AIMS Mathematics

http://www.aimspress.com/journal/Math

AIMS Mathematics, 7(5): 7657–7684.
DOI:10.3934/math.2022430
Received: 01 December 2021
Revised: 19 January 2022
Accepted: 24 January 2022
Published: 16 February 2022

Research article

Weighted composite asymmetric Huber estimation for partial functional linear models

Juxia Xiao[1,2], **Ping Yu**[2] **and Zhongzhan Zhang**[1,*]

[1] Faculty of Science, Beijing University of Technology, Beijing 100124, China
[2] School of Mathematics and Computer Science, Shanxi Normal University, Taiyuan 030000, China

* **Correspondence:** Email: zzhang@bjut.edu.cn.

Abstract: In this paper, we first investigate a new asymmetric Huber regression (AHR) estimation procedure to analyze skewed data with partial functional linear models. To automatically reflect distributional features as well as bound the influence of outliers effectively, we further propose a weighted composite asymmetric Huber regression (WCAHR) estimation procedure by combining the strength across multiple asymmetric Huber loss functions. The slope function and constant coefficients are estimated through minimizing the combined loss function and approximating the slope function with principal component analysis. The asymptotic properties of the proposed estimators are derived. To realize the WCAHR estimation, we also develop a practical algorithm based on pseudo data. Numerical results show that the proposed WCAHR estimators can well adapt to the different error distributions, and thus are more useful in practice. Two real data examples are presented to illustrate the applications of the proposed methods.

Keywords: functional principal component analysis; partial functional linear model; asymmetric Huber regression; weighted composite asymmetric Huber regression; functional data analysis
Mathematics Subject Classification: 62G05, 62G20

1. Introduction

Functional data analysis (FDA) (e.g., [1]) has drawn considerable attention over recent years, owing to a great deal of flexibilities and universal applications in handling high-dimensional data sets. A fundamental and important tool for FDA is functional linear models.

There are a lot of researches in literature on the inference of functional linear models and their extensions, see, among others, [2–4] for earlier works, and [5–10] for recent works. As is well known, in the estimation of regression models, the choice of loss function is essential to obtain a highly efficient and robust estimator. Most of earlier works employed the square loss function and obtained ordinary least squares (OLS) estimators. In recent years, many other loss functions have been considered in the estimation of functional linear models and their extensions. Kato [6], Tang and Cheng [11] studied the quantile regression (QR) with functional linear models and partial functional linear models, respectively. Yu et al. [12] proposed a robust exponential squared loss estimation procedure (ESL) and established the asymptotic properties of the proposed estimators. Cai et al. [13] introduced a new robust estimation procedure by employing a modified Huber function, whose tail function is replaced by the exponential squared loss (H-ESL) in the partial functional linear model.

It is well known that the square loss function pays attention to reflect the distributional features of the entire distribution, whereas QR, ESL and H-ESL methods focus on bounding the influence of outliers when the data are heavy-tailed, respectively. Thus, developing a method, which can both reflect distributional features and bound outliers effectively, is highly desirable in data analysis. We note that, in the context of principal component analysis (PCA), Lim and Oh [14] proposed a new approach using a weighted linear combination of asymmetric Huber loss functions to demonstrate the distributional features of data as well as keep robust to outliers. The asymmetric Huber loss functions are defined as

$$\rho_\tau(u) = \begin{cases} (\tau - 1)(u + 0.5c^*) & \text{for } u < -c^* \\ 0.5(1-\tau)u^2/c^* & \text{for } -c^* \le u < 0 \\ 0.5\tau u^2/c^* & \text{for } 0 \le u < c^* \\ \tau(u - 0.5c^*) & \text{for } c^* \le u, \end{cases} \tag{1.1}$$

with $c^* = 1.345$, and $\tau \in (0, 1)$ being a parameter to control the degree of skewness. The function $\rho_\tau(\cdot)$ is equivalent to the Huber loss function (see, [15]) when τ is equal to 0.5 and is most exactly the same as the quantile loss function when c^* is small enough.

Motivated by the appealing characteristics of the asymmetric Huber functions, in this paper, we first investigate a new asymmetric Huber regression (AHR) estimation procedure to analyse skewed data for the partial functional linear model, based on the functional principal component analysis. To improve the estimation accuracy for single AHR estimation, we develop a weighted composite asymmetric Huber regression (WCAHR) estimation by combining the strength across multiple asymmetric Huber regression models. A practical algorithm for WCAHR estimators based on pseudo data is developed to implement the estimation method. The asymptotic properties of the proposed estimators are also derived. Extensive simulations are carried out to show the superiority of the proposed estimators.

Finally, we apply the proposed methods to two data sets. In the first example, we analyze the electricity data. Figure 1 presents the estimated density of the residuals and the residual diagnostic plot obtained by fitting the model (4.1) in Section 4.1 via the OLS method. The distribution of the residuals is skewed, bimodal, and there are some outliers in the dataset. Given that the WCAHR can effectively manage such data, we use the proposed method to conduct an analysis to this data set. Another example in Section 4 considers the Tecator data set. Similarly, Figure 2 presents the density of the residuals and the residual diagnostic plot obtained by fitting the model (4.2) in Section 4.2 via the OLS method, which demonstrates that the distribution of the residuals is skewed and far from normality. Undoubtedly, WCAHR regression is applicable to analyzing this data set on account of its appealing features.

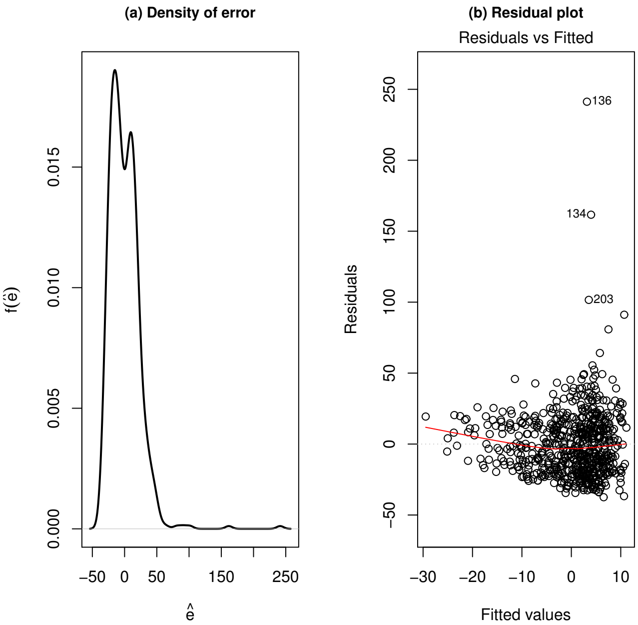


Figure 1. (a) The density of estimated errors for Electricity data; (b) the residual plot for Electricity data.

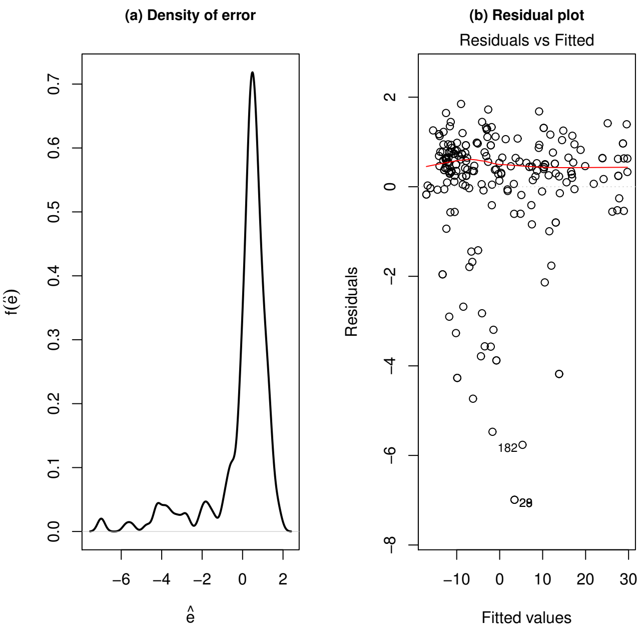


Figure 2. (a) The density of estimated errors for Tecator data; (b) the residual plot for Tecator data.

To our knowledge, it is the first to discuss the asymmetric Huber regression problems under functional models framework. The proposed WCAHR method possesses advantages that include the robustness to outliers as well as reflecting the relationships between potential explanatory variables and the entire distribution of response. It retains the advantages in analysing skewed data and the obtained estimators rely on the shape of the entire distribution rather than merely on the data nearby a specific quantile level or skewness level of the asymmetric Huber loss, thereby avoiding the limitations of these methods. These advantages are revealed by both theoretical conclusions and numerical results. The relevant algorithm is data-adaptive, and capable of reflecting the distributional features of the data without prior information, and is robust to outliers.

The rest of this paper is organized as follows. In Section 2, we formally describe the estimation procedures, and develop a new algorithm. We also establish the asymptotic behaviors of the proposed estimators as well as a list of technical assumptions needed in the theoretical research. In Section 3, the finite performances of the proposed estimators are evaluated through simulations. Section 4 illustrates the use of the proposed methods in the analyses of electricity data and Tecator data. Brief conclusions on the proposed methods are made in Section 5. All technique proofs are provided in Section A.

2. Methodology and main results

2.1. Proposed methods

Let Y be a real value random variable, $\mathbf{Z} = (Z_1, \cdots, Z_p)^{\mathrm{T}}$ be a p dimensional random vector with zero mean and finite second moment. Let $\{X(t) : t \in \mathscr{T}\}$ be a zero mean, second-order stochastic process with sample paths in $L_2(\mathscr{T})$, which consists of square integrable functions with inner product $\langle x, y\rangle = \int_{\mathscr{T}} x(t)y(t)dt$ and norm $\|x\| = \langle x, x\rangle^{1/2}$, respectively, here $\mathscr{T}$ is a bounded closed interval. Without loss of generality, we suppose $\mathscr{T} = [0, 1]$ throughout the paper. The dependence between Y and $(X, \mathbf{Z})$ is expressed by the partial functional linear regression as following,

$$Y = \mathbf{Z}^{\mathrm{T}}\boldsymbol{\alpha} + \int_0^1 \beta(t)X(t)dt + e. \tag{2.1}$$

Here, random error e is assumed to be independent of $\mathbf{Z}$ and X, $\boldsymbol{\alpha} = (\alpha_1, \cdots, \alpha_p)^{\mathrm{T}}$ is an unknown p-dimensional parameter vector, and the slope function $\beta(\cdot)$ is an unknown square integrable function on $[0, 1]$.

Let $(\mathbf{Z}_i, X_i(\cdot), Y_i), i = 1, \cdots, n$, be independent observations generated by model (2.1) and let $e_i = Y_i - \mathbf{Z}_i^{\mathrm{T}}\boldsymbol{\alpha} - \int_0^1 \beta(t)X_i(t)dt, i = 1, \cdots, n$. The covariance and empirical covariance functions for $X(\cdot)$ are defined as $c_X(t, s) = \mathrm{Cov}(X(t), X(s))$, $\hat{c}_X(t, s) = \frac{1}{n}\sum_{i=1}^n X_i(t)X_i(s)$ respectively. Based on the Mercer's Theorem, c_X and $\hat{c}_X$ can be represented as following,

$$c_X(t, s) = \sum_{i=1}^{\infty} \lambda_i v_i(t)v_i(s), \qquad \hat{c}_X(t, s) = \sum_{i=1}^{\infty} \hat{\lambda}_i\hat{v}_i(t)\hat{v}_i(s),$$

where $\lambda_1 > \lambda_2 > \cdots > 0$ and $\hat{\lambda}_1 \geq \hat{\lambda}_2 \geq \cdots \geq \hat{\lambda}_{n+1} = \cdots = 0$ are each the ordered eigenvalue sequences of the covariance operator C_X and its estimator $\hat{C}_X$ with kernels c_X and $\hat{c}_X$, which are defined by $C_X f(s) = \int_0^1 c_X(t, s)f(t)dt$ and $\hat{C}_X f(s) = \int_0^1 \hat{c}_X(t, s)f(t)dt$ with C_X being assumed strictly positive, and $\{v_i(\cdot)\}$ and $\{\hat{v}_i(\cdot)\}$ are the corresponding orthonormal eigenfunction sequences. Besides, $(\hat{v}_i(\cdot), \hat{\lambda}_i)$ is treated as an estimator of $(v_i(\cdot), \lambda_i)$.

Similarly, we can define $c_{YX}(\cdot) = \text{Cov}(Y, X(\cdot))$, $c_{\boldsymbol{Z}} = \text{Var}(\boldsymbol{Z}) = E[\boldsymbol{Z}\boldsymbol{Z}^{\text{T}}]$, $c_{\boldsymbol{Z}Y} = \text{Cov}(\boldsymbol{Z}, Y)$, $c_{\boldsymbol{Z}X}(\cdot) = \text{Cov}(\boldsymbol{Z}, X(\cdot)) = (c_{Z_1X}(\cdot), \cdots, c_{Z_pX}(\cdot))^{\text{T}}$. And the corresponding empirical counterparts defined below can be used as their estimators,

$$\hat{c}_{YX} = \frac{1}{n}\sum_{i=1}^{n} Y_i X_i, \qquad \hat{c}_{\boldsymbol{Z}} = \frac{1}{n}\sum_{i=1}^{n} \boldsymbol{Z}_i \boldsymbol{Z}_i^{\text{T}},$$
$$\hat{c}_{\boldsymbol{Z}Y} = \frac{1}{n}\sum_{i=1}^{n} \boldsymbol{Z}_i Y_i, \qquad \hat{c}_{\boldsymbol{Z}X} = \frac{1}{n}\sum_{i=1}^{n} \boldsymbol{Z}_i X_i.$$

By the Karhunen-Loève representation, $X_i(t)$ and $\beta(t)$ can be expanded into

$$\beta(t) = \sum_{j=1}^{\infty} \gamma_j v_j(t), \qquad X_i(t) = \sum_{j=1}^{\infty} \xi_{ij} v_j(t), \qquad i = 1, \cdots, n, \tag{2.2}$$

here $\gamma_j = \langle \beta(\cdot), v_j(\cdot) \rangle = \int_0^1 \beta(t) v_j(t) dt$, and $\xi_{ij} = \langle X_i(\cdot), v_j(\cdot) \rangle$.

