Statistica Sinica Preprint No: SS-2023-0092	
Title	Optimal Conditional Quantile Prediction via Model Averaging of Partially Linear Additive Models
Manuscript ID	SS-2023-0092
URL	http://www.stat.sinica.edu.tw/statistica/
DOI	10.5705/ss.202023.0092
Complete List of Authors	Jing Lv and Shujie Ma
Corresponding Authors	Shujie Ma
E-mails	shujie.ma@ucr.edu

Statistica Sinica: Preprint
doi:10.5705/ss.202023.0092

Optimal Conditional Quantile Prediction via Model Averaging of Partially Linear Additive Models

Jing Lv[1] and Shujie Ma[2]

[1]*School of Mathematics and Statistics,*

Southwest University, Chongqing, China

[2]*Department of Statistics,*

University of California, Riverside, USA

Abstract: Partially linear additive models (PLAMs) have been considered one of the most popular semiparametric models for prediction, as they enjoy model flexibility and interpretability. However, choosing the linear and nonlinear parts in PLAMs is always a challenging task. In the literature, there are a few studies that propose choosing the linear part by using a regularization method. As a result, they can identify a single optimal PLAM. We propose a novel strategy based on model averaging to obtain an optimal weighted combination of a series of partially linear additive candidate models. Our approach provides a new perspective on accounting for the structure uncertainty of PLAMs. It improves prediction accuracy compared to the estimation method based on each single PLAM, and reduces the risk of model mis-specification. Moreover, we consider a conditional quantile process setting that provides a more comprehensive analysis of the relationships between the response and covariates as well as a more robust prediction. Theoretically, we show that the proposed method of choosing the weights is asymptotically optimal in terms of minimizing the out-of-sample quantile prediction error by allowing misspecification of

Statistica Sinica: Preprint
doi:10.5705/ss.202023.0092

each candidate model. The numerical results demonstrate that our method yields smaller prediction errors than the conventional regularization methods of selecting a single PLAM.

Key words and phrases: Asymptotic optimality, B-splines, Model averaging, Partially linear additive models, Quantile prediction error.

1. Introduction

Partially linear additive models (PLAMs) have been considered as one of the most popular semiparametric models for prediction, as they enjoy model flexibility as well as interpretability, see, for example, Wang et al. (2011); Ma (2012); Ma et al. (2013); Wang et al. (2014); Wong et al. (2019). In recent years, estimation in PLAMs has also been investigated in quantile regression which can provide a more robust prediction as well as a more comprehensive analysis of the relationship between the response and covariates, see Lian (2012); Sherwood and Wang (2016), and the reference therein. As far as we know, it is usually a challenging task to choose between the linear and nonlinear parts for each covariate when one fits a PLAM, which is called the *model structure uncertainty*. In practice, people often deal with the model structure uncertainty according to their empirical experiences. For example, a commonly adopted strategy is to put continuous covariates in the nonparametric part and discrete predictors in the linear part of a PLAM. Another way is to choose a few covariates which may have nonlinear effects on the response based on prior knowledge, and place them in the nonparametric additive part.

Statistica Sinica: Preprint
doi:10.5705/ss.202023.0092

However, one rarely knows a *priori* which covariates have linear effects and which ones have nonlinear effects. Over the past several years, various regularization methods have been applied to automatically identify linear and nonlinear components, including Zhang et al. (2011); Lian (2012); Huang et al. (2012); Lian et al. (2015); Lou et al. (2016), and so forth. These methods are proposed for model selection, but they are not for optimal prediction. When all candidate models are not correctly specified, using a single model without considering the information from others can result in poor predictive performance.

Model averaging, as a well-known ensemble technique, takes all potential models into account and assigns a weight to each model. It can markedly reduce the risk of model misspecification, and generally results in more accurate predictions than using a single model. Over the last two decades, we have witnessed a booming development of parametric and semiparametric model averaging methods; see Wan et al. (2010); Hansen and Racine (2012); Li et al. (2018); Zhang and Wang (2019); Zhang et al. (2020); Seng and Li (2022); Li et al. (2021, 2022), and the references therein. However, these aforementioned research findings mainly focus on conditional mean prediction. Under the framework of quantile regression, Lu and Su (2015) proposed a jackknife model averaging of a series of linear models to handle *the uncertainty of variable selection*. Theoretically, they demonstrated its asymptotic optimality in terms of minimizing the out-of-sample prediction error. Wang et al. (2023) investigated the jackknife model averaging for linear quantile

Statistica Sinica: Preprint
doi:10.5705/ss.202023.0092

regression with high-dimensional covariates. Unlike Lu and Su (2015) and Wang et al. (2023), our goal is to address the PLAM model uncertainty problem and to achieve an optimal conditional quantile prediction when each candidate model may be misspecified.

In this paper, we propose a novel strategy for optimal quantile prediction in PLAMs using both model averaging and quantile regression coefficient modeling (QRCM). Specifically, we obtain an optimal weighted conditional quantile estimate from a series of partially linear additive sub-models, where quantile regression coefficients of each sub-model are modeled as smooth functions of the quantile levels and estimated by minimizing an integrated loss, and the model weights are chosen by using an integrated loss with a penalty term. Our approach provides a new perspective of accounting for the structure uncertainty of PLAMs via model averaging, and it improves prediction accuracy compared to the regularization-based method which selects an optimal single PLAM.

Our contributions are three-fold. First, instead of selecting an optimal model, we average a series of semiparametric partially linear additive sub-models, each with different nonlinear components, and the model weights automatically adjust the relative importance of these sub-models. Compared to the popular regularization methods for model selection such as Lian (2012), our model averaging approach provides a different way to circumvent the problem of artificially setting linear and nonlinear index sets. Furthermore, our approach leads to a more

Statistica Sinica: Preprint
doi:10.5705/ss.202023.0092

accurate prediction than a single selected optimal PLAM by regularization, as it combines useful information from different semiparametric candidate models. Second, we consider the regression coefficients as an unknown smooth function of the quantile levels, called QRCM, so that we can simultaneously estimate the entire conditional quantile process of the response rather than at a given individual quantile level. The QRCM can provide a full range of quantile analyses, leading to more appropriate and comprehensive findings, and it can also increase statistical efficiency; see Frumento and Bottai (2016) for applications and discussions of QRCM. However, previous works on QRCM (Frumento and Bottai, 2016; Frumento et al., 2021) use a parametric model to model the quantile levels, while we consider the quantile process as an unknown smooth function approximated by B-splines for more flexibility. Third, we show that the proposed model averaging estimator enjoys the asymptotic optimality (Lu and Su, 2015), in the sense that its out-of-sample average quantile prediction error is asymptotically identical to that of the best but infeasible model averaging estimator. This theoretical property ensures that the new procedure has good predictive capability in our considered setting.

The remainder of this article is organized as follows. In Section 2, we introduce the proposed method for model averaging of PLAMs for conditional quantile prediction. Section 3 focuses on the theoretical properties of the proposed procedure. In Section 4, implementation is discussed. Section 5 presents empirical evidence

Statistica Sinica: Preprint
doi:10.5705/ss.202023.0092

from simulation studies and a real data example. We provide concluding remarks in Section 6. Finally, additional simulation results and all proofs are left in the Supplementary Material.

2. Methodology

2.1 Model and Estimation

Let $Y \in \mathbb{R}^1$ be the response variable and $\boldsymbol{X} = (X_1, \cdots, X_p)^\top \in \mathbb{R}^p$ be the covariate vector, where $^\top$ is the transpose of a vector or matrix. The τth $(0 < \tau < 1)$ conditional quantile of Y given $\boldsymbol{X}$ is denoted by $\mu(\boldsymbol{X}, \tau) \triangleq Q(Y|\boldsymbol{X}, \tau) = \inf\{t : F_Y(t|\boldsymbol{X}) \geq \tau\}$, where $F_Y(t|\boldsymbol{X})$ is the conditional distribution function of Y given $\boldsymbol{X}$. In this study, what we are most interested in is to predict $\mu(\boldsymbol{X}, \tau)$. We all know that, when the dimension of $\boldsymbol{X}$ is high, modeling the conditional quantile function $\mu(\boldsymbol{X}, \tau)$ by purely multivariate nonparametric approaches without any structure specification is infeasible because of the curse of dimensionality. Earlier authors estimate $\mu(\boldsymbol{X}, \tau)$ by using popular semiparametric models such as the PLAM. However, we do not always know the true model in practice. Using a single model, despite its sophistication, may have a high risk of model misspecification, and thus result in poor prediction performance. Instead of using one model, aggregating S partially linear additive sub-models with a weighted average effectively might offer a better approximation to $\mu(\boldsymbol{X}, \tau)$, where S is allowed to diverge to infinity as the sample size tends to infinity.

Statistica Sinica: Preprint
doi:10.5705/ss.202023.0092

Specifically, suppose that a total of S candidate models are given as $\mathbb{M}_1, \cdots, \mathbb{M}_S$. For each $s = 1, \cdots, S$, we assume that $\mathcal{A}_s$ and $\mathcal{A}_s^c$ are mutually exclusive and complementary subsets of $\{1, \cdots, p\}$ with the cardinality p_s and $(p - p_s)$ respectively. Define $\boldsymbol{X}_{\mathcal{A}_s} = \left(\overline{X}_1^{(s)}, \cdots, \overline{X}_{p_s}^{(s)}\right)^\top$ and $\boldsymbol{X}_{\mathcal{A}_s^c} = \left(\underline{X}_1^{(s)}, \cdots, \underline{X}_{p-p_s}^{(s)}\right)^\top$, where $\overline{X}_j^{(s)}$, $\underline{X}_{j'}^{(s)}$ are two different components of $\boldsymbol{X}$ for $j = 1, \cdots, p_s$ and $j' = 1, \cdots, p - p_s$. Then, the sth candidate model with a partially linear additive structure is defined by

$$\mathbb{M}_s : \mu^{(s)}(\boldsymbol{X}, \tau) \triangleq \alpha^{(s)}(\tau) + \sum_{j=1}^{p_s} g_j^{(s)}\left(\overline{X}_j^{(s)}, \tau\right) + \boldsymbol{X}_{\mathcal{A}_s^c}^\top \boldsymbol{\beta}^{(s)}(\tau), \tag{2.1}$$

where $\alpha^{(s)}(\tau)$ is the unknown intercept, $\boldsymbol{\beta}^{(s)}(\tau) = \left(\beta_j^{(s)}(\tau) : 1 \le j \le p - p_s\right)^\top$ and $g_j^{(s)}(\cdot, \tau)$ are the unknown parameter vector and smooth function at the τ-th quantile. To ensure identifiability, for a given τ, we usually assume $E\left\{g_j^{(s)}\left(\overline{X}_j^{(s)}, \tau\right)\right\} = 0$, $1 \le j \le p_s$. The features in $\mathcal{A}_s$ contribute to the model in a nonlinear fashion while the features in $\mathcal{A}_s^c$ contribute in a linear fashion. Here $\mu^{(s)}(\boldsymbol{X}, \tau)$ can be regarded as the τth conditional quantile function under the sth sub-model $\mathbb{M}_s$. In Section 4.2, we will show how to prepare S semiparametric sub-models.

To offer an optimal weighting scheme, it is necessary to accurately estimate $\mu^{(s)}(\boldsymbol{X}, \tau)$ and the model average weights. Therefore, we first should estimate all unknown parameters and nonparametric functions of the sth candidate model $\mathbb{M}_s$.

