007	0.0003	-0.0025	0.0013	-0.0027	0.0014
	std	0.0044	0.0039	0.0092	0.0090	0.0100	0.0097
0.6	bias	-0.0003	0.0002	-0.0007	0.0005	-0.0007	0.0006
	std	0.0047	0.0048	0.0083	0.0077	0.0091	0.0085
$n = 1000$							
τ		γ_1	γ_2	γ_1	γ_2	γ_1	γ_2
0.4	bias	0.0006	0.0010	-0.0003	0.0027	-0.0003	0.0030
	std	0.0146	0.0130	0.0355	0.0305	0.0393	0.0337
0.5	bias	-0.0004	-0.0006	-0.0010	-0.0027	-0.0011	-0.0029
	std	0.0146	0.0157	0.0372	0.0374	0.0404	0.0406
0.6	bias	0.0004	0.0002	-0.0002	0.0017	-0.0002	0.0019
	std	0.0156	0.0155	0.0351	0.0323	0.0388	0.0362
τ		β_2	β_3	β_2	β_3	β_2	β_3
0.4	bias	-0.0011	0.0006	-0.0037	0.0020	-0.0041	0.0021
	std	0.0021	0.0017	0.0046	0.0040	0.0051	0.0044
0.5	bias	-0.0010	0.0006	-0.0036	0.0021	-0.0040	0.0022
	std	0.0021	0.0018	0.0053	0.0046	0.0057	0.0050
0.6	bias	-0.0002	-0.0001	-0.0007	0.0004	-0.0008	0.0004
	std	0.0023	0.0020	0.0050	0.0042	0.0056	0.0047