Owing to the orthogonality of $\{v_1(\cdot), ..., v_m(\cdot)\}$ and Eq (2.2), Model (2.1) can be transformed into

$$Y_i = \boldsymbol{Z}_i^{\text{T}}\boldsymbol{\alpha} + \sum_{j=1}^{m} \gamma_j \xi_{ij} + \widetilde{e}_i = \boldsymbol{Z}_i^{\text{T}}\boldsymbol{\alpha} + \boldsymbol{U}_i^{\text{T}}\boldsymbol{\gamma} + \widetilde{e}_i, \quad i = 1, \cdots, n,$$

where $\boldsymbol{U}_i = (\xi_{i1}, ..., \xi_{im})^{\text{T}}$, $\boldsymbol{\gamma} = (\gamma_1, ..., \gamma_m)^{\text{T}}$, $\widetilde{e}_i = \sum_{j=m+1}^{\infty} \gamma_j \xi_{ij} + e_i$, and the tuning parameter m may increase with the sample size n.

Replacing $v_j(\cdot)$ with its estimator $\hat{v}_j(\cdot)$, the τth AHR estimators $\bar{\boldsymbol{\alpha}}$ and $\bar{\beta}(t) = \sum_{j=1}^{m} \bar{\gamma}_j \hat{v}_j(t)$ can be obtained by minimizing the loss function over $\boldsymbol{\alpha}$, $\boldsymbol{\gamma}$ and b_τ as follows:

$$(\bar{b}, \bar{\boldsymbol{\alpha}}, \bar{\boldsymbol{\gamma}}) \triangleq \underset{(b_\tau, \boldsymbol{\alpha}, \boldsymbol{\gamma})}{\text{argmin}} \sum_{i=1}^{n} \rho_\tau\Big(Y_i - b_\tau - \boldsymbol{Z}_i^{\text{T}}\boldsymbol{\alpha} - \hat{\boldsymbol{U}}_i^{\text{T}}\boldsymbol{\gamma}\Big),$$

where the asymmetric Huber loss function $\rho_\tau(u)$ is defined in (1.1), and $\hat{\boldsymbol{U}}_i = (\hat{\xi}_{i1}, \cdots, \hat{\xi}_{im})^{\text{T}}$ with $\hat{\xi}_{ij} = \langle X_i(\cdot), \hat{v}_j(\cdot) \rangle$. Here the true value of b_τ is defined as the solution that minimizes the loss function $E\{\rho_\tau(e - \theta)\}$ over $\theta \in \mathbb{R}$, and we call it the τth number of e with respect to (1.1).

Remark 1. *In model* (2.1)*, we suppose the intercept term is zero. In fact, if there is an intercept, we then may absorb it into the distribution of e. Thus, the main impact of model* (2.1) *is finding the contribution of the predictors to the response, and the zero mean assumption for e is not needed.*

Noting that the regression coefficients are the same across different skewness asymmetric Huber regression models, and being inspirited by [14] and [16], we combine the strength across multiple asymmetric Huber regression models and propose a WCAHR method. Specifically, the WCAHR estimators $\hat{\boldsymbol{\alpha}}$ and $\hat{\beta}(t) = \sum_{j=1}^{m} \hat{\gamma}_j \hat{v}_j(t)$ can be obtained by minimizing the following loss function with respect to $(\boldsymbol{b}, \boldsymbol{\alpha}, \boldsymbol{\gamma})$:

$$Q_n(\boldsymbol{b}, \boldsymbol{\alpha}, \boldsymbol{\gamma}) \triangleq \sum_{i=1}^{n} \sum_{k=1}^{K} w_k \rho_{\tau_k}\Big(Y_i - b_k - \boldsymbol{Z}_i^{\text{T}}\boldsymbol{\alpha} - \hat{\boldsymbol{U}}_i^{\text{T}}\boldsymbol{\gamma}\Big),$$

where $\{\tau_k\}_{k=1}^{K}$ are predetermined levels over (0,1), $b_k = b_{\tau_k}$ for brevity, $\boldsymbol{b} = (b_1, \cdots, b_K)^{\text{T}}$, and the weights $w_1, \ldots, w_K$, which control the contribution of each loss function, are positive constants satisfying $\sum_k w_k = 1$.

Remark 2. *Generally speaking, one can choose the equidistant levels as $\tau_k = k/(K+1)$ for $k = 1, 2, \ldots, K$ for a given K, similar to what often has been done in composite quantile regression. Although one can also apply data-adaptive methods, such as cross validation, to select K, we do not pursue this topic here. As for the weights, we consider two choices. The first is using the equal weights $w_1 = \cdots = w_K = 1/K$. The obtained estimators are called composite asymmetric Huber regression (CAHR) estimators. As the second choice in this study, to preferably describe the distribution information of the data, we consider a K-dimensional weight vector $(w_1, \cdots, w_K) = (f(b_{01}), \cdots, f(b_{0K}))\Big/\sum_{k=1}^K f(b_{0k})$, where $\boldsymbol{b}_0 = (b_{01}, \cdots, b_{0K})$ is the true value vector of $\boldsymbol{b}$, and $f(\cdot)$ is the density function of the random error. In practice, we estimate $f(\cdot)$ by kernel density estimation method.*

Denote $\mathcal{S} = \{(Y_i, \boldsymbol{Z}_i, X_i(\cdot)) : 1 \le i \le n\}$, and given a new copy of $(\boldsymbol{Z}, X)$, namely the predictor variables $(\boldsymbol{Z}_{n+1}, X_{n+1}(\cdot))$, once we gain the estimated $\boldsymbol{\alpha}$ and $\beta(t)$, the mean squared prediction error (MSPE) can be obtained, take asymmetric Huber regression for example,

$$\begin{aligned}&\mathrm{MSPE}_{AHR}\\ =&E\left[\left\{\left(\bar{b} + \boldsymbol{Z}_{n+1}^{\mathrm{T}}\bar{\boldsymbol{\alpha}} + \int_0^1 \bar{\beta}(t)X_{n+1}(t)dt\right) - \left(b_\tau + \boldsymbol{Z}_{n+1}^{\mathrm{T}}\boldsymbol{\alpha}_0 + \int_0^1 \beta_0(t)X_{n+1}(t)dt\right)\right\}^2 \middle| \mathcal{S}\right],\end{aligned}$$

where $\boldsymbol{\alpha}_0$ and $\beta_0(t)$ are the true values of $\boldsymbol{\alpha}$ and $\beta(t)$, respectively. The MSPEs of CAHR and WCAHR have analoguos definitions, and denoted by MSPE_{CAHR} and MSPE_{WCAHR}, respectively.

2.2. Computational algorithm

Note that the minimization problems for AHR and CAHR estimators are special cases of WCAHR method. Here, we just present the practical algorithm for WCAHR based on pseudo data. A similar argument can be found in [14] to implement the data-adaptive principal component analysis. The algorithm is simply described as following.

Step 1. Given initial estimators $\hat{\boldsymbol{\alpha}}^{(0)}$ and $\hat{\boldsymbol{\gamma}}^{(0)}$ for $\boldsymbol{\alpha}_0$ and γ_0, respectively.

Step 2. Iterate, until convergence, following these three steps for $L = 0, 1, \ldots$

(a) Compute the residuals as $\hat{e}_i^{(L)} = Y_i - \boldsymbol{Z}_i^{\mathrm{T}}\hat{\boldsymbol{\alpha}}^{(L)} - \hat{\boldsymbol{U}}_i^{\mathrm{T}}\hat{\boldsymbol{\gamma}}^{(L)}$. (b) Calculate the empirical pseudo data vector $\boldsymbol{G}^{(L)} = (G_1^{(L)}, \cdots, G_n^{(L)})^{\mathrm{T}}$ in the element-wise way, $G_i^{(L)} = \boldsymbol{Z}_i^{\mathrm{T}}\hat{\boldsymbol{\alpha}}^{(L)} + \hat{\boldsymbol{U}}_i^{\mathrm{T}}\hat{\boldsymbol{\gamma}}^{(L)} + \sum_{k=1}^K w_k\psi_{\tau_k}(\hat{e}_i^{(L)} - \hat{b}_k^{(L)})$, for given weights $(w_1, \cdots, w_K)$ and $\hat{b}_k^{(L)} = \mathrm{argmin}_\mu \sum_{i=1}^n \rho_{\tau_k}\left(\hat{e}_i^{(L)} - \mu\right)$ at each k. Here $\psi_{\tau_k}(u) = \rho'_{\tau_k}(u) = (\tau_k - 1)I(u < -c^*) + \frac{(1-\tau_k)}{c^*}uI(-c^* \le u < 0) + \frac{\tau_k}{c^*}uI(0 \le u < c^*) + \tau_k I(u \ge c^*)$. (c) Obtain next iterative estimates $\hat{\boldsymbol{\alpha}}^{(L+1)}$ and $\hat{\boldsymbol{\gamma}}^{(L+1)}$ by using the OLS method for response variable $\tilde{Y}_i = G_i^{(L)}$ and covariates $\boldsymbol{Z}_i, \hat{\boldsymbol{U}}_i$.

2.3. Asymptotic properties

In this section, we will establish the asymptotic properties of the estimators defined in the previous section. We shall first present some notations, suppose $\boldsymbol{\gamma}_0 = (\gamma_{01}, ..., \gamma_{0m})^{\mathrm{T}}$ is the true values of $\boldsymbol{\gamma}$, $F(\cdot)$ is the cumulative distribution function of the random error. In addition, the notation $\|\cdot\|$ represents the $\mathcal{L}^2$ norm of a function or the Euclidean norm of a vector, and $a_n \sim b_n$ indicates that a_n/b_n is bounded away from zero and infinity as $n \to \infty$. For simplicity, in this paper, C represents a generic positive constant whose value may change from line to line. Next, to obtain the asymptotic properties, some technical assumptions are listed as follows.

C1. The random process $X(\cdot)$ and score $\xi_j = \langle X(\cdot), v_j(\cdot)\rangle$ satisfy the following condition: $\mathrm{E}\|X(\cdot)\|^4 < \infty$ and $\mathrm{E}[\xi_j^4] \le C\lambda_j^2,\ j \ge 1$.

C2. The eigenvalues of C_X and the score coefficients fulfil the conditions below:
(a) There exist constants C and $a > 1$ such that $C^{-1} j^{-a} \le \lambda_j \le C j^{-a}, \lambda_j - \lambda_{j+1} \ge C j^{-a-1},\ j \ge 1$;
(b) There exist constants C and $b > a/2 + 1$ such that $|\gamma_j| \le C j^{-b},\ j \ge 1$.

C3. The random vector $\boldsymbol{Z}$ satisfies $\mathrm{E}\|\boldsymbol{Z}\|^4 < \infty$.

C4. There is some constant C such that $|\langle c_{Z_l X}, v_j\rangle| \le C j^{-(a+b)},\ l = 1, \cdots, p,\ j \ge 1$.

C5. Let $\eta_{il} = Z_{il} - \langle g_l, X_i\rangle$ with $g_l = \sum_{j=1}^{\infty} \lambda_j^{-1}\langle c_{Z_l X}, v_j\rangle v_j$, and $\boldsymbol{\eta}_i = (\eta_{i1}, \cdots, \eta_{ip})^{\mathrm{T}}$, then $\boldsymbol{\eta}_1, \cdots, \boldsymbol{\eta}_n$ are i.i.d random vectors. We further assume that $\mathrm{E}[\boldsymbol{\eta}_i|X_i] = 0$, $\mathrm{E}[\boldsymbol{\eta}_i\boldsymbol{\eta}_i^{\mathrm{T}}|X_i] \overset{a.s.}{=} \boldsymbol{\Sigma}$ is a constant positive definite matrix.

C6. b_τ is the unique solution of $\mathrm{E}\left[\rho'_\tau(e - b_\tau)\right] = 0$, and $h_\tau(b_\tau) = (1-\tau) f(b_\tau - c^*) + \frac{1-\tau}{c^*}(F(b_\tau) - F(b_\tau - c^*)) + \frac{\tau}{c^*}(F(b_\tau + c^*) - F(b_\tau)) + \tau f(b_\tau + c^*)$ is continuous at b_τ. Furthermore, we suppose that $h_\tau(b_\tau) > 0$.