Suppose we have a random sample $\{(Y_i, \boldsymbol{X}_i)\}_{i=1}^n$ of size n from the population

Statistica Sinica: Preprint
doi:10.5705/ss.202023.0092

$(Y, \boldsymbol{X})$. Because of the centering constraint on $g_j^{(s)}(\cdot, \tau)$, we approximate the nonparametric functions $g_j^{(s)}(\cdot, \tau)$'s by the centralized B-spline basis functions (Ma, 2012). Using spline expansions, we can approximate $g_j^{(s)}(\cdot, \tau)$ by

$$g_j^{(s)}(\cdot, \tau) \approx \boldsymbol{\psi}_j^{\top}(\cdot)\boldsymbol{\gamma}_j^{(s)}(\tau), j = 1, \cdots, p_s, \tag{2.2}$$

where $\boldsymbol{\psi}_j(\cdot) = (\psi_{jk}(\cdot) : 1 \leq k \leq J_n = N_n + d)^{\top}$ is a vector of centralized B-spline basis functions of order d with N_n internal knots and $\boldsymbol{\gamma}_j^{(s)}(\tau) = \left(\gamma_{jk}^{(s)}(\tau) : 1 \leq k \leq J_n\right)^{\top}$ is the spline coefficient vector. Then, substituting (2.2) into the model (2.1), we can get

$$\begin{aligned}\mathbb{M}_s : \mu^{(s)}(\boldsymbol{X}, \tau) \approx & \alpha^{(s)}(\tau) + \sum_{j=1}^{p_s} \boldsymbol{\psi}_j^{\top}\left(\overline{X}_j^{(s)}\right)\boldsymbol{\gamma}_j^{(s)}(\tau) + \boldsymbol{X}_{\mathcal{A}_s^c}^{\top}\boldsymbol{\beta}^{(s)}(\tau) \\ \triangleq & \boldsymbol{Z}^{(s)\top}\boldsymbol{\xi}^{(s)}(\tau),\end{aligned} \tag{2.3}$$

where $\boldsymbol{Z}^{(s)} = \left(\boldsymbol{\psi}_1^{\top}\left(\overline{X}_1^{(s)}\right), \cdots, \boldsymbol{\psi}_{p_s}^{\top}\left(\overline{X}_{p_s}^{(s)}\right), \left(1, \boldsymbol{X}_{\mathcal{A}_s^c}^{\top}\right)/\sqrt{J_n}\right)^{\top}$ and $\boldsymbol{\xi}^{(s)}(\tau) = \Big(\boldsymbol{\gamma}_1^{(s)\top}(\tau), \cdots, \boldsymbol{\gamma}_{p_s}^{(s)\top}(\tau), \sqrt{J_n}\big(\alpha^{(s)}(\tau), \boldsymbol{\beta}^{(s)\top}(\tau)\big)\Big)^{\top}$.

Clearly, $\boldsymbol{\xi}^{(s)}(\tau)$ can be taken as a set of quantile regression coefficient functions describing how each regression coefficient replies on the quantile level τ. The large majority of references on standard quantile regression (e.g., Koenker (2005)) focused on the problem of estimating a single quantile of interest. However, we usually need to obtain the entire quantile process over $\tau \in (0, 1)$ rather than only

Statistica Sinica: Preprint
doi:10.5705/ss.202023.0092

obtaining a discrete set of quantiles in practice. An alternative approach, introduced by Frumento and Bottai (2016), is to model quantile regression coefficients as parametric functions of the quantile level, which has advantages in terms of parsimony and efficiency, and may expand the potential of statistical modeling. Thus, we model $\boldsymbol{\xi}^{(s)}(\tau)$ by a series of K known basis functions

$$\boldsymbol{\xi}^{(s)}(\tau) = \boldsymbol{\theta}^{(s)}\boldsymbol{b}(\tau), \tag{2.4}$$

where $\boldsymbol{b}(\tau) = (b_j(\tau) : 1 \leq j \leq K)^\top$ is a set of K known basis functions of $\tau \in (0,1)$, and $\boldsymbol{\theta}^{(s)}$ is a $\phi_s \times K$ matrix with $\phi_s \triangleq 1+p_sJ_n+p-p_s$. Let $\theta_{uv}^{(s)}$ be the (u,v)th element of $\boldsymbol{\theta}^{(s)}$ for $u = 1, \cdots, \phi_s$ and $v = 1, \cdots, K$. Under model (2.4), it is easy to see that the lth component of $\boldsymbol{\xi}^{(s)}(\tau)$ is $\xi_l^{(s)}(\tau) = \sum_{k=1}^{K} \theta_{lk}^{(s)} b_k(\tau), l = 1, \cdots, \phi_s$. The model (2.4) extracts the common features of $\boldsymbol{\xi}^{(s)}(\tau)$ over $\tau \in (0,1)$ via the K-dimensional known basis function vector $\boldsymbol{b}(\tau)$. So this modeling strategy presents numerous superiorities including a simpler computation, increased statistical efficiency, and easy interpretability of the results. To obtain an estimate of $\boldsymbol{\theta}^{(s)}$, we should specify $\boldsymbol{b}(\tau)$ in advance. We give detailed discussions on selection of $\boldsymbol{b}(\tau)$ in Section 4.1.

To facilitate the presentation, some notations should be introduced. Let $\boldsymbol{D}^{(s)}(\tau) = \boldsymbol{b}(\tau) \otimes \boldsymbol{Z}^{(s)}$ and $\boldsymbol{\zeta}^{(s)} = \left(\boldsymbol{\theta}_1^{(s)\top}, \cdots, \boldsymbol{\theta}_K^{(s)\top}\right)^\top$, where $\boldsymbol{\theta}_k^{(s)} = \left(\theta_{1k}^{(s)}, \cdots, \theta_{\phi_s k}^{(s)}\right)^\top$ is the k-th column of the parameter matrix $\boldsymbol{\theta}^{(s)}$ and $\otimes$ is the Kronecker product

Statistica Sinica: Preprint
doi:10.5705/ss.202023.0092

of two matrices. By a simple calculation, we have $\boldsymbol{Z}^{(s)\top}\boldsymbol{\xi}^{(s)}(\tau) = \boldsymbol{D}^{(s)\top}(\tau)\boldsymbol{\zeta}^{(s)}$. Then, the formula (2.3) can be rewritten as

$$\mathbb{M}_s : \mu^{(s)}(\boldsymbol{X}, \tau) \approx \boldsymbol{D}^{(s)\top}(\tau)\boldsymbol{\zeta}^{(s)}. \tag{2.5}$$

To integrate information from different quantile levels, we propose to estimate $\boldsymbol{\zeta}^{(s)}$ in (2.5) by minimizing the following integrated loss function

$$\begin{aligned}\hat{\boldsymbol{\zeta}}^{(s)} &= \arg\min_{\boldsymbol{\zeta}^{(s)}} \bar{\mathcal{L}}_n^{(s)}\left(\boldsymbol{\zeta}^{(s)}\right) \\ &= \arg\min_{\boldsymbol{\zeta}^{(s)}} \int_0^1 \mathcal{L}_n^{(s)}\left(\boldsymbol{\zeta}^{(s)}\right) d\tau \\ &= \arg\min_{\boldsymbol{\zeta}^{(s)}} \int_0^1 \sum_{i=1}^n \rho_\tau\left(Y_i - \boldsymbol{D}_i^{(s)\top}(\tau)\boldsymbol{\zeta}^{(s)}\right) d\tau,\end{aligned} \tag{2.6}$$

where $\boldsymbol{D}_i^{(s)}(\tau)$ is the sample version of $\boldsymbol{D}^{(s)}(\tau)$ and $\rho_\tau(u) = u(\tau - I(u<0))$ is the quantile check function. The objective function $\bar{\mathcal{L}}_n^{(s)}\left(\boldsymbol{\zeta}^{(s)}\right)$ can be regarded as an average loss function, achieved by marginalizing $\mathcal{L}_n^{(s)}\left(\boldsymbol{\zeta}^{(s)}\right)$ over the entire interval $(0,1)$. So far, the solution of minimizing (2.6) can be implemented by the "iqr" function in the R package " qrcm" (Frumento and Bottai, 2016). Let $\hat{\mu}^{(s)}(\boldsymbol{X}_i, \tau) = \boldsymbol{D}_i^{(s)\top}(\tau)\hat{\boldsymbol{\zeta}}^{(s)}$ be the estimator of $\mu(\boldsymbol{X}_i, \tau)$ under the sth sub-model $\mathbb{M}_s$ for $s = 1, \cdots, S$.

With the estimators $\hat{\mu}^{(s)}(\boldsymbol{X}_i, \tau)$ of each sub-model readily available, we can define the model average estimate of $\mu(\boldsymbol{X}_i, \tau)$ as a weighted average of $\hat{\mu}^{(s)}(\boldsymbol{X}_i, \tau)$,

Statistica Sinica: Preprint
doi:10.5705/ss.202023.0092

i.e.,

$$\hat{\mu}^{[\boldsymbol{w}]}\left(\boldsymbol{X}_i, \tau\right) = \sum_{s=1}^{S} w_s \hat{\mu}^{(s)}\left(\boldsymbol{X}_i, \tau\right),$$

where $\boldsymbol{w} = (w_s, 1 \leq s \leq S)^\top$ is the vector of model weights belonging to the set $\mathbb{W} = \left\{ \boldsymbol{w} \in [0,1]^S : \sum_{s=1}^{S} w_s = 1 \right\}$.

Remark 1. In fact, it would be difficult to find an optimal parametric or semiparametric model for a dataset of interest. Basically, all sub-models $\mathbb{M}_1, \cdots, \mathbb{M}_S$ under investigation might be wrong, but aggregating S semiparametric sub-models with a weighted average effectively offers a close approximation to the reality. Thus, we may achieve improved prediction accuracy by combining $\mathbb{M}_1, \cdots, \mathbb{M}_S$ in an effective manner. Furthermore, it is worth emphasizing that we provide another attractive strategy to solve model structure uncertainty, which distinguishes from existing model averaging literature that mainly considers the problem of covariate uncertainty (Wan et al., 2010; Hansen and Racine, 2012; Lu and Su, 2015; Zhang et al., 2020; Wang et al., 2023).