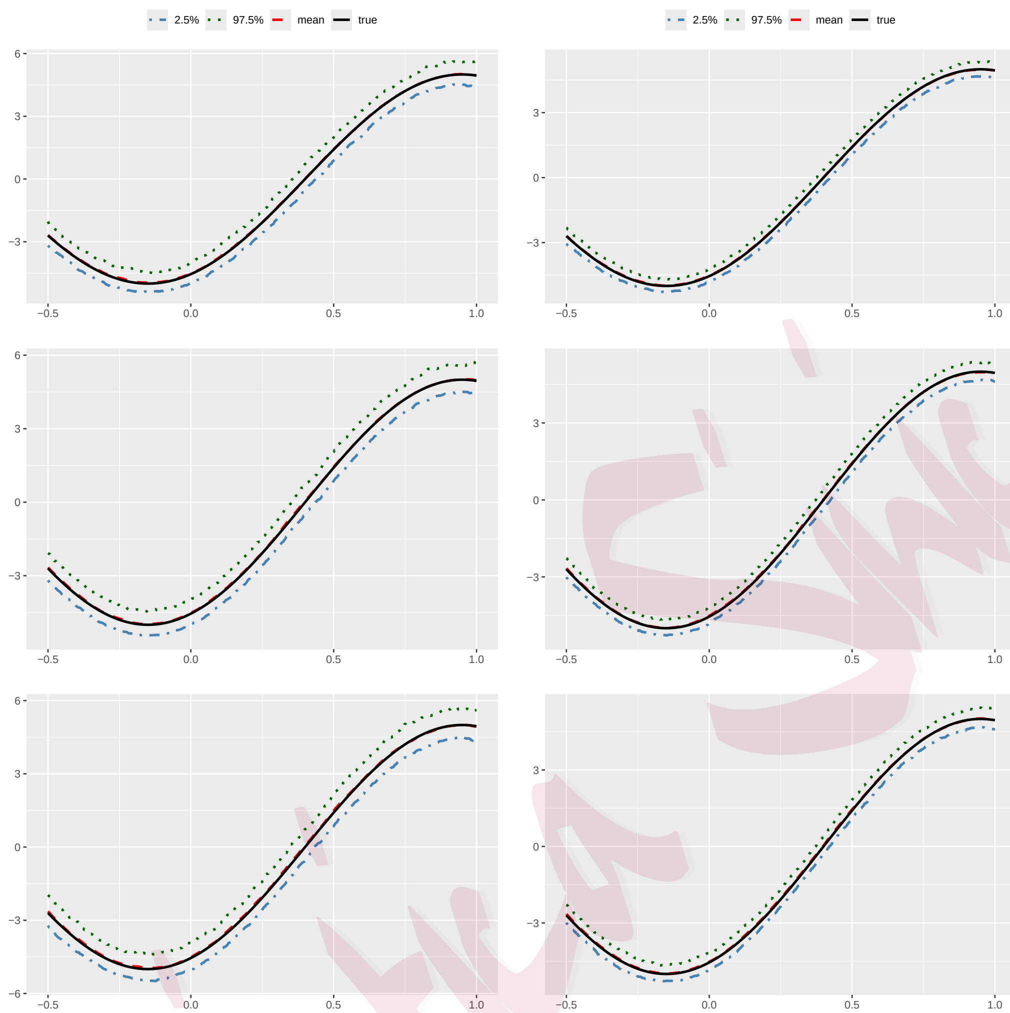

Figure 7: Plots of estimators for $g(\cdot)$ in Simulation 2, Setting 1. The true error $\epsilon_i = \sigma(\mathbf{x}_i, \mathbf{z}_i)\{\widetilde{\epsilon}_i - F^{-1}(\tau)\}$, where $\widetilde{\epsilon}_i = F(\cdot)$ and $\sigma(\mathbf{x}_i, \mathbf{z}_i) = 1$. Estimation is based on our proposed method with the correct model $\widetilde{\epsilon}_i \sim SN(0, 1, 3)$. Top to bottom panels: $\tau = 0.4, 0.5, 0.6$. Left panels: $n = 400$, right panels: $n = 1000$. The black sold lines are the true, the dashed red lines are the mean curve, the blue dot-dashed lines are 2.5% quantile, and the darkgreen dotted lines are 97.5% quantile.