C6′. b_{0k} is the unique solution of $E\left[\rho'_{\tau_k}(e_i - b_{0k})\right] = 0$, $h_{\tau_k}(b_{0k}) = (1-\tau_k) f(b_{0k} - c^*) + \frac{1-\tau_k}{c^*}(F(b_{0k}) - F(b_{0k} - c^*)) + \frac{\tau_k}{c^*}(F(b_{0k} + c^*) - F(b_{0k})) + \tau_k f(b_{0k} + c^*)$ is continuous at b_{0k}, $k = 1, \cdots, K$. Furthermore, there exist some positive constants C_1, C_2 such that $0 < C_1 \le \min_{1\le k\le K} h_{\tau_k}(b_{0k}) \le \max_{1\le k\le K} h_{\tau_k}(b_{0k}) \le C_2 < +\infty$.

C1 is the commonly used condition for establishing the consistency of the the empirical covariance operator of X in functional linear model and partial functional regression model. For example, it has been adopted in [3, 17, 18] (mean regression), [6, 11] (quantile regression), [12, 13] (robust estimation procedure), among others. C2(a) is used to identify the slope function $\beta(t)$ via preventing the spacings between eigenvalues being too small, and C2(b) ensures the sufficiently smooth of slope function $\beta(t)$. C3–C5 are needed to deal with the vector-type covariate in the model (2.1) (see [19]). More concretely, C3 is for the asymptotic behaviors of $\hat{c}_{ZX}$ and $\hat{c}_{Z}$. C4 is used to ensure the effect of truncation on the estimator of $\beta(\cdot)$ be sufficiently small. C5 is a commonly used condition in the literature on partial functional regression models (see for example, [4,20,21]). The assumptions on $\mathrm{E}[\boldsymbol{\eta}_i|X_i]$ and $\mathrm{E}[\boldsymbol{\eta}_i\boldsymbol{\eta}_i^{\mathrm{T}}|X_i]$ are slightly strong, and are used to fix the identifiability of $\boldsymbol{\alpha}$ and simplify the proof of the theorems. It is easy to see that $\langle g_l, X_i\rangle$ is the projection of Z_l onto X_i, and $\mathrm{E}(\boldsymbol{\eta}_i) = 0$, $\mathrm{C}ov(\eta_{il}, \langle g_l, X_i\rangle) = 0$, $\mathrm{E}\|\boldsymbol{\eta}_i\|^4 < \infty$ even without the assumptions. The facts can be used to obtain similar results to the following theorems with more complicated technics (see, for example, [6]) and more conditions to ensure the identifiability. Other type conditions on $\boldsymbol{\eta}_i$ can be found in [11,22]. C6 and C6′ are specific to the AHR and WCAHR (CAHR) cases respectively, which are primarily used to ensure the asymptotic behaviors of our estimators.

The following theorems discuss the convergence rate of the estimated $\beta(\cdot)$, the asymptotic normality of the estimated $\boldsymbol{\alpha}$ and the convergence rate of the mean squared prediction error. To obtain this, we further assume $(\boldsymbol{Z}_{n+1}, X_{n+1}(\cdot))$ is independent of $\mathcal{S}$ in this paper.

The next theorem establishes the large sample properties of the AHR estimators.

Theorem 1. *Suppose that the Conditions C1–C6 are satisfied, and the tuning parameter* $m \sim n^{1/(a+2b)}$, *then*

$$\|\bar{\beta}(\cdot) - \beta_0(\cdot)\|^2 = O_p(n^{-\frac{2b-1}{a+2b}}),$$
$$\sqrt{n}(\bar{\boldsymbol{\alpha}} - \boldsymbol{\alpha}_0) \xrightarrow{d} N\left(0, \frac{V}{\{h_\tau(b_\tau)\}^2}\boldsymbol{\Sigma}^{-1}\right),$$
$$MSPE_{AHR} = O_p(n^{-\frac{a+2b-1}{a+2b}}),$$

where $\xrightarrow{d}$ *represents convergence in distribution, and* $V = E\left[\psi_\tau^2(e - b_\tau)\right]$ *with* $\psi_\tau(u) = \rho_\tau'(u) = (\tau - 1)I(u < -c^*) + \frac{(1-\tau)}{c^*}uI(-c^* \le u < 0) + \frac{\tau}{c^*}uI(0 \le u < c^*) + \tau I(u \ge c^*)$.

The asymptotic properties of the proposed WCAHR estimators are presented in the following theorem.

Theorem 2. *Under the Conditions C1–C5 and C6′, if the tuning parameter is taken as* $m \sim n^{1/(a+2b)}$, *then*

$$\|\hat{\beta}(\cdot) - \beta_0(\cdot)\|^2 = O_p(n^{-\frac{2b-1}{a+2b}}),$$
$$\sqrt{n}(\hat{\boldsymbol{\alpha}} - \boldsymbol{\alpha}_0) \xrightarrow{d} N\left(0, \frac{\boldsymbol{w}^T\boldsymbol{V}\boldsymbol{w}}{\left\{\sum_{k=1}^K w_k h_{\tau_k}(b_{0k})\right\}^2}\boldsymbol{\Sigma}^{-1}\right),$$
$$MSPE_{WCAHR} = O_p(n^{-\frac{a+2b-1}{a+2b}}),$$

where $\boldsymbol{w} = (w_1, \ldots, w_K)^T$ *and* $\boldsymbol{V} = (V_{kl})_{1\le k,\, l\le K}$, *here* $V_{kl} = E\left[\psi_{\tau_k}(e - b_{0k})\psi_{\tau_l}(e - b_{0l})\right]$ *with* $1 \le k, l \le K$.

Remark 3. *The results illustrate that the slope function estimator has the same convergence rate as the estimators in [6] and [4], which are optimal in the minimax sense. Note that it is similar to quantile regression that no moment condition on error term is needed here. In addition, we notice that the rate attained in predicting* Y_{n+1} *is faster than the rate attained in estimating* $\beta(t)$. *Trace its root and use* $MSPE_{AHR}$ *as an example, it is for the integral operator providing additional smoothness in computing* $\int_0^1 \bar{\beta}(t)X_{n+1}(t)dt$ *from* $\bar{\beta}(t)$.

Remark 4. *If all* w_k*s are equal, then Theorem 2 reduces to the asymptotic properties of the CAHR estimators. Taking* $\tau_1 = \tau$, *it is easy to see that there is a weight vector* $\boldsymbol{w}$ *such that* $\frac{\boldsymbol{w}^T\boldsymbol{V}\boldsymbol{w}}{\left\{\sum_{k=1}^K w_k h_{\tau_k}(b_{0k})\right\}^2} < \frac{V}{\{h_\tau(b_{01})\}^2}$. *Note that the right hand side of the inequality is just the asymptotic variance given in Theorem 1.*

3. Simulation study

In this section, a Monte Carlo simulation is used to investigate the finite sample properties of the proposed estimation approaches. The data sets used in the simulation are generated from the following model,

$$Y = \mathbf{Z}^{\mathrm{T}}\boldsymbol{\alpha} + \int_0^1 X(t)\beta(t)dt + \sigma(\mathbf{Z}, X)e,$$

where the slope function $\beta(t) = \sqrt{2}\sin(\pi t/2) + 3\sqrt{2}\sin(3\pi t/2)$, and $X(t) = \sum_{j=1}^{50} \xi_j \phi_j(t)$, here $\phi_j(t) = \sqrt{2}\sin((j-0.5)\pi t)$, and ξ_js are mutually independent normal random variables with mean 0 and variance $\lambda_j = ((j-0.5)\pi)^{-2}$. The true values of parameters are set as $\boldsymbol{\alpha} = (\alpha_1, \alpha_2)^{\mathrm{T}} = (10, 5)^{\mathrm{T}}$, and $\mathbf{Z} \sim N(0, \Sigma_{\mathbf{Z}})$ with $(\Sigma_{\mathbf{Z}})_{i,j} = 0.75^{|i-j|}$ for $i, j = 1, 2$.

Five different distributions for e are considered as follows: (a) standard normal distribution $N(0, 1)$; (b) positively skewed normal distribution $SN(0, 1, 15)$; (c) positively skewed t-distribution $St(0, 1, 5, 3)$; (d) mixture of normals (MN) $0.95N(0, 1) + 0.05N\left(0, 10^2\right)$, which produces a distribution with outliers of response; (e) bimodal distribution ($Bimodal$) $\tilde{\eta}N(-1.2, 1) + (1 - \tilde{\eta})N(1.2, 1)$ with $\tilde{\eta} \sim Binomial(1, 0.5)$. The multiplier $\sigma(\mathbf{Z}, X)$ can be generated from either of the following two models:

(A) (homoscedastic) $\sigma(\mathbf{Z}, X) = 1$;

(B) (heteroscedastic) $\sigma(\mathbf{Z}, X) = \left|1 + 0.1\left(Z_1\alpha_1^* + Z_2\alpha_2^* + \int_0^1 X(t)\beta^*(t)dt\right)\right|$, where $\alpha_1^* = \alpha_2^* = 1$, and $\beta^*(t) = \sqrt{2}\sin(\pi t/2) + \sqrt{2}\sin(3\pi t/2)$.

Implementing the proposed estimation method requires the predetermined levels over (0,1), i.e., $\{\tau_k\}_{k=1}^K$. Similar to the setting in [14], we take $K = 19$, and choose the equidistant levels $\tau_k = k/(K + 1)$, $k = 1, 2, \ldots, K$. In addition, for the WCAHR estimator, we employ the adaptive weights given in Remark 2.

For comparison, we also calculate the OLS estimator, the least absolute deviation (LAD) estimator, the ESL estimator, the H-ESL estimator, the Huber estimator (which corresponds to the case of AHR estimator at $\tau = 0.5$), the CAHR estimator, and the AHR estimators at $\tau = 0.25$ and 0.75. In this study, the sample size n is set as 200 or 400.

To implement these methods, we need to choose the tuning parameter m. In this paper, m is selected by the cumulative percentage of total variability (CPV) method, that is,

$$m = \underset{p}{\operatorname{argmin}}\left\{ \sum_{i=1}^{p} \hat{\lambda}_i \Big/ \sum_{i=1}^{\infty} \hat{\lambda}_i \geqslant \delta \right\},$$

where δ equals 85%. Other criterion, such as AIC, BIC, can be employed.

For each setting and different methods, bias (Bias), standard deviation (Sd) of the estimated α_1 and α_2, and the mean squared error (MSE) of the estimated $\boldsymbol{\alpha}$ with MSE $= \frac{1}{S}\sum_{s=1}^{S}\sum_{j=1}^{2}(\hat{\alpha}_j^s - \alpha_j)^2$, as well as the mean integrated squared error (MISE) of the estimated $\beta(t)$ over S=500 repetitions are summarized, where MISE $= \left\{\frac{1}{100S}\sum_{s=1}^{S}\sum_{i=1}^{100}\left(\hat{\beta}^s(t_i) - \beta(t_i)\right)^2\right\}$ with t_is being 100 equally spaced grids in [0,1], here $\hat{\alpha}_j^s, \hat{\beta}^s(\cdot)$ are the estimates of $\hat{\alpha}_j$ and $\hat{\beta}(\cdot)$ from the sth sampling, $j = 1, 2$.

Table 1. Simulation results under different homoscedastic error distributions (A).

Errors	n	Method	MISE	MSE($\hat{\alpha}$)	$\hat{\alpha}_1$		$\hat{\alpha}_2$	
					Bias	Sd	Bias	Sd
		OLS	0.2690	0.0229	−0.0048	0.1074	0.0096	0.1059
		LAD	0.3294	0.0354	−0.0099	0.1320	0.0115	0.1333
		ESL	0.3434	0.0394	0.0009	0.1405	−0.0013	0.1403
		H-ESL	0.2824	0.0259	0.0055	0.1141	−0.0059	0.1133
	200	AHR(0.25)	0.3294	0.0786	−0.0077	0.2014	0.0105	0.1948