2.2 Weight-Choosing Criterion

Until now, the weight vector $\boldsymbol{w}$ in $\hat{\mu}^{[\boldsymbol{w}]}\left(\boldsymbol{X}_i, \tau\right)$ is left unspecified. Clearly, each element of $\boldsymbol{w}$ measures the relative importance of each candidate model in the final prediction. For example, the sth weight with zero value implies that the final prediction removes the sth candidate model. Thus, the key question here is how

Statistica Sinica: Preprint
doi:10.5705/ss.202023.0092

to choose a suitable weight for each candidate model as it plays a critical role in producing good and reliable prediction performance. We propose to select the weight vector by minimizing the following criterion

$$\mathcal{Q}_n(\boldsymbol{w}) = n^{-1}\int_0^1 \sum_{i=1}^{n} \rho_\tau\left(Y_i - \sum_{s=1}^{S} w_s \boldsymbol{D}_i^{(s)\top}(\tau)\,\hat{\boldsymbol{\zeta}}^{(s)}\right) d\tau + n^{-1}\lambda_n \sum_{s=1}^{S} w_s\phi_s, \quad (2.7)$$

that is, $\hat{\boldsymbol{w}} = \arg\min_{\boldsymbol{w}\in\mathbb{W}} \mathcal{Q}_n(\boldsymbol{w})$, where λ_n is a positive scale depending on the sample size n. We take $\lambda_n = \log(n)/2$ in numerical experiments. Note that the first term in (2.7) measures the goodness of fit and the second term controls the model complexity and prevents assigning relative weights to the overfitting candidate models. The second term $n^{-1}\lambda_n \sum_{s=1}^{S} w_s\phi_s$ is related to the information criterion. We use a special case for illustration. If we only consider the sth candidate model $\mathbb{M}_s$ $(1 \le s \le S)$ for prediction and thus take the sth element of $\boldsymbol{w}$ to be one and others to be zeros, then $n^{-1}\lambda_n \sum_{s=1}^{S} w_s\phi_s$ reduces to $\phi_s\log(n)/(2n)$ when $\lambda_n = \log(n)/2$, which is the same as the second term of Schwarz information criterion (SIC) in Lian (2012). Previous studies such as Lian (2012) used SIC to select tuning parameters (i.e., knots selection and regularization parameter selection), but we use a similar form to estimate the weight vector $\boldsymbol{w}$. Such a similar penalty term has also been used in semiparametric model averaging works, see for example Fang et al. (2022).

Furthermore, it is easy to understand that the model weight estimator $\hat{\boldsymbol{w}}$ does not depend on the quantile level τ because of using the integrated loss function.

Statistica Sinica: Preprint
doi:10.5705/ss.202023.0092

Compared with the standard quantile loss function, the penalized integrated loss function (2.7) utilizing the common characteristics of multiple quantiles can gain better estimation efficiency of the model weight $\boldsymbol{w}$. Therefore, our weight-choosing criterion is different form the jackknife criterion of Lu and Su (2015) and Wang et al. (2023), which focuses on the problem of estimating a single quantile of interest. Given $\hat{\boldsymbol{w}}$, the model averaging estimator of $\mu\left(\boldsymbol{X}_i, \tau\right)$, termed as the *semiparametric model averaging quantile prediction* (SMAQP), is defined by

$$\hat{\mu}^{[\hat{\boldsymbol{w}}]}\left(\boldsymbol{X}_i, \tau\right) = \sum_{s=1}^{S} \hat{w}_s \hat{\mu}^{(s)}\left(\boldsymbol{X}_i, \tau\right) = \sum_{s=1}^{S} \hat{w}_s \boldsymbol{D}_i^{(s)\top}\left(\tau\right) \hat{\boldsymbol{\zeta}}^{(s)}. \tag{2.8}$$

Suppose that $(\mathrm{y}, \mathbf{x})$ is an independent copy of $(Y_i, \boldsymbol{X}_i)$. Define $\mathcal{D}_n = \{(Y_i, \boldsymbol{X}_i)\}_{i=1}^n$, $\mathbf{z}^{(s)} = \left(\boldsymbol{\psi}_1^\top\left(\overline{\mathrm{x}}_1^{(s)}\right), \cdots, \boldsymbol{\psi}_{p_s}^\top\left(\overline{\mathrm{x}}_{p_s}^{(s)}\right), \left(1, \mathbf{x}_{\mathcal{A}_s^c}^\top\right)/\sqrt{J_n}\right)^\top$, $\mathbf{x}_{\mathcal{A}_s} = \left(\overline{\mathrm{x}}_1^{(s)}, \cdots, \overline{\mathrm{x}}_{p_s}^{(s)}\right)^\top$ and $\mathbf{x}_{\mathcal{A}_s^c} = \left(\underline{\mathrm{x}}_1^{(s)}, \cdots, \underline{\mathrm{x}}_{p-p_s}^{(s)}\right)^\top$ where $\overline{\mathrm{x}}_j^{(s)}$, $\underline{\mathrm{x}}_{j'}^{(s)}$ are two different components of $\mathbf{x}$ for $j = 1, \cdots, p_s$ and $j' = 1, \cdots, p - p_s$. The definitions of $\mathcal{A}_s$, $\mathcal{A}_s^c$ and $\boldsymbol{\psi}_j\left(\cdot\right)$ are given in subsection 2.1. Let $\mathbb{D}^{(s)}\left(\tau\right) = \boldsymbol{b}\left(\tau\right) \otimes \mathbf{z}^{(s)}$ and $\hat{\mu}^{(s)}\left(\mathbf{x}, \tau\right) = \mathbb{D}^{(s)\top}\left(\tau\right) \hat{\boldsymbol{\zeta}}^{(s)}$ for $s = 1, \cdots, S$. Define the out-of-sample average quantile prediction error (denoted as OAQPE_n) as follows

$$\mathrm{OAQPE}_n\left(\boldsymbol{w}\right) = E\left\{\int_0^1 \rho_\tau\left(\mathrm{y} - \sum_{s=1}^{S} w_s \hat{\mu}^{(s)}\left(\mathbf{x}, \tau\right)\right) d\tau \,|\mathcal{D}_n\right\}. \tag{2.9}$$

Next, we will show that the weight vector selected by (2.7) is asymptotically optimal in the sense of achieving the lowest possible $\mathrm{OAQPE}_n\left(\boldsymbol{w}\right)$ under some reg-

Statistica Sinica: Preprint
doi:10.5705/ss.202023.0092

ularity conditions.

Remark 2. It is understood that the selection of the loss function is closely related to the characteristic of the response variable's distribution that one wants to predict. For example, the traditional quadratic (or quantile) loss function corresponds to the conditional mean (or quantile) of the distribution of the response. Here the object of our interest is the average of quantile prediction over the interval $(0, 1)$, and thus it is natural to define the risk function (2.9) which can be regarded as a beneficial extension of the criterion (2.13) in Lu and Su (2015).

Remark 3. Of course, we can utilize standard quantile regression to estimate the parameters and nonparametric functions of each candidate model and choose the optimal model weight. Specifically, for a given quantile level $\tau \in (0, 1)$, we obtain the estimator of $\mu(\boldsymbol{X}_i, \tau)$ in the sth sub-model $\mathbb{M}_s$ by $\tilde{\mu}^{(s)}\big(\boldsymbol{X}_i, \tau\big) = \boldsymbol{Z}_i^{(s)\top}\tilde{\boldsymbol{\xi}}^{(s)}(\tau)$, where $\tilde{\boldsymbol{\xi}}^{(s)}(\tau) = \underset{\boldsymbol{\xi}^{(s)}(\tau)}{\arg\min} \sum_{i=1}^{n} \rho_\tau\big(Y_i - \boldsymbol{Z}_i^{(s)\top}\boldsymbol{\xi}^{(s)}(\tau)\big)$ and $\boldsymbol{Z}_i^{(s)}$ is the sample version of $\boldsymbol{Z}^{(s)}$ in (2.3). Then, the model weight estimator can be obtained by minimizing the penalized standard quantile loss function, i.e., $\tilde{\boldsymbol{w}}(\tau) = \underset{\boldsymbol{w}(\tau)\in\mathbb{W}}{\arg\min}\, \tilde{\mathcal{Q}}_n(\boldsymbol{w}, \tau)$, where $\tilde{\mathcal{Q}}_n\big(\boldsymbol{w}, \tau\big) = n^{-1}\sum_{i=1}^{n} \rho_\tau\left\{Y_i - \sum_{s=1}^{S} w_s(\tau)\tilde{\mu}^{(s)}(\boldsymbol{X}_i, \tau)\right\} + n^{-1}\lambda_n \sum_{s=1}^{S} w_s(\tau)\phi_s$. Then, similar to (2.8), the final model averaging estimator of $\mu(\boldsymbol{X}_i, \tau)$, termed as *standard semiparametric model averaging quantile prediction* (SSMAQP), is defined by $\tilde{\mu}^{[\tilde{\boldsymbol{w}}]}(\boldsymbol{X}_i, \tau) = \sum_{s=1}^{S} \tilde{w}_s(\tau)\tilde{\mu}^{(s)}(\boldsymbol{X}_i, \tau)$.

Remark 4. We note that the standard quantile regression fitted at each given quantile level can be viewed as a special case of the quantile regression with the

Statistica Sinica: Preprint
doi:10.5705/ss.202023.0092

integrated loss. Specifically, we carry out the integral in a sufficiently small interval $(\tau-\Delta, \tau+\Delta)$, and let the regression coefficients and model weights be constant on this interval. Then, the estimator based on the integrated loss approaches to the one obtained from the standard quantile regression when Δ approaches zero, which reasonably clarifies in what sense this approach can use more information from the data than the standard quantile regression, so that SMAQP can be more efficient than SSMAQP. The numerical evidence for the above discussion is provided in Section S1.2 of the Supplementary Material.

3. Theoretical results

In this section, we investigate the asymptotic properties of the SMAQP estimator. For notational simplicity, let $\mu_i \triangleq \mu\left(\boldsymbol{X}_i, \tau\right)$ and $\varepsilon_i \triangleq \varepsilon_i\left(\tau\right) = Y_i - \mu_i$, so we have $P\left\{\varepsilon_i \leq 0 \,|\boldsymbol{X}_i\right\} = \tau$. Suppose that $f\left(\cdot\,|\boldsymbol{X}_i\right)$ and $F\left(\cdot\,|\boldsymbol{X}_i\right)$ are the conditional probability density function (PDF) and cumulative distribution function (CDF) of ε_i given $\boldsymbol{X}_i$, respectively. Since all models are potentially misspecified in the model averaging literature, we define the pseudo-true parameter value for the sth sub-model

$$\boldsymbol{\zeta}_0^{(s)} = \arg\min_{\boldsymbol{\zeta}^{(s)}} \bar{Q}^{(s)}\left(\boldsymbol{\zeta}^{(s)}\right) \triangleq \arg\min_{\boldsymbol{\zeta}^{(s)}} E\left\{\int_0^1 \rho_\tau\left(Y_i - \boldsymbol{D}_i^{(s)\top}\left(\tau\right)\boldsymbol{\zeta}^{(s)}\right) d\tau\right\}.$$

Statistica Sinica: Preprint
doi:10.5705/ss.202023.0092

For $s = 1, \cdots, S$, we define $\boldsymbol{A}^{(s)} = E\left\{\int_0^1 f\left(-u_i^{(s)} | \boldsymbol{X}_i\right) \boldsymbol{D}_i^{(s)}(\tau) \boldsymbol{D}_i^{(s)\top}(\tau) d\tau\right\}$, $\boldsymbol{C}^{(s)} = Cov\left\{\int_0^1 \psi_\tau\left(\varepsilon_i + u_i^{(s)}\right) \boldsymbol{D}_i^{(s)}(\tau) d\tau\right\}$ and $\boldsymbol{\Sigma}^{(s)} = E\left\{\int_0^1 \boldsymbol{D}_i^{(s)}(\tau) \boldsymbol{D}_i^{(s)\top}(\tau) d\tau\right\}$, where $\psi_\tau(u) = \tau - I\{u < 0\}$ and $u_i^{(s)} \triangleq u_i^{(s)}(\tau) = \mu_i - \boldsymbol{D}_i^{(s)\top}(\tau) \boldsymbol{\zeta}_0^{(s)}$ can be regarded as the approximation bias for the sth candidate model. Let $\lambda_{\min}(\boldsymbol{A})$ and $\lambda_{\max}(\boldsymbol{A})$ be the smallest eigenvalue and largest eigenvalue of a symmetric matrix $\boldsymbol{A}$, respectively. For a vector $\boldsymbol{\xi} = (\xi_1, \cdots, \xi_s)^\top \in \mathbb{R}^s$, denote its L_2 norm by $\|\boldsymbol{\xi}\| = (\xi_1^2 + \cdots + \xi_s^2)^{1/2}$. Let $\bar{\phi} = \max_{1 \le s \le S} \phi_s$. To build the theoretical properties of $\hat{\boldsymbol{\zeta}}^{(s)}$ and $\hat{\boldsymbol{w}}$, the following regularity conditions are needed.