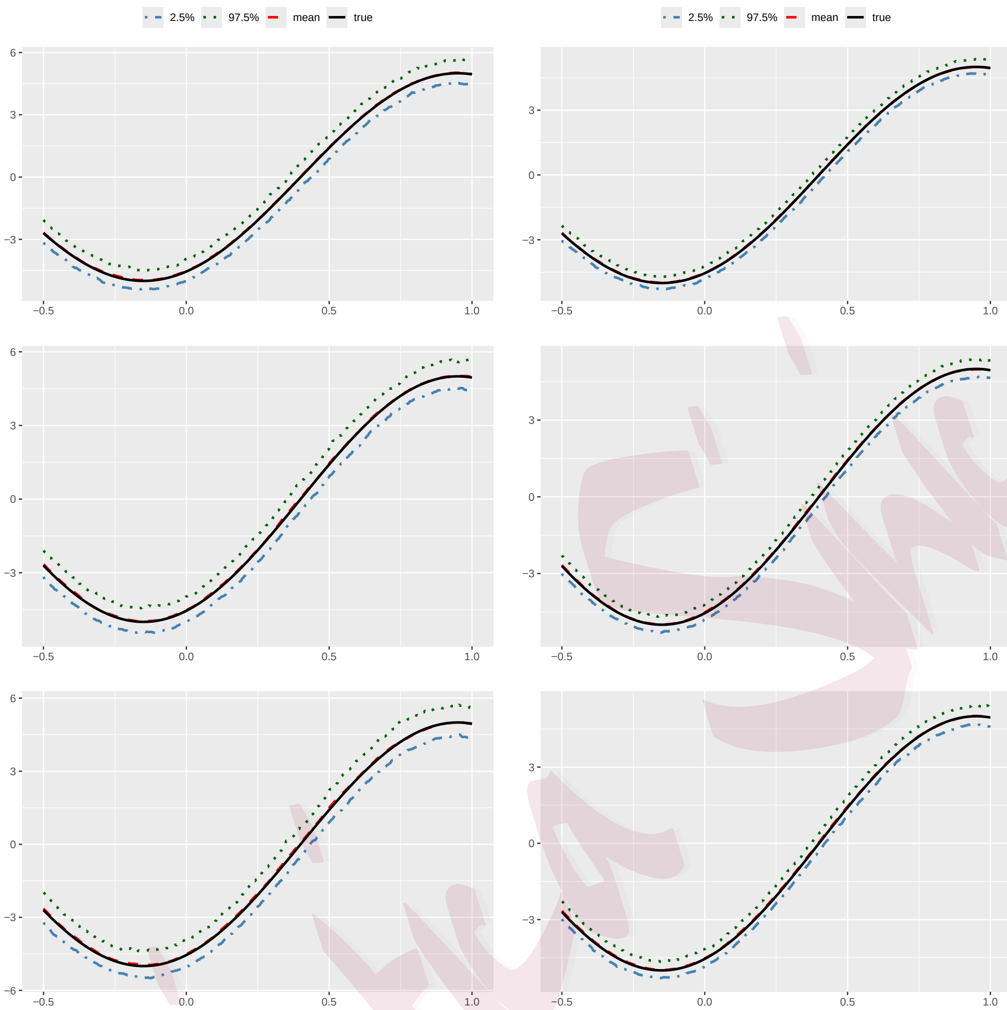

Figure 8: Plots of estimators for $g(\cdot)$ in Simulation 2, Setting 1. The true error $\epsilon_i = \sigma(\mathbf{x}_i, \mathbf{z}_i)\{\widetilde{\epsilon}_i - F^{-1}(\tau)\}$, where $\widetilde{\epsilon}_i = F(\cdot)$ and $\sigma(\mathbf{x}_i, \mathbf{z}_i) = 1$. Estimation is based on our proposed method with the first misspecified error model $\widetilde{\epsilon}_i^* \sim N(0, 1)$. Top to bottom panels: $\tau = 0.4, 0.5, 0.6$. Left panels: $n = 400$, right panels: $n = 1000$. The black sold lines are the true, the dashed red lines are the mean curve, the blue dot-dashed lines are 2.5% quantile, and the darkgreen dotted lines are 97.5% quantile.

