

# Statistical Inference for Partially Nonlinear Single-index Models with Application to Correlated Medical Costs\*

Xiaobing Zhao

School of Mathematics and Statistics, Zhejiang University  
of Finance and Economics, Hangzhou, Zhejiang Province, China

Lei Liu

Department of Preventive Medicine, Northwestern University,  
Chicago, IL, U.S.A.

## Abstract

Medical costs are often to be skewed and heteroscedastic, and having a complicated relation with covariates. The mean function regression models with low-dimensional covariates have been extensively considered in literature, however, it is an important issue to find a robust alternative to find the underlying relationship between the medical costs and the high-dimensional covariates. In this paper, we propose a partially nonlinear single-index model to analyze the medical costs, the minimizing average check loss estimation with modified average quantile regression procedure is adapted to conduct quantile regression of this model, and an adaptive lasso penalized variable selection method is considered to find some significant factors of the covariates. The asymptotic properties of the above estimators are established as well. Some simulation studies and the Medical Expenditure Panel Survey dataset are carried to assess the performance of the proposed model and methods.

*Keywords:* Medical costs, high-dimensional covariates, partially nonlinear, single-index, quantile regression, minimizing average check loss estimation, variable selection, adaptive lasso.

---

\*This work was partially supported by the NSFC under Grant No.11271317 and Zhejiang Provincial Planning Projects of Philosophy and Social Science under Grant No. 12JCJJ17YB.

# 1 Introduction

Modeling medical costs is of great interest in health economics and health policy, such as in the USA, the health care costs per person reached \$US8,100, for a total of \$US2.5 trillion, which accounted for 17% of the gross domestic product in 2009 (cf. Chen *et al.*, 2014). Medical cost data are routinely collected by hospitals, disease registries, claims of health insurers, etc. The major task of the analysis of the medical costs is identified the risks of covariates associated with the corresponding to the medical costs. The analysis of the medical costs, however, complicate and pose major challenges in their statistical analysis duo to their characters including severe skewness, heteroscedasticity and non-normality. Hence it's very important to find an innovative statistical approach to analyze the medical costs.

Various statistical methods to analyze the medical costs can be found in literature, One can refer to a comprehensive overview in a special issue devoted to this topic in Medical Care (Vol. 47, No.7 supplement 1, July 2009). In literature, the approaches can be divided into two classes for a brief review: to analysis the total medical costs ( see such as Lin *et al.*, 1997, Zhao and Zhou, 2014 ), and to analyze the longitudinal (panel survey) medical costs (see such as Liu, 2009, Chen *et al.*, 2014 ).

For the analysis of longitudinal (panel survey) medical costs, it is great challenge to consider appropriately their complicated correlation of the medical costs. The first correlation is between the medical costs and the terminal events (such as death) and/or the observation processes, see such as Liu *et al.* (2008a), Zhao *et al.* (2014), Sun *et al.* (2012) and among others. The second is the longitudinal dependency of the repeated observations among the same subject. In literature, there are three main approaches (see Zeger and Liang, 1992) to account for the within-subject correlation (longitudinal dependency) of the medical costs including marginal models (see Chen, *et al.*, 2014, Chen *et al.*, 2012), transition (Markove) models (see such as Castelli *et al.*, 2007; Zhao and Zhou, 2012) and random or fixed effects models (see Liu *et al.*, 2008b, Liu, 2009).

Most recently, an natural analogue for correlated responses of marginal models, the generalised linear models (GLM) have been extensively discussed by many authors, such as Blough *et al.* (1999), Manning (2001,2005), in which the regression and within-subject correlation are modelled separately, however, the disadvantages have been found that this approach could not account conveniently for the correation of the medical costs, and the effects of the time-

varying covariates can not be enough addressed (Zeger and Liang, 1992), Most recently, Chen *et al.* (2013, 2104) extended the classic generalised linear models to another one by adding unspecified nonlinear function to account for the time-varying covariates, but still remianing the link function to be known function.

To account for the within-subject correlation to be more flexible and relax the link function to be unspecified in generalised linear model, in this paper, we proposed an partially nonlinear single-index model (PNSIM), the nonlinear compenent is used to deal with the time-varying covariate, and the single-index model is applied to deal with the mean function of the medical costs but with unspecified link function, and furthermore, single-index model is usfull model for the high-dimensional covariates. For longitudinal or panel data, a fixed effect variable is used to account for the correlation among observations within the same subject or cluster, it's different from that of Chen *et al.* (2014) through a covariate matrix of response with the same subject.

The above proposed model is a sub-model of general partially nonlinear models which have been researched by authors including Wahba (1990), Liang (1995), Gao and Liang (1997), Li and Nie (2008), Li and Mei (2013) and Song *et al.* (2010) with a pre-specified link function of single-index in their paper. However, litter has been done on the partially nonlinear single-index model with unspecified link funtion of the single-index model.

In the analysis of the medical costs, the mean cost is the mian interest in health economics studies, however, the mean function might be unrobust. Recently, some robust approach can be found in liternature, such as the median cost is proposed by Bang and Tsiatis (2002), Dominici and Zeger (2005) and Dominic *et al.* (2005) developed a smooth quantile ratio approach to estimate the mean cost, Wang and Zhou (2010) proposed a quantile regression approach to estimate the transformed cost data. In this paper, the quantile regression is applied to the partially nonlinear sing-index model, quantile regression might offer a systematic strategy for testing how covariates influence the location, scale, and shape of the entrie response distribution (Tang *et al.*, 2013), and hence the quantile regression can accommodate non-normal errors, as often seen in the analysis of medical cost data.

Single-index is a widely used dimensional reduction approach to avoid “curse of dimensionality”, which also has been used in the analysis of the medical cost data (see Zhou and Liang, 2006, Chen *et al.*, 2014), however, it's not sufficient for the covariates which may in-

clude many irrelevant components. In these cases, sparse model is often considered superior, due to the enhancements of the model predictability and interpretability (Lv *et al.*, 2014). In this paper, variable selection method will be considered to the quantile regression model of the partially nonlinear single-index model. It is challenging to perform variable selection in the proposed model with unspecified link function of the single-index and unknown nonlinear function to be estimated. There are many authors developed variable selection for (partially linear) single-index model, or quantile regression with (partially linear) single-index model, such as Liang *et al.* (2010) proposed an variable selection method for single-index model, an variable selection for the composite quantile regression of single-index model has been considered by Fan *et al.* (2013). And most recently, Lv *et al.* (2014) investigated an variable selection method for quantile regression of partially linear single-index model. However, there has been little research on (quantile regression of) partially nonlinear single-index model, let alone the single-index with unspecified link function.

In this paper, we present a comprehensive study of partially nonlinear single-index model and consider its application to the analysis of medical cost data. To the best of our knowledge, the proposed model with unspecified link function of the single-index and the statistical inference of variable selection for quantile regression of partially nonlinear single-index model appears to be the first paper in the literature.

There are three main contributions of this study. The first is to construct a partially nonlinear single-index model to analyze medical cost data which also motivated by the works of Chen *et al.* (2014). The model, which has been researched by many authors, however, is the first time to be discussed with unspecified link function of single-index, and this poses a challenge in statistical inference as it has at least two nonparametric components to be estimated. The second is that quantile regression method is applied to make statistical inference to the proposed model, this will give a systematic strategy for testing how covariates influence, and provide some robust statistical results. The third, an variable selection method is used for quantile regression of partially nonlinear single-index model, it's an extension of Xia and Härdle (2006) and Li and Liang (2008) for semiparametric setting.

We will assess the proposed model together with their statistical inference methods by some simulation results, in which various non-normal errors can be considered, and just as done in Chen *et al.* (2014), some combinations for nonparametric components of the link function

of single-index model and the nonlinear function will be checked as well. The real data of the Medical Expenditure Panel Survey (MEPS) data will be analyzed. In simulation study and real data analysis, the minimizing average check loss estimation (MACLE) with modified average quantile regression (MAQR) procedure is adapted to conduct quantile regression of this PNSIM, and an adaptive lasso penalized variable selection method is considered.

The rest of paper is organized as follows. In Section 2, we review existing models and specify our models. Section 3 we introduce the estimation method and the calculation procedure of quantile regression for partially nonlinear single-index model, the asymptotic properties of the estimators will be established as well. In Section 4, adaptive lasso penalized quantile regression method is proposed and its oracle property is presented. Some simulation results are reported in Section 5, and an example of application is illustrated in Section 6, followed by concluding remarks in Section 7.

## 2 Model Specification

The single-index model is a natural generalization of the classic linear regression models and a restrictive version of a completed nonparametric model, it is one of effective tools to overcome one major difficult so-called “curse of dimensionality” in a general nonparametric regression model, which can lead to the difficulty and inefficiency of smoothing in high-dimensional settings. The single-index model has been extensively discussed by many authors, such as Powell *et al.* (1989), Duan and Li (1991), Liang and Wang (2005), One can refer to Horowitz (1998) for the detailed discussion of model.

In the analysis of medical cost data, the single-index model also has been considered by Zhou and Liang (2006), where a single-index two-part regression models were researched to analyze skewed health care cost data, Chen *et al.* (2014) proposed a partial linear single-index to model repeated measures data of medical cost data. However, The single-index model is not sufficient to explain the variation of responses via covariates, such as in the analysis of MEPS dataset of Chen *et al.* (2014), their study found that there may exist nonlinear covariate effect for some covariates of interest, for example, age (Chen *et al.*, 2013, 2014), and hence, a partially linear model was proposed in their paper to model the mean function of the medical cost data.

For MEPS dataset, we include 2,139 individuals aged between 65 and 84 years with 1,556

households from the MEPS 2010 Full Year Consolidation Data File of Household Survey. the histogram of the medical costs of Chen *et al.* (2014) implies that the medical costs are skewed to right and possible heteroscedastic. Furthermore, medical costs of individuals in the elderly households are likely to be correlated for some reasons including such as health behaviors, attitudes and beliefs among individuals in the same household.

As mentioned in the above, the partially linear model might not be sufficient to model the medical cost dataset including skewed, heteroscedasticity, severe skewness and non-normality, since it may encounter the difficulty and inefficiency of smoothing in high-dimensional settings. In this paper, we propose an alternative to model the medical cost dataset with longitudinal observations (panel survey) as follows:

$$\begin{aligned} Y_{ij} &= g(X_{ij}^\top \beta) + f(t_{ij}) + \alpha_i + \varepsilon_{ij} \\ &= q_1(Z_1(ij)) + q_2(Z_2(ij)) + \alpha_i + \varepsilon_{ij}, \end{aligned} \quad (2.1)$$

where  $Z = (Z_1, Z_2)^\top$  with  $Z_1(ij) = X_{ij}^\top \beta$  and  $Z_2(ij) = t_{ij}$ ,  $Y_{ij}$  is the medical cost of the  $j^{th}$  observation of the subject belongs to the cluster  $i^{th}$  ( $j = 1, \dots, n_i; i = 1, \dots, n$ ),  $X_{ij} \in R^p$  and  $t_{ij} \in R^1$  to be the linear covariate and nonlinear covariates corresponding to the outcome  $Y_{ij}$ . Due to the curse of dimensionality, it is assumed through this paper that  $t_{ij}$  is a univariate continuous random variable (Li and Nie, 2008),  $\alpha_i$  is unobservable time-invariant effect distributed independently across cluster  $i^{th}$ , and  $\varepsilon_{ij}$  are the error terms, and are assumed to be independently identically distributed with mean 0 and finite variance  $\sigma^2 > 0$ , and  $(X_{ij}, t_{ij}), \varepsilon_{ij}$  are independent. And the  $g(\cdot)$  and  $f(\cdot)$  are unspecified functions to be estimated,  $\beta \in R^p$  is unknown parameter. For identifiability, we assume  $\|\beta\| = 1$  and the first nonzero element of  $\beta$  is positive, where  $\|\cdot\|$  denotes the Euclidean norm. For convenience, we call  $\beta$  index parameter,  $g(\cdot)$  index function and  $f(\cdot)$  nonlinear function. Model (2.1) is referred to be partially nonlinear single-index model (PNSIM) in this paper.

Many existing semiparametric and nonparametric models are special cases of model (2.1). For instance, without the fixed effect  $\alpha_i$ , model (2.1) reduces to partially nonlinear model, which has been discussed by Song *et al.* (2010) when  $g(\cdot)$  is assumed to be known. Provided that the index function  $g(\cdot) = 1$ , model (2.1) reduces to partially linear model of such as Hardle *et al.* (2000), and references therein, and which has been discussed recently by Chen *et al.* (2014). It also includes nonparametric model with  $g(\cdot) = 0$  (see Hardle, 1990) and single-index

model (see Horowitz, 1998) with  $f(\cdot) = 0$ . If including the fixed effect  $\alpha_i$ , model (2.1) were discussed by Lamarch (2010), Galvao *et al.* (2013) with  $f(\cdot) = 0$  and  $g(\cdot) = 1$ , Chen *et al.* (2013) and Lai *et al.* (2013) with  $f(\cdot) = 0$ ; and Ai *et al.* (2014) with  $g(\cdot) = 1$ .

The next subsection, the minimizing average check loss estimation (MACLE) and modified average quantile regression (MAQR) procedure and the adaptive lasso penalized MACLE variable selection are proposed to model (2.1). The asymptotic properties of the estimators will be established as well and the proofs can be found in the Appendix.

**Remark 1:** The fixed effect  $\alpha_i$  including in (2.1) can be explained to be individual intercepts, it is intended to capture individual specific sources of variability, or unobservable heterogeneity that was not adequately controlled by other covariates (Galvo *et al.*, 2013). As done in Lamarch (2010), we also assume that the individual effects does not represent a distributional shift, since it is unrealistic to estimate it when the number of observations on each cluster is small (Koenker, 2004). The individual effect is then a pure location shift effect  $\alpha_i$  of the response, it implies that the distribution of  $Y_{ij}$  for each cluster conditional on covariates  $X_{ij}$  and  $t_{ij}$  have the same shape, but different location as long as the  $\alpha_i$ 's are different (Lamarch, 2010).