		Huber	0.2758	0.0234	−0.0063	0.1090	0.0102	0.1067
		AHR(0.75)	0.5251	0.0754	−0.0092	0.1894	0.0150	0.1981
		CAHR	0.2689	0.0233	−0.0051	0.1095	0.0091	0.1060
$N(0,1)$		WCAHR	0.2693	0.0230	−0.0055	0.1080	0.0101	0.1058
		OLS	0.1004	0.0105	−0.0011	0.0710	0.0015	0.0738
		LAD	0.1304	0.0163	−0.0027	0.0880	0.0045	0.0924
		ESL	0.1349	0.0175	0.0010	0.0918	0.0008	0.0955
		H-ESL	0.1031	0.0113	0.0011	0.0727	0.0010	0.0773
	400	AHR(0.25)	0.1304	0.0358	−0.0048	0.1313	−0.0013	0.1361
		Huber	0.1048	0.0107	−0.0018	0.0720	0.0027	0.0742
		AHR(0.75)	0.2337	0.0405	0.0043	0.1454	0.0019	0.1390
		CAHR	0.1006	0.0107	−0.0014	0.0721	0.0020	0.0743
		WCAHR	0.1005	0.0105	−0.0014	0.0709	0.0019	0.0736
		OLS	0.2929	0.0241	−0.0037	0.1127	−0.0012	0.1065
		LAD	0.2452	0.0137	−0.0001	0.0844	−0.0007	0.0813
		ESL	0.3665	0.0387	−0.0031	0.1412	−0.0023	0.1368
		H-ESL	0.3377	0.0305	−0.0009	0.1244	−0.0028	0.1225
	200	AHR(0.25)	0.2260	0.0086	0.0012	0.0652	−0.0022	0.0655
		Huber	0.1998	0.0098	−0.0011	0.0698	−0.0007	0.0702
		AHR(0.75)	0.2314	0.0172	−0.0038	0.0949	−0.0001	0.0903
		CAHR	0.2122	0.0099	−0.0007	0.0717	−0.0010	0.0691
$SN(0,1,15)$		WCAHR	0.1884	0.0086	0.0002	0.0659	−0.0017	0.0652
		OLS	0.0994	0.0122	−0.0024	0.0794	0.0052	0.0769
		LAD	0.0718	0.0065	−0.0004	0.0569	0.0001	0.0568
		ESL	0.1372	0.0193	−0.0012	0.1003	0.0031	0.0962
		H-ESL	0.1022	0.0139	−0.0043	0.0832	0.0056	0.0832
	400	AHR(0.25)	0.0796	0.0037	0.0028	0.0418	−0.0010	0.0437
		Huber	0.0688	0.0045	0.0005	0.0468	0.0003	0.0483
		AHR(0.75)	0.0912	0.0082	−0.0010	0.0627	0.0019	0.0652
		CAHR	0.0712	0.0043	0.0007	0.0454	0.0009	0.0471
		WCAHR	0.0567	0.0035	−0.0009	0.0418	0.0018	0.0424
		OLS	0.4781	0.0695	0.0047	0.1815	−0.0185	0.1902
		LAD	0.2461	0.0215	0.0057	0.1052	−0.0094	0.1015
		ESL	0.3793	0.0451	0.0023	0.1538	−0.0062	0.1462
		H-ESL	0.3682	0.0443	0.0059	0.1533	−0.0078	0.1437
	200	AHR(0.25)	0.2541	0.0203	0.0014	0.0987	0.0039	0.1025
		Huber	0.2829	0.0284	0.0082	0.1204	−0.0013	0.1175
		AHR(0.75)	1.8541	0.3745	0.0128	0.4566	−0.0243	0.4065
		CAHR	0.3606	0.0286	0.0037	0.1188	−0.0001	0.1202
$St(0,1,5,3)$		WCAHR	0.2205	0.0169	0.0018	0.0920	−0.0089	0.0915
		OLS	0.2310	0.0325	−0.0021	0.1296	0.0008	0.1252
		LAD	0.1004	0.0109	0.0006	0.0742	−0.0001	0.0735

	400	ESL	0.1563	0.0212	−0.0053	0.1002	0.0027	0.1054
		H-ESL	0.1516	0.0178	−0.0045	0.0917	0.0011	0.0966
		AHR(0.25)	0.1000	0.0088	0.0028	0.0671	−0.0004	0.0659
		Huber	0.1108	0.0116	0.0019	0.0781	0.0021	0.0743
		AHR(0.75)	1.5565	0.3644	−0.0416	0.4269	0.0216	0.4242
		CAHR	0.1496	0.0153	0.0016	0.0873	0.0007	0.0874
		WCAHR	0.0838	0.0076	−0.0015	0.0616	−0.0000	0.0618
MN	200	OLS	0.8806	0.1464	−0.0134	0.2675	−0.0016	0.2732
		LAD	0.3783	0.0358	−0.0005	0.1320	−0.0013	0.1355
		ESL	0.3719	0.0363	0.0025	0.1331	−0.0044	0.1361
		H-ESL	0.3101	0.0280	−0.0018	0.1175	−0.0028	0.1189
		AHR(0.25)	0.3297	0.1148	−0.0120	0.2346	0.0105	0.2439
		Huber	0.3685	0.0499	−0.0042	0.1613	0.0071	0.1543
		AHR(0.75)	0.7037	0.1060	0.0078	0.2321	−0.0033	0.2281
		CAHR	0.7857	0.1035	0.0031	0.2292	0.0035	0.2257
		WCAHR	0.3252	0.0340	−0.0046	0.1289	−0.0033	0.1316
	400	OLS	0.3822	0.0715	0.0063	0.1887	−0.0099	0.1892
		LAD	0.1307	0.0190	0.0055	0.0983	−0.0060	0.0965
		ELS	0.1268	0.0180	0.0032	0.0963	−0.0043	0.0932
		H-ESL	0.1032	0.0129	0.0048	0.0807	−0.0067	0.0793
		AHR(0.25)	0.1491	0.0604	0.0052	0.1731	0.0055	0.1742
		Huber	0.1332	0.0156	−0.0007	0.0880	0.0033	0.0887
		AHR(0.75)	0.3391	0.0505	−0.0049	0.1597	0.0040	0.1581
		CAHR	0.3435	0.0419	-0.0030	0.1442	0.0074	0.1449
		WCAHR	0.1127	0.0157	0.0055	0.0894	−0.0077	0.0872
Bimodal	200	OLS	0.4317	0.0560	−0.0001	0.1690	−0.0018	0.1657
		LAD	0.8634	0.1201	−0.0040	0.2438	0.0008	0.2463
		ESL	2.2921	0.4163	−0.0148	0.4541	−0.0002	0.4582
		H-ESL	0.4417	0.0558	−0.0004	0.1687	−0.0017	0.1653
		AHR(0.25)	0.9240	0.3169	−0.0258	0.3973	0.0352	0.3964
		Huber	0.5694	0.0776	−0.0056	0.2018	0.0129	0.1914
		AHR(0.75)	1.8171	0.3150	0.0110	0.4044	−0.0107	0.3889
		CAHR	0.5191	0.0546	−0.0075	0.1652	0.0116	0.1647
		WCAHR	0.4215	0.0552	0.0016	0.1679	−0.0016	0.1642
	400	OLS	0.1861	0.0296	0.0032	0.1170	−0.0017	0.1262
		LAD	0.4069	0.0670	−0.0068	0.1765	0.0061	0.1891
		ESL	1.5317	0.2511	−0.0185	0.3481	0.0033	0.3600
		H-ESL	0.1860	0.0298	0.0032	0.1175	−0.0021	0.1265
		AHR(0.25)	0.4420	0.1620	−0.0081	0.2835	−0.0025	0.2855
		Huber	0.2369	0.0454	0.0023	0.1483	−0.0082	0.1529
		AHR(0.75)	0.8504	0.1523	0.0142	0.2774	−0.0021	0.2742
		CAHR	0.2047	0.0296	0.0025	0.1195	−0.0014	0.1239

		WCAHR	0.1825	0.0291	0.0033	0.1162	−0.0019	0.1249

Table 1 presents the results in the homoscedastic case. From Table 1, we can see the following facts: (a) The Sd, MSE and MISE decrease as the sample size n increases from 200 to 400. (b) The proposed estimators are almost unbiased, which further illustrates the consistency combining with the fact (a). (c) The proposed adaptively weighted estimator performs similarly to the OLS estimator under the normal error, and is comparable to the corresponding H-ESL estimator for the mixture of normal distributions, but is significantly better than the other estimators considered when the distribution of model error is skewed or bimodal, and still enjoys the favoured being robust to outliers. This demonstrates that the proposed WCAHR estimator can well adapt to different error distributions, thus is more useful in practice.

Table 2. Simulation results under different heteroscedastic error distributions (B).

Errors	n	Method	MISE	MSE($\hat{\alpha}$)	$\hat{\alpha}_1$		$\hat{\alpha}_2$	
					Bias	Sd	Bias	Sd
$N(0,1)$	200	OLS	0.2560	0.0237	0.0049	0.1105	−0.0052	0.1069
		LAD	0.2988	0.0300	0.0054	0.1228	−0.0009	0.1222
		ELS	0.2990	0.0316	0.0041	0.1272	−0.0021	0.1240
		H-ESL	0.2655	0.0258	0.0060	0.1163	−0.0052	0.1104
		AHR(0.25)	0.2988	0.1065	−0.0860	0.2178	−0.0965	0.2058
		Huber	0.2531	0.0234	0.0050	0.1099	−0.0045	0.1062
		AHR(0.75)	0.5960	0.0911	0.0958	0.1880	0.0836	0.1990
		CAHR	0.2562	0.0236	0.0055	0.1099	−0.0051	0.1070
		WCAHR	0.2519	0.0227	0.0051	0.1081	−0.0045	0.1047
	400	OLS	0.1056	0.0119	−0.0029	0.0767	−0.0002	0.0777
		LAD	0.1269	0.0153	−0.0027	0.0865	0.0008	0.0882
		ELS	0.1257	0.0157	−0.0026	0.0884	0.0019	0.0888
		H-ESL	0.1061	0.0116	−0.0036	0.0760	0.0018	0.0762
		AHR(0.25)	0.1269	0.0588	−0.1020	0.1418	−0.0889	0.1427
		Huber	0.1060	0.0117	−0.0032	0.0764	0.0001	0.0766
		AHR(0.75)	0.2695	0.0581	0.0990	0.1428	0.0859	0.1434
		CAHR	0.1064	0.0120	−0.0023	0.0769	−0.0005	0.0777
		WCAHR	0.1046	0.0115	−0.0030	0.0754	0.0002	0.0762
$SN(0,1,15)$	200	OLS	0.3164	0.0357	0.0862	0.1070	0.0748	0.1057
		LAD	0.2417	0.0209	0.0695	0.0789	0.0586	0.0803
		ELS	0.3578	0.0306	0.0140	0.1234	0.0119	0.1228
		H-ESL	0.3390	0.0316	0.0308	0.1211	0.0304	0.1228
		AHR(0.25)	0.2417	0.0139	0.0555	0.0651	0.0485	0.0650
		Huber	0.2344	0.0209	0.0780	0.0711	0.0684	0.0713
		AHR(0.75)	0.2809	0.0427	0.1122	0.1026	0.0969	0.1010
		CAHR	0.2454	0.0211	0.0774	0.0725	0.0684	0.0718
		WCAHR	0.2155	0.0174	0.0727	0.0644	0.0623	0.0642
		OLS	0.1180	0.0242	0.0859	0.0754	0.0776	0.0718

	400	LAD	0.0776	0.0149	0.0683	0.0573	0.0622	0.0553
		ELS	0.1267	0.0143	0.0180	0.0841	0.0001	0.0833
		H-ESL	0.1199	0.0179	0.0527	0.0831	0.0393	0.0819
		AHR(0.25)	0.0776	0.0096	0.0532	0.0478	0.0493	0.0453
		Huber	0.0763	0.0163	0.0753	0.0511	0.0740	0.0500
		AHR(0.75)	0.1060	0.0330	0.1043	0.0689	0.1117	0.0700
		CAHR	0.0813	0.0158	0.0732	0.0491	0.0755	0.0483
		WCAHR	0.0672	0.0130	0.0678	0.0451	0.0668	0.0441
$St(0,1,5,3)$	200	OLS	0.5695	0.0947	0.1071	0.1861	0.1158	0.1876
		LAD	0.3044	0.0287	0.0704	0.0970	0.0738	0.0941
		ELS	0.4205	0.0400	−0.0035	0.1404	−0.0162	0.1415
		H-ESL	0.3927	0.0405	0.0125	0.1409	0.0038	0.1431
		AHR(0.25)	0.3044	0.0219	0.0477	0.0944	0.0466	0.0925
		Huber	0.3450	0.0355	0.0764	0.1077	0.0784	0.1093
		AHR(0.75)	2.6347	0.5711	0.1826	0.5139	0.1919	0.4867
		CAHR	0.4548	0.0423	0.0776	0.1227	0.0826	0.1200
		WCAHR	0.2863	0.0246	0.0667	0.0868	0.0703	0.0875
	400	OLS	0.2663	0.0577	0.1027	0.1286	0.1196	0.1278
		LAD	0.1092	0.0199	0.0732	0.0694	0.0738	0.0657
		ELS	0.1423	0.0190	−0.0083	0.0973	−0.0090	0.0968
		H-ESL	0.1429	0.0202	0.0184	0.0971	0.0177	0.1003
		AHR(0.25)	0.1092	0.0148	0.0456	0.0725	0.0507	0.0698
		Huber	0.1447	0.0237	0.0777	0.0762	0.0784	0.0755
		AHR(0.75)	2.4856	0.6166	0.2100	0.5037	0.1899	0.5317
		CAHR	0.1858	0.0264	0.0726	0.0833	0.0831	0.0852
		WCAHR	0.0932	0.0160	0.0645	0.0596	0.0694	0.0592
MN	200	OLS	0.9780	0.1653	−0.0191	0.2838	0.0090	0.2904
		LAD	0.3672	0.0409	−0.0111	0.1436	0.0065	0.1420
		ELS	0.3644	0.0390	−0.0092	0.1407	0.0091	0.1381
		H-ESL	0.3186	0.0321	−0.0072	0.1260	0.0047	0.1271
		AHR(0.25)	0.3672	0.1579	−0.1021	0.2647	−0.0797	0.2665
		Huber	0.3700	0.0407	−0.0134	0.1412	0.0106	0.1431
		AHR(0.75)	0.7890	0.1350	0.0802	0.2419	0.0878	0.2498
		CAHR	0.8570	0.0989	−0.0114	0.2229	0.0077	0.2214
		WCAHR	0.3324	0.0352	−0.0109	0.1326	0.0071	0.1322
	400	OLS	0.4215	0.0781	0.0126	0.2002	−0.0081	0.1944
		LAD	0.1244	0.0189	0.0000	0.0980	0.0028	0.0964
		ELS	0.1247	0.0173	0.0004	0.0940	0.0028	0.0921
		H-ESL	0.1073	0.0131	0.0010	0.0808	0.0019	0.0808
		AHR(0.25)	0.1244	0.0759	−0.0849	0.1701	−0.0953	0.1752
		Huber	0.1222	0.0147	0.0054	0.0861	−0.0011	0.0854
		AHR(0.75)	0.3546	0.0735	0.1014	0.1642	0.0908	0.1673