(C1) (i) $\{(Y_i, \boldsymbol{X}_i)\}_{i=1}^n$ are independent and identically distributed and $E[\mu_i^4] < \infty$. (ii) The support of each predictor, $\bar{X}_j^{(s)}$, is $[a, b]$, where a and b are finite real numbers. The marginal densities for $\bar{X}_j^{(s)}$ are all bounded from below and above by two fixed positive constants on $[a, b]$, for $j = 1, \cdots, p_s$, $s = 1, \cdots, S$. (iii) There are constants $0 < C_1 < C_2 < \infty$, such that $C_1 \le \lambda_{\min}\left[E\left\{\left(1, \boldsymbol{X}_{\mathcal{A}_s^c}^\top\right)^\top \left(1, \boldsymbol{X}_{\mathcal{A}_s^c}^\top\right)\right\}\right] \le \lambda_{\max}\left[E\left\{\left(1, \boldsymbol{X}_{\mathcal{A}_s^c}^\top\right)^\top \left(1, \boldsymbol{X}_{\mathcal{A}_s^c}^\top\right)\right\}\right] \le C_2$.

(C2) For $s = 1, \cdots, S, j = 1, \cdots, p_s$ and sufficiently large n, $J_n \sum_{i=1}^n \boldsymbol{\psi}_j\left(\bar{X}_{ij}^{(s)}\right) \boldsymbol{\psi}_j^\top\left(\bar{X}_{ij}^{(s)}\right)$ is nonsingular, and the eigenvalues of $J_n n^{-1} \sum_{i=1}^n \boldsymbol{\psi}_j\left(\bar{X}_{ij}^{(s)}\right) \boldsymbol{\psi}_j^\top\left(\bar{X}_{ij}^{(s)}\right)$ are bounded away from zero and infinity.

(C3) The conditional cumulative distribution function $F(\cdot | \boldsymbol{X}_i)$ is absolutely continuous, with continuous conditional density function $f(\cdot | \boldsymbol{X}_i)$ uniformly bounded, and its first-order derivative $f'(\cdot | \boldsymbol{X}_i)$ exists and is uniformly bounded.

(C4) For $s = 1, \cdots, S$, there exist constants $\underline{c}_{\boldsymbol{A}^{(s)}}$, $\bar{c}_{\boldsymbol{A}^{(s)}}$, $\underline{c}_{\boldsymbol{C}^{(s)}}$, $\bar{c}_{\boldsymbol{C}^{(s)}}$ that may de-

Statistica Sinica: Preprint
doi:10.5705/ss.202023.0092

pend on ϕ_s such that (i) $0 < \underline{c}_{\boldsymbol{A}^{(s)}} \leq J_n\lambda_{\min}\left(\boldsymbol{A}^{(s)}\right) \leq J_n\lambda_{\max}\left(\boldsymbol{A}^{(s)}\right) \leq C_0 J_n\lambda_{\max}\left(\boldsymbol{\Sigma}^{(s)}\right) \leq \bar{c}_{\boldsymbol{A}^{(s)}} < \infty$ for some positive constant C_0; (ii) $0 < \underline{c}_{\boldsymbol{C}^{(s)}} \leq J_n\lambda_{\min}\left(\boldsymbol{C}^{(s)}\right) \leq J_n\lambda_{\max}\left(\boldsymbol{C}^{(s)}\right) \leq \bar{c}_{\boldsymbol{C}^{(s)}} < \infty$; (iii) $(\bar{c}_{\boldsymbol{A}^{(s)}} + \bar{c}_{\boldsymbol{C}^{(s)}})/\phi_s = O(\underline{c}^2_{\boldsymbol{A}^{(s)}})$.

(C5) Let $\underline{c}_{\boldsymbol{A}} = \min_{1\leq s\leq S} \underline{c}_{\boldsymbol{A}^{(s)}}$, $\underline{c}_{\boldsymbol{C}} = \min_{1\leq s\leq S} \underline{c}_{\boldsymbol{C}^{(s)}}$, $\bar{c}_{\boldsymbol{A}} = \max_{1\leq s\leq S} \bar{c}_{\boldsymbol{A}^{(s)}}$ and $\bar{c}_{\boldsymbol{C}} = \max_{1\leq s\leq S} \bar{c}_{\boldsymbol{C}^{(s)}}$. (i)As $n \to \infty$, $p^2 J_n^2/(n\underline{c}^2_{\boldsymbol{C}^{(s)}}) \to 0$ and $p\bar{\phi}J_n \log n/n \to 0$; (ii) $Sn^{-L^2\bar{\phi}\underline{c}^3_{\boldsymbol{A}}/(4\bar{c}_{\boldsymbol{A}}\bar{c}_{\boldsymbol{C}})} = o(1)$ for a sufficiently large constant L.

(C6) $\lambda_n\bar{\phi} = o(n)$.

Conditions (C1)(i) specifies the data are IID, which is commonly imposed in the quantile model averaging literature, see Lu and Su (2015); Lee and Shin (2023); Wang et al. (2023). Condition (C1)(ii) requires a boundedness condition on the covariates, and it is often assumed in the asymptotic analysis of nonparametric regression problems. Condition (C1)(iii) can be found in Condition (C2) of Huang et al. (2007) and Assumption (A2) of Ma et al. (2013). Condition (C2) is similar to Condition (C6) in Sun et al. (2017). This condition can be derived from the properties of B-splines. It can also be written in another way such that $c_1 J_n^{-1} \leq \lambda_{\min}\left[E\{\boldsymbol{\psi}_j(\bar{X}_{ij}^{(s)})\boldsymbol{\psi}_j^\top(\bar{X}_{ij}^{(s)})\}\right] \leq \lambda_{\max}\left[E\{\boldsymbol{\psi}_j(\bar{X}_{ij}^{(s)})\boldsymbol{\psi}_j^\top(\bar{X}_{ij}^{(s)})\}\right] \leq J_n^{-1}C_1$ for $0 < c_1 \leq C_1 < \infty$, which holds by using the conclusion of Lemma S.2 in Ma and He (2016). Condition (C3) is a common assumption for the conditional density function of the random error. Moreover, Condition (C4) (i) and (ii) can be derived from Condition (C1) (iii) and Condition (C2). Condition (C4)(i)-(ii) imply that the eigenvalues of $J_n\boldsymbol{A}^{(s)}$ and $J_n\boldsymbol{C}^{(s)}$ are bounded and bounded away from zero

Statistica Sinica: Preprint
doi:10.5705/ss.202023.0092

for a given s. However, these bounds ($\underline{c}_{\boldsymbol{A}_{(s)}}$, $\underline{c}_{\boldsymbol{C}_{(s)}}$) can converge to zero and ($\bar{c}_{\boldsymbol{A}_{(s)}}$, $\bar{c}_{\boldsymbol{C}_{(s)}}$) diverge to infinity, both at slow rates when $\phi_s \to \infty$. The rates are restricted in conditions (C4)(iii) so that the usual consistency $\sqrt{\phi_s J_n/n}$ for the estimator $\hat{\boldsymbol{\zeta}}^{(s)}$ in $\mathbb{M}_s$ is not affected. Condition (C5) imposes restrictions on the dimension of the covariates p, J_n, the largest dimension of the models $\bar{\phi}$, the potential number of models under investigation S, and the constants $\underline{c}_{\boldsymbol{A}}$, $\underline{c}_{\boldsymbol{C}}$, $\bar{c}_{\boldsymbol{A}}$ and $\bar{c}_{\boldsymbol{C}}$. Condition (C6) is satisfied when we use $\lambda_n = 1$ or $\log(n)/2$.

Proposition 1. *Under conditions (C1)–(C5), we have*

$$\max_{1\le s\le S} \left\| \hat{\boldsymbol{\zeta}}^{(s)} - \boldsymbol{\zeta}_0^{(s)} \right\| = O_p\left(\sqrt{J_n \bar{\phi} \log n/n} \right).$$

Remark 5. The prediction of the conditional quantile function is of our main interest. We provide the uniform convergence rate of $\hat{\boldsymbol{\zeta}}^{(s)}$ for the purpose of establishing the desirable asymptotic optimality of SMAQP.

Theorem 1. *Suppose conditions (C1)–(C6) hold, and* $p^{3/2}\bar{\phi}^{3/2} S \log n/n \to 0$ *as* $n \to \infty$. *Then* $\hat{\boldsymbol{w}}$ *is asymptotically optimal in the sense that*

$$\frac{\mathrm{OAQPE}_n\left(\hat{\boldsymbol{w}}\right)}{\inf_{\boldsymbol{w}\in\mathbb{W}} \mathrm{OAQPE}_n\left(\boldsymbol{w}\right)} = 1 + o_p(1).$$

Remark 6. Theorem 1 presents the asymptotic optimality of the SMAQP estimator even if all candidate models are misspecified, implying that the estimated

Statistica Sinica: Preprint
doi:10.5705/ss.202023.0092

weight vector $\hat{\boldsymbol{w}}$ by minimizing (2.7) is optimal in the sense of making the OAQPE_n as small as possible among all feasible weight vectors $\boldsymbol{w}$ lying in the set $\mathbb{W}$. This result is similar to Theorem 3.3 of Lu and Su (2015), but our proof may be more cumbersome due to the different model weight choice criterion (2.7) and complicated semi-parametric candidate models. In addition, we should be aware that $\hat{\boldsymbol{w}}$ will change if the set of semi-parametric sub-models under consideration are changed. Any SMAQP estimator based on the given set of candidate models may not outperform an estimator that is not considered by the given candidate models.

4. Implementation details

4.1 Selection of turning parameters and basis functions

Selection of the order and knots for B-spline plays important roles in numerical studies. Following Ma (2012) and Huang et al. (2012), we used cubic B-splines ($d = 4$) to approximate each nonparametric function in all numerical studies. For convenience, the number of interior knots (N_n) is selected as the integer part of $n^{1/(2d+1)}$, which is also adopted in Xue et al. (2010) and Wang et al. (2014). One often uses $\lambda_n = 1$ or $\log(n)/2$ in practice (Fang et al., 2022). Both of them satisfy Condition (C6). To investigate whether there is a significant difference in the numerical results for $\lambda_n = 1$ or $\log(n)/2$, we conduct a sensitivity analysis of the proposed SMAQP. Figure S3 in the Supplementary Material shows that the two values of λ_n lead to similar prediction results. Therefore, both values can be used

Statistica Sinica: Preprint
doi:10.5705/ss.202023.0092

in practice. In our numerical studies, we set $\lambda_n = \log(n)/2$.