**Remark 2:** For identifiability of model (2.1), it has been discussed by many authors. Obviously, it's required to have different conditions of identification for the different statistical inference procedures. Song *et al.* (2010) apply sieve least squares estimation to the coefficient of index and the nonparametric component by assuming that the index function  $g(\cdot)$  is a strictly monotone function for identifiability. Fan and *et al.* (1998) assume that  $E[g(X_{ij})] = E[f(t_{ij})] = 0$  in their average regression surface method, Yu and Lu (2004) and Cheng *et al.* (2011) proposed to use the condition of  $E[q_{u,\tau}(Z_u)] = 0$  (see (3.1) below) for model identification in quantile regression, which is based on the back-fitting algorithm. If  $E[g(X_{ij})] \neq 0$  and  $E[f(t_{ij})] \neq 0$ , we can rewrite  $g^*(X_{ij}) = g(X_{ij}) - E[g(X_{ij})]$  and  $f^*(t_{ij}) = f(t_{ij}) - E[f(t_{ij})]$ , this implies that  $E[g^*(X_{ij})] = E[f^*(t_{ij})] = 0$  and suppressed  $E[g(X_{ij})] + E[f(t_{ij})]$  (constant) into  $\alpha_i$ . Then, we will follow the identification of Cheng *et al.* (2011) for their modified average quantile regression. It can be seen from the Remark 1 of Yu and Lu (2004) that  $E[g(X_{ij})] = E[f(t_{ij})] = 0$  implies that  $E[q_{u,\tau}(Z_u)] = 0$ .

### 3 Quantile Regression for PNSIM

This subsection, we develop the local linear quantile regression to PNSIM. Given  $\tau \in (0, 1)$ , we define the quantile regression model as follows:

$$\begin{aligned} y_{ij} &= g_\tau(X_{ij}^\top \beta(\tau)) + f_\tau(t_{ij}) + \alpha_i(\tau) + \varepsilon_{ij}(\tau) \\ &= q_{1,\tau}(Z_1(ij)) + q_{2,\tau}(Z_2(ij)) + \alpha_i(\tau) + \varepsilon_{ij}(\tau), \end{aligned} \quad (3.1)$$

where  $\beta(\tau)$  is a  $p \times 1$  vector of parameters,  $\alpha_i(\tau)$  is a scalar individual effect for each cluster  $i$ , and  $\varepsilon_{ij}(\tau)$  is the innovation term whose  $\tau^{th}$  conditional quantile is zero given  $(X_{ij}, t_{ij}, \alpha_i)$ . The conditional  $\tau$ -quantile of  $y_{ij}$  given  $(X_{ij}, t_{ij}, \alpha_i)$  is denoted by  $Q_{y_{ij}}(\tau|X_{ij}, t_{ij}, \alpha_i) = \inf\{y : Pr(y_{ij} < y|X_{ij}, t_{ij}, \alpha_i)\}$ . The partially nonlinear single-index model assumes that the  $\tau^{th}$  conditional quantile function of  $y_{ij}$  can be expressed as

$$Q_{y_{ij}}(\tau|X_{ij}, t_{ij}, \alpha_i) = g_\tau(X_{ij}^\top \beta(\tau)) + f_\tau(t_{ij}) + \alpha_i(\tau). \quad (3.2)$$

For notational similitude, we omit  $\tau$  from such as  $\beta(\tau)$  in model (3.1) wherever clear from the context, but it is important to keep in mind that those quantities are  $\tau$ -specific. Then, the estimator of the true parameter vector  $(\beta(\tau), \alpha(\tau))^\top$  is defined as the minimizer of the following objective function:

$$\mathcal{L}(\beta(\tau), \alpha(\tau)) = \sum_{i=1}^n \sum_{j=1}^{n_i} \rho_\tau(y_{ij} - g(X_{ij}^\top \beta) - f(t_{ij}) - \alpha_i), \quad (3.3)$$

where  $\alpha = (\alpha_1, \dots, \alpha_n)^\top$ , and  $\rho_\tau(u) = |u| + (2\tau - 1)u$  is the check loss function at  $\tau \in (0, 1)$ .

To estimate the parameters and nonparametric components of model (3.3), the classic two-steps procedure can be applied, the first to estimate the nonparametric components  $g(\cdot)$ ,  $f(\cdot)$  and  $\alpha_i$  by replacing  $\beta$  with its estimators. The second is to estimate the parameters  $\beta$  with known or estimated nonparametric components  $g(\cdot)$ ,  $f(\cdot)$  and  $\alpha_i$ .

It's well known that the first step is based on a standard additive model, which has been researched by many authors, such as *backfitting algorithm* of Yu and Lu (2004), however, it is no guarantee for convergence and its iterative algorithm make it difficult to get asymptotic results (Cheng *et al.*, 2011); *average quantile* estimation of Gooijer and Zerom (2003), it's a modification of backfitting estimation by marginal integration, but it also has some shortcomings including computational inconvenience and not robustness to correlated covariates (Cheng



*et al.*, 2011). Most recently, Cheng *et al.* (2011) proposed an efficient estimation method to the additive models, which is referred to *modified average quantile regression* estimation, and have some advantages such as it also computationally more attractive and more robust for correlated covariates than the average quantile estimation.

In this paper, we will adapt modified average quantile regression estimation of Chen *et al.* (2011) to the estimators of the nonparametric components. We will begin the statistical inference of our model by introducing the modified average quantile regression estimation in the next.

### 3.1 Modified average quantile regression

Recall that  $Z = (Z_1, Z_2)^\top$  with  $Z_1(ij) = X_{ij}^\top \beta$  and  $Z_2(ij) = t_{ij}$ . Let  $Z_u$  the  $u$ th element of  $Z$  and  $W_u$  the set of all  $Z$  variables excluding  $Z_u$ , that is,  $W_u = (Z_1, \dots, Z_{u-1}, Z_{u+1}, \dots, Z_r)^\top$  and  $Z = (Z_u, W_u)$ . Assume that  $f_u(\cdot)$  and  $f_W(\cdot)$  are the density functions of variables  $X_u$  and  $W_u$ . We define

$$\Phi(z_u, w_u) = \frac{f_u(z_u)f_W(w_u)}{f(z_u, w_u)}. \quad (3.4)$$

One can see from the function of (3.4) that it has the following features

$$E[\Phi(Z_u, W_u)|Z_u = z_u] = 1, \quad \text{and} \quad E[\Phi(Z_u, W_u)q_{k,\tau}(Z_k)|Z_k = z_k] = 0 \quad \text{for} \quad k \neq u. \quad (3.5)$$

Conditional on  $Z_u = z_u$ , it implies from (3.2) and (3.5) that

$$E[\Phi(Z_u, W_u)Q_y(\tau|Z(ij), \alpha)|Z_u = z_u] \stackrel{\text{def}}{=} q_{u,\tau}^*(z_u) = \alpha_i(\tau) + q_{u,\tau}(z_u), u = 1, 2. \quad (3.6)$$

Hence, we can estimate  $q_{u,\tau}(z_u)$  by the modified average quantile estimation of Cheng *et al.* (2011) as follows

$$\hat{q}_{u,\tau}(z_u) = \hat{q}_{u,\tau}^*(z_u) - \hat{\alpha}_i(\tau) \quad (3.7)$$

provided that we can give the estimators  $\hat{q}_{u,\tau}^*(z_u)$  and  $\hat{\alpha}_i(\tau)$ , in fact, these estimators can be defined as below.

$$\hat{\alpha}_i(\tau) = \frac{1}{n_i} \sum_{j=1}^{n_i} \hat{Q}_{y_{ij}}(\tau|Z(ij), \alpha), i = 1, 2, \dots, n, \quad (3.8)$$

where  $\hat{Q}_{y_{ij}}(\tau|Z(ij), \alpha)$  is a consistent estimator of  $Q_{y_{ij}}(\tau|Z(ij), \alpha)$ , which will be given (3.11) below. It is easily known that (3.8) is an empirical version of  $\alpha_i(\tau) = E[Q_{y_{ij}}(\tau|Z(ij), \alpha)]$

under the condition of identification  $E[q_{u,\tau}(Z_u)] = 0$ . An estimator  $\hat{q}_{u,\tau}^*(z_u)$  can be obtained from Jones *et al.* (1994) as follows

$$\hat{q}_{u,\tau}^*(z_u) = \frac{1}{Nh_1} \sum_{i=1}^n \sum_{j=1}^{n_i} K\left(\frac{z_u - Z_u(ij)}{h_1}\right) \frac{\hat{f}_w(W_u(ij))}{\hat{f}(Z(ij))} \hat{Q}_{y_{ij}}(\tau|Z(ij), \alpha), \quad (3.9)$$

where  $N = \sum_{i=1}^n n_i$ ,  $K(\cdot)$  is a kernel function,  $h_1$  is a bandwidth,  $\hat{f}_w(\cdot)$  and  $\hat{f}(\cdot)$  are the kernel estimators of the corresponding density functions. Obviously, (3.9) is also an empirical version of the left hand of (3.6).

Now we can discuss the estimation of  $Q_y(\tau|z, \alpha)$  which has already used in (3.8) and (3.9) to estimate  $q_{u,\tau}^*(z_u)$  and  $\alpha_i(\tau)$ . Assume that  $Q_y(\tau|z, \alpha)$  is  $m$  times ( $m \geq 2$ ) continuously differentiable in neighbourhood of  $z \in R^2$ , this will allow us to apply the local quantile regression of Honda (2000). By Taylor expansion for multivariable function, we have

$$\begin{aligned} q(Z_1(ij), Z_2(ij)) &= q(z_1, z_2) + \left[ (Z_1(ij) - z_1) \frac{\partial}{\partial z_1} + (Z_2(ij) - z_2) \frac{\partial}{\partial z_2} \right]^{(1)} q(z_1, z_2) \\ &\quad + \left[ (Z_1(ij) - z_1) \frac{\partial}{\partial z_1} + (Z_2(ij) - z_2) \frac{\partial}{\partial z_2} \right]^{(2)} q(z_1, z_2) + \dots, \end{aligned}$$

where the  $m$ -order mixed partial derivative can be further specified as follows,

$$\begin{aligned} &\left[ (Z_1(ij) - z_1) \frac{\partial}{\partial z_1} + (Z_2(ij) - z_2) \frac{\partial}{\partial z_2} \right]^{(m)} q(z_1, z_2) \\ &= \sum_{s=0}^m C_m^s \frac{\partial^s q(z_1, z_2)}{\partial z_1^s \partial z_2^{m-s}} (Z_1(ij) - z_1)^s (Z_2(ij) - z_2)^{m-s}. \end{aligned} \quad (3.10)$$

Taking into account the observations of  $(Z_1(ij), Z_2(ij))$  in the neighbourhood  $z = (z_1, z_2)$ , we define the vector  $\hat{\theta}_z$  minimizes

$$(Nh^2)^{-1} \sum_{i=1}^n \sum_{j=1}^{n_i} \rho_\tau \left( y_{ij} - \theta_z^\top V \left( \frac{Z(ij) - z}{h} \right) \right) L \left( \frac{z - Z(ij)}{h} \right),$$

where  $\rho_\tau = \rho_\tau(u) = |u| + (2\tau - 1)u$ ,  $V \left( \frac{Z(ij) - z}{h} \right)$  and  $\theta_z$  are constructed from the elements  $h^{-m} \prod (Z_1(ij) - z_1)^s (Z_2(ij) - z_2)^{m-s}$  and  $h^{-m} \partial^s q(z_1, z_2) / \partial z_1^s \partial z_2^{m-s}$  with  $m \leq p - 1$ ,  $L(\cdot)$  is a kernel function with  $h$  to be a bandwidth. Then,  $\hat{Q}_y(\tau|z, \alpha)$  can be defined by

$$\hat{Q}_y(\tau|z, \alpha) = e_1^\top \hat{\theta}_z, \quad (3.11)$$

where  $e_1$  is a  $p$ -dimensional unit vector with the first element 1 and all other elements 0.

**Remark: 3** An estimator of  $q'_{u,\tau}(z_u)$  of the first-order partial derivative of  $q(z_1, z_2)$  can be defined as below:

$$\hat{q}'_{u,\tau}(z_u) = \frac{1}{Nh_1} \sum_{i=1}^n \sum_{j=1}^{n_i} K\left(\frac{z_u - Z_u(ij)}{h_1}\right) \frac{\hat{f}_w(W_u(ij))}{\hat{f}(Z(ij))} \hat{Q}_{y_{ij}}^{(1)}(\tau|Z(ij), \alpha), \quad (3.12)$$

where  $\hat{Q}_{y_{ij}}^{(1)}(\tau|z(ij), \alpha)$  is an estimator of the first-order partial derivative of  $q(z_1, z_2)$  with respect to  $z_u$  ( $u=1,2$ ), which is a component of an estimator of  $\hat{\theta}_z$ . Obviously, (3.12) follows from that fact that  $E\left[\Phi(Z_u, W_u)Q_y^{(1)}(\tau|Z(ij), \alpha)|Z_u = z_u\right] = q'_{u,\tau}(z_u), u = 1, 2$ .