		CAHR	0.3496	0.0531	0.0116	0.1647	−0.0069	0.1604
		WCAHR	0.1114	0.0148	0.0038	0.0860	0.0001	0.0857
Bimodal	200	OLS	0.4259	0.0604	−0.0118	0.1733	0.0020	0.1738
		LAD	0.7261	0.1323	−0.0248	0.2573	0.0096	0.2558
		ELS	1.9087	0.3505	−0.0283	0.4120	0.0078	0.4241
		H-ESL	0.4337	0.0614	−0.0118	0.1758	0.0031	0.1743
		AHR(0.25)	0.7261	0.5567	−0.1960	0.5100	−0.1893	0.4716
		Huber	0.4839	0.0763	−0.0180	0.1974	0.0064	0.1924
		AHR(0.75)	2.4130	0.4776	0.1823	0.4534	0.1885	0.4509
		CAHR	0.6745	0.1187	−0.0223	0.2449	0.0087	0.2411
		WCAHR	0.4361	0.0642	−0.0096	0.1797	0.0068	0.1783
	400	OLS	0.1934	0.0295	0.0087	0.1251	−0.0113	0.1169
		LAD	0.3626	0.0578	0.0145	0.1715	−0.0169	0.1670
		ELS	1.0701	0.1718	0.0200	0.2893	−0.0187	0.2955
		H-ESL	0.1948	0.0299	0.0098	0.1256	−0.0113	0.1178
		AHR(0.25)	0.3626	0.3363	−0.2023	0.3440	−0.2239	0.3562
		Huber	0.2316	0.0383	0.0093	0.1438	−0.0130	0.1320
		AHR(0.75)	1.2586	0.3302	0.2280	0.3504	0.1848	0.3484
		CAHR	0.3292	0.0506	0.0163	0.1645	−0.0174	0.1516
		WCAHR	0.2058	0.0313	0.0119	0.1298	−0.0092	0.1193

Table 2 presents the results in the more challenged heteroscedastic case, which violates the condition in this paper. The proposed WCAHR estimator outperforms the other estimators considered for the normal, skewed normal and skewed t error distributions, and is comparable to the corresponding H-ESL estimator for the mixture of normal distribution and bimodal distribution. This further illustrates that the proposed WCAHR estimator may be more applicable. Although the simulation results show the appealing performance for the considered heteroscedastic errors, general theoretical results are still challenging, and more conditions on the conditional moments of e may be helpful.

In order to detect the effect of the level choice to the performance of the WCAHR estimators, especially for the skewed error distributions, we also change in the simulation the number K_1 of levels over (0,0.5) given the total level number K. Specifically, for the given $K = 19$ and different values of K_1, we set $\tau_i = \frac{i}{2K_1}$, for $i = 1, \cdots, K_1$ and $\tau_i = \frac{K+1-2K_1+i}{2(K+1-K_1)}$, for $i = K_1 + 1, \cdots, K$. Table 3 presents the estimation results. We find from the results that the choice of the levels does not destroy the performance of the WCAHR estimators, although less number of levels over (0, 0.5) leads to slightly larger MSE and MISE for the positively skewed error distributions. In addition, the MISEs and MSEs decrease as K_1 increases, and stabilize eventually. This may motivate that one can take more levels appropriately over (0,0.5) in dealing with the positively skewed error distributions.

Table 3. Simulation results under skewed error distributions, with different numbers of τs over $(0, 0.5]$.

Errors	n	K_1	MISE	MSE($\hat{\alpha}$)	$\hat{\alpha}_1$		$\hat{\alpha}_2$	
					Bias	Sd	Bias	Sd
$SN(0,1,15)$	200	4	0.2321	0.0088	−0.0034	0.0663	0.0044	0.0665
		6	0.2295	0.0084	−0.0036	0.0646	0.0042	0.0645
		8	0.2279	0.0081	−0.0038	0.0635	0.0042	0.0633
		10	0.2269	0.0079	−0.0040	0.0629	0.0041	0.0626
		12	0.2264	0.0078	−0.0041	0.0627	0.0041	0.0621
		14	0.2262	0.0078	−0.0043	0.0627	0.0041	0.0620
	400	4	0.0626	0.0039	0.0071	0.0450	-0.0054	0.0425
		6	0.0611	0.0037	0.0068	0.0440	−0.0050	0.0415
		8	0.0601	0.0036	0.0065	0.0434	−0.0046	0.0410
		10	0.0595	0.0036	0.0063	0.0432	−0.0043	0.0409
		12	0.0591	0.0036	0.0060	0.0432	−0.0039	0.0409
		14	0.0589	0.0036	0.0059	0.0433	−0.0038	0.0411
$St(0,1,5,3)$	200	4	0.2445	0.0178	−0.0027	0.0961	−0.0023	0.0927
		6	0.2388	0.0169	−0.0025	0.0938	−0.0024	0.0900
		8	0.2349	0.0163	−0.0025	0.0923	−0.0023	0.0880
		10	0.2321	0.0158	−0.0025	0.0911	−0.0023	0.0866
		12	0.2301	0.0155	−0.0024	0.0903	−0.0022	0.0856
		14	0.2287	0.0153	−0.0023	0.0898	−0.0022	0.0849
	400	4	0.0925	0.0090	−0.0079	0.0678	0.0076	0.0652
		6	0.0899	0.0085	−0.0078	0.0659	0.0076	0.0637
		8	0.0880	0.0082	−0.0076	0.0646	0.0076	0.0626
		10	0.0867	0.0080	−0.0076	0.0636	0.0076	0.0619
		12	0.0857	0.0078	−0.0075	0.0629	0.0077	0.0615
		14	0.0850	0.0077	−0.0074	0.0623	0.0077	0.0611

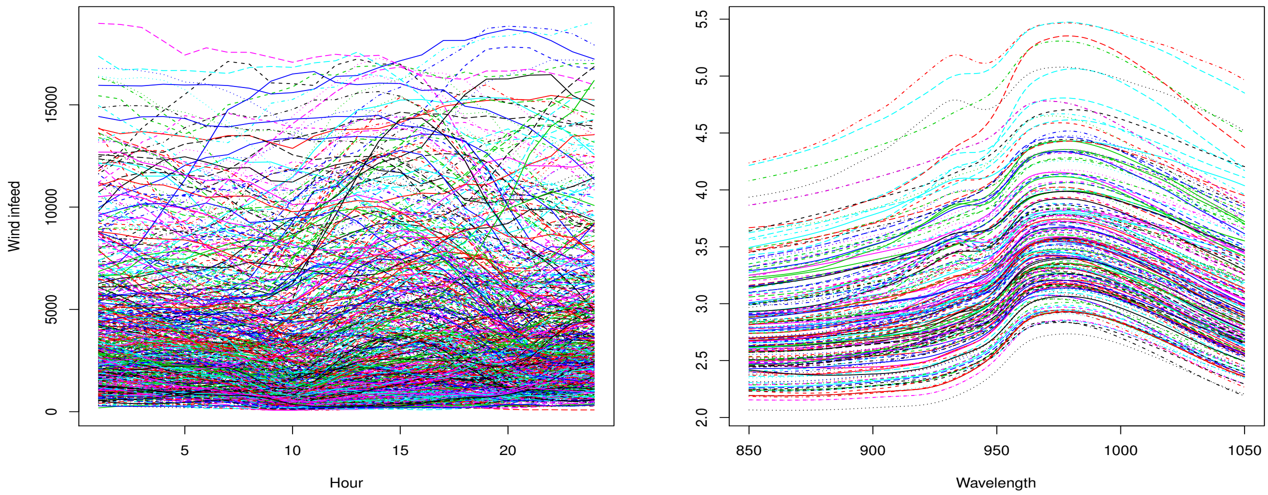
4. Real data application

In this section, we use the proposed estimation methods to the Electricity data and the Tecator data set, and the competing methods mentioned in Section 3. In the applications, all the observations are centralized before the regression analysis.

4.1. Electricity data

The data set consists of the daily average hourly electricity spot prices of the German electricity market (Y), the hourly values of Germany's wind power infeed ($X(t)$), the precipitation height (Z_1) and the sunshine duration (Z_2). Here we only consider the working days span from January 1, 2006 to September 30, 2008. The hourly values of Germany's wind power infeed curves are shown in the left panel of Figure 3. The data set can be obtained from the online supplements of Liebl [23]. Now we adopt the following partial functional linear regression model to fit the data:

$$Y = Z_1\alpha_1 + Z_2\alpha_2 + \int_1^{24} X(t)\beta(t)dt + e. \tag{4.1}$$



(a) Germany's wind power infeed **(b)** Spectrum curves

Figure 3. Plots of wind infeed and spectrum curves.

Firstly, the OLS method is applied to fit model (4.1). Then Shapiro-Wilk test is applied to test the normality of the residuals and the p-value is less than 2.2×10^{-16}. In addition, we also give the estimated density of the residuals and the residual diagnostic plot (see Figure 1). Both the test and plot clearly indicate that e follows a skewed distribution with outliers. Notice that the density of the residuals is similar to the error distribution *Bimodal* discussed in Section 3, and the simulation results illustrate that the proposed method can be applied and provide reliable inference for this kind of data.

To evaluate the predictions obtained with different estimation methods, we randomly divide data set into a training sample of size 478 subjects and a testing sample with the remaining 160 subjects (indexed by $\mathcal{J}$). The data are split for $N = 100, 200, 400$ times, respectively. We use the median quadratic error of prediction (MEDQEP) defined below as the criterion to evaluate the performances:

$$\text{MEDQEP} = \frac{1}{N}\sum_{i=1}^{N} \text{MEDIAN}\left\{(Y_{ij} - \hat{Y}_{ij})^2/\mathsf{Var}_{\mathcal{J}}(Y_{ij}),\, j \in \mathcal{J}\right\}.$$

The left 3 columns of Table 4 present the MEDQEPs of different methods mentioned above. According to the results of calculation, the WCAHR method is uniformly superior to the other estimators.

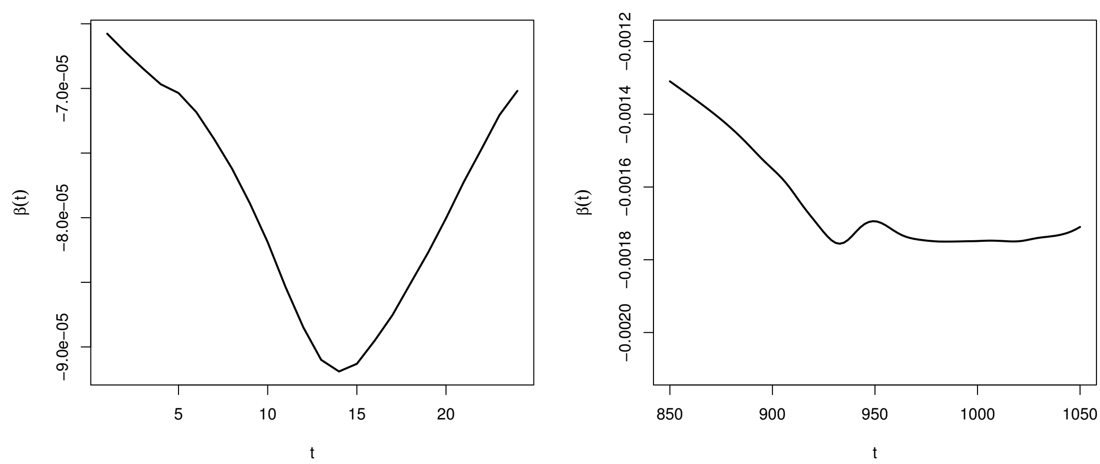
Table 4. MEDQEPs of different methods in the two applications.

	Electricity prices			Tecator		
Methods	N=100	N=200	N=400	N=100	N=200	N=400
OLS	0.4269	0.4132	0.4153	2.7824×10^{-3}	2.7182×10^{-3}	2.7257×10^{-3}
LAD	0.4094	0.4005	0.4068	2.9388×10^{-3}	2.8611×10^{-3}	2.8253×10^{-3}
ESL	0.5751	0.5626	0.5578	2.7268×10^{-3}	2.6701×10^{-3}	2.6142×10^{-3}
H-ESL	0.4104	0.4026	0.4064	2.7395×10^{-3}	2.7336×10^{-3}	2.6476×10^{-3}
AHR(0.25)	0.4052	0.3985	0.4032	2.9056×10^{-3}	2.7658×10^{-3}	2.7753×10^{-3}
Huber	0.4077	0.4027	0.4074	2.7636×10^{-3}	2.6491×10^{-3}	2.6278×10^{-3}
AHR(0.75)	0.4296	0.4198	0.4280	2.7409×10^{-3}	2.6494×10^{-3}	2.6063×10^{-3}
CAHR	0.4103	0.4019	0.4089	2.8442×10^{-3}	2.8106×10^{-3}	2.7457×10^{-3}
WCAHR	0.4030	0.3967	0.4008	2.7236×10^{-3}	2.6152×10^{-3}	2.5799×10^{-3}

Table 5 (the first 2 columns) presents the estimates of the parametric part by the estimation methods based on the whole data set. According to the results, both the precipitation height and the sunshine duration have negative effects on the daily average hourly electricity spot prices. In addition, Figure 4(a) plots the estimated slope function obtained by the WCAHR method, the estimates for $\beta(\cdot)$ obtained by other methods mentioned above exhibit similar patterns and thus omitted here. From the figure, we can see the prices have a larger (in the absolute value) linkage with the wind power infeed in the daytime, which reflects the economic phenomena of price sensitivity, and more specifically, the market is active during the daytime and thus there is more correlation between the prices and the wind power infeed in the daytime. Secondly, the Germany's wind power infeed has a negative effect on the daily average hourly electricity spot prices, which reflects supply-demand balance, that is, more wind infeed creates the oversupply of electricity and thus reduces the price.