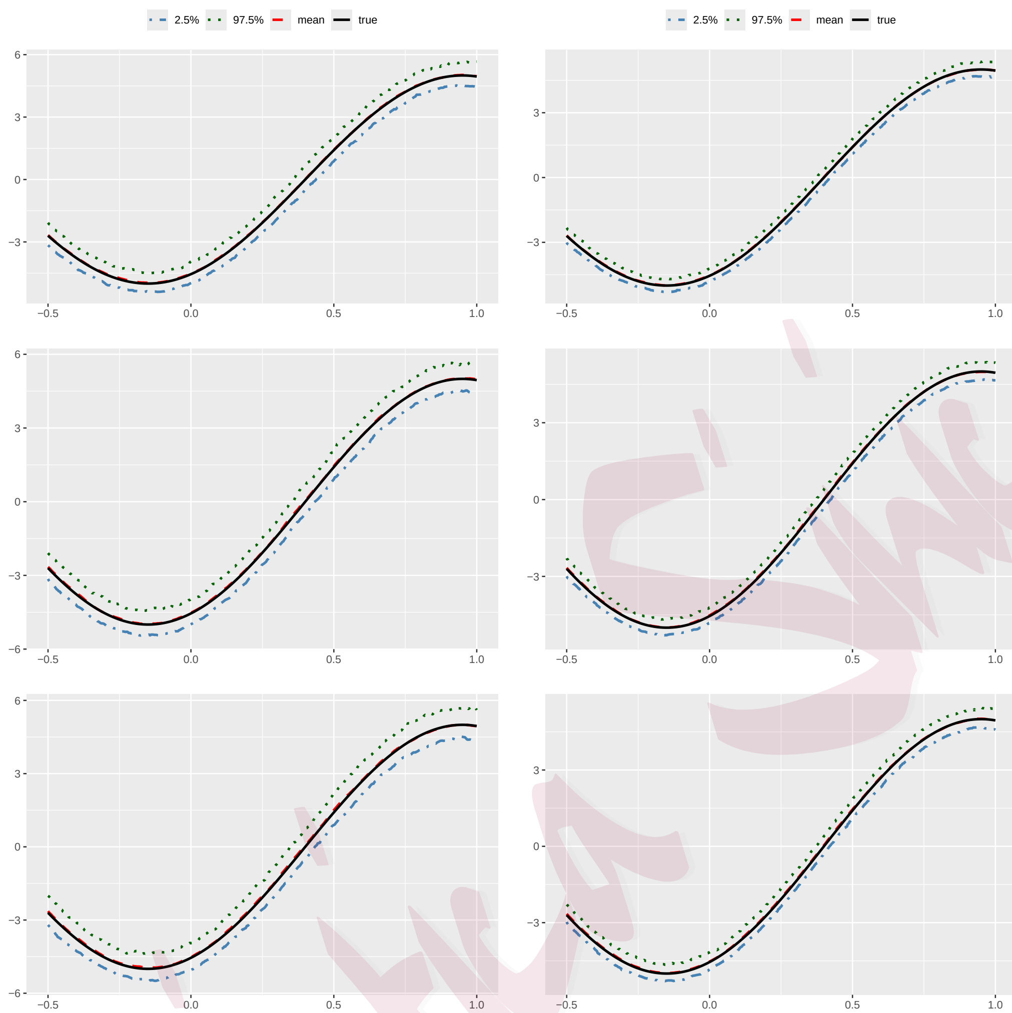

Figure 9: Plots of estimators for $g(\cdot)$ in Simulation 2, Setting 1. The true error $\epsilon_i = \sigma(\mathbf{x}_i, \mathbf{z}_i)\{\widetilde{\epsilon}_i - F^{-1}(\tau)\}$, where $\widetilde{\epsilon}_i = F(\cdot)$ and $\sigma(\mathbf{x}_i, \mathbf{z}_i) = 1$. Estimation is based on our proposed method with the second misspecified error model $\widetilde{\epsilon}_i^* \sim t_{(3)}$. Top to bottom panels: $\tau = 0.4, 0.5, 0.6$. Left panels: $n = 400$, right panels: $n = 1000$. The black sold lines are the true, the dashed red lines are the mean curve, the blue dot-dashed lines are 2.5% quantile, and the darkgreen dotted lines are 97.5% quantile.

5. Real Data Application

We consider an application regarding Boston Housing Price data. The data set can be downloaded from the R package MASS(`https://cran.r-project.org/web/packages/MASS`). It contains 506 observations on 14 variables, including the dependent variable of interest named

MEDV (median value of owner occupied homes in \$1000s in Boston suburb area in 1970's) and 13 independent variables. We apply the partially linear single index quantile regression model to analyze this data set. Zhang, Lian and Yu (2017) considered estimation and variable selection for quantile partially linear single-index models and applied their method to Boston Housing Price data. Following their analysis, we included the covariates RM (average number of rooms per dwelling), LSTAT (percentage of lower status of the population), and DIS (weighted distances to five Boston employment centers) into the single index part, and used BLACK (BLACK$= 1000(B-0.63)^2$, B is the proportion of blacks by town) in the linear part (Harrison and Rubinfeld, 1978). Specifically, we consider the partially linear single index quantile regression model,

$$\begin{aligned} Q_\tau(\text{MEDV} \mid \mathbf{x}, \mathbf{z}) &= \mathbf{x}^{\mathrm{T}}\boldsymbol{\gamma} + g(\mathbf{z}^{\mathrm{T}}\boldsymbol{\beta}) \qquad (5.8)\\ &= \gamma\text{BLACK} + g(\text{RM} + \beta_2\log(\text{LSTAT}) + \beta_3\text{DIS}), \end{aligned}$$

for $\tau = 0.5$. In (5.8), LSTAT is in its logarithm scale, and all covariates are standardized. To implement our estimator, we considered two working models. Specifically, we assumed the error $\epsilon_i^* = \widetilde{\epsilon}_i^* - F^{*-1}(\tau)$ where

- Model 1: $\widetilde{\epsilon}_i^* \sim N(0, 1)$,
- Model 2: $\widetilde{\epsilon}_i^* \sim t(5)$.