### 3.2 Iterative algorithm for estimating parameter

We note from (3.3) that the estimator of parameter  $\beta$  can be obtained by using the classic two-step estimation method for single-index model. For unknown function  $g(\cdot)$ , the  $\tau^{th}$  conditional quantile at  $X_{ij}^\top \beta$  can be approximated linearly by

$$g(X_{ij}^\top \beta) \approx g(u) + g'(u)(X_{ij}^\top \beta - u), \quad (3.13)$$

where  $X_{ij}^\top \beta$  close to  $u$ . Given the estimators of  $g(\cdot)$ ,  $g'(\cdot)$ ,  $f(\cdot)$  and  $\alpha_i$ , following (3.3), we can employ the local linear regression technique to estimate the parameter  $\beta$  by minimizing

$$\sum_{i=1}^n \sum_{j=1}^{n_i} \rho_\tau \left( Y_{ij} - g(u) - g'(u)(X_{ij}^\top \beta - u) - f(t_{ij}) - \alpha_i \right) w_{ij0}, \quad (3.14)$$

where  $w_{ij0} = w_{h_2}(X_{ij}^\top \beta - u)$  is non-negative weights with  $\sum_{i=1}^n \sum_{j=1}^{n_i} w_{ij0} = 1$ ,  $w_{h_2}(\cdot) = w(\cdot/h_2)/h_2$  is a one-dimensional continuous symmetric kernel density function with bandwidth  $h_2$ . By averaging (3.14) over  $u_{ls} = X_{ls}^\top \beta$ , we can get an empirical approximation as below

$$\sum_{l=1}^n \sum_{s=1}^{n_i} \sum_{i=1}^n \sum_{j=1}^{n_i} \rho_\tau \left( Y_{ij} - g(X_{ls}^\top \beta) - g'(X_{ls}^\top \beta)(X_{ij}^\top \beta - X_{ls}^\top \beta) - f(t_{ij}) - \alpha_i \right) w_{ij,ls}, \quad (3.15)$$

with

$$w_{ij,ls} = \frac{w_{h_2}(X_{ij}^\top \beta - X_{ls}^\top \beta)}{\sum_{l_1=1}^n \sum_{s_1=1}^{n_i} w_{h_2}(X_{l_1 s_1}^\top \beta - X_{ls}^\top \beta)}. \quad (3.16)$$

The quantile regression estimators of  $\beta$  can be then given by

$$\arg \min_{\substack{\beta \\ \|\beta\|=1, \beta_1 > 0}} \sum_{l=1}^n \sum_{s=1}^{n_i} \sum_{i=1}^n \sum_{j=1}^{n_i} \rho_\tau \left( Y_{ij} - g(X_{ls}^\top \beta) - g'(X_{ls}^\top \beta)(X_{ij}^\top \beta - X_{ls}^\top \beta) - f(t_{ij}) - \alpha_i \right) w_{ij,ls}$$

with  $g(\cdot), g'(\cdot), f(\cdot)$  and  $\alpha_i$  replaced by their consistent estimators defined in subsection 3.1.

The above quantile regression procedure is referred to be the minimizing average check loss estimation (MACLE) in Wu *et al.* (2010), which can be regarded as parallel to the minimizing average variance estimation (MAVE) method in mean regression of partially linear single-index (PLSIM) model (Xia and Hardle, 2006). In practice, the following computational procedure can be applied to obtain the estimators by iteratively solving two simplicity problems, one for nonparametric components and the other is for parameters, respectively.

**Step 1:** Standardized the “initial value” of  $\beta^{(1)}$  such that  $\|\beta^{(1)}\| = 1$  and  $\beta^{(1)} > 0$ , that is,  $\beta^{(1)} = \text{sign}(\beta_1^{(1)})\beta^{(1)}/\|\beta^{(1)}\|$  with  $\text{sign}(\beta_1^{(1)})$  to be the first component of  $\beta^{(1)}$ . Then the estimators of  $g(X_{ls}^\top \beta^{(1)}), g'(X_{ls}^\top \beta^{(1)}), f(t_{ls})$  and  $\alpha$  can be obtained from the results of Subsection 3.1.

**Step 2:** Given their estimators of  $g(X_{ls}^\top \beta^{(1)}), g'(X_{ls}^\top \beta^{(1)}), f(t_{ls})$  and  $\alpha$ , the estimators  $\hat{\beta}$  can be obtained by

$$\arg \min_{\substack{\beta \\ \|\beta\|=1, \beta_1>0}} \sum_{l=1}^n \sum_{s=1}^{n_i} \sum_{i=1}^n \sum_{j=1}^{n_i} \rho_\tau \left( Y_{ij} - g(X_{ls}^\top \beta^{(1)}) - g'(X_{ls}^\top \beta^{(1)})(X_{ij}^\top \beta - X_{ls}^\top \beta) - f(t_{ij}) - \alpha_i \right) w_{ij,ls}$$

**Step 3:** Repeat Step 1 and Step 2 until convergence, the estimators  $\beta(\tau)$  from the final procedure are also denoted by  $\hat{\beta}$ .

The estimation of the nonparametric components in the step 1 is referred *modified average quantile estimator* in Cheng *et al.* (2011), and it also shows that it estimates the additive components can attain a one-dimensional non-parametric optimal rate regardless of the dimensional number of the covariates. However, the asymptotic variance of the estimator has the second term  $\Gamma_{2,\tau}(z_u)$  (see Theorem ?? below). We can also refine this estimator by using *oracle efficient estimator* of Cheng *et al.* (2011), of which has the same variance as it would have if the other additive components were known. We will give the oracle efficient estimator for the nonparametric components. Define

$$\hat{Q}_y^*(\tau|Z_u, \alpha) = \hat{q}_{s,\tau}^*(z_s), \quad (3.17)$$

where  $u, s = 1, 2$  with  $u \neq s$ , and  $\hat{q}_{s,\tau}^*(z_s)$  given in (3.9). Also assume that

$$Y_{ij}^* = Y_{ij} - \hat{\alpha}_i(\tau) - \hat{Q}_{y_{ij}}^*(\tau|Z_u(ij), \alpha), \quad (3.18)$$

where  $\hat{\alpha}_i(\tau)$  is defined in (3.8). We can hence then apply local linear smoothing to define the oracle efficient estimator as follows

$$\hat{q}_{u,\tau}^e(z_u) = e_1^\top \hat{\theta}_z^e \quad (3.19)$$

with  $\hat{\theta}_z^e$  to be obtained from an estimator by minizing the following

$$(Nh_e^2)^{-1} \sum_{i=1}^n \sum_{j=1}^{n_i} \rho_\tau \left( y_{ij}^* - \theta_z^{e\top} V_e \left( \frac{Z(ij) - z}{h_e} \right) \right) L_e \left( \frac{z - Z(ij)}{h_e} \right),$$

where  $V_e(t)$  is a  $p$ -dimensional vector with its  $j^{\text{th}}$  element to be  $t^{j-1}$ , and  $L_e(\cdot)$  is a kernel function with  $h_e$  be the bandwidth.

Upon getting the oracle efficient estimators of the nonparametric components using the above method, we then propose the iterative algorithm to estimating parameter  $\beta$ , but with the modified quantile estimator defined in subsection 3.1 with the oracle effeicient estimators given by (3.19).

**Remark: 4** We can consider an alternative as follows.  $Q_y(\tau|z, \alpha) = \alpha_i + q_1(z_1) + q_2(z_2)$ , by Taylor expansion of  $q_1(z)$  we have

$$Q_{y_{ij}}(\tau|Z(ij), \alpha) \approx \left( \alpha_i + q_1(z_1) + q_1'(z_1)(Z_1(ij) - z_1) + q_2(z_2) \right) w_{ij}, \quad (3.20)$$

by (3.5) we have

$$E[\Phi(Z_1, W_1)Q_y(\tau|Z, \alpha)|Z_1 = z_1] \stackrel{\text{def}}{=} q_{1,\tau}^*(z_1) = \alpha_i(\tau) + q_{1,\tau}(z_1) + q_{1,\tau}'(z_1)(E(Z_1(ij) - z_1)). \quad (3.21)$$

Then we can obtained the estimators of  $q_{1,\tau}(z_1)$  and  $q_{1,\tau}'(z_1)$  as done in subsection 3.1 provided that  $\alpha_i$  is known (since we can not assume simultaneously  $E[q_1(Z_i)] = 0$  and  $E[q_1'(Z_1)] = 0$ ). However, an estimator of  $\alpha_i$  can be also given in the first step, that is,  $(\alpha^\top, \beta)^\top$  can be obtained simultaneously through the local linear regression step. The difference between this method and the proposed in Subsection 3.2 is the parametric or nonparametric approach to estimate the fixed effect  $\alpha$ .

### 3.3 Asymptotic properties of estimators

The following will present the asymptotic properties of the estimators for the parametric and nonparametric components, respectively. Conditional on  $Z = z$ , the density function and

cumulative distribution function of  $Y$  are denoted by  $f_Y(y|z)$  and  $F_Y(y|z)$ . Assume  $f_u(z_u)$  to be the marginal density function of  $Z_u$  and  $d(y|z)$  to be the conditional density function of  $\varepsilon_\tau$  given  $Z = z$ . Defined

$$\mu_j = \int \mu^j K(u) du, \quad \nu_j = \int \mu^j K^2(u) du, \quad j = 0, 1, 2, \dots, \quad (3.22)$$

where  $K(\cdot)$  is a symmetric kernel density function. Then, we can established the following asymptotic properties for parametric part:

**Theorem 3.1** *Provided that the conditions A.1 – A.3, A.6 – A.7 and (v) of A.5 given in Appendix holds, we have*

$$\sqrt{n} \left( \hat{\beta} - \beta(\tau) \right) \xrightarrow{D} N \left( 0, \tau(1 - \tau) \mathcal{D}_1^{-1} \mathcal{D}_0 \mathcal{D}_1^{-1} \right), \quad (3.23)$$

where  $\xrightarrow{D}$  present convergence in distribution,  $\mathcal{D}_0 = E[\mathcal{D}]$ ,  $\mathcal{D}_1 = E[f_Y(q_\tau(X, t) | X^\top \beta(\tau)) \mathcal{D}]$ ,  $\mathcal{D} = [g'(X^\top \beta(\tau))]^2 \tilde{X} \tilde{X}^\top$ ,  $\tilde{X} = X - E[X | X^\top \beta(\tau)]$ .

The final estimators of the nonparametric components is obtained from step 1 of Subsection 3.2 and can be efficient if treat the  $\beta$  is known or replaced by a consistent estimator (Wu *et al.*, 2010), which also have been proved by the above Theorem. The asymptotic properties of the nonparametric parts can be given below:

**Theorem 3.2** *Suppose that  $u$  is an interior point of  $f_u(z_u)$  for  $u = 1, 2$ , under the regularity conditions A.1 – A.5 (except for (v) of A.5), we also have*

$$\sqrt{nh_1} \left\{ \hat{q}_i(z_u, h_1, \hat{\beta}) - q_i(z_u) - \frac{1}{2} q_i''(z_u) \mu_2 h_1^2 \right\} \xrightarrow{D} N(0, \Gamma_\tau(z_u)) \quad (3.24)$$

where  $\Gamma_\tau(z_u) = \Gamma_{1,\tau}(z_u) + \Gamma_{2,\tau}(z_u)$  with  $\Gamma_{1,\tau}(z_u)$  and  $\Gamma_{2,\tau}(z_u)$  are given below:

$$\Gamma_{1,\tau}(z_u) = \frac{\tau(1 - \tau) \nu_2}{f_u(z_u)} E \left( \frac{\Phi^2(Z)}{d^2(0|Z)} | Z_u = z_u \right) \quad (3.25)$$

and

$$\Gamma_{2,\tau}(z_u) = \frac{\nu_2}{f_u(z_u)} E \left( \Phi^2(Z) Q_y(\tau | Z, \alpha) | Z_u = z_u \right) \quad (3.26)$$

**Remark 5:** Bandwidth selection is always important in local reregression and likelihood methods. In theorem, the optimal bandwidth could be attained by minimizing the asymptotic mean squared error (AMSE) from Theorem 3.2 when the sample size is large, that is

$$h_1^{opt} = \left( \frac{2\Gamma_\tau^{1/2}(z_u)}{q_{u,\tau}(z_u) \mu_2} \right)^{2/5} n^{-1/5}. \quad (3.27)$$

However, the optimal bandwidth can not be gotten directly due to several unknown values, such as the index  $Z$  which may include unknown parameter  $\beta$ , the  $g(\cdot)$ ,  $f(\cdot)$  and the conditional quantile distribution function of  $Y$ . Following the argument of Yu and Jones (1998), we adapt the following rule-of-thumb bandwidth  $h_\tau$  in  $w_{ij,ls}$  of step 2 for estimating  $\beta$ :

$$h_\tau = h_m \left\{ \tau(1 - \tau) / \phi(\Phi^{-1}(\tau))^2 \right\}^{1/5}, \quad (3.28)$$

where  $\phi(\cdot)$  and  $\Phi(\cdot)$  are the density function and cumulative distribution function of the standard normality, respectively.  $h_m$  is the optimal bandwidth used in mean regression which can be obtained by the plug-in-method (see Ruppert *et al.*, 1995). The detail discussion of this approximation can be found in Yu and Jones (1998).

**Theorem 3.3** *Under the conditions of Theorem 3.2, that is, conditions A.1 – A.5 (except for (v) of  $A_5$ ), we have the asymptotic results of  $\hat{\alpha}$  defined in (3.8):*

$$\hat{\alpha}_i - \alpha_i = o_p(n^{-2/5}), \quad (3.29)$$

## 4 Variable Selection for QR of PNSIM

In this Subsection, we apply the variable selection method to the setting of quantile regression model with single-index and nonparametric components, in which the number of variables in single-index model is very large, but the number of significant variables is small relative to  $n$ . This motivates us to employ the variable selection in QR of PNSIM.

In this paper, we adopt the adaptive lasso penalized of Zou (2006). Theorem (3.1) implies that the estimator  $\hat{\beta}$  is  $\sqrt{n}$ -consistent, we can then use  $\hat{\beta}$  to construct the adaptively weighted lasso penalized objective function as follows

$$\begin{aligned} G_n(\beta) = & \sum_{l=1}^n \sum_{s=1}^{n_l} \sum_{i=1}^n \sum_{j=1}^{n_i} \rho_\tau \left( Y_{ij} - g(X_{ls}^\top \hat{\beta}) - g'(X_{ls}^\top \hat{\beta})(X_{ij}^\top \beta - X_{ls}^\top \beta) - f(t_{ij}) - \alpha_i \right) w_{ij,ls} \\ & + \lambda \sum_{j=1}^p \frac{|\beta_j|}{|\hat{\beta}_j|^2}. \end{aligned} \quad (4.30)$$

For the given tuning parameter  $\lambda$ , we can get the penalized estimators by minimizing  $G_n(\beta)$  with respect to  $\beta$  under the constraint condition of  $\|\beta\| = 1$  and the first nonzero element of  $\beta$  is positive. We denote the estimator from the above by  $\hat{\beta}^\lambda$ , which will be used to discuss

its asymptotic properties below. For selecting the tuning parameter  $\lambda$ , we follow the idea of Fan and Li (2004) and set  $\lambda = \lambda_1 \text{SE}(X^\top \hat{\beta})$  with the other tuning parameter  $\lambda_1$  to be chosen optimally by BIC criteria defined in below, where  $\text{SE}(X^\top \hat{\beta})$  of the standard errors of the unpenalized estimator of  $\beta$ . The BIC criteria of Wang and Leng (2007) is given by

$$\text{BIC}(\lambda_1) = \log P_\tau(\lambda_1) + \log(n)/n\text{DF}_{\lambda_1}, \quad (4.31)$$

where  $P_\tau(\lambda_1) = \sum_{l=1}^n \sum_{s=1}^{n_i} \sum_{i=1}^n \sum_{j=1}^{n_i} \rho_\tau \left( Y_{ij} - \hat{g}(X_{ls}^\top \hat{\beta}^\lambda) - \hat{f}(t_{ij}) - \hat{\alpha}_i \right) w_{ij,ls}$  with  $\hat{\beta}^\lambda$  to be an estimator from (4.30).  $\text{DF}_{\lambda_1}$  is the number of nonzero coefficient of  $\hat{\beta}^\lambda$ . Then, the optimal tuning parameter is defined by  $\hat{\lambda}_1 = \text{argmin}_{\lambda_1} \text{BIC}(\lambda_1)$ .