Table 5. Estimators of coefficients for different methods.

	Electricity prices		Tecator	
	$\hat{\alpha}_1$	$\hat{\alpha}_2$	$\hat{\alpha}_1$	$\hat{\alpha}_2$
OLS	−0.5983	−0.4672	−1.1056	−0.6894
LAD	−0.7095	−0.9438	−1.0828	−0.7611
ESL	−0.5125	−0.4629	−1.0894	−0.7455
H-ESL	−0.6007	−0.7212	−1.0983	−0.7367
AHR(0.25)	−0.5618	−0.4725	−1.1122	−0.7026
Huber	−0.5799	−0.4416	−1.0981	−0.7235
AHR(0.75)	−0.5812	−0.4302	−1.0854	−0.7576
CAHR	−0.5394	−0.4582	−1.0990	−0.7274
WCAHR	−0.5924	−0.6182	−1.0964	−0.7270



(a) Electricity prices

(b) Tecator data

Figure 4. The slope function estimators based on WCAHR method.

4.2. Tecator data

The Tecator data set consists of 215 meat samples. For each sample, moisture, fat and protein are recorded in percent, and a 100-channel spectrum of absorbances is measured by the spectrometer. The data set is available from the R package fda.usc (see [24]). The right panel of Figure 3 shows the spectral trajectories. In this paper, the objective is to investigate the effects of the spectral trajectories $X(t)$, water content Z_1 and protein content Z_2 on the fat content Y by fitting the following model:

$$Y = Z_1\alpha_1 + Z_2\alpha_2 + \int_{850}^{1050} X(t)\beta(t)dt + e \tag{4.2}$$

The density of the residuals and the residual diagnostic plot in Figure 2 illustrate the error follows a skewed distribution with outliers. Similarly, to assess the prediction accuracy, the 215 meat samples are randomized into training set with 180 subjects and testing set with 35 subjects. We also randomly split the data set for $N = 100, 200, 400$ times and use MEDQEP as criterion to evaluate the finite sample performances of different estimation procedures. The comparison results are shown in the right 3 columns of Table 4, from which we know the proposed method performs better than the competing estimation procedures in view of prediction accuracy.

The estimated coefficients $\hat{\alpha}_1$, $\hat{\alpha}_2$ using various methods based on the whole data set are also shown in the last 2 columns of Table 5. Both the protein content and water content have negative effects on the fat content. Next, the right panel of Figure 4 demonstrates the estimated slope function curves based on the WCAHR method. It is obvious that the spectrum curve of absorbance has negative impact on the quantity of fatty. In addition, the estimated slope functions by other methods mentioned above exhibit similar patterns and thus omitted here.

5. Conclusions

In this paper, we study the WCAHR estimation in the partial functional linear regression model. We use the functional principal component basis to approximate the functions, and obtain the estimators of the unknown parameter vector and the slope function through minimizing the weighted asymmetric Huber loss function. The asymptotic normality of the estimated parameter vector and the convergence rate of the estimated slope function are presented.

The proposed approach is designed for automatically reflecting distributional features as well as bounding outliers effectively without requiring prior information of the data. Simulation results show that the proposed method is almost as efficient as OLS when the error follows a normal distribution, while keeps robust to outliers and still works well when the error follows skewed or bimodal distributions. That is to say, the method is adaptive to the distribution of the error in the regression model. The analyses of two examples further illustrate that the utility of the proposed methods in modelling and forecasting.

The novelty of the method is that it focuses on the extraction of major features as well as shielding the estimator from outliers. The proposed WCAHR estimation procedure can be extended to more general situations, including dependent functional data, sparse modeling, partially observed functional data, and high dimension setting. In addition, we still need to try objective way to select K.

Acknowledgments

Authors are highly grateful to anonymous referees and editor in chief for their valuable comments and suggestions for improving this paper. This research was supported by National Natural Science Foundation of China (Grant No. 11771032 and No. 11971045), Natural Science Foundation of Beijing (Grant No. 1202001), and Natural Science Foundation of Shanxi Province, China (Grant No. 20210302124530).

Conflict of interest

The authors declare that they have no competing interests.

References

1. J. Ramsay, B. Silverman, *Functional data analysis*, 2 Eds., New York: Springer, 2005. http://dx.doi.org/10.1007/b98888
2. F. Yao , H. Müller, J. Wang, Functional linear regression analysis for longitudinal data, *Ann. Statist.*, **33** (2005), 2873–2903. http://dx.doi.org/10.1214/009053605000000660
3. P. Hall, J. Horowitz, Methodology and convergence rates for functional linear regression, *Ann. Statist.*, **35** (2007), 70–91. http://dx.doi.org/10.1214/009053606000000957
4. H. Shin, Partial functional linear regression, *J. Stat. Plan. Infer.*, **139** (2009), 3405–3418. http://dx.doi.org/10.1016/j.jspi.2009.03.001
5. P. Hall, G. Hooker, Truncated linear models for functional data, *J. Roy. Stat. Soc. Ser. B*, **78** (2016), 637–653. http://dx.doi.org/10.1111/rssb.12125

6. K. Kato, Estimation in functional linear quantile regression, *Ann. Statist.*, **40** (2012), 3108–3136. http://dx.doi.org/10.1214/12-AOS1066
7. D. Kong, K. Xue, F. Yao, H. Zhang, Partially functional linear regression in high dimensions, *Biometrika*, **103** (2016), 147–159. http://dx.doi.org/10.1093/biomet/asv062
8. P. Yu, Z. Zhang, J. Du, A test of linearity in partial functional linear regression, *Metrika*, **79** (2016), 953–969. http://dx.doi.org/10.1007/s00184-016-0584-x
9. J. Zhou, Z. Chen, Q. Peng, Polynomial spline estimation for partial functional linear regression models, *Comput. Stat.*, **31** (2016), 1107–1129. http://dx.doi.org/10.1007/s00180-015-0636-0
10. M. Imaizumi, K. Kato, PCA-based estimation for functional linear regression with functional responses, *J. Multivariate Anal.*, **163** (2018), 15–36. http://dx.doi.org/10.1016/j.jmva.2017.10.001
11. Q. Tang, L. Cheng, Partial functional linear quantile regression, *Sci. China Math.*, **57** (2014), 2589–2608. http://dx.doi.org/10.1007/s11425-014-4819-x
12. P. Yu, Z. Zhu, Z. Zhang, Robust exponential squared loss-based estimation in semi-functional linear regression models, *Comput. Stat.*, **34** (2019), 503–525. http://dx.doi.org/10.1007/s00180-018-0810-2
13. X. Cai, L. Xue, F. Lu, Robust estimation with a modified Huber's loss for partial functional linear models based on splines, *J. Korean Stat. Soc.*, **49** (2020), 1214–1237. http://dx.doi.org/10.1007/s42952-020-00052-x
14. Y. Lim, H. Oh, A data-adaptive principal component analysis: use of composite asymmetric Huber function, *J. Comput. Graph. Stat.*, **25** (2015), 1230–1247. http://dx.doi.org/10.1080/10618600.2015.1067621
15. P. Huber, Robust estimation of a location parameter, *Ann. Math. Statist.*, **35** (1964), 73–101. http://dx.doi.org/10.1214/aoms/1177703732
16. H. Zou, M. Yuan, Composite quantile regression and the oracle model selection theory, *Ann. Statist.*, **36** (2008), 1108–1126. http://dx.doi.org/10.1214/07-AOS507
17. T. Cai, P. Hall, Prediction in functional linear regression, *Ann. Statist.*, **34** (2006), 2159–2179. http://dx.doi.org/10.1214/009053606000000830
18. H. Ma, Y. Bai, Z. Zhu, Dynamic single-index model for functional data, *Sci. China Math.*, **59** (2016), 2561–2584. http://dx.doi.org/10.1007/s11425-016-0051-3
19. Y. Lu, J. Du, Z. Sun, Functional partially linear quantile regression model, *Metrika*, **77** (2014), 317–332. http://dx.doi.org/10.1007/s00184-013-0439-7
20. H. Shin, M. Lee, On prediction rate in partial functional linear regression, *J. Multivariate Anal.*, **103** (2012), 93–106. http://dx.doi.org/10.1016/j.jmva.2011.06.011
21. H. Zhu, R. Zhang, Z. Yu, H. Lian, Y. Liu, Estimation and testing for partially functional linear errors-in-variables models, *J. Multivariate Anal.*, **170** (2019), 296–314. http://dx.doi.org/10.1016/j.jmva.2018.11.005
22. X. Cui, H. Lin, H. Lian. Partially functional linear regression in reproducing kernel Hilbert spaces, *Comput. Stat. Data Anal.*, **150** (2020), 106978. http://dx.doi.org/10.1016/j.csda.2020.106978

23. D. Liebl, Modeling and forecasting electricity spot prices: a functional data perspective, *Ann. Appl. Stat.*, **7** (2013), 1562–1592. http://dx.doi.org/10.1214/13-AOAS652

24. M. Febrero-bande, M. Fuente, Statistical computing in functional data analysis: The R Package fda.usc, *J. Stat. Softw.*, **51** (2012), 1–28. http://dx.doi.org/10.18637/jss.v051.i04

A. Proofs of the Theorems

We just prove Theorem 2. The Theorem 1 is a special case of Theorem 2.

Proof of Theorem 2:

Let $\delta_n = \sqrt{\frac{m}{n}}$, $P_n = \frac{1}{n}\sum_{i=1}^n \boldsymbol{\eta}_i\boldsymbol{\eta}_i^{\mathrm{T}}$, $\boldsymbol{V}_n = \delta_n^{-1}P_n^{1/2}(\hat{\boldsymbol{\alpha}}-\boldsymbol{\alpha}_0)$, $\tilde{\boldsymbol{\eta}}_i = P_n^{-1/2}\boldsymbol{\eta}_i$, $\Lambda = \mathrm{diag}\{\lambda_1,\cdots,\lambda_m\}$, $\boldsymbol{A}_i = \Lambda^{-1/2}\boldsymbol{U}_i$, $\hat{\boldsymbol{A}}_i = \Lambda^{-1/2}\hat{\boldsymbol{U}}_i$, $H_m = \left(\lambda_1^{-1}\langle c_{\boldsymbol{Z},X}, v_1\rangle,\ldots,\lambda_m^{-1}\langle c_{\boldsymbol{Z},X}, v_m\rangle\right)^{\mathrm{T}}$, $\boldsymbol{W}_n = \delta_n^{-1}\Lambda^{1/2}[(\hat{\boldsymbol{\gamma}}-\boldsymbol{\gamma}_0)+H_m(\hat{\boldsymbol{\alpha}}-\boldsymbol{\alpha}_0)]$, $r_i = \int_0^1\beta_0(t)X_i(t)dt - \hat{\boldsymbol{U}}_i^{\mathrm{T}}\boldsymbol{\gamma}_0$, $B_i = H_m^{\mathrm{T}}(\boldsymbol{U}_i-\hat{\boldsymbol{U}}_i)+\sum\limits_{j=m+1}^{\infty}\lambda_j^{-1}\langle c_{\boldsymbol{Z}X}, v_j\rangle\xi_{ij}$, $\tilde{B}_i = \delta_n^{-1}(\hat{\boldsymbol{\alpha}}-\boldsymbol{\alpha}_0)^{\mathrm{T}}B_i$, $S_{nk} = \delta_n^{-1}\left(\hat{b}_k - b_{0k}\right)$, $\boldsymbol{S}_n = \left(\sqrt{w_1}S_{n1},\ldots,\ \sqrt{w_K}S_{nK}\right)^{\mathrm{T}}$, $\mathcal{F}_n = \left\{(\boldsymbol{V}_n,\boldsymbol{W}_n,\boldsymbol{S}_n) : \left\|(\boldsymbol{V}_n^{\mathrm{T}},\boldsymbol{W}_n^{\mathrm{T}},\boldsymbol{S}_n^{\mathrm{T}})^{\mathrm{T}}\right\| \le L\right\}$, $T_n = \{(\boldsymbol{Z}_1, X_1(\cdot)),\ldots,(\boldsymbol{Z}_n, X_n(\cdot))\}$.