Another key problem in the implementation of our model averaging is how to specify the basis function $\boldsymbol{b}(\tau)$. In the absence of prior knowledge, one may define $\boldsymbol{b}(\tau)$ by using polynomials, splines, trigonometric functions, known quantile functions, and combinations of the above. For instance, we consider the three types of basis functions $\boldsymbol{b}(\tau)^{(1)} = (1, \Phi^{-1}(\tau))^\top$, $\boldsymbol{b}(\tau)^{(2)} = (1, \tau, \tau^2)^\top$, $\boldsymbol{b}(\tau)^{(3)} = (1, \log(\tau/(1-\tau)))^\top$, where $\Phi(\cdot)$ denotes the distribution function of the standard normal distribution. The basis set $\boldsymbol{b}(\tau)^{(1)}$ relates to the quantile function of the standard normal, $\boldsymbol{b}(\tau)^{(2)}$ simply consists of polynomials of increasing orders and $\boldsymbol{b}(\tau)^{(3)}$ relates to a logistic distribution. In general, any set of functions could be utilized, with the only requirement that $\boldsymbol{b}(\tau)$ should induce a well-defined quantile function for some $\boldsymbol{\theta}^{(s)}$. More examples can be found in Frumento and Bottai (2016) and Frumento et al. (2021). In Section 5.1, a sensitivity analysis has been made to investigate whether our method is sensitive to the selection of basis function.

4.2 Building partially linear additive sub-models

Prior to model averaging, we have to prepare suitable candidate models for producing good predictions. So far, many model averaging studies focus on dealing with the covariate uncertainty, see Wan et al. (2010); Hansen and Racine (2012); Lu and Su (2015); Zhang and Liu (2018); Zhang et al. (2020); Fang et al. (2023). In this paper, we introduce another attractive scheme to build a series of semipara-

Statistica Sinica: Preprint
doi:10.5705/ss.202023.0092

metric sub-models for solving the model structure uncertainty. Without loss of generality, we assume that the first S^* covariates are continuous and the last $p-S^*$ covariates are discrete. Let $\boldsymbol{X}_{\mathcal{A}} = (X_1, \cdots, X_{S^*})^\top$ and $\boldsymbol{X}_{\mathcal{A}^c} = (X_{S^*+1}, \cdots, X_p)^\top$. Usually, discrete covariates are only taken as the linear parts. However, we still do not know which continuous covariates have linear effects on the conditional quantile function. In the absence of prior information, we suppose that all continuous covariates have nonlinear effects on the conditional quantile function, and thus the following PLAM is used initially

$$\mu^{(0)}(\boldsymbol{X}, \tau) \triangleq \alpha^{(0)}(\tau) + \sum_{j=1}^{S^*} g_j^{(0)}(X_j, \tau) + \sum_{l=1}^{p-S^*} X_{l+S^*} \beta_l^{(0)}(\tau), \tag{4.1}$$

where $\alpha^{(0)}(\tau)$ is the unknown intercept, $\beta_l^{(0)}(\tau)$ and $g_j^{(0)}(\cdot, \tau)$ are the unknown parameter and smooth function at the τ-th quantile for $1 \leq l \leq p - S^*$ and $1 \leq j \leq S^*$. It is often assumed $E\left\{g_j^{(0)}(X_j, \tau)\right\} = 0$ with $j = 1, \cdots, S^*$ for identification purposes.

For a bivariate function $g(x, \tau)$, let $g'(x, \tau) = \partial g(x, \tau)/\partial x$ be the first partial derivative of $g(x, \tau)$ with respect to x. Let $b_{ij}(\tau) = g_j'^{(0)}(X_{ij}, \tau) - g_j'^{(0)}(X_{i-1,j}, \tau)$, $\boldsymbol{b}_j(\tau) = (b_{2j}, \cdots, b_{nj})^\top \in \mathbb{R}^{n-1}$ and $\|\boldsymbol{b}_j(\tau)\| = \left(\sum_{i=2}^n b_{ij}^2(\tau)\right)^{1/2}$ for $2 \leq i \leq n$ and $1 \leq j \leq S^*$. It is easy to understand that the jth covariate has a linear effect (i.e., $g_j^{(0)}(x, \tau)$ is a linear function of x) if and only if its first partial derivative with respect to x is a constant, then $\|\boldsymbol{b}_j(\tau)\| = 0$ (Hu and Xia, 2012). We still adopt the

Statistica Sinica: Preprint
doi:10.5705/ss.202023.0092

strategy of subsection 2.1 to estimate all unknown parameters and nonparametric functions in (4.1). Thus, we can easily obtain $\hat{g}_j'^{(0)}(x,\tau)$, $\|\hat{\boldsymbol{b}}_j(\tau)\|$ and $\|\hat{\boldsymbol{b}}_j\| = \int_0^1 \|\hat{\boldsymbol{b}}_j(\tau)\| d\tau$ for $j = 1, \cdots, S^*$. To build a series of semiparametric sub-models, an explicit ordering of the S^* continuous covariates is needed. Specifically, we can determine the order of the covariates entering the nonlinear set according to the decreasing order of $\|\hat{\boldsymbol{b}}_1\|, \cdots, \|\hat{\boldsymbol{b}}_{S^*}\|$ (called as S^* threshold values). Suppose that $\{j_1, \cdots, j_{S^*}\}$ is a permutation of $\{1, \cdots, S^*\}$ such that $\|\hat{\boldsymbol{b}}_{j_1}\| \geq \cdots \geq \|\hat{\boldsymbol{b}}_{j_{S^*}}\|$. Clearly, the larger $\|\hat{\boldsymbol{b}}_j\|$ is, the more reasonable it is to take the jth covariate as a nonlinear component. Moreover, the model (4.1) that includes all continuous covariates in the nonlinear part is only used as an initial model to obtain $\|\hat{\boldsymbol{b}}_j\|$ based on which one can rank the continuous covariates and build the series of semiparametric sub-models.

Given S^* available continuous covariates, we all know that the total number of semiparametric sub-models is $(2^{S^*} - 1)$ if we require that at least one continuous covariate is taken as the nonlinear component. Applying our model averaging procedure in such a case is computationally infeasible when S^* is large or moderate. So we provide two versions of constructing semiparametric sub-models for obtaining good numerical results.

Strategy A: Let $S = S^*$, $\mathcal{A}_s = \{j_l : 1 \leq l \leq s\}$ and $\mathcal{A}_s^c$ be the complementary set of $\mathcal{A}_s$ for $s = 1, \cdots, S$. All the nonlinear sets are nested, that is, $\mathcal{A}_1 \subset \mathcal{A}_2 \subset \cdots \subset \mathcal{A}_S$. In the sth sub-model $\mathbb{M}_s$, we take $\boldsymbol{X}_{\mathcal{A}_s} = (X_{j_1}, \cdots, X_{j_s})^\top \triangleq$

Statistica Sinica: Preprint
doi:10.5705/ss.202023.0092

$\left(\overline{X}_1^{(s)}, \cdots, \overline{X}_s^{(s)}\right)^\top$ as the nonlinear parts and regard the remaining covariates $\boldsymbol{X}_{\mathcal{A}_s^c} = \left(\underline{X}_1^{(s)}, \cdots, \underline{X}_{p-s}^{(s)}\right)^\top$ as the linear parts, where $\overline{X}_j^{(s)}$, $\underline{X}_{j'}^{(s)}$ are two different components of $\boldsymbol{X}$ for $j = 1, \cdots, s$ and $j' = 1, \cdots, p-s$.

Strategy B: We assume that the first q covariates $\left\{X_{j_1}, \cdots, X_{j_q}\right\}$ may have nonlinear effects on the conditional quantile function. Let $\mathcal{A}_1, \cdots, \mathcal{A}_S$ be all possible nonempty subsets of $\{j_1, \cdots, j_q\}$, which indicates that we take all the possible $(S = 2^q - 1)$ sub-models into consideration. Here q is often small, and thus we take $q = 5$ in our numerical studies. In the sth sub-model $\mathbb{M}_s$, we regard $\boldsymbol{X}_{\mathcal{A}_s}$ and $\boldsymbol{X}_{\mathcal{A}_s^c}$ as the nonlinear and linear components respectively.

5. Numerical studies

5.1 Simulation examples

In this section, we conduct two simulation examples to evaluate the finite-sample performance of the proposed approach and compare it with the following model selection and model averaging approaches.

(1) QRCM: Frumento and Bottai (2016) described an alternative approach for linear quantile regression through modeling quantile regression coefficients as parametric functions of the order of the quantile, which can be implemented by the `iqr` function in the package `qrcm` of R software.

(2) PAQRM: Lian (2012) proposed a penalized polynomial spline-based procedure that simultaneously eliminates irrelevant predictors and identifies nonlinear

Statistica Sinica: Preprint
doi:10.5705/ss.202023.0092

components for additive quantile regression models. This procedure aims to identify a single optimal model to make predictions.

(3) JQLMA: Lu and Su (2015) introduced a jackknife model averaging estimator for linear quantile regression. This procedure considers all candidate models to be fully linear.

(4) It is known that the performance of model averaging heavily relies on the construction of candidate models. Thus, two versions are considered for SMAQP. Let SMAQP1 and SMAQP2 be the respective optimal model averaging methods with candidate models constructed by *Strategy A* and *Strategy B*.

Next, we examine the numerical performances of the above methods by two simulated examples. For each simulation, we estimate unknown parameters, functions and model weights with $n = 500$ in-sample observations, and then generate additional 100 out-of-sample observations to calculate prediction performances. We utilize the sample version of OAQPE_n, defined in (2.9), to measure the average of out-of-sample quantile prediction error

$$\mathrm{OAQPE} = \sum_{k=1}^{K} \sum_{i \in \mathcal{I}} \frac{\rho_{\tau_k} \left(Y_i - \hat{\mu} \left(\boldsymbol{X}_i, \tau_k \right) \right)}{K \left| \mathcal{I} \right|},$$

where $\tau_k = k/(K+1), k = 1, \cdots, K = 100$, $\hat{\mu} \left(\boldsymbol{X}_i, \tau \right)$ is an estimator of the τth conditional quantile function $\mu \left(\boldsymbol{X}_i, \tau \right)$ and $\mathcal{I}$ stands for the testing set with the size $|\mathcal{I}| = 100$. Following Lee and Shin (2023), we adopt the following criteria to

Statistica Sinica: Preprint
doi:10.5705/ss.202023.0092

assess the performance of different methods

$$\begin{aligned}
\text{Average OAQPE}_A =& R^{-1}\sum_{r=1}^{R}\text{OAQPE}\left(r\right)_A,\\
\text{Winning Ratio}_A =& R^{-1}\sum_{r=1}^{R} I\big\{\text{OAQPE}\left(r\right)_A < \text{OAQPE}\left(r\right)_B,\cdots,\\
&\qquad \text{OAQPE}\left(r\right)_A < \text{OAQPE}\left(r\right)_E\big\},\\
\text{Loss to SMAQP2}_A =& R^{-1}\sum_{r=1}^{R} I\left\{\text{OAQPE}\left(r\right)_{\mathsf{SMAQP2}} < \text{OAQPE}\left(r\right)_A\right\},
\end{aligned}$$

where $I\left(\mathcal{C}\right)$ is an indicator function for an event $\mathcal{C}$ and OAQPE(r) is the value of OAQPE in the rth replication for $r = 1, \cdots, R$. Similar to Lee and Shin (2023), the subscript A in $\text{OAQPE}\left(r\right)_A$ denotes a generic notation for an estimation approach. Note that the loss to SMAQP2 ratio gives us a more direct binary comparison of each approach to SMAQP2. It is easy to see that the smaller OAQPE and the bigger winning ratio, the method is better. We generate $R = 200$ replications for each simulation example.