The following computational procedure can be applied to proceed the variable selection of the parameter  $\beta$  by iteratively solving two simplicity problems.

**Step 1:** Standarded the “initial value” of  $\hat{\beta}$  such that  $\|\hat{\beta}\| = 1$  and  $\hat{\beta} > 0$ , that is,  $\hat{\beta} = \text{sign}(\hat{\beta}_1) \hat{\beta} / \|\hat{\beta}\|$  with  $\text{sign}(\hat{\beta}_1)$  to be the first component of  $\hat{\beta}$ . Then the estimators of  $g(X_{ls}^\top \hat{\beta})$ ,  $g'(X_{ls}^\top \hat{\beta})$ ,  $f(t_{ls})$  and  $\alpha$  can be obtained from the resluts of Subsection 3.1.

**Step 2:** Given their estimators of  $g(X_{ls}^\top \hat{\beta})$ ,  $g'(X_{ls}^\top \hat{\beta})$ ,  $f(t_{ls})$  and  $\alpha$ , the estimators  $\hat{\beta}^\lambda$  can be obtained by the penalized likelihood function defined in (4.30).

**Step 3:** Repeat Step 1 and Step 2 untill convergence, the estimators  $\beta(\tau)$  from the final procedure are also denoted by  $\hat{\beta}^\lambda$ .

Let  $\mathcal{A} = \{j : \beta_j \neq 0\}$ , we also assume that the correct model has regression coefficient  $\beta(1, \tau) \in R^{p_0}$  to be nonzero components of  $\beta(\tau)$ , and  $\beta(2, \tau) \in R^{p-p_0}$  to be zeros. It implses that  $\mathcal{A} = \{1, 2, \dots, p_0\}$ , we define  $X^1$  in such way that they consist of the first  $p_0$  elements of  $X$ , and let  $\tilde{X}^1 = X^1 - E[X^1 | X^{1\top} \beta(1, \tau)]$ . Now we give the oracle properties of the estimator  $\hat{\beta}^\lambda$  from (4.30).

**Theorem 4.1** (*Oracle properties*) Under the reguarity A.1 – A.8 in Appendix, if  $\lambda/\sqrt{n} \rightarrow 0$ , then the adaptive lasso penalized estimator  $\hat{\beta}^\lambda$  must satisfy

1. *Consistency in selection:*  $\Pr \left( \{j : \hat{\beta}_j^\lambda \neq 0\} = \mathcal{A} \right) \rightarrow 1$ .
  2. *Asymptotic normality:*  $\sqrt{n} \left( \hat{\beta}^{1,\lambda} - \beta^{1,\lambda} \right) \xrightarrow{D} N \left( 0, \tau(1-\tau), \mathcal{D}_{1*}^{-1} \mathcal{D}_{0*} \mathcal{D}_{1*}^{-1} \right)$ .
- $\mathcal{D}_{1*} = E[f_Y(0) | X^{1\top} \beta(1, \tau) \mathcal{D}_*]$ ,  $\mathcal{D}_* = \left[ g'(X^{1\top} \beta(1, \tau)) \right]^2 (\tilde{X}^1)(\tilde{X}^1)^\top$ , and  $\mathcal{D}_{0*} = E[\mathcal{D}_*]$ .



## 5 Simulation Studies

In this section, we present two examples in simulation studies to assess the finite sample performance of the proposed model and method in this paper.

**Example 1:** In this example, there are 1000 datasets generated from it, each consisting of  $n = 200, 400$  subjects and 5 observations with each subject. We assume that the sample from the model

$$y_{ij} = \exp(X_{ij}^\top \beta) - e^{1/2} + \sin(2\pi t_{ij}) + \alpha_i + \epsilon_{ij} \quad (5.32)$$

where  $X_{ij}$  is from independent standard normal  $N(0, 1)$ , and  $\beta = (1/\sqrt{2}, 0, 1/\sqrt{2}, 0, 0, 0, 0, 0)^\top \in R^8$ ,  $t_{ij}$  is independently generated from a uniform  $[0, 1]$ . This implies that  $E[g(X^\top \beta)] = E[\exp(X^\top \beta) - e^{1/2}] = 0$  and  $E[f(t)] = E[\sin(2\pi t)] = 0$ . The fixed effect  $\alpha_i$ , is generated as  $\alpha_i = v_i + \frac{1}{N} \sum_{i=1}^n \sum_{j=1}^{n_i} (X_{ij} + t_{ij})$  with  $v_i$  from  $N(0, 1)$ . The error  $\epsilon_{ij}$  is independent from  $N(0, 1)$ . We also assume that  $(X, t)$  is independent of  $\epsilon$ . The simulation results are displayed by the following:

- A) Figure 1 shows the estimated  $g(\cdot)$  and  $f(t)$  through quantile regression with  $\tau = 0.25, 0.5$  and  $0.75$ , respectively. And also, the confidence intervals of the estimated nonparametric components are listed in Figure at the 95% per cent level. The bias and standard deviation (Std) of the estimated  $\beta$  are given in Table 1 as well.
- B) We also conducted the variable selection by adaptive lasso method for each  $\tau$ , the results over 250 simulations are summarized in Table 2. The results are reported by
  - (i) The median relative model error, which is labeled by  $MRME(\hat{\beta}) = (\hat{\beta} - \beta)^\top \Sigma(\hat{\beta} - \beta)$ , where  $\Sigma$  is the variance/covariance matrix of the regressors.
  - (ii) The average number of the true zero coefficients of  $\beta$  that are correctedly set to zero, which is denoted by  $C(\beta)$ .
  - (iii) The average number of the true nonzero coefficients of  $\beta$  that are incorrectly set to zero, which is given by  $IC(\beta)$ .
  - (iv) The proportion of trials excluding any nonzero coefficient in 250 replications, denoted by  $U - \text{fit}$ .

(v) The probability of trials selection the exact subset model (correct-fit) and the probability of trial including all the variables (over-fit) are given in Tables, they are labeled by  $C$  – fit and  $O$  – fit, respectively.

**Example 2:** In this simulation, there are 1000 datasets generated from model (5.33), each consisting of  $n = 200, 400$  subjects and 5 observations with each subject. We assume that the sample from the model

$$y_{ij} = X_{ij}^\top \beta + \sin(2\pi t_{ij}) + \alpha_i + \epsilon_{ij}, \quad (5.33)$$

we keep all the other assumptions are the same in Example 1, but the covariates  $X_{ij}$  and  $t_{ij}$  are assumed to be correlated. Let  $\tilde{W} = (X^\top t)^\top \in R^{p+1}$  with  $\tilde{W} \sim N(\mathbf{0}, \Sigma)$ ,  $\Sigma(j, j) = 1$ ,  $\Sigma(i, k) = \gamma$  if  $i \neq k$ . We consider the cases of  $\gamma = 0.2$  and  $0.8$  with low and high correlation between the covariates. Then, it's easily found that  $E[g(X^\top \beta)] = 0$  and  $E[f(t)] = 0$  with a standard normal sample  $\tilde{W}$  since  $\sin(2\pi t)$  is a odd function of  $t$ . All the questions of A) and B) are reported Figure 2 and Tables 3-4 as well.

In the simulation studies of step 1 of Subsection 3.1, we use the bandwidth values just as done in De Gooijer and Zerom (2003). We consider a varying bandwidth  $h_1$  for different covariates. For the covariate  $X$  and  $t$ , we take  $h_1 = 3s_1N^{-1/5}$  and  $h_1 = s_2N^{-1/5}$ , respectively, where  $s_1$  and  $s_2$  are the sample standard deviations of the covariates  $X$  and  $t$ . We also choose  $h_1$  and  $h_e$  are the same as  $h$  for different covariates. For estimation of  $\beta$ , we need a bandwidth  $h_2$  in the weight function  $w_{ij,ls}$ , we follow the idea of (3.28) to choose this bandwidth in the simulation.

**Simulation results:** We can see from the above simulation results of Tables 1,3 and Figure 1-2 that the estimated parameters and the nonparametric components perform well when the parameters varying with  $\tau$ . The estimated parameters are centered around the true values, and the estimated nonparametric components is close to the true curves. These estimators are also robust even if for the correlated covariates. From Tables 2,4, we can also see that the variable selection procedure can efficiently choose the true submodel. ....

**Table 1. Simulation results of the estimated parameters for Example 1**

$\tau$	Bias and Std	$\beta_1$	$\beta_2$	$\beta_3$	$\beta_4$	$\beta_5$	$\beta_6$	$\beta_7$	$\beta_8$
n=200									
0.25	Bias	0.582	0.285	0.574	0.138	0.879	0.276	0.862	0.121
	Std	0.582	0.285	0.574	0.138	0.879	0.276	0.862	0.121
0.50	Bias	0.582	0.285	0.574	0.138	0.879	0.276	0.862	0.121
	Std	0.582	0.285	0.574	0.138	0.879	0.276	0.862	0.121
0.75	Bias	0.582	0.285	0.574	0.138	0.879	0.276	0.862	0.121
	Std	0.582	0.285	0.574	0.138	0.879	0.276	0.862	0.121
n=400									
0.25	Bias	0.582	0.285	0.574	0.138	0.879	0.276	0.862	0.121
	Std	0.582	0.285	0.574	0.138	0.879	0.276	0.862	0.121
0.50	Bias	0.582	0.285	0.574	0.138	0.879	0.276	0.862	0.121
	Std	0.582	0.285	0.574	0.138	0.879	0.276	0.862	0.121
0.75	Bias	0.582	0.285	0.574	0.138	0.879	0.276	0.862	0.121
	Std	0.582	0.285	0.574	0.138	0.879	0.276	0.862	0.121

**Table 2. Variable section results for Example 1**

$\tau$	$MRME(\beta)$	$C(\beta)$	$IC(\beta)$	U-fit	O-fit	C-fit
n=200						
0.25	0.582	0.582	0.285	0.574	0.138	0.879
0.50	0.582	0.582	0.285	0.574	0.138	0.879
0.75	0.582	0.582	0.285	0.574	0.138	0.879
n=400						
0.25	0.582	0.582	0.285	0.574	0.138	0.879
0.50	0.582	0.582	0.285	0.574	0.138	0.879
0.75	0.582	0.582	0.285	0.574	0.138	0.879

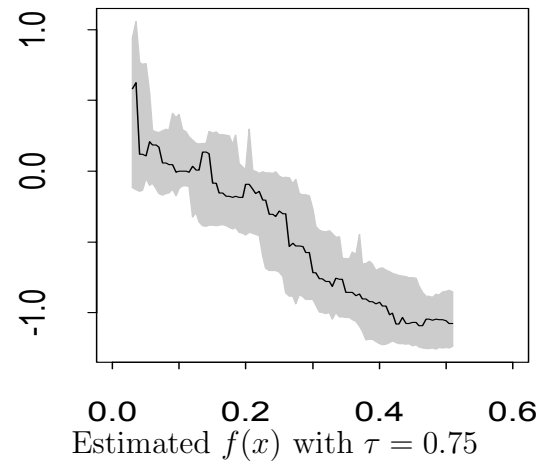
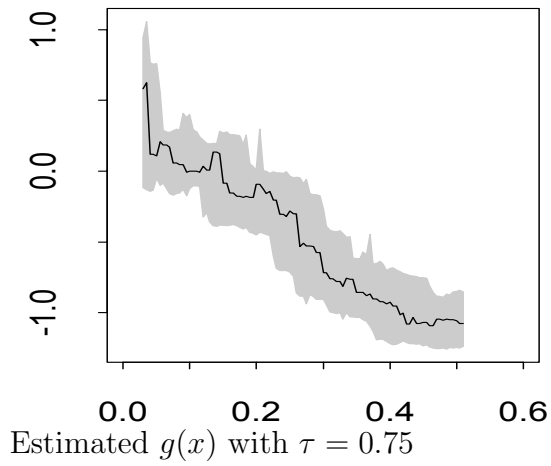
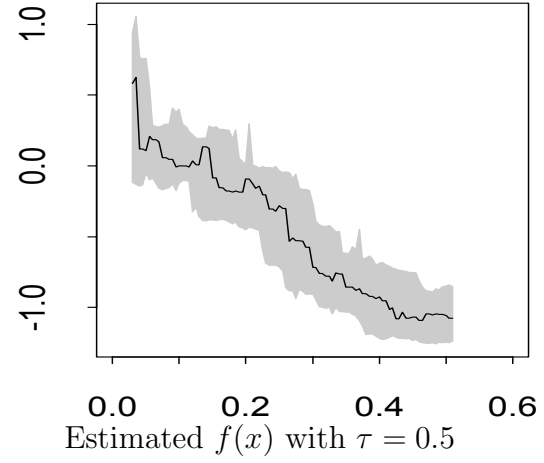
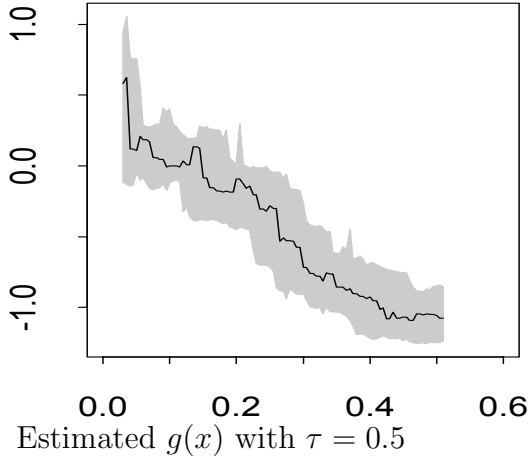
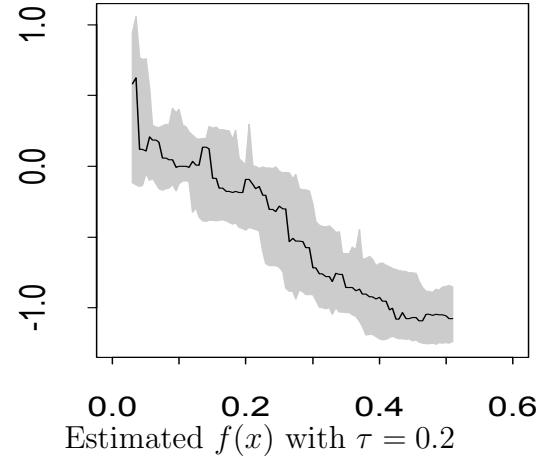
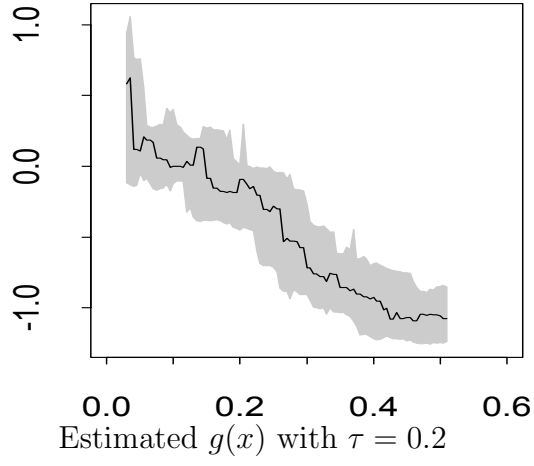
**Table 3. Simulation results of the estimated parameters for Example 2**