Next, we will show that, for any given $\eta > 0$, there exists a sufficiently large constant L, such that

$$\begin{aligned} P\Big\{ \inf_{(\boldsymbol{V}_n,\boldsymbol{W}_n,\boldsymbol{S}_n)\in\mathcal{F}_n} &\sum_{i=1}^n\sum_{k=1}^K w_k\rho_{\tau_k}\Big(Y_i - b_{0k} - \boldsymbol{Z}_i^{\mathrm{T}}\boldsymbol{\alpha}_0 - \hat{\boldsymbol{U}}_i^{\mathrm{T}}\boldsymbol{\gamma}_0 - \delta_n\left(\boldsymbol{V}_n^{\mathrm{T}}\tilde{\boldsymbol{\eta}}_i + \boldsymbol{W}_n^{\mathrm{T}}\hat{\boldsymbol{A}}_i + \tilde{B}_i + S_{nk}\right)\Big) \\ &> \sum_{i=1}^n\sum_{k=1}^K w_k\rho_{\tau_k}\Big(Y_i - b_{0k} - \boldsymbol{Z}_i^{\mathrm{T}}\boldsymbol{\alpha}_0 - \hat{\boldsymbol{U}}_i^{\mathrm{T}}\boldsymbol{\gamma}_0\Big)\Big\} \ge 1-\eta. \end{aligned} \tag{A.1}$$

This means that there is a local minimizer $\hat{\boldsymbol{\alpha}}$, $\hat{\boldsymbol{b}}$ and $\hat{\boldsymbol{\gamma}}$ in the ball $\left\{(\boldsymbol{V}_n,\boldsymbol{W}_n,\boldsymbol{S}_n) : \left\|(\boldsymbol{V}_n^{\mathrm{T}},\boldsymbol{W}_n^{\mathrm{T}},\boldsymbol{S}_n^{\mathrm{T}})^{\mathrm{T}}\right\| \le L\right\}$ such that $\|\hat{\boldsymbol{\alpha}}-\boldsymbol{\alpha}_0\| = O_p(\delta_n)$, $\left|\hat{b}_k - b_{0k}\right| = O_p(\delta_n)$, $\|\Lambda^{1/2}(\hat{\boldsymbol{\gamma}}-\boldsymbol{\gamma}_0)\| = O_p(\delta_n)$ with a probability at least $1-\eta$.

First, by $\|v_j - \hat{v}_j\|^2 = O_p(n^{-1}j^2)$ (see, [4]), one has

$$\begin{aligned} |r_i|^2 &= \left|\int_0^1 \beta_0(t)X_i(t)dt - \hat{\boldsymbol{U}}_i^{\mathrm{T}}\boldsymbol{\gamma}_0\right|^2 \\ &\le 2\left|\sum_{j=1}^m\langle X_i, \hat{v}_j - v_j\rangle\gamma_{0j}\right|^2 + 2\left|\sum_{j=m+1}^{\infty}\langle X_i, v_j\rangle\gamma_{0j}\right|^2 \triangleq 2\mathrm{D}_1 + 2\mathrm{D}_2. \end{aligned}$$

For D_1, using Conditions C1, C2, and the Hölder inequality, one can obtain

$$\begin{aligned} \mathrm{D}_1 &= \left|\sum_{j=1}^m\langle X_i, v_j - \hat{v}_j\rangle\gamma_{0j}\right|^2 \\ &\le Cm\sum_{j=1}^m\|v_j - \hat{v}_j\|^2|\gamma_{0j}|^2 \le Cm\sum_{j=1}^m O_p(n^{-1}j^{2-2b}) = O_p\left(\frac{m}{n}\right) = O_p(\delta_n^2). \end{aligned}$$

As for D_2, due to $\mathrm{E}\left\{\sum_{j=m+1}^{\infty}\langle X_i, v_j\rangle\gamma_{0j}\right\} = 0$, $\mathrm{Var}\left\{\sum_{j=m+1}^{\infty}\langle X_i, v_j\rangle\gamma_{0j}\right\} = \sum_{j=m+1}^{\infty}\lambda_j{\gamma_{0j}}^2 \le C\sum_{j=m+1}^{\infty}j^{-(a+2b)} = O(n^{-\frac{a+2b-1}{a+2b}})$, one has $D_2 = O_p(n^{-\frac{a+2b-1}{a+2b}}) = O_p(\delta_n^2)$. To sum up, we have $|r_i|^2 = O_p(\delta_n^2)$.

Now consider B_i. Due to

$$\|B_i\|^2 = \left\| H_m^T(\boldsymbol{U}_i - \hat{\boldsymbol{U}}_i) + \sum_{j=m+1}^{\infty} \lambda_j^{-1}\langle c_{ZX}, v_j\rangle \xi_{ij} \right\|^2$$
$$\leq 2\sum_{l=1}^{d}\left\{ \left\| \sum_{j=1}^{m} \lambda_j^{-1}\langle c_{Z_l X}, v_j\rangle\langle X_i, \hat{v}_j - v_j\rangle \right\|^2 + \left\| \sum_{j=m+1}^{\infty} \lambda_j^{-1}\langle c_{Z_l X}, v_j\rangle\langle X_i, v_j\rangle \right\|^2 \right\},$$

by Conditions C1, C2, C4, and the Hölder inequality, we obtain

$$\left\| \sum_{j=1}^{m} \lambda_j^{-1}\langle c_{Z_l X}, v_j\rangle\langle X_i, \hat{v}_j - v_j\rangle \right\|^2 \leq Cm\sum_{j=1}^{m} \|v_j - \hat{v}_j\|^2 |\lambda_j^{-1}\langle c_{Z_l X}, v_j\rangle|^2$$
$$\leq Cm\sum_{j=1}^{m} O_p(n^{-1} j^{2-2b}) = O_p\left(\frac{m}{n}\right) = O_p\left(\delta_n^2\right).$$

In addition, noting that

$$\mathrm{E}\Big\{ \sum_{j=m+1}^{\infty} \lambda_j^{-1}\langle c_{Z_l X}, v_j\rangle\langle X_i, v_j\rangle \Big\} = 0,$$
$$\mathrm{Var}\Big\{ \sum_{j=m+1}^{\infty} \lambda_j^{-1}\langle c_{Z_l X}, v_j\rangle\langle X_i, v_j\rangle \Big\} = \sum_{j=m+1}^{\infty} \lambda_j^{-1}\langle c_{Z_l X}, v_j\rangle^2 = O(n^{-\frac{a+2b-1}{a+2b}}),$$

together with the above inequality, one has

$$\|B_i\|^2 = O_p\left(\frac{m}{n}\right) = O_p(\delta_n^2). \tag{A.2}$$

Recall that $\psi_{\tau_k}(u) = \rho'_{\tau_k}(u) = (\tau_k - 1)I(u < -c^*) + \frac{(1-\tau_k)}{c^*}uI(-c^* \leq u < 0) + \frac{\tau_k}{c^*}uI(0 \leq u < c^*) + \tau_k I(u \geq c^*)$, and denote $\widetilde{Q}_n(\boldsymbol{V}_n, \boldsymbol{W}_n, \boldsymbol{S}_n) = \sum_{i=1}^{n}\sum_{k=1}^{K} w_k\rho_{\tau_k}\Big(Y_i - b_{0k} - \boldsymbol{Z}_i^{\mathrm{T}}\boldsymbol{\alpha}_0 - \hat{\boldsymbol{U}}_i^{\mathrm{T}}\boldsymbol{\gamma}_0 - \delta_n(\boldsymbol{V}_n^{\mathrm{T}}\tilde{\boldsymbol{\eta}}_i + \boldsymbol{W}_n^{\mathrm{T}}\hat{\boldsymbol{A}}_i + \tilde{B}_i + S_{nk})\Big) - \sum_{i=1}^{n}\sum_{k=1}^{K} w_k\rho_{\tau_k}\Big(Y_i - b_{0k} - \boldsymbol{Z}_i^{\mathrm{T}}\boldsymbol{\alpha}_0 - \hat{\boldsymbol{U}}_i^{\mathrm{T}}\boldsymbol{\gamma}_0\Big)$. Then $\widetilde{Q}_n(\boldsymbol{V}_n, \boldsymbol{W}_n, \boldsymbol{S}_n)$ can be transformed into

$$\widetilde{Q}_n(\boldsymbol{V}_n, \boldsymbol{W}_n, \boldsymbol{S}_n) = E[\widetilde{Q}_n(\boldsymbol{V}_n, \boldsymbol{W}_n, \boldsymbol{S}_n)|T_n] + \sum_{k=1}^{K} w_k \sum_{i=1}^{n} R_{nik}(\boldsymbol{V}_n, \boldsymbol{W}_n, \boldsymbol{S}_n)$$
$$- \sum_{k=1}^{K} w_k \sum_{i=1}^{n} \delta_n\left(\boldsymbol{V}_n^{\mathrm{T}}\tilde{\boldsymbol{\eta}}_i + \boldsymbol{W}_n^{\mathrm{T}}\hat{\boldsymbol{A}}_i + \tilde{B}_i + S_{nk}\right)\psi_{\tau_k}(e_i - b_{0k}))$$
$$\triangleq D_1^* + D_2^* + D_3^*,$$

where

$$R_{nik}(\boldsymbol{V}_n, \boldsymbol{W}_n, \boldsymbol{S}_n)$$
$$= \rho_{\tau_k}\Big(r_i + e_i - b_{0k} - \delta_n(\boldsymbol{V}_n^{\mathrm{T}}\tilde{\boldsymbol{\eta}}_i + \boldsymbol{W}_n^{\mathrm{T}}\hat{\boldsymbol{A}}_i + \tilde{B}_i + S_{nk})\Big) - \rho_{\tau_k}(r_i + e_i - b_{0k})$$

$$- E\left[\left\{\rho_{\tau_k}(r_i + e_i - b_{0k} - \delta_n(\boldsymbol{V}_n^{\mathrm{T}}\tilde{\boldsymbol{\eta}}_i + \boldsymbol{W}_n^{\mathrm{T}}\hat{\boldsymbol{A}}_i + \tilde{B}_i + S_{nk})) - \rho_{\tau_k}(r_i + e_i - b_{0k})\right\}\middle| T_n\right]$$
$$+ \delta_n(\boldsymbol{V}_n^{\mathrm{T}}\tilde{\boldsymbol{\eta}}_i + \boldsymbol{W}_n^{\mathrm{T}}\hat{\boldsymbol{A}}_i + \tilde{B}_i + S_{nk})\psi_{\tau_k}(e_i - b_{0k}).$$

Consider D_1^*. According to $(A.2)$, we have

$$\boldsymbol{V}_n^{\mathrm{T}}\tilde{\boldsymbol{\eta}}_i + \boldsymbol{W}_n^{\mathrm{T}}\hat{\boldsymbol{A}}_i + \tilde{B}_i + S_{nk} = \left(\boldsymbol{V}_n^{\mathrm{T}}\tilde{\boldsymbol{\eta}}_i + \boldsymbol{W}_n^{\mathrm{T}}\hat{\boldsymbol{A}}_i + S_{nk}\right)(1 + o_p(1)). \tag{A.3}$$

The proof of Theorem 3.1 in [6] indicates that $\left|\frac{1}{n}\sum_{i=1}^{n}\left[\left(\boldsymbol{A}_i^{\mathrm{T}}h^m\right)^2\right] - 1\right| = O_P\left(n^{-1/4}m^{1/2}(\log n)^{1/2}\right) = o_P(1)$ with $h^m = \boldsymbol{W}_n/\|\boldsymbol{W}_n\|$, which leads to

$$\frac{1}{n}\sum_{i=1}^{n}\left(\hat{\boldsymbol{A}}_i^{\mathrm{T}}\boldsymbol{W}_n\right)^2 = \|\boldsymbol{W}_n\|^2(1 + o_P(1)). \tag{A.4}$$

Observe that $\sum_{i=1}^{n} h_{\tau_k}(b_{0k})\boldsymbol{V}_n^{\mathrm{T}}\tilde{\boldsymbol{\eta}}_i\boldsymbol{A}_i^{\mathrm{T}}\boldsymbol{W}_n = \boldsymbol{V}_n^{\mathrm{T}}P_n^{-1/2}\sum_{i=1}^{n} h_{\tau_k}(b_{0k})\boldsymbol{\eta}_i\boldsymbol{U}_i^{\mathrm{T}}\Lambda^{-1/2}\boldsymbol{W}_n$, then by Conditions C1–C3, C5, $E\left[\sum_{i=1}^{n} h_{\tau_k}(b_{0k})\eta_{il}\boldsymbol{U}_i^{\mathrm{T}}\Lambda^{-1/2}\boldsymbol{W}_n\right] = 0$ and $E\left(\left[\sum_{i=1}^{n} h_{\tau_k}(b_{0k})\eta_{il}\boldsymbol{U}_i^{\mathrm{T}}\Lambda^{-1/2}\boldsymbol{W}_n\right]^2\right) = O(nm)$. Hence,

$$\sum_{i=1}^{n} h_{\tau_k}(b_{0k})\boldsymbol{V}_n^{\mathrm{T}}\tilde{\boldsymbol{\eta}}_i\boldsymbol{A}_i^{\mathrm{T}}\boldsymbol{W}_n = O_p\left(n^{1/2}m^{1/2}\right). \tag{A.5}$$