The estimation results are given in Table 7, where the results are based on 200 bootstrap samples. In addition, the comparison estimation results based on Zhang, Lian and Yu (2017) are provided in table 7 as well. As seen in Tables 7, our estimators based on the specified error models (Model 1 and Model 2) have better performance than that by Zhang, Lian and Yu (2017). The performance of the nonparametric function estimates, with mean and confidence bands, are presented in Figure 10 together with scatter plots of $MEDV - x\hat{\gamma}^{bk}$ and the estimated indices $(1, \hat{\beta}_2^{bk}, \hat{\beta}_3^{bk})z$, where $\hat{\gamma}^{bk}$, $\hat{\beta}_2^{bk}$ and $\hat{\beta}_3^{bk}$ are the bootstrap estimation results, for Model $k = 1, 2$. One can see that the estimated curve depicts the trend of the scattered points very well.

Combining (5.8), the estimation in Table 7 and Figure 10, we can find the following results. The number of rooms per house (RM) has significant positive effects on housing prices. Black also has significant positive effect on housing prices. LSTAT has significant negative effect on housing prices. Percentage of lower status of the population (LSTAT) may impact the cultural environment. A less negative effect on housing prices is caused by DIS.

Table 7: Estimation ("est") and bootstrap standard deviation ("bstd") for γ and β_{-1}. Estimation results are provided by our proposed method consisting of those based on $\widetilde{\epsilon}_i^* \sim N(0,1)$ and $\widetilde{\epsilon}_i^* \sim t_{(5)}$, and by Zhang, Lian and Yu (2017) ("ZH").

		γ	β_2	β_3
Model 1	est	1.5957	-1.3603	-0.2590
	bstd	0.2103	0.4203	0.1080
Model 2	est	1.5964	-1.3620	-0.2584
	bstd	0.2095	0.4191	0.1080
ZH	est	1.5990	-1.2916	-0.2663
	bstd	0.2432	0.4446	0.1090

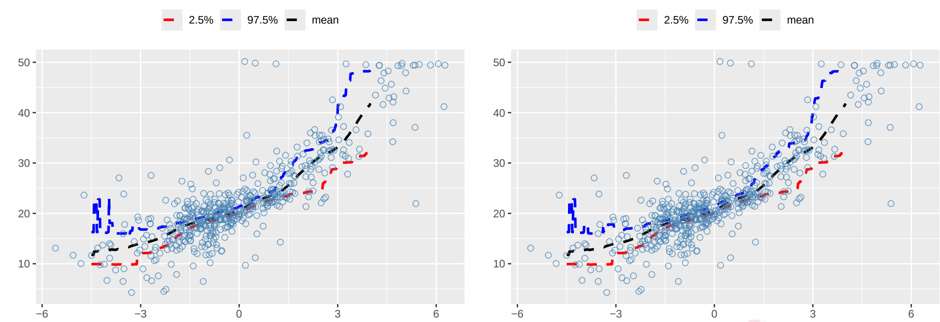

Figure 10: Plots of estimators for $g(\cdot)$ in Estimation is based on our proposed method with Model 1 and Model 2. Left panels: Model 1, right panels: Model 2. The black dashed lines are the mean curve, the red dashed lines are 2.5% quantile, and the blue dashed lines are 97.5% quantile.

6. Conclusion

Our main contribution in this work is to derive semiparametric efficiency theory for partially linear single index quantile regression, to propose a family of locally efficient estimators, and to rigorously establish the asymptotic properties of the proposed estimators. The asymptotic properties turn out to be very challenging to establish mathematically, where we engaged data splitting and high-order kernel to by pass the difficulties caused by discontinuity and unsmoothness. We also establish that our estimators show a robustness property against the misspecified error model for $f_\epsilon(\epsilon, \mathbf{x}, \mathbf{z})$. If $f_\epsilon(\epsilon, \mathbf{x}, \mathbf{z})$ is correctly specified, the asymptotic variances of our estimators achieve the optimal semiparametric efficiency bound. We illustrate that the finite sample performance of our estimators in simulations is generally better than that of the existing methods. In real data, we also find our estimators continue to have good performance.