$\tau$	Bias and Std	$\beta_1$	$\beta_2$	$\beta_3$	$\beta_4$	$\beta_5$	$\beta_6$	$\beta_7$	$\beta_8$
n=200									
0.25	Bias	0.582	0.285	0.574	0.138	0.879	0.276	0.862	0.121
	Std	0.582	0.285	0.574	0.138	0.879	0.276	0.862	0.121
0.50	Bias	0.582	0.285	0.574	0.138	0.879	0.276	0.862	0.121
	Std	0.582	0.285	0.574	0.138	0.879	0.276	0.862	0.121
0.75	Bias	0.582	0.285	0.574	0.138	0.879	0.276	0.862	0.121
	Std	0.582	0.285	0.574	0.138	0.879	0.276	0.862	0.121
n=400									
0.25	Bias	0.582	0.285	0.574	0.138	0.879	0.276	0.862	0.121
	Std	0.582	0.285	0.574	0.138	0.879	0.276	0.862	0.121
0.50	Bias	0.582	0.285	0.574	0.138	0.879	0.276	0.862	0.121
	Std	0.582	0.285	0.574	0.138	0.879	0.276	0.862	0.121
0.75	Bias	0.582	0.285	0.574	0.138	0.879	0.276	0.862	0.121
	Std	0.582	0.285	0.574	0.138	0.879	0.276	0.862	0.121

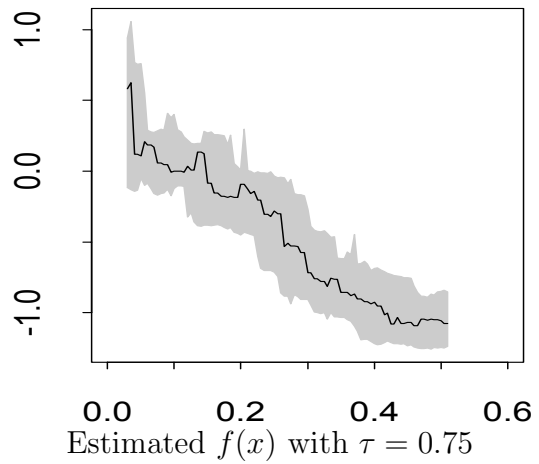
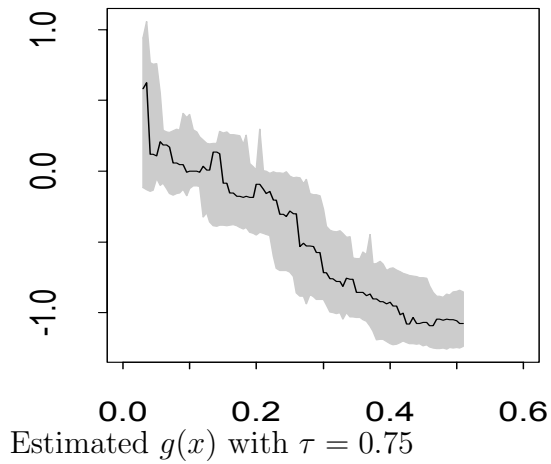
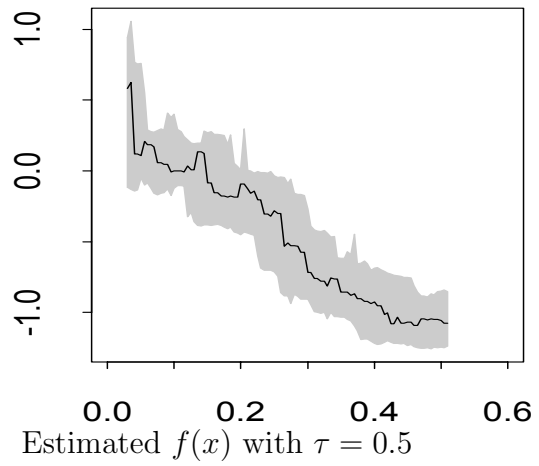
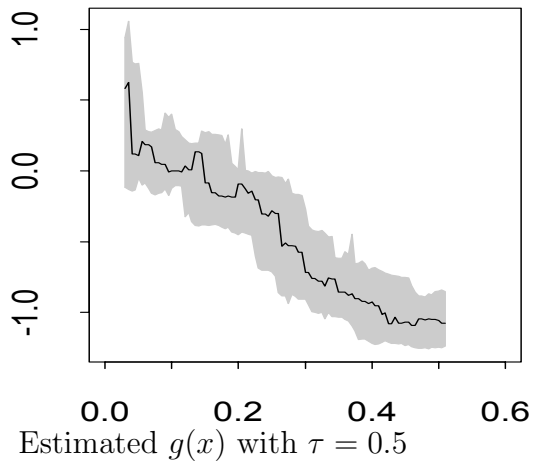
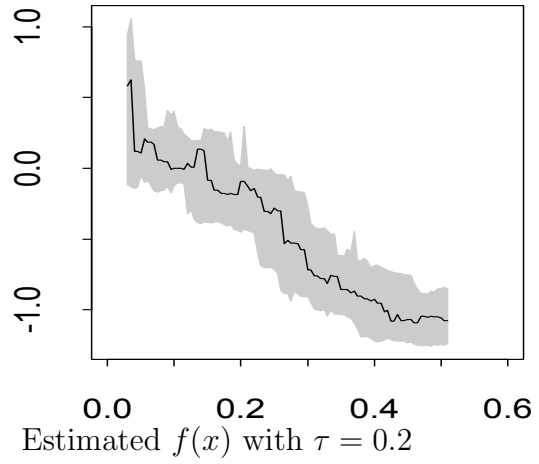
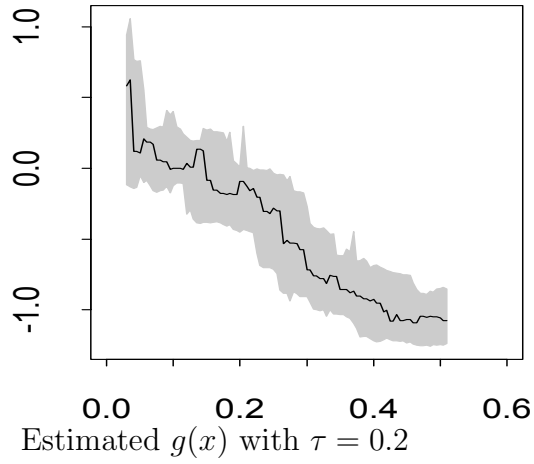
**Table 4. Variable section results for Example 2**

$\tau$	$MRME(\beta)$	$C(\beta)$	$IC(\beta)$	U-fit	O-fit	C-fit
n=200						
0.25	0.582	0.582	0.285	0.574	0.138	0.879
0.50	0.582	0.582	0.285	0.574	0.138	0.879
0.75	0.582	0.582	0.285	0.574	0.138	0.879
n=400						
0.25	0.582	0.582	0.285	0.574	0.138	0.879
0.50	0.582	0.582	0.285	0.574	0.138	0.879
0.75	0.582	0.582	0.285	0.574	0.138	0.879

**Figures 1** : Estimated nonparametric components with different  $\tau$  for Example 1.



**Figures 2 :** Estimated nonparametric components with different  $\tau$  for Example 2.



## 6 Application

In this Subsection, we will re-analyzed the data of elderly households from the MEPS 2010 Full Year Consolidated Data File of the Household Survey, the dataset include households in which every member was 65 years of age or over in the survey year. This data has been discussed by many authors including Chen *et al.* (2014) by using the partially linear marginal model with semiparametric covariance structure for correlated data. Chen *et al.* (2014) applied a subset of the full MEPS data by excluding the zero values (because very few elderly individuals would have zero medical expenditures) and endogeneity associated with choices of insurance (because Medicare eligibility starts at 65, except for those with end-stage renal disease or distributed). We also follow their idea and take this sample which includes 2139 individuals aged between 65 and 84 years within 1556 households to waive the erroneously attributing age effects of individuals over 85 to those at 85, since the age in MEPS data is truncated if the age is beyond 85.

We apply the partially nonparametric single-index model to re-analyzed this subset of MEPS data, the most advantage of this model is that it can accommodate the high-dimensional covariates through single-index function and time-varying covariates by nonparametric component. The response variable of model is annual medical costs in US dollars. We can see from the analysis of Chen *et al.* (2014) that medical costs are highly skewed to the right with an evident by a much larger mean (US \$9235) and small median (US \$3955). In the study sample, 41.1% male in the observations have hospitalization experience and have 15.0% mortality rate, 79.5% white race have hospitalization experience and have 1.4% mortality rate. For underlying disease(s) on his/her medical costs, we also consider five disease categories: heart and blood vessel disease (82.7%), respiratory disease (16.9%), body movement disorder (76.6%), cancer (28.9%), diabetes (21.9%). Then we consider the following 10 covariates: white ( $X_1$ ), male ( $X_2$ ), death ( $X_3$ ), hospitalization ( $X_4$ ), hbv disease ( $X_5$ ), respiration disease ( $X_6$ ), body movement disorder ( $X_7$ ), cancer ( $X_8$ ), diabetes ( $X_9$ ), family ( $X_{10}$ ). In this dataset, we take the response variable  $Y = 0$  if the individual have zero value of the medical cost, which has a small proportion 3.6% of subjects. The continuous covariate  $t$  in  $f(t)$  is age. In order to get a comprehensive analysis and find significant variables, in this paper we will study the dataset by using partially nonlinear single-index model to re-analyze the dataset.

The estimated  $\beta$  based on the variable selection method with different  $\tau$  are displayed in

Table 5. From Table 5, we can see that.....

The estimated nonparametric components  $g(x)$  and  $f(t)$  are given in Figure 3 for some different  $\tau$ , from Figure, we can see that.....

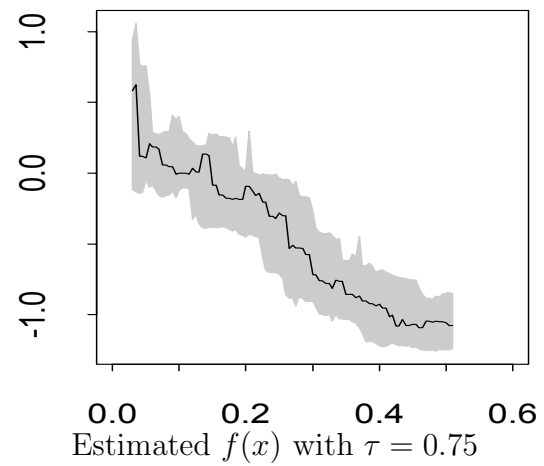
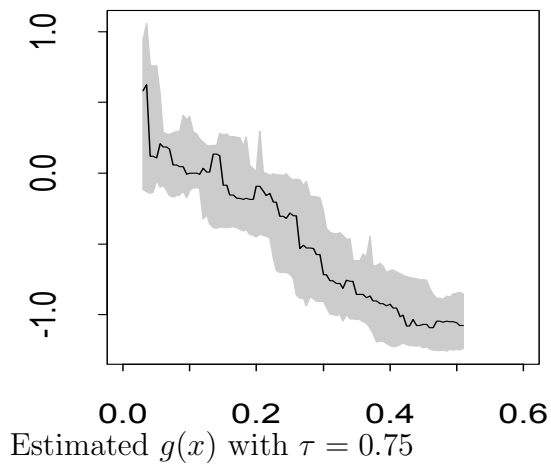
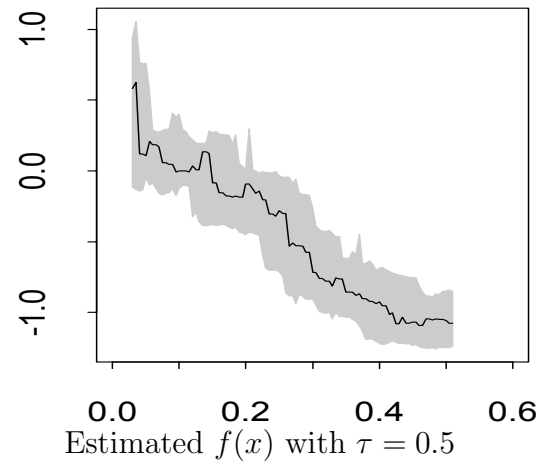
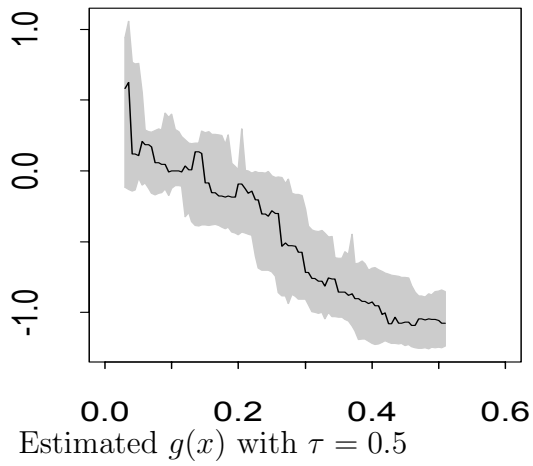
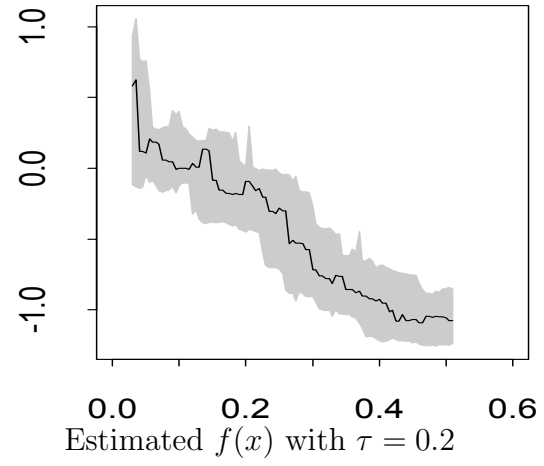
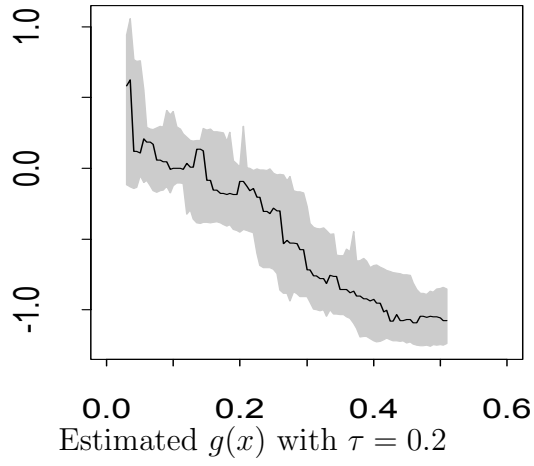
To identify the effects of each covariate  $\beta_j$  on the different amounts of the medical costs, we also give the estimated of  $\beta$  in Figure 4 with different  $\tau$ , we can see from Figure 4 that .....