Example 1. Similar to the simulation setups of Lian (2012), we generate data from an additive model $Y_i = \sum_{j=1}^{p} m_j\left(X_{ij}\right) + \sigma_i\epsilon_i, i = 1, \cdots, n$, where $m_1\left(u\right) = 3\sin\left(2\pi u\right)$, $m_2\left(u\right) = 2\cos\left(2\pi u\right) - 3u$, $m_3\left(u\right) = 6u\left(1-u\right)$, $m_4\left(u\right) = \exp\left(-2u\right) + 3u^3$, $m_5\left(u\right) = u$, $m_6\left(u\right) = -u$ and $m_j\left(u\right) = 0$ for $j = 7, \cdots, p$. The covariates $\boldsymbol{X}_i = \left(X_{i1}, \cdots, X_{ip}\right)^\top$ are generated by $X_{il} = \left(W_{il} + tU_i\right)/\left(1+t\right)$ for $1 \le l \le p$, where W_{il} and U_i are generated independently from $Uniform(0,1)$. Thus, the

Statistica Sinica: Preprint
doi:10.5705/ss.202023.0092

covariates have an exchangeable correlation structure with $\rho_x \triangleq Corr(X_{il}, X_{il'}) = t^2/(1+t^2)$ for $l \neq l'$. The parameter t is taken as 0, 1 and 3, representing uncorrelated ($\rho_x = 0$), moderate ($\rho_x = 0.5$) and high ($\rho_x = 0.9$) correlations between covariates. We take $\sigma_i = 1$ (*case a*) or $0.5 + |X_{i1}|$ (*case b*), representing homoscedasticity and heteroscedasticity respectively. We also consider $p = 6$ or 10, corresponding to different sparsity levels. Three different distributions of the error term ϵ_i are considered: (case i) standard normal distribution; (case ii) t distribution with 3 degrees of freedom; (case iii) mixture distribution which is a mixture of $N(0,1)$ and $N(0,25)$ with weights 95% and 5% respectively.

Example 2. To reflect the flexibility of our procedure, we consider two more complicated data-generating processes. In the first case (Model I), we generate data from the previous model in Example 1, only except that $m_j(u) = \cos(2\kappa_j u)/\sqrt{j}$ with $\kappa_j \overset{i.i.d}{\sim} Uniform(1,4)$ for $j = 1, \cdots, 1000$. Our setting is similar to the infinite-order regression in Lu and Su (2015) except that regression parameters are replaced by nonlinear additive functions. In this case, we only use the first 10 covariates for prediction, which indicates that some covariates are excluded. In the second case (Model II), we assume that the true data-generating process is from the multivariate nonparametric regression model $Y_i = 4\sin(2\pi X_{i1}X_{i3}X_{i5})X_{i7} - \exp(X_{i9}X_{i10}) + \sigma_i\epsilon_i, i = 1, \cdots, n$. In this example, we consider $p = 10$. Other specifications are the same as those in Example 1. It is easy to see that all candidate models are wrongly specified in both two cases.

Statistica Sinica: Preprint
doi:10.5705/ss.202023.0092

An intuitive approach is to specify basis functions by a collection of polynomials, known quantile functions, splines, and combinations of the above as long as the regression coefficient can be approximated reasonably well by the specified basis functions. We conduct a sensitivity analysis of the proposed method SMAQP1 using $\boldsymbol{b}(\tau)^{(1)} = (1, \Phi^{-1}(\tau))^\top$, $\boldsymbol{b}(\tau)^{(2)} = (1, \tau, \tau^2)^\top$ and $\boldsymbol{b}(\tau)^{(3)} = (1, \log(\tau/(1-\tau)))^\top$. We attain 200 replicates and corresponding results are given in Table S2 of the Supplementary Material. We only consider $t = 1$, $p = 10$ and homoscedasticity (*case a*), and the results for other cases are similar. To save space, we don't report those results. From Table S2, we can see that there is essentially little difference in the numerical results when different bias functions are used, indicating that our method is not sensitive to the selection of basis function Thus, we fix $\boldsymbol{b}(\tau) = (1, \Phi^{-1}(\tau))^\top$ in these two simulation examples.

Table 1 and Table S3 in the Supplementary Material summarize the simulation results of various estimators in all designs. From Table 1 and Table S3, we observe that the proposed SMAQP2 enjoys the smallest OAQPEs and the highest winning ratio among all compared methods. The results in the tables show that SMAQP2 slightly outperforms SMAQP1 which typically has the second-best prediction performance. The good performances of SMAQP1 and SMAQP2 are partly because the optimality of the proposed semiparametric model averaging does not depend on the correct specification of candidate models.

Interestingly, the popular model selection method PAQRM is inferior to SMAQP1

Statistica Sinica: Preprint
doi:10.5705/ss.202023.0092

and SMAQP2 in all scenarios, showing the benefits of model averaging. One possible reason is that identifying the best semiparametric model is usually difficult as it often involves complicated tuning steps, resulting in model selection exhibiting unsatisfactory prediction results in most practical applications. Furthermore, it is not surprising that QRCM and JQLMA yield the worst outcomes as they ignore the nonlinearity between the response and covariates.

Finally, it is not difficult to understand that all approaches use misspecified models in example 2. Even in the misspecified scenarios, SMAQP1 and SMAQP2 still dominate the other methods in all cases. This can be partly explained by the fact that aggregating more flexible partially linear additive sub-models with a weighted average effectively provides a close approximation to the true model which is unknown and thus leads to a smaller risk of model misspecification and more accurate prediction. Generally speaking, we confirm that our proposal has a satisfactory finite sample performance via two simulation examples.

5.2 A real data example

In this section, we will visit the empirical example in Harrison and Rubinfeld (1978) and apply the proposed model averaging procedure to predict the quantiles of the Boston housing price. The empirical data can be downloaded freely from the R package "mlbench", which contains $n = 506$ observations taken from 506 census tracts of Boston from the 1970 census. This is a widely used dataset, see,

Statistica Sinica: Preprint
doi:10.5705/ss.202023.0092

Table 1: Simulation results over different settings in Example 1.

error	t,p	homoscedasticity (*case a*)					heteroscedasticity (*case b*)				
		QRCM	PAQRM	JQLMA	SMAQP1	SMAQP2	QRCM	PAQRM	JQLMA	SMAQP1	SMAQP2
						Average OAQPE					
	0,6	0.657	0.333	0.662	0.328	0.327	0.667	0.335	0.671	0.331	0.330
	1,6	0.529	0.350	0.529	0.343	0.337	0.524	0.348	0.525	0.343	0.336
	3,6	0.525	0.346	0.524	0.342	0.334	0.526	0.357	0.524	0.354	0.346
	0,10	0.667	0.335	0.668	0.328	0.327	0.664	0.334	0.664	0.327	0.326
	1,10	0.531	0.354	0.529	0.346	0.340	0.529	0.357	0.527	0.349	0.342
	3,10	0.526	0.344	0.523	0.339	0.332	0.518	0.354	0.515	0.350	0.342
						Winning Ratio					
	0,6	0%	25.5%	0%	20.0%	54.5%	0%	32.0%	0%	18.5%	49.5%
	1,6	0%	16.5%	0%	6.0%	77.5%	0%	22.0%	0.5%	4.0%	73.5%
case i	3,6	0%	10.5%	0%	7.5%	82.0%	0%	13.0%	0%	5.0%	81.5%
	0,10	0%	24.5%	0%	19.0%	56.0%	0%	29.5%	0%	18.5%	51.5%
	1,10	0.5%	14.5%	0%	9.0%	75.5%	0%	15.5%	0%	3.0%	80.5%
	3,10	0%	10.0%	0%	6.0%	84.0%	0%	16.0%	1.5%	4.0%	78.5%
						Loss to SMAQP2					
	0,6	100%	73.5%	100%	68.5%	NA	100%	66.0%	100%	69.5%	NA
	1,6	100%	83.5%	100%	90.5%	NA	99.5%	77.5%	99.5%	94.5%	NA
	3,6	100%	88.5%	100%	91.0%	NA	100%	86.0%	100%	94.0%	NA
	0,10	100%	74.5%	100%	76.0%	NA	100%	70.0%	100%	72.5%	NA
	1,10	99%	85.0%	99.0%	91.0%	NA	100%	84.0%	100%	95.0%	NA
	3,10	100%	89.5%	100%	93.0%	NA	98.5%	84.0%	98.5%	94.5%	NA
						Average OAQPE					
	0,6	0.738	0.467	0.742	0.456	0.455	0.754	0.478	0.758	0.469	0.468
	1,6	0.633	0.486	0.634	0.472	0.466	0.619	0.485	0.621	0.472	0.465
	3,6	0.623	0.475	0.622	0.467	0.460	0.617	0.477	0.617	0.468	0.461
	0,10	0.748	0.476	0.749	0.463	0.461	0.755	0.479	0.754	0.468	0.467
	1,10	0.632	0.481	0.632	0.464	0.458	0.619	0.486	0.617	0.471	0.464
	3,10	0.630	0.482	0.627	0.473	0.466	0.629	0.479	0.625	0.468	0.462
						Winning Ratio					
	0,6	0%	16.5%	0%	25%	58.5%	0%	18.5%	0%	22.0%	59.0%
	1,6	0%	6.5%	0%	12.5%	81.0%	0%	5.0%	0%	5.0%	89.0%
case ii	3,6	0%	7.5%	0%	9.5%	83.0%	0%	2.0%	0.5%	8.0%	88.5%
	0,10	0%	17.5%	0%	24.0%	58.0%	0%	17.0%	0%	26.5%	56.0%
	1,10	0%	6.5%	0%	10.0%	83.5%	0%	9.5%	0%	6.0%	84.5%
	3,10	0%	8.0%	0%	11.0%	81.0%	0%	8.0%	0%	9.2%	82.5%
						Loss to SMAQP2					
	0,6	100%	83.0%	100%	71.0%	NA	100%	80.5%	100%	69.0%	NA
	1,6	100%	93.5%	100%	86.5%	NA	100%	94.5%	100%	93.0%	NA
	3,6	100%	91.0%	100%	89.0%	NA	98.5%	97.0%	98.5%	91.0%	NA
	0,10	100%	79.0%	100%	70.0%	NA	100%	80.5%	100%	67.5%	NA
	1,10	100%	93.5%	100%	89.0%	NA	100%	89.5%	100%	93.0%	NA
	3,10	100%	92.0%	100%	85.5%	NA	100%	92.0%	100%	91.0%	NA
						Average OAQPE					
	0,6	0.705	0.401	0.709	0.395	0.394	0.714	0.408	0.719	0.400	0.399
	1,6	0.586	0.427	0.588	0.416	0.409	0.588	0.423	0.589	0.411	0.405
	3,6	0.580	0.412	0.579	0.402	0.395	0.580	0.422	0.580	0.416	0.408
	0,10	0.716	0.420	0.717	0.408	0.407	0.716	0.416	0.717	0.405	0.403
	1,10	0.582	0.423	0.581	0.412	0.406	0.582	0.429	0.581	0.414	0.408
	3,10	0.590	0.423	0.587	0.414	0.407	0.583	0.432	0.58	0.423	0.416
						Winning Ratio					
	0,6	0%	26.5%	0%	22.5%	51.0%	0%	22.0%	0%	20.0%	58.0%
	1,6	0%	11.0%	0%	6.0%	83.0%	0%	8.5%	0%	9.5%	82.0%
case iii	3,6	0%	7.5%	0%	6.0%	86.5%	0%	8.5%	0%	4.0%	87.0%
	0,10	0%	13.5%	0%	27.0%	59.0%	0%	17.5%	0%	23.5%	58.5%
	1,10	0%	12.0%	0%	10.0%	80.0%	0%	7.5%	0%	6.5%	85.5%
	3,10	0%	5.0%	0%	9.0%	86.0%	0%	4.0%	0%	5.0%	91.0%
						Loss to SMAQP2					
	0,6	100%	71.5%	100%	67.5%	NA	100%	77.0%	100%	70.5%	NA
	1,6	100%	88.5%	100%	94.0%	NA	100%	91.0%	100%	88.5%	NA
	3,6	100%	92.5%	100%	93.5%	NA	100%	91.0%	100%	93.5%	NA
	0,10	100%	85.5%	100%	67.5%	NA	100%	81.0%	100%	70.5%	NA
	1,10	100%	87.0%	100%	88.5%	NA	100%	92.0%	100%	92.0%	NA
	3,10	100%	93.5%	100%	90.5%	NA	100%	96.0%	100%	94.0%	NA