Many further developments are possible following our work, although we foresee challenges and technical difficulties as well. For example, high-dimensional data is common in practice, and to broaden the scope of our method to encompass this case is very meaningful and important. However, although this problem might be methodologically easy, the theoretical establishment is no easy task due to the combination of the unsmoothness/discontinuity, various assumptions in terms of reducing the dimension and various machine learning methods that need to be considered depending on the size of the reduced dimension. Other possible examples of extension include partially linear multiple index quantile regression models and partially linear single index additive quantile regression models. One can of course also extend the partially linear structure to partially nonlinear framework, which could be similar or more complex than the partially linear setting depending on what are the specific nonlinear assumptions. These are all interesting research problems that could be challenging yet fruitful.

Supplementary Materials

The supplementary materials provide the proofs of Proposition 1, Theorem 1, Theorem 2 and Theorem 3 and additional figures.

Acknowledgements

Luo's work was supported by China Scholarship Council and Shanxi Scholarship Council of China(grants 2022-012).

References

Harrison, D. and Rubinfeld, D. L. (1978). Hedonic price and the demand for clean air. *Journal of Environmental Economics and Management* **5**, 81–102.

He, X. M. and Shi, P. D. (1998). Monotone B-Spline smoothing. *Journal of the American Statistical Association* **93**, 643–650.

Horowitz, J. L. and Shi, S. (2005). Nonparametric estimation of an additive quantile regression model. *Journal of the American Statistical Association* **100**, 1238–1249.

Kong, E. and Xia, Y. C. (2012). A single-index quantile regression model and its estimation. *Econometric Theory* **28**, 730–768.

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

Li, L. X., Zhu, L. P. and Zhu, L. X. (2011). Inference on primary parameter of interest with the aid of dimension reduction estimation. *Journal of the Royal Statistical Society B* **73**, 59–80.

Lv, Y. Z., Zhang, R. Q., Zhao, W. H. and Liu, J. C. (2015). Quantile regression and variable selection of partial linear single-index model. *Ann Inst Stat Math* **67**, 375–409.

Ma, S. J. and He, X. M. (2016). Inference for single-index quantile regression models with profile optimization. *Annals of Statistics* **44**, 1234–1268.

Ma, Y. Y. and Zhu, L. P. (2013a). Efficient Estimation in Sufficient Dimension Reduction. *Annals of Statistics* **41**, 250–268.

Ma, Y. Y. and Zhu, L. P. (2013b). Efficiency Loss Caused by Linearity Condition in Dimension Reduction. *Biometrika* **100**, 371–383.

Ma, Y. Y. and Zhu, L. P. (2013c). Doubly Robust and Efficient Estimators for Heteroscedastic Partially Linear Single-index Model Allowing High Dimensional Covariates. *Journal of the Royal Statistical Society, Series B* **75**, 305–322.

Ma, Y. Y. and Zhu, L. P. (2014). On Estimation Efficiency of the Central Mean Subspace. *Journal of the Royal Statistical Society, Series B.* **76**, 885–901.

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.

Sherwood, B. and Wang, L. (2016). Partially linear additive quantile regression in ultra-high dimension. *Annals of Statistics* **44**, 288–317.

Wu, W. Z., Yu, K. M. and Yu, Y. (2010). Single-index quantile regression. *Journal of Multivariate Analysis* **101**, 1607–1621.

Zhang, Y. K., Lian, H. and Yu, Y. (2017). Estimation and variable selection for quantile partially linear single-index models. *Journal of Multivariate Analysis* **162**, 215–234.

School of Mathematics and Statistics, Shanxi University, Taiyuan, P.R. China.

E-mail: (luorm955@sxu.edu.cn)

Department of Mathematics and Statistics, University of Maryland, Baltimore County, MD 21250 USA.

E-mail: (baek@umbc.edu)

Department of Statistics, Penn State University, University Park, PA 16802, USA.

E-mail: (yzm63@psu.edu)