**Table 5. Variable selection for Medical Expenditure Pandel Survey data**

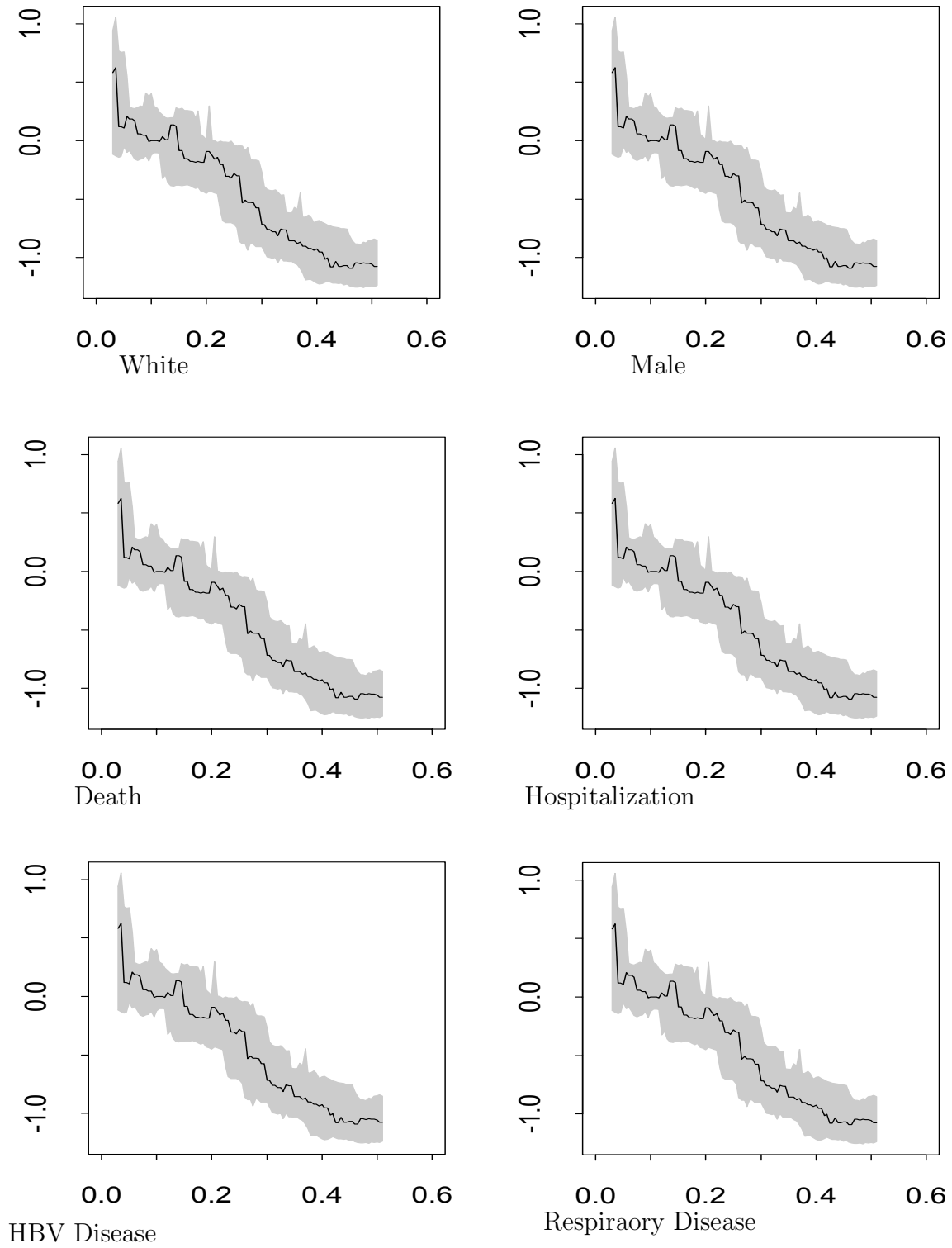
Covariates	$\tau$	0.1	0.3	0.5	0.7	0.9
White		0.582	0.582	0.285	0.574	0.138
Male		0.582	0.582	0.285	0.574	0.138
Death		0.582	0.582	0.285	0.574	0.138
Hospitalization		0.582	0.582	0.285	0.574	0.138
HBV Disease		0.582	0.582	0.285	0.574	0.138
Respiraory Disease		0.582	0.582	0.285	0.574	0.138
Boyd Movement Disorder		0.582	0.582	0.285	0.574	0.138
Canccer		0.582	0.582	0.285	0.574	0.138
Diabetes		0.582	0.582	0.285	0.574	0.138
Family		0.582	0.582	0.285	0.574	0.138



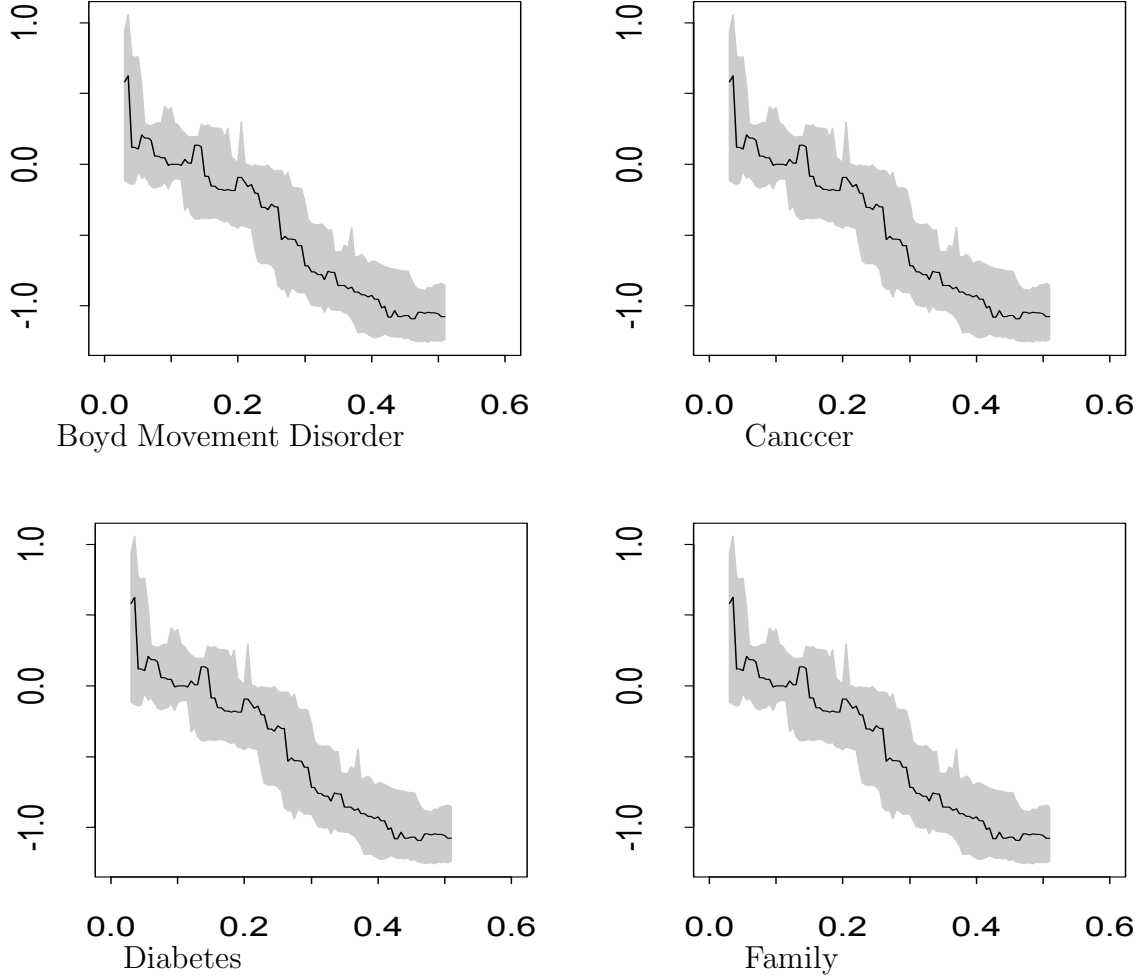
**Figures 3 :** Estimated nonparametric components for MEPS data.



**Figures 4 :** Estimated  $\{\hat{\beta}_j(\tau), j = 1, \dots, 10\}$  for MEPS data with different  $\tau$ .



**Figures 4 :** Continuous



## 7 Concluding Remarks

In this paper we proposed a partially nonlinear single-index model, which can accommodate the high-dimensional covariates in single-index function, and also have the other nonparametric component for the time-varying covariates. In order to get a comprehensive and robust estimator of the coefficients, a quantile regression is proposed to obtain their estimators of nonparametric and parametric parts, furthermore, the variable selection method is applied to identify some significant factors. In both quantile regression and variable selection procedure, the two-step estimation method is used, the first is to estimate the nonparametric components by using the modified average quantile regression based on the given single-index parameters, and

the second is to estimate the single-index parameter by using local regression method with the nonparametric components replaced by their estimators from the first step. We also established the asymptotic properties of the estimators as well.

Simulations are carried out to assess the proposed method in this paper, and the Medical Expenditure Panel Survey data are re-analyzed to demonstrate its effectiveness. The simulation results and the example of application show several advantages of the proposed method in the analysis of medical costs: (i) it has a high estimation efficiency for high-dimensional covariates through a single-index model; (ii) it's more flexible by using the other nonparametric part for time-varying covariates; (iii) it has a strong ability to detect and identify covariate effects through the variable selection procedure; and (iv) we can get a comprehensive and robust estimation for the partially nonlinear model with quantile regression.

The sufficient dimension reduction is another approach for high-dimensional covariates, it doesn't need any model assumption in dimension reduction. This is a very interesting topic to be studied in the future.

## References

- [1] Ai, C. R., You, J. H. and Zhou, Y. (2014). Estimation of fixed effects panel data partially linear additive regression models.. *Econometrics Journal*, 17, 83-106.
- [2] Bang, H. and Tsiatis, A. A. (2002). Median regression with censored cost data. *Biometrics*, 58, 643-649.
- [3] Blough, D. K., Madden, C. W. and Hornbrook, M. C. (1999). Modeling risk using generalized linear models. *Journal of Health Economics*, 18, 153-171.
- [4] Castelli, C., Combescure, C., Foucher, Y. and Daures, J.P. (2007). Cost-effectiveness analysis in colorectal cancer using a semi-Markov model. *Statistics in Medicine*, 26, 5557-5571.
- [5] Chen, J. S., Liu, L., Shih, Y-C. T., Zhang, D. W. and Severini, T.A. (2014). A flexible model for correlated medical costs with application to medical expenditure panel survey data. *Statistics in Medicine*, DOI 10. 1002/sim.0000.
- [6] Chen, J. S., Liu, L., Zhang, Daowen and Shih, Y-C. (2013). A flexible model for the mean and variable functions, with application to medical cost data. *Statistics in Medicine*, 32, 4306-4318.
- [7] Chen, J. S., Kimb, I. Terrellb, G. R. and Liu, L. (2014). Generalised partial linear single-index mixed models for repeated measures data. *Journal of Nonparametric Statistics*, DOI: 10.1080/10485252.2014.891029.