Statistica Sinica: Preprint
doi:10.5705/ss.202023.0092

e.g., Hu and Xia (2012), but existing literature focused on studying parametric or nonparametric function estimation and hypothesis testing based on a single model. However, the goal of this paper is to analyze how the listed factors affect the quantiles of house prices in Boston and obtain accurate predictions. Following the previous studies, we take *medv* as the response and use the following 13 explanatory variables as covariates, including *crim*, *indus*, *nox*, *rm*, *age*, *dis*, *tax*, *ptratio*, *b*, *lstat*, *zn*, *chas* and *rad*. Note that the first ten variables are continuous and the last three are discrete. More descriptions of this data can refer to the R package "mlbench". Each continuous covariate is standardized with mean zero and variance one before any analysis is implemented.

It is well known that any continuous covariate can serve as a nonparametric component when one fits a PLAM. To address model structure uncertainty, we adopt the model averaging idea to address this issue. The strategy described in the subsection 4.2 is utilized to prepare semiparametric candidate models, yielding 10 and 31 candidate models for SMAQP1 and SMAQP2 respectively.

To investigate the numerical performance of our approach for different basis functions, we still consider $\boldsymbol{b}(\tau)^{(1)}$, $\boldsymbol{b}(\tau)^{(2)}$ and $\boldsymbol{b}(\tau)^{(3)}$ in the subsection 4.1. Tables S4 and S5 in the Supplementary Material report the estimates of the model weights, their standard errors, and the 95% confidence intervals for the model weights under different choices of basis functions. The standard errors and confidence intervals are computed based on a bootstrap method given in Section S1.1 of the

Statistica Sinica: Preprint
doi:10.5705/ss.202023.0092

Supplementary Material with 400 bootstrap samples. From Tables S4 and S5, we see that the model weights obtained by different basis functions might be different. For example, SMAQP2 with $\boldsymbol{b}(\tau)^{(1)}$ assigns relatively large weights to the 1st, 7th, 24th and 28th sub-models but the estimated weights with $\boldsymbol{b}(\tau)^{(3)}$ for the 6th, 7th, 21th and 24th sub-models are relatively large. Though our proposed model averaging obtained by different basis functions yields different model weights, they are robust against misspecification of the basis functions, as they have similar prediction performances (see Figure S4). In what follows, we fix the basis function as $\boldsymbol{b}(\tau) = (1, \Phi^{-1}(\tau))^{\top}$ to evaluate the out-of-sample predictive performance.

We also plot the estimated regression coefficients and the corresponding 95% pointwise confidence intervals for the coefficients of the 5th candidate model and the 28th candidate model in Figures S5 and S6 of the Supplementary Material, respectively. We can clearly see that almost all estimated coefficients are quite smooth over the quantile levels, indicating that modeling the quantile regression coefficients as unknown functions of the quantile levels is reasonable in practice. The regression coefficients of other candidate models show similar patterns, so we omit their plots to save space.

To evaluate the predictive performance of various methods, following Lu and Su (2015), we randomly choose n_{train} samples from the data set to estimate all parameters and nonparametric functions in candidate models, as well as the model weights, and calculate the OAQPE based on the remaining n_{test} test samples. We

Statistica Sinica: Preprint
doi:10.5705/ss.202023.0092

Table 2: The average OAQPE, winning ratio and loss to SMAQP2 of various approaches for the Boston housing price data.

n_{test}	QRCM	PAQRM	JQLMA	SSMAQP1	SSMAQP2	SMAQP1	SMAQP2
				Average OAQPE			
100	1.240	1.269	1.247	1.098	1.067	1.077	1.041
200	1.249	1.131	1.255	1.042	1.021	1.022	0.997
300	1.269	1.204	1.275	1.085	1.070	1.068	1.043
				Winning Ratio			
100	5.0%	0.0%	2.0%	2.5%	2.0%	0.0%	78.0%
200	0.0%	1.0%	0.0%	2.5%	2.5%	1.5%	82.0%
300	0.0 %	0.5%	0.0%	3.0%	5.0%	2.5%	66.0%
				Loss to SMAQP2			
100	93.0%	100%	93.0%	95.5%	96.5%	88.0%	NA
200	100%	97.5%	100%	93.5%	94.5%	86.0%	NA
300	100%	99.5%	100%	88.0%	92.0%	71.5%	NA

repeat the random splitting procedure 200 times, and compute the mean of the obtained 200 OAQPEs for each method. To make a comprehensive comparison, we still compare the proposed SMAQP1 and SMAQP2 with the same approaches in simulations, including QRCM, PAQRM, JQLMA, SSMAQP1 and SSMAQP2, where SSMAQP1 (or SSMAQP2) is the model averaging approach based on the standard quantile loss function for a single quantile with candidate models constructed by *Strategy A* (or *Strategy B*). We set the size n_{test} of the test set to be 100, 200, and 300, respectively.

Table 2 summarizes prediction results at the entire quantile process over 200 repetitions. It is clear that SMAQP1 (or SMAQP2) is uniformly better than SSMAQP1 (or SSMAQP2) in terms of OAQPEs, suggesting that it is beneficial for improving the accuracy of prediction to consider a combination of modeling quantile functions parametrically and the integrated loss. Moreover, Table 2 also shows that SMAQP2 outperforms other compared methods uniformly due to the relatively high winning ratio and Loss to SMAQP2. Figure 1 describes the standard

Statistica Sinica: Preprint
doi:10.5705/ss.202023.0092

Table 3: The average FPE_τ, winning ratio and loss to SMAQP2 of various approaches at $\tau = 0.1, 0.3, 0.5, 0.7, 0.9$ for the Boston housing price data.

n_{test}	τ	QRCM	PAQRM	JQLMA	SSMAQP1	SSMAQP2	SMAQP1	SMAQP2
					Average FPE_τ			
100	0.1	0.599	0.900	0.601	0.675	0.635	0.667	0.628
	0.3	1.268	1.378	1.281	1.231	1.188	1.200	1.159
	0.5	1.629	1.579	1.633	1.441	1.398	1.394	1.354
	0.7	1.635	1.510	1.669	1.313	1.281	1.301	1.263
	0.9	1.075	1.073	1.100	0.877	0.848	0.828	0.786
200	0.1	0.597	0.737	0.597	0.636	0.614	0.617	0.583
	0.3	1.257	1.207	1.266	1.141	1.119	1.135	1.107
	0.5	1.615	1.406	1.616	1.335	1.318	1.33	1.311
	0.7	1.621	1.341	1.654	1.247	1.225	1.244	1.222
	0.9	1.068	0.909	1.096	0.843	0.822	0.794	0.761
300	0.1	0.616	0.792	0.617	0.646	0.630	0.627	0.596
	0.3	1.293	1.264	1.301	1.179	1.166	1.166	1.140
	0.5	1.658	1.461	1.667	1.364	1.361	1.378	1.361
	0.7	1.670	1.421	1.707	1.294	1.288	1.297	1.281
	0.9	1.128	1.011	1.136	0.902	0.882	0.849	0.820
					Winning Ratio			
100	0.1	27.0%	1.0%	21.5%	3.0%	14.5 %	0.5%	25.5%
	0.3	11.5%	5.5%	5.5%	9.0%	20.0%	1.0%	40.5%
	0.5	3.0%	5.0%	6.5%	8.0%	8.5%	2.0%	58.5%
	0.7	1.0%	5.5%	2.0%	12.5%	23.0%	3.0%	48.0%
	0.9	2.0%	7.0%	0.0%	11.0%	13.5%	0.5%	55.0%
200	0.1	18.0%	2.5%	18.0%	2.0%	8.0%	0.5%	41.0%
	0.3	1.5%	5.5%	3.0%	14.5%	21.0%	0.0%	43.5%
	0.5	0.0%	6.0%	0.5%	17.0%	22.5%	3.5%	42.5%
	0.7	0.0%	5.5%	0.5%	12.0%	29.5%	6.0%	40.5%
	0.9	0.0%	10.5%	0.0%	9.5%	8.0%	0.5%	60.0%
300	0.1	11.5%	0.0%	10.0%	0.5%	11.5%	0.5%	51.5%
	0.3	0.5%	4.0%	2.5%	10.0%	15.5%	2.0%	53.0%
	0.5	0.0%	8.0%	0.0%	27.5%	24.0%	3.5%	27.0%
	0.7	0.0%	6.5%	0.0%	20.5%	24.5%	5.0%	32.5%
	0.9	0.5%	4.5%	0.0%	7.0%	6.5%	3.5%	58.5%
					Loss to SMAQP2			
100	0.1	43.5%	93.0%	45.0%	77.5%	62.5%	84.0%	NA
	0.3	81.0%	92.5%	82.5%	80.0%	66.5%	84.5%	NA
	0.5	94.0%	92.5%	91.0%	85.0%	81.5%	83.0%	NA
	0.7	97.5%	93.5%	98.0%	77.5%	63.5%	79.5%	NA
	0.9	97.5%	93.0%	98.5%	83.5%	83.0%	84.5%	NA
200	0.1	64.5%	95.0%	65.5%	89.5%	81.0%	83.0%	NA
	0.3	95.5%	90.5%	95.5%	78.0%	66.5%	77.5%	NA
	0.5	99.5%	86.0%	99.0%	63.0%	59.0%	69.0%	NA
	0.7	99.5%	90.5%	99.5%	71.5%	53.0%	73.0%	NA
	0.9	99.5%	89.0%	100%	86.5%	88.0%	81.5%	NA
300	0.1	79.0%	99.5%	77.0%	91.0%	84.0%	80.0%	NA
	0.3	98.0%	92.5%	96.5%	76.5%	73.5%	75.5%	NA
	0.5	99.5%	82.0%	100%	49.5%	50.5%	65.5%	NA
	0.7	100%	87.5%	100%	61.5%	54.0%	64.5%	NA
	0.9	98.5%	94.0%	99.5%	86.0%	86.5%	69.5%	NA