- [8] Chen, J., Gao, J.T. and Li, D.G. (2013). Estimation in partially linear single-index panel data models with fixed effects. *Journal of the American Statistical Association*, <http://dx.doi.org/10.1080/07350015.2013.775093>.
- [9] Cheng, Y. B., Gooijer, J. D. and Zerom, D. (2011). Efficient estimation of an additive quantile regression model. *Scandinavian Journal of Statistics*, 38, 46-62.
- [10] Dominici, F. and Zeger, S. L. (2005). Smooth quantile ratio estimation with regression: estimating medical expenditures for smoking-attributable disease. *Biostatistics*, 6, 505-519.
- [11] Dominici, F., Cope, L., Naiman, D. Q. and Zeger, S. L. (2005). Smooth quantile ratio estimation. *Biometrika*, 6, 505-519.
- [12] Duan, N. H. and Li, K. C. (1991). Slicing regression: A link-free regression method. *The Annals of Statistics*, 19, 505-530.
- [13] Fan, J. Q. and Li, R. Z. (2004). New estimation and model selection procedures for semiparametric modeling in longitudinal data analysis. *Journal of the American Statistical Association*, 99, 701-723.
- [14] Fan, Y., Harlde, W., Wang, W. Zhu, L. (2013). Composite quantile regression for single-index model. In *Sonderforschungsbereich 649: Okonomisches Risiko-(SFB 649 Papers)*. Humboldt-Universität zu Berlin, Wirtschaftswissenschaftliche Fakultät.
- [15] Fan, J. Q., Hardle, W. and Mammen, E. (1998). Direct estimation of low-dimensional components in additive models. *The Annals of Statistics*, 26, 943-971.
- [16] Gao, J. T and Liang, H. (1997). Statistical inference in single-index and partially nonlinear models. *Annals of the Institute of Statistical Mathematics*, 49, 493-517.
- [17] Galvao, A. F., Lamarch, C. and Lima, L. R. (2013). Estimation of censored quantile regression for panel data with fixed effects. *Journal of the American Statistical Association*, DOI:10.1080/01621459.2013.818002.
- [18] Gooijer, J. D. and Zerom, D. (2003). On additive conditional quantile with high-dimensional covariates. *Journal of the American Statistical Association*, 98, 135-146.
- [19] Horowitz, J. (1998). Lecture Notes in Statistics. *Springer; New York: 1998. Semiparametric Methods in Econometrics*.
- [20] Honda, T. (2000). Nonparametric estimation of a conditional quantile for mixing processes. *Annals of the Institute of Statistical Mathematics*, 52, 459-470.
- [21] Hardle, W., Liang, H. and Gao, J. T. (2000). Partially linear models *Physica-Verlag, Heidelberg*.
- [22] Hardle, W. (1990). Application nonparametric regression *Cambridge University Press, Boston*.
- [23] Koenker, R. (2004). Quantile analysis for longitudinal data. *Journal of Multivariate Analysis*, 91, 74-89.

- [24] Lai, P., Li, G. R. and Lian, H. (2013). Semiparametric estimation of fixed effects panel data single-index model. *Statistics and Probability Letters*, 83, 1595-1602.
- [25] Li, T. Z. and Mei, C.L. (2013). Estimation and inference for varying coefficient partially nonlinear models. *Journal of Statistical Planning and Inference*, 143, 2023-2037.
- [26] Li, R.Z. and Nie, L. (2007). A new estimation procedure for a partially nonlinear model via a mixed-effects approach. *The Canadian Journal of Statistics*, 35, 399-411.
- [27] Li, R. Z and Liu, N. (2008). Efficient Statistical inference procedures for partially nonlinear models and their applications. *Biometrics*, 64, 904-911.
- [28] Li, R. Z. and Liang, H. (2008). Variable selection in semiparametric regression modeling. *The Annals of Statistics*, 36, 261-286.
- [29] Liu, L., Huang, X.L. and O'Quigley, J. (2008a). Analysis of longitudinal data in the presence of informative observational times and a Dependent Terminal Event, with application to medical cost data. *Biometrics*, 64, 950-958.
- [30] Liu, L., Conaway, M., Knaus, W.A. and Bergin, J. D. (2008b). A random effects four-part model, with application to correlated medical costs. *Computational Statistics and Data Analysis*, 52, 4458C4473.
- [31] Liu, L. (2009). Joint modeling longitudinal semi-continuous data and survival, with application to longitudinal medical cost data. *Statistics in Medicine*, 28, 972-986.
- [32] Lin, D.Y., Feuer, E.J., Etzioni, R. and Wax, Y. (1997). Estimating medical costs from incomplete follow-up data. *Biometrics*, 53, 419-434.
- [33] Liu, L. (2009). Joint modeling longitudinal semi-continuous data and survival, with application to longitudinal medical cost data. *Statistics in Medicine*, 28, 972-986.
- [34] Liang, H., Li, R. Z., Liu, X. and Tsai, C. L. (2010). Estimation and testing for partially linear single-index models. *The Annals of Statistics*, 38, 3811-3836.
- [35] Liang, H. (1995). Second-order asymptotic efficiency of PMLE in generalized linear models. *Statistics and Probability Letters*, 24, 271-279.
- [36] Liang, H. and Wang, N. (2005). Partially linear single-index measurement error models. *Statistica Sinica*, 15, 99-116.
- [37] Lamarch, C. (2013). Robust penalized quantile regression estimation for panel data. *Journal of Economics*, 157, 396-408.
- [38] Lv, Y. Z., Zhang, R. Q., Zhao, W. H. and Liu, J. C. (2014). Quantile regression and variable selection of partial linear single-index model. *Annals of the Institute of Statistical Mathematics*, DOI 10.1007/s10463-014-0457-x.

- [39] Manning, W. G., Mullahy, J. (2001). Estimating log models: to transform or not to transform?. *Journal of Health Economics*, 20, 461C494.
- [40] Manning, W. G., Mullahy, J. (2005). Generalized modeling approaches to risk adjustment of skewed outcomes data. *Journal of Health Economics*, 24, 465C488.
- [41] Powell, J.L., Stock, J.H. and Stoker, T.M. (1989). Semiparametric estimation of index coefficient. *Econometrica*, 51, 1403-1430.
- [42] Ruppert, D. and Sheather, S. J., Wand, M. P. (1995). An effective bandwidth selector for local least squares regression. *Journal of the American Statistical Association*, **90**, 1257-1270.
- [43] Song, L. X., Zhao, Y. and Wang, X. G. (2010). Sieve least squares estimation for partially nonlinear models. *Statistics and Probability Letters*, 80, 1271-1283.
- [44] Sun, L.Q., Song, X. Y., Zhou, J. and Liu, L. (2012). Joint Analysis of longitudinal data With informative observation times and a dependent terminal event. *Journal of the American Statistical Association*, **107**, 688-700.
- [45] Tang, Y.L., Wang, H.X. and Zhu, Z.Y. (2013). Variable selection in quantile varying coefficient models with longitudinal data. *Computational Statistics and Data Analysis*, 57, 435-449.
- [46] Wahba, G. (1990). *Spline models for observational data*. SIAM, Philadelphia.
- [47] Wang, H.X. and Zhou, X.H. (2010). Estimation of the retransformed conditional mean in health care cost studies. *Biometrika*, 97, 147-158.
- [48] Wu, T.Z., Yu, K. M. and Y. Y. (2010). Single-index quantile regression. *Journal of Multivariate Analysis*, 101, 1607-1621.
- [49] Xia, Y.C. and Hardle, W. (2006). Semiparametric estimation of partially linear single-index models. *Journal of Multivariate Analysis*, 97, 1162-1184.
- [50] Xia, Y.C. and Harlde, W. (2006). Semi-parametric estimation of partially linear single-index models. *Journal of Multivariate Analysis*, 97, 1162-1184.
- [51] Yu, K. M. and Lu, Z. D. (2004). Local linear additive quantile regression. *The Scandinavian Journal of Statistics*, 31, 333-346.
- [52] Yu, K. and Jones, M. C. (1998). Local linear quantile regression. *Journal of the American Statistical Association*, **93**, 228-237.
- [53] Zeger, S. L. and Liang, K. Y. (1992). An overview of methods for the analysis of longitudinal. *Statistics in Medicine*, 11, 1825-1839.
- [54] Zhao, X.B. and Zhou, X. (2014). Estimation of copula based models for lifetime medical costs. *Annals of the Institute of Statistical Mathematics*, DOI 10. 1007/s10463-014-0477-6.
- [55] Zhao, X.Q., Deng, S. R., Liu, L. and Liu, L. (2014). Sieve estimation in semiparametric modeling of longitudinal data with informative observation times. *Biostatistics*, 15, 140-153.

- [56] Zhao, X. B. and Zhou, X. (2012). (2012). Estimation of medical costs by copula models with dynamic change of health status. *Insurance: Mathematics and Economics*, 51, 480-491.
- [57] Zhou, X. H. and Liang, H. (2006). Semiparametric single-index two-part regression models. *Computational Statistics and Data Analysis*, 50, 1378-1390.
- [58] Zou, H. (2006). The adaptive lasso and its oracle properties. *Journal of the American Statistical Association*, 101, 1418-1429.



## Appendix II: Proofs of Asymptotic Properties

We need the following regularity conditions to derive the asymptotic properties of the estimators in this paper.

**A<sub>1</sub>** : The additive function  $g(x)$  and  $f(t)$  are  $p$  times ( $p \geq 2$ ) continuously differentiable in the neighbourhood of  $x, t \in R$ . The full-dimensional conditional quantile  $Q_y(x, t, \alpha)$  is also  $p$  times continuously differentiable in the neighbourhood of  $(x, t) \in R^2$ . The density function of  $(X, t)$  is bounded from above and has  $\tilde{p}$  derivatives on their support set, where  $\tilde{p} > 2p/(p+1)$ .

**A<sub>2</sub>** : The conditional probability density of  $\varepsilon_\tau$  given  $X = x$ , say,  $d(y|x)$ , has the first continuous derivative with respect to the argument  $y$  in the neighbourhood of 0.

**A<sub>3</sub>** :  $K(\cdot)$  is a  $p$ th order kernel function and satisfies  $\int K(t_1)dt_1 = 1$ ,  $\int t_1^j K(t_1)dt_1 = 0$  for  $j = 1, \dots, p-1$  and  $\int t_1^p K(t_1)dt_1 \neq 0$ . For  $i = 1, 2$ ,  $L_i(\cdot)$  is  $\tilde{p}$ th order kernel function and satisfies  $\int L_i(s)ds = 1$ ,  $\int s^j L_i(s)ds = 0$  for  $j = 1, \dots, \tilde{p}-1$  and  $\int s^{\tilde{p}} L_i(s)ds \neq 0$ .  $L(t)$  is a second-order kernel which has bounded and continuous partial derivatives of order 1.

**A<sub>4</sub>** : (i) There exist two constant  $\delta > 2$  and  $\gamma > 0$  such that  $\delta > \gamma + 2$  and the function  $E \left\{ \left| \frac{f_w(W_u)}{f(Z)} Q_y(\tau|Z, \alpha) \right|^\delta | Z_u = z'_u \right\}$  is bounded in the neighbourhood of  $z'_u = z_u$ . (ii) The mixing coefficients  $\pi(i) = O(i^{-\theta})$  with  $\theta > \max \left\{ p + \frac{4}{p} + 6, \frac{2(p+1)\delta}{\delta-2+1} \right\}$ , and the  $\pi(i)$  is defined in Cheng *et al.* (2011).

**A<sub>5</sub>** : (i) The bandwidth satisfies that  $n^{-\gamma/4} h_1^{(2+\gamma)/\delta-1-\gamma/4} = O(1)$  and  $\limsup_n n h_1^{2p+1} < \infty$ . (ii) Assume that there exists a sequence of positive integers  $s_n$  such that  $s_n \rightarrow \infty$ ,  $s_n = o((n h_1)^{1/2})$  and  $(n h_1)^{1/2} \pi(s_n) \rightarrow 0$  as  $n \rightarrow \infty$ . (iii) For some sufficiently small constant  $\epsilon > 0$ ,  $h = C n^{-k}$  with constant  $k$  satisfying  $1/(2p+1) < k < \{2p+3 - (2p+1)(1/\theta - 2\epsilon)\} / \{6(2p+1)(1+2/3\theta - 1/3\theta^2)\}$  and  $h/h_1 \rightarrow 0$ . (iv)  $h_1^{\theta(1-2/\delta)} h^{2/\delta-2} \rightarrow 0$ ,  $n h^2 (h_1 h^2)^{3/\theta+\epsilon} \rightarrow \infty$ . (v) The bandwidth  $h_2$  satisfies  $h_2 \sim n^{-\tilde{\delta}}$  with  $1/6 < \tilde{\delta} < 1/4$ .

**A<sub>6</sub>** : The conditional expectations  $E[X|X^\top \beta = u]$ ,  $E[XX^\top f_Y(q_\tau(X, t)|X^\top \beta)|X^\top \beta = u]$  are twice differentiable with respect to  $u$ .

**A<sub>7</sub>** :  $E[g'(X^\top \beta(\tau))^2 [X - E(X)][X - E(X)]^\top]$  and  $E[g'(X^\top \beta(\tau))^2 \tilde{X} \tilde{X}^\top]$  are positive-definite matrix's, where  $g'(\cdot)$  is the first derivative of  $g(\cdot)$ .

The conditions **A<sub>1</sub>** to **A<sub>5</sub>** are used to show the asymptotic results of Theorems 3.2, 3.3,

the condition **C6** of Cheng *et al.* (2011) has been included in condition **A<sub>1</sub>** for the our setting with these two additive components. We also need the conditions to show Theorem 3.1, which is similar that of Lv *et al.* (2014), however, their conditions have also been included into conditions of **A<sub>1</sub>** to **A<sub>3</sub>**, **A<sub>6</sub>**, **A<sub>7</sub>** and condition (v) of **A<sub>5</sub>** of the above.

**Proof of Theorem 3.1:** *Proof.* The proof of this theorem is similar to that of Lv *et al.* (2014), but with the asymptotic presentations of  $\hat{g}(X_{ls}^\top \tilde{\beta})$ ,  $\hat{g}'(X_{ls}^\top \tilde{\beta})$ ,  $\hat{f}(t_{ls})$  and  $\hat{\alpha}_i$  from the modified average quantile regression of Cheng *et al.* (2011) in Subsection 3.1 rather than from the minimizing average check loss estimation of Lv *et al.* (2014), where  $\tilde{\beta}$  is estimator of  $\beta$  in iterative algorithm. And hence then, we only outline the proof of Theorem below.

From the second step we can know that  $\hat{\beta}$  by minimizing the following

$$\arg \min_{\substack{\beta \\ \|\beta\|=1, \beta_1 > 0}} \sum_{l=1}^n \sum_{s=1}^{n_i} \sum_{i=1}^n \sum_{j=1}^{n_i} \rho_\tau \left( Y_{ij} - \hat{g}(X_{ls}^\top \tilde{\beta}) - \hat{g}'(X_{ls}^\top \tilde{\beta}) X_{ij,ls}^\top \beta - \hat{f}(t_{ij}) - \hat{\alpha}_i \right) w_{ij,ls},$$

where  $X_{ij,ls} = X_{ij} - X_{ls}$ , the estimators  $\hat{g}(X_{ls}^\top \tilde{\beta})$ ,  $\hat{g}'(X_{ls}^\top \tilde{\beta})$ ,  $\hat{f}(t_{ls})$  and  $\hat{\alpha}_i$  are given in Subsection 3.1. Let  $\hat{\gamma}^* = \sqrt{n}(\hat{\beta} - \beta)$ ,  $M_{ij,ls} = \hat{g}'(X_{ls}^\top \tilde{\beta}) X_{ij,ls}$  and

$$r_{ij,ls} = - \left( g_\tau(X_{ls}^\top \beta(\tau)) + f_\tau(t_{ls}) + \alpha_i(\tau) \right) + \left( \hat{g}(X_{ls}^\top \tilde{\beta}) + \hat{g}'(X_{ls}^\top \tilde{\beta}) X_{ij,ls}^\top \beta(\tau) + \hat{f}(t_{ij}) + \hat{\alpha}_i \right),$$

then  $\hat{\gamma}^*$  minimizes

$$\aleph_n = \sum_{l=1}^n \sum_{s=1}^{n_i} \sum_{i=1}^n \sum_{j=1}^{n_i} w_{ij,ls} \left[ \rho_\tau(\varepsilon_i - r_{ij,ls} - M_{ij,ls} \gamma^* / \sqrt{n}) - \rho_\tau(\varepsilon_i - r_{ij,ls}) \right].$$

Then,  $\aleph_n$  can be written as follows by using Kinght's identify

$$\begin{aligned} \aleph_n &= -\frac{1}{\sqrt{N}} \sum_{l=1}^n \sum_{s=1}^{n_i} \sum_{i=1}^n \sum_{j=1}^{n_i} w_{ij,ls} \psi_\tau(\varepsilon_i) M_{ij,ls} \gamma^* \\ &\quad + \sum_{l=1}^n \sum_{s=1}^{n_i} \sum_{i=1}^n \sum_{j=1}^{n_i} w_{ij,ls} \int_{r_{ij}}^{r_{ij} + M_{ij,ls} \gamma^* / \sqrt{n}} [I(\varepsilon_i \leq s) - I(\varepsilon_i \leq 0)] ds \\ &= \aleph_{1n}(\gamma^*) + \aleph_{2n}(\gamma^*), \end{aligned}$$

where

$$\begin{aligned}\aleph_{1n}(\gamma^*) &= -\frac{1}{\sqrt{N}} \sum_{l=1}^n \sum_{s=1}^{n_i} \sum_{i=1}^n \sum_{j=1}^{n_i} w_{ij,ls} \psi_\tau(\varepsilon_i) M_{ij,ls} \gamma^*, \\ \aleph_{2n}(\gamma^*) &= \sum_{l=1}^n \sum_{s=1}^{n_i} \sum_{i=1}^n \sum_{j=1}^{n_i} w_{ij,ls} \int_{r_{ij}}^{r_{ij} + M_{ij,ls} \gamma^* / \sqrt{n}} [I(\varepsilon_i \leq s) - I(\varepsilon_i \leq 0)] ds.\end{aligned}$$

Proceeding along the same way of Lv *et al.* (2014) we have the following presentation by taking the expectation of  $\aleph_{2n}(\gamma^*)$ :

$$\aleph_{2n}(\gamma^*) = \aleph_{2n1}(\gamma^*) + \aleph_{2n2}(\gamma^*) + o_p(1),$$

where

$$\begin{aligned}\aleph_{2n1}(\gamma^*) &= \frac{1}{2N} (\gamma^*)^2 \left( \sum_{l=1}^n \sum_{s=1}^{n_i} \sum_{i=1}^n \sum_{j=1}^{n_i} w_{ij,ls} f_Y(q_\tau(X, t) | X_{ij}^\top \tilde{\beta}) M_{ij,ls}^2 \right), \\ \aleph_{2n2}(\gamma^*) &= \frac{1}{\sqrt{N}} \gamma^* \left( \sum_{l=1}^n \sum_{s=1}^{n_i} \sum_{i=1}^n \sum_{j=1}^{n_i} w_{ij,ls} f_Y(q_\tau(X, t) | X_{ij}^\top \tilde{\beta}) M_{ij,ls} r_{ij,ls} \right).\end{aligned}$$

By Lemma 3 of Lv *et al.* (2014), we have

$$\aleph_{2n1}(\gamma^*) = (\gamma^*)^2 \mathcal{W}_{2n} + o_p(1), \quad (7.1)$$

where  $\mathcal{W}_{2n} = E[f_Y(q_\tau(X, t) | X^\top \beta(\tau)) \{g'(X^\top \beta(\tau))\}^2 \times (X - \mu_{\beta(\tau)}(X))(X - \mu_{\beta(\tau)}(X))^\top]$  with  $\mu_\beta(x) = E[X | X^\top \beta = x^\top \beta]$ . We also note that

$$\begin{aligned}r_{ij,ls} &= g(X_{ij}^\top \tilde{\beta}) - g(X_{ij}^\top \tilde{\beta}(\tau)) - \left( g(X_{ij}^\top \tilde{\beta}) - g(X_{ls}^\top \tilde{\beta}) - g'(X_{ls}^\top \tilde{\beta}) X_{ij,ls}^\top \tilde{\beta} \right) \\ &\quad + \left[ \hat{g}(X_{ls}^\top \tilde{\beta}) - g(X_{ls}^\top \tilde{\beta}) \right] + \left[ \hat{g}'(X_{ls}^\top \tilde{\beta}) - g'(X_{ls}^\top \tilde{\beta}) X_{ij,ls}^\top \right] \tilde{\beta} - \hat{g}'(X_{ls}^\top \tilde{\beta}) X_{ij,ls}^\top (\tilde{\beta} - \beta(\tau)) \\ &\quad + (\hat{f}(t_{ij}) - f(t_{ij})) + ((\hat{\alpha}_i - \alpha_i(\tau))) \\ &= \left( 1 \quad X_{ij,ls}^\top \tilde{\beta} / h \right) \left( \begin{array}{c} \hat{g}(X_{ls}^\top \tilde{\beta}) - g(X_{ls}^\top \tilde{\beta}) \\ h \left( \hat{g}'(X_{ls}^\top \tilde{\beta}) - g'(X_{ls}^\top \tilde{\beta}) \right) \end{array} \right) + (\hat{f}(t_{ij}) - f(t_{ij})) + ((\hat{\alpha}_i - \alpha_i(\tau))) \\ &\quad + g'(X_{ij}^\top \tilde{\beta}) X_{ij}^\top \beta_d - \hat{g}'(X_{ls}^\top \tilde{\beta}) X_{ij,ls}^\top \beta_d - \frac{1}{2} g''(X_{ls}^\top \tilde{\beta}) (X_{ij,ls}^\top \tilde{\beta})^2 + O \left[ \delta_\beta + (X_{ij,ls}^\top \tilde{\beta})^3 \right],\end{aligned}$$

where  $\beta_d = \tilde{\beta} - \beta(\tau)$  and  $\delta_\beta = |\tilde{\beta} - \beta(\tau)|$ . Then  $\aleph_{2n2}(\gamma^*)$  can be represented as

$$\aleph_{2n2}(\gamma^*) = (\aleph_{2n21}(\gamma^*) + \aleph_{2n22}(\gamma^*)) \gamma^* + O(\delta_\beta^2 + h^3),$$

where

$$\begin{aligned} \aleph_{2n21}(\gamma^*) &= \frac{1}{\sqrt{N}} \sum_{l=1}^n \sum_{s=1}^{n_i} \sum_{i=1}^n \sum_{j=1}^{n_i} f_Y(q_\tau(X, t) | X_{ij}^\top \tilde{\beta}) w_{ij,ls} M_{ij,ls} \\ &\times \begin{pmatrix} 1 & X_{ij,ls}^\top \tilde{\beta}/h \end{pmatrix} \begin{pmatrix} \hat{g}(X_{ls}^\top \tilde{\beta}) - g(X_{ls}^\top \tilde{\beta}) \\ h \left( \hat{g}'(X_{ls}^\top \tilde{\beta}) - g'(X_{ls}^\top \tilde{\beta}) \right) \end{pmatrix}, \end{aligned}$$

$$\begin{aligned} \aleph_{2n22}(\gamma^*) &= \frac{1}{\sqrt{N}} \sum_{l=1}^n \sum_{s=1}^{n_i} \sum_{i=1}^n \sum_{j=1}^{n_i} f_Y(q_\tau(X, t) | X_{ij}^\top \tilde{\beta}) w_{ij,ls} M_{ij,ls} \\ &\times \left( g'(X_{ij}^\top \tilde{\beta}) X_{ij}^\top \beta_d - \hat{g}'(X_{ls}^\top \tilde{\beta}) X_{ij,ls}^\top \beta_d - \frac{1}{2} g''(X_{ls}^\top \tilde{\beta}) (X_{ij,ls}^\top \tilde{\beta})^2 (\hat{f}(t_{ij}) - f(t_{ij})) + ((\hat{\alpha}_i - \alpha_i(\tau)) \right). \end{aligned}$$

By the asymptotic results of Theorems 3.2 and 3.3, via direct calaulating we further have

$$\aleph_{2n21}(\gamma^*) = T_1 + T_2 + o_p(1),$$

where

$$\begin{aligned} T_1 &= \frac{1}{\sqrt{N}} \sum_{l=1}^n \sum_{s=1}^{n_i} \sum_{i=1}^n \sum_{j=1}^{n_i} \psi_\tau(\varepsilon_i) w_{ij,ls} \left( \hat{g}'(X_{ls}^\top \tilde{\beta}) (\mu_{\beta(\tau)}(X_{ls}) - X_{ls}) \right), \\ T_2 &= \frac{1}{\sqrt{N}} \sum_{l=1}^n \sum_{s=1}^{n_i} \sum_{i=1}^n \sum_{j=1}^{n_i} f_Y(q_\tau(X, t) | X_{ij}^\top \tilde{\beta}) w_{ij,ls} M_{ij,ls} \times \left( \frac{1}{2} g''(x^\top \tilde{\beta}) \mu_2 h^2 - g'(x^\top \tilde{\beta}) \mu_{\beta(\tau)}(x) \beta_d \right). \end{aligned}$$

By incorporating  $T_1$  with  $\aleph_{1n}(\gamma^*)$ ,  $T_2$  with  $\aleph_{2n22}(\gamma^*)$ , respectively, we have

$$\aleph_{1n}(\gamma^*) + T_1 \gamma^* = -\sqrt{n} \mathcal{W}_{1n} \gamma^* + o_p(1), \quad (7.2)$$

$$\aleph_{2n22}(\gamma^*) + T_2 = -\sqrt{n} \mathcal{W}_{2n} \beta_d + o_p(1), \quad (7.3)$$

where  $\mathcal{W}_{1n} = \frac{1}{\sqrt{N}} \sum_{l=1}^n \sum_{s=1}^{n_i} \sum_{i=1}^n \sum_{j=1}^{n_i} \psi_\tau(\varepsilon_{ij}) w_{ij,ls} \left( \hat{g}'(X_{ls}^\top \tilde{\beta}) [X_{ij} - \mu_{\beta(\tau)}(X_{ls})] \right)$  and  $\mathcal{W}_{2n}$  is defined in (7.1). Combinating (7.1), (7.2) and (7.3), we have

$$\aleph_n(\gamma^*) = (\gamma^*)^2 \mathcal{W}_{2n} + [\mathcal{W}_{1n} + \sqrt{n} \mathcal{W}_{2n} \beta_d] \gamma^* + o_p(1).$$

Following Lemma 2 of Lv *et al.* (2014),  $\hat{\gamma}^*$ , which minimizes  $\aleph_n(\gamma^*)$ , can be expressed as

$$\hat{\gamma}^* = \mathcal{W}_{2n}^{-1} \mathcal{W}_{1n} + \sqrt{N}(\tilde{\beta} - \beta(\tau)) + o_p(1), \quad (7.4)$$

it also implies that

$$\hat{\beta} - \beta(\tau) = \mathcal{W}_{2n}^{-1} \mathcal{W}_{1n} + (\tilde{\beta} - \beta(\tau)) + o_p(1/\sqrt{N}), \quad (7.5)$$

By the Cramer-Wald device and CLT, using similar analysis as Xia and Härdle (2006), we can complete the proof. ■

**Proof of Theorem 3.2:**

*Proof.* The proof this Theorem is similar to that of Cheng *et al.* (2011), we can see the difference between  $\hat{q}_{u,\tau}^*(z_u)$  defined in (3.9) and the estimator of Cheng *et al.* (2011) is only that the covariate  $Z_u(ij) = X_{ij}^\top \beta$  is replaced by its estimator  $\hat{Z}_u(ij) = X_{ij}^\top \hat{\beta}$ , which also depends on the kernel function  $K((z_u - Z_u(ij))/h_1), \hat{f}_w(W_u(ij)), \hat{f}(Z(ij))$  and  $\hat{Q}_{y_{ij}}(\tau|Z(ij), \alpha)$ , and where  $\hat{Q}_{y_{ij}}(\tau|Z(ij), \alpha)$  is related to the kernel functions  $V((Z(ij) - z)/h)$  and  $L((z - Z(ij))/h)$  and By Taylor expansion for multivariable function of  $q(Z_1(ij), Z_2(ij))$ . Then, the proof of Theorem 3.2 can be completed by the results of Cheng *et al.* (2011) and Taylor expansion for kernel function, such as  $K_{h_1}(z_u - \hat{Z}_u(ij))$ :

$$K_{h_1}(z_u - \hat{Z}_u(ij)) = K_{h_1}(z_u - Z_u(ij)) + K'_{h_1}(z_u - Z_u(ij)) X_{ij}^\top (\hat{\beta} - \beta) + o_p(1) \quad (7.6)$$

**Proof of Theorem 3.3:**

*Proof.* The proof is similar to that of Theorem 3.2. ■

**Proof of Theorem 4.1:**

*Proof.* This can be completed directly from the proof of Theorem 3.1 and the results of Theorem 3 of Lv *et al.* (2014). ■