Statistica Sinica: Preprint
doi:10.5705/ss.202023.0092

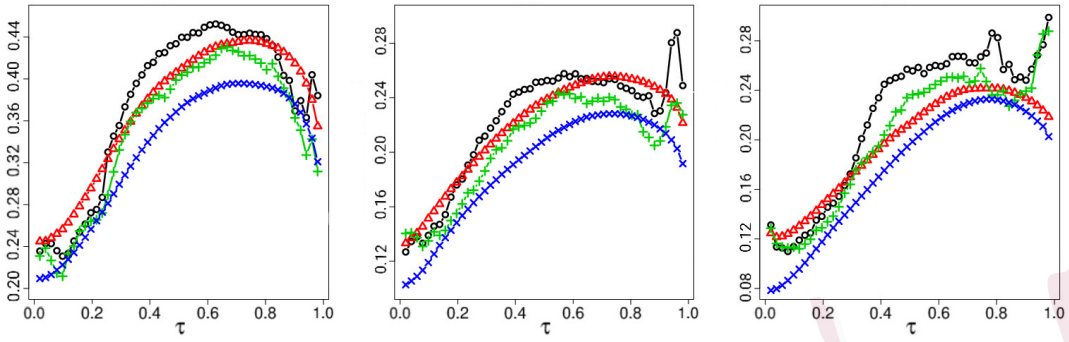

Figure 1: The standard deviations of FPE_τ for SSMAQP1 (black), SSMAQP2(green), SMAQP1(red) and SMAQP2(blue) at various quantiles with the test sample size $n_{test} = 100$ (left), 200 (middle) and 300 (right).

deviations of FPE_τ defined in S1.1 in the Supplementary Material for SSMAQP1, SSMAQP2, SMAQP1 and SMAQP2 at different quantiles. Figure 1 reveals that SMAQP1 (or SMAQP2) produces more stable and reliable predictions than that of SSMAQP1 (or SSMAQP2) as it has smaller standard deviations. Table 3 displays the average FPE_τ, winning ratio and loss to SMAQP2 of various approaches at given quantile levels $\tau = 0.1, 0.3, 0.5, 0.7, 0.9$. We can see that the proposed approach outperforms other methods at a single quantile level. Overall, the empirical example confirms that our semiparametric model averaging provides a more accurate quantile prediction of house prices in Boston.

6. Concluding Remarks

We develop a novel model averaging procedure of PLAMs for optimal conditional quantile prediction. The proposed approach is more robust against model misspecification and leads to a better prediction than the estimation method based

Statistica Sinica: Preprint
doi:10.5705/ss.202023.0092

on a single PLAM. Different from popular regularization methods that focus on identifying a single optimal PLAM, we propose a new strategy via model averaging to account for the model structure uncertainty of PLAMs and improve the accuracy of the conditional quantile prediction. We have shown the asymptotic optimality of the proposed model averaging method when the number of sub-models diverges to infinity as the sample size increases. Extensive simulation studies illustrate that our proposed method performs better than competing methods in most situations. The empirical application of quantile prediction of the Boston housing price also demonstrates that the new method has a good numerical performance in practice.

Note that the proposed SMAQP may have the quantile crossing problem, i.e., the estimator $\hat{\mu}^{[\hat{\boldsymbol{w}}]}(\boldsymbol{X}_i, \tau)$ given in Section 2.2 may not be nondecreasing in τ. The quantile crossing is a well-known problem in quantile regression and special treatment is often needed to solve it. Note that we treat each unknown coefficient as a smooth function of τ, which is approximated by a B-spline function. By a property given in Section 3 of Ma (2016), the B-spline function is monotone nondecreasing if the spline coefficients are ordered. Therefore, one possible way to address the quantile crossing problem in our setting is to estimate the spline coefficients subject to an ordering constraint or rearrange the estimated coefficients. Alternatively, we can also take the rearrangement strategy given in Chernozhukov et al. (2010). The theoretical, computational, and numerical investigation of quantile

Statistica Sinica: Preprint
doi:10.5705/ss.202023.0092

non-crossing deserves further thorough investigation and thus it is considered as a future research topic to explore.

In numerical analysis, we use a bootstrap method to calculate the standard errors of the weight estimator $\hat{\boldsymbol{w}}$ and construct confidence intervals for the model weights, since it is difficult to derive the asymptotic distribution of the weight estimator. Bootstrapping is also used in other model averaging studies for inference, see, for example, Li et al. (2018). Although the bootstrap method has reliable numerical performance, as demonstrated by our simulation studies, its theoretical properties in our setting need further investigation in future research studies.

Supplementary Material

The supplementary material provides the bootstrap procedure for the model weight vector, a comprehensive comparison of SMAQP and SSMAQP, a sensitivity analysis of SMAQP for $\lambda_n = 1$ and $\log(n)/2$, additional simulation results, and three technical lemmas and proofs of the lemmas, Proposition 1 and Theorem 1.

Acknowledgments

The authors sincerely thank the editor, the associate editor, and the referees for their constructive suggestions and comments. The research of Ma is supported in part by U.S. NSF grants DMS-20-14221, DMS-23-10288, and DMS-24-12922. The research of Lv is supported by Natural Science Foundation of Chongqing Grant

Statistica Sinica: Preprint
doi:10.5705/ss.202023.0092

(Grants No. cstc2021jcyj-msxmX0502 and CSTB2022NSCQ-MSX0852), Fundamental Research Funds for the Central Universities (Grant No. SWU-KU24002) and National Statistical Science Research Program (Grant No. 2022LY019).

References

Chernozhukov, V., Fernández-Val, I., and Galichon, A. (2010). Quantile and probability curves without crossing. *Econometrica*, 78:1093–1125.

Fang, F., Li, J., and Xia, X. (2022). Semiparametric model averaging prediction for dichotomous response. *Journal of Econometrics*, 229:219–245.

Fang, F., Yuan, C., and Tian, W. (2023). An asymptotic theory for least squares model averaging with nested models. *Econometric Theory*, 39:412–441.

Frumento, P. and Bottai, M. (2016). Parametric modeling of quantile regression coefficient functions. *Biometrics*, 72:74–84.

Frumento, P., Bottai, M., and Fernandez-Val, I. (2021). Parametric modeling of quantile regression coefficient functions with longitudinal data. *Journal of the American Statistical Association*, 116:783–797.

Hansen, B. and Racine, J. (2012). Jackknife model averaging. *Journal of Econometrics*, 167:38–46.

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

Hu, T. and Xia, Y. (2012). Adaptive semi-varying coefficient model selection. *Statistica Sinica*, 22:575–599.

Huang, J., Wei, F., and Ma, S. (2012). Semiparametric regression pursuit. *Statistica Sinica*, 22:1403–

Statistica Sinica: Preprint
doi:10.5705/ss.202023.0092

1426.

Huang, J., Zhang, L., and Zhou, L. (2007). Efficient estimation in marginal partially linear models for longitudinal/clustered data using splines. *Scand. J. Statist.*, 34:451–477.

Koenker, R. (2005). *Quantile regression*. Cambridge University Press, Cambridge.

Lee, J. and Shin, Y. (2023). Complete subset averaging for quantile regressions. *Econometric Theory*, 39:146–188.

Li, J., Lv, J., Wan, A., and Liao, J. (2022). Adaboost semiparametric model averaging prediction for multiple categories. *Journal of the American Statistical Association*, 117:495–509.

Li, J., Xia, X., Wong, W., and Nott, D. (2018). Varying-coefficient semiparametric model averaging prediction. *Biometrics*, 74:1417–1426.

Li, J., Yu, T., Lv, J., and Lee, M. (2021). Semiparametric model averaging prediction for lifetime data via hazards regression. *Journal of the Royal Statistical Society, Series C*, 70:1187–1209.

Lian, H. (2012). Semiparametric estimation of additive quantile regression models by two-fold penalty. *Journal of Business and Economic Statistics*, 30:337–350.

Lian, H., Liang, H., and Ruppert, D. (2015). Separation of covariates into nonparametric and parametric parts in high-dimensional partially linear additive models. *Statistica Sinica*, 25:591–607.

Lou, Y., Bien, J., Caruana, R., and Gehrke, J. (2016). Sparse partially linear additive models. *Journal of Computational and Graphical Statistics*, 25:1126–1140.

Lu, X. and Su, L. (2015). Jackknife model averaging for quantile regressions. *Journal of Econometrics*, 188:40–58.

Statistica Sinica: Preprint
doi:10.5705/ss.202023.0092

Ma, S. (2012). Two-step spline estimating equations for generalized additive partially linear models with large cluster sizes. *The Annals of Statistics*, 40:2943–2972.

Ma, S. (2016). Estimation and inference in functional single-index models. *Annals of the Institute of Statistical Mathematics*, 68:181–208.

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

Ma, S., Song, Q., and Wang, L. (2013). Simultaneous variable selection and estimation in semiparametric modeling of longitudinal/clustered data. *Bernoulli*, 19:252–274.

Seng, L. and Li, J. (2022). Structural equation model averaging: methodology and application. *Journal of Business and Economic Statistics*, 40:815–828.

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

Sun, Z., Sun, L., Lu, X., Zhu, J., and Li, Y. (2017). Frequentist model averaging estimation for the censored partial linear quantile regression model. *Journal of Statistical Planning and Inference*, 189:1–15.

Wan, A., Zhang, X., and Zou, G. (2010). Least squares model averaging by mallows criterion. *Journal of Econometrics*, 156:277–283.

Wang, L., Liu, X., Liang, H., and Carroll, R. (2011). Estimation and variable selection for generalized additive partial linear models. *The Annals of Statistics*, 39:1827–1851.

Wang, L., Xue, L., Qu, A., and Liang, H. (2014). Estimation and model selection in generalized additive partial linear models for correlated data with diverging number of covariates. *The Annals of*

Statistica Sinica: Preprint
doi:10.5705/ss.202023.0092

Statistics, 42:592–624.

Wang, M., Zhang, X., Wan, A.-K., and You, K. (2023). Jackknife model averaging for high-dimensional quantile regression. *Biometrics*, 79:178–189.

Wong, R. K. W., Li, Y., and Zhu, Z. (2019). Partially linear functional additive models for multivariate functional data. *Journal of the American Statistical Association*, 114:406–418.

Xue, L., Qu, A., and Zhou, J. (2010). Consistent model selection for marginal generalized additive model for correlated data. *Journal of the American Statistical Association*, 105:1518–1530.

Zhang, H., Cheng, G., and Liu, Y. (2011). Linear or nonlinear? automatic structure discovery for partially linear models. *Journal of the American Statistical Association*, 106:1099–1112.

Zhang, X. and Liu, C.-A. (2018). Inference after model averaging in linear regression models. *Econometric Theory*, 35:816–841.

Zhang, X. and Wang, W. (2019). Optimal model averaging estimation for partially linear models. *Statistica Sinica*, 29:693–718.

Zhang, X., Zou, G., Liang, H., and Carroll, R. (2020). Parsimonious model averaging with a diverging number of parameters. *Journal of the American Statistical Association*, 115:972–984.

School of Mathematics and Statistics, Southwest University, Chongqing, China

E-mail: (lvjing@swu.edu.cn)

Department of Statistics, University of California, Riverside, USA

E-mail: (shujie.ma@ucr.edu)