Similarly, $\sum_{i=1}^{n} h_{\tau_k}(b_{0k})\boldsymbol{V}_n^{\mathrm{T}}\tilde{\boldsymbol{\eta}}_iS_{nk} = O_p\left(n^{1/2}\right)$, $\sum_{i=1}^{n} h_{\tau_k}(b_{0k})\boldsymbol{A}_i^{\mathrm{T}}\boldsymbol{W}_nS_{nk} = O_p\left(n^{1/2}m^{1/2}\right)$. Then, together with formulas $(A.3)$–$(A.5)$, we have

$$\begin{aligned}
D_1^* &= E[\widetilde{Q}_n(\boldsymbol{V}_n, \boldsymbol{W}_n, \boldsymbol{S}_n)\,|T_n] \\
&= \sum_{k=1}^{K} w_k\sum_{i=1}^{n}\int_{r_i}^{r_i-\delta_n\left(\boldsymbol{V}_n^{\mathrm{T}}\tilde{\boldsymbol{\eta}}_i+\boldsymbol{W}_n^{\mathrm{T}}\hat{\boldsymbol{A}}_i+\tilde{B}_i+S_{nk}\right)} E[\psi_{\tau_k}(e_i - b_{0k} + t)|T_n]dt \\
&= \frac{1}{2}\sum_{k=1}^{K} w_k\sum_{i=1}^{n} h_{\tau_k}(b_{0k})\left\{\left(r_i - \delta_n\left(\boldsymbol{V}_n^{\mathrm{T}}\tilde{\boldsymbol{\eta}}_i + \boldsymbol{W}_n^{\mathrm{T}}\hat{\boldsymbol{A}}_i + \tilde{B}_i + S_{nk}\right)\right)^2 - {r_i}^2\right\}(1 + o_p(1)) \\
&\geq Cn\delta_n^2\left(\|\boldsymbol{V}_n\|^2 + \|\boldsymbol{W}_n\|^2 + \|\boldsymbol{S}_n\|^2\right)(1 + o_p(1)).
\end{aligned}$$

As for D_2^*, due to the continuity of $\psi_{\tau_k}(\cdot)$, we have

$$Var\left(\sum_{i=1}^{n} R_{nik}(\boldsymbol{V}_n, \boldsymbol{W}_n, \boldsymbol{S}_n)\middle| T_n\right) = o_p\left(n\delta_n^2(\|\boldsymbol{V}_n\|^2 + \|\boldsymbol{W}_n\|^2 + |S_{nk}|^2)\right),$$

then

$$Var\left(\sum_{k=1}^{K} w_k\sum_{i=1}^{n} R_{nik}(\boldsymbol{V}_n, \boldsymbol{W}_n, \boldsymbol{S}_n)\middle| T_n\right) = o_p\left(n\delta_n^2\left(\|\boldsymbol{V}_n\|^2 + \|\boldsymbol{W}_n\|^2 + \|\boldsymbol{S}_n\|^2\right)\right),$$

from which we get

$$\sup_{\|(\boldsymbol{V}_n, \boldsymbol{W}_n, \boldsymbol{S}_n)\|\leq L}\left|D_2^*\right| = \sup_{\|(\boldsymbol{V}_n, \boldsymbol{W}_n, \boldsymbol{S}_n)\|\leq L}\left|\sum_{k=1}^{K} w_k\sum_{i=1}^{n} R_{nik}(\boldsymbol{V}_n, \boldsymbol{W}_n, \boldsymbol{S}_n)\right| = o_p(\sqrt{n}\delta_n L).$$

For the term D_3^*, it is easy to show that

$$E\left[\sum_{k=1}^{K} w_k \sum_{i=1}^{n} \delta_n \left(\boldsymbol{V}_n^{\mathrm{T}}\tilde{\boldsymbol{\eta}}_i + \boldsymbol{W}_n^{\mathrm{T}}\boldsymbol{A}_i + S_{nk}\right)\psi_{\tau_k}(e_i - b_{0k})\middle|T_n\right] = 0,$$

and

$$E\left[\left\{\sum_{k=1}^{K} w_k \sum_{i=1}^{n} \delta_n \left(\boldsymbol{V}_n^{\mathrm{T}}\tilde{\boldsymbol{\eta}}_i + \boldsymbol{W}_n^{\mathrm{T}}\boldsymbol{A}_i + S_{nk}\right)\psi_{\tau_k}(e_i - b_{0k})\right\}^2 \middle|T_n\right]$$
$$\leq Cn\delta_n^2\left(\|\boldsymbol{V}_n\|^2 + \|\boldsymbol{W}_n\|^2 + \|\boldsymbol{S}_n\|^2\right)(1 + o_p(1)).$$

Combining with the equation $(A.4)$, we can obtain

$$\sup_{\|(\boldsymbol{V}_n,\boldsymbol{W}_n,\boldsymbol{S}_n)\|\leq L}\left|D_3^*\right| = \sup_{\|(\boldsymbol{V}_n,\boldsymbol{W}_n,\boldsymbol{S}_n)\|\leq L}\left|\sum_{k=1}^{K} w_k \sum_{i=1}^{n} \delta_n \left(\boldsymbol{V}_n^{\mathrm{T}}\tilde{\boldsymbol{\eta}}_i + \boldsymbol{W}_n^{\mathrm{T}}\hat{\boldsymbol{A}}_i + \tilde{B}_i + S_{nk}\right)\psi_{\tau_k}(e_i - b_{0k})\right| = O_p(\delta_n n^{1/2} L).$$

From the results about D_1^*, D_2^* and D_3^*, it is easy to obtain that $\widetilde{Q}_n(\boldsymbol{V}_n, \boldsymbol{W}_n, \boldsymbol{S}_n)$ is dominated by the positive quadratic term $Cn\delta_n^2\left(\|\boldsymbol{V}_n\|^2 + \|\boldsymbol{W}_n\|^2 + \|\boldsymbol{S}_n\|^2\right)$. Hence, Eq $(A.1)$ is established, and there exists local minimizer $\hat{\boldsymbol{\gamma}}$ such that

$$\|\hat{\boldsymbol{\alpha}} - \boldsymbol{\alpha}_0\| = O_p(\delta_n), \quad \left|\hat{b}_k - b_{0k}\right| = O_p(\delta_n), \quad \|\Lambda^{1/2}(\hat{\boldsymbol{\gamma}} - \boldsymbol{\gamma}_0)\|^2 = O_p(\delta_n^2). \tag{A.6}$$

Now we consider the convergence rate of $\hat{\beta}$. Since

$$\|\Lambda^{\frac{1}{2}}(\hat{\boldsymbol{\gamma}} - \boldsymbol{\gamma}_0)\|^2 = \sum_{j=1}^{m} \lambda_j(\hat{\gamma}_j - \gamma_{0j})^2 \geq \lambda_m \sum_{j=1}^{m}(\hat{\gamma}_j - \gamma_{0j})^2,$$

and based on the basic Condition C2, we have

$$\|\hat{\boldsymbol{\gamma}} - \boldsymbol{\gamma}_0\|^2 \leq O_p(\lambda_m^{-1}\frac{m}{n}) = O_p(m^{a+1}n^{-1}) = O_p(n^{-\frac{2b-1}{a+2b}}). \tag{A.7}$$

Note that

$$\begin{aligned}\|\hat{\beta} - \beta_0\|^2 &= \left\|\sum_{j=1}^{m}\hat{\gamma}_j\hat{v}_j - \sum_{j=1}^{\infty}\gamma_{0j}v_j\right\|^2 \\ &\leq 4\left\|\sum_{j=1}^{m}(\hat{\gamma}_j - \gamma_{0j})\hat{v}_j\right\|^2 + 4\left\|\sum_{j=1}^{m}\gamma_{0j}(\hat{v}_j - v_j)\right\|^2 + 2\sum_{j=m+1}^{\infty}\gamma_{0j}^2 \\ &\triangleq 4D_1^{**} + 4D_2^{**} + 2D_3^{**}.\end{aligned}$$

Based on the Condition C2, Eq $(A.7)$ and the orthogonality of $\{\hat{v}_j\}$, as well as $\|v_j - \hat{v}_j\|^2 = O_p(n^{-1}j^2)$, we can obtain

$$D_1^{**} = \left\|\sum_{j=1}^{m}(\hat{\gamma}_j - \gamma_{0j})\hat{v}_j\right\|^2 \leq \sum_{j=1}^{m}\left|\hat{\gamma}_j - \gamma_{0j}\right|^2 = \|\hat{\boldsymbol{\gamma}} - \boldsymbol{\gamma}_0\|^2 = O_p(n^{-\frac{2b-1}{a+2b}}),$$

$$D_2^{**} = \Big\|\sum_{j=1}^{m}\gamma_{0j}(\hat{v}_j - v_j)\Big\|^2 \le m\sum_{j=1}^{m}\|\hat{v}_j - v_j\|^2\,\gamma_{0j}^2 \le \frac{m}{n}O_p(\sum_{j=1}^{m} j^2\gamma_{0j}^2)$$
$$= O_p\Big(n^{-1}m\sum_{j=1}^{m} j^{2-2b}\Big) = O_p(n^{-1}m) = o_p(n^{-\frac{2b-1}{a+2b}}), \tag{A.8}$$
$$D_3^{**} = \sum_{j=m+1}^{\infty}\gamma_{0j}^2 \le C\sum_{j=m+1}^{\infty} j^{-2b} = O_p(n^{-\frac{2b-1}{a+2b}}).$$

These lead to

$$\|\hat{\beta} - \beta_0\|^2 = O_p(n^{-\frac{2b-1}{a+2b}}).$$

Next, we turn to the asymptotic normality of $\hat{\alpha}$. Note that $Q_n(\boldsymbol{\alpha}, \boldsymbol{\gamma}, \boldsymbol{b})$ attains the minimal value at $(\hat{\boldsymbol{\alpha}}, \hat{\boldsymbol{\gamma}}, \hat{\boldsymbol{b}})$ with probability tending to 1 as n tends to infinity. Then, we have the following score equations

$$\frac{1}{n}\sum_{k=1}^{K} w_k \sum_{i=1}^{n} \boldsymbol{Z}_i\psi_{\tau_k}\left(Y_i - \hat{b}_k - \boldsymbol{Z}_i^{\mathrm{T}}\hat{\boldsymbol{\alpha}} - \hat{\boldsymbol{U}}_i^{\mathrm{T}}\hat{\boldsymbol{\gamma}}\right) = 0, \tag{A.9}$$

$$\frac{1}{n}\sum_{k=1}^{K} w_k \sum_{i=1}^{n} \hat{\boldsymbol{U}}_i\psi_{\tau_k}\left(Y_i - \hat{b}_k - \boldsymbol{Z}_i^{\mathrm{T}}\hat{\boldsymbol{\alpha}} - \hat{\boldsymbol{U}}_i^{\mathrm{T}}\hat{\boldsymbol{\gamma}}\right) = 0. \tag{A.10}$$

Further, we can write $(A.9)$ as $H_n + \sum_{k=1}^{K} w_k B_{n1}^{(k)} + \sum_{k=1}^{K} w_k B_{n2}^{(k)} = 0$, with

$$H_n = \frac{1}{n}\sum_{k=1}^{K} w_k \sum_{i=1}^{n} \boldsymbol{Z}_i\psi_{\tau_k}(e_i - b_{0k}),$$
$$B_{n1}^{(k)} = \frac{1}{n}\sum_{i=1}^{n}\boldsymbol{Z}_i E\Big[\psi_{\tau_k}\Big(e_i - b_{0k} + r_i - (\hat{b}_k - b_{0k}) - \boldsymbol{Z}_i^{\mathrm{T}}(\hat{\boldsymbol{\alpha}} - \boldsymbol{\alpha}_0) - \hat{\boldsymbol{U}}_i^{\mathrm{T}}(\hat{\boldsymbol{\gamma}} - \boldsymbol{\gamma}_0)\Big) - \psi_{\tau_k}(e_i - b_{0k})\Big|T_n\Big],$$

$$B_{n2}^{(k)} = \frac{1}{n}\sum_{i=1}^{n}\boldsymbol{Z}_i\Big[\Big\{\psi_{\tau_k}\Big(e_i - b_{0k} + r_i - (\hat{b}_k - b_{0k}) - \boldsymbol{Z}_i^{\mathrm{T}}(\hat{\boldsymbol{\alpha}} - \boldsymbol{\alpha}_0) - \hat{\boldsymbol{U}}_i^{\mathrm{T}}(\hat{\boldsymbol{\gamma}} - \boldsymbol{\gamma}_0)\Big) - \psi_{\tau_k}(e_i - b_{0k})\Big\}$$
$$- E\Big\{\psi_{\tau_k}\Big(e_i - b_{0k} + r_i - (\hat{b}_k - b_{0k}) - \boldsymbol{Z}_i^{\mathrm{T}}(\hat{\boldsymbol{\alpha}} - \boldsymbol{\alpha}_0) - \hat{\boldsymbol{U}}_i^{\mathrm{T}}(\hat{\boldsymbol{\gamma}} - \boldsymbol{\gamma}_0)\Big) - \psi_{\tau_k}(e_i - b_{0k})\Big|T_n\Big\}\Big].$$

By simple calculations, we have $B_{n1}^{(k)} = -\frac{1}{n}\sum_{i=1}^{n} h_{\tau_k}(b_{0k})\left[\boldsymbol{Z}_i\boldsymbol{Z}_i^{\mathrm{T}}(\hat{\boldsymbol{\alpha}} - \boldsymbol{\alpha}_0) + \boldsymbol{Z}_i\hat{\boldsymbol{U}}_i^{\mathrm{T}}(\hat{\boldsymbol{\gamma}} - \boldsymbol{\gamma}_0)\right](1 + o_p(1))$. Through calculating the mean and variance directly, we can obtain $B_{n2}^{(k)} = o_p(\delta_n)$. Then, Eq $(A.9)$ can be written as

$$\begin{aligned}&\frac{1}{n}\sum_{k=1}^{K} w_k \sum_{i=1}^{n} \boldsymbol{Z}_i \psi_{\tau_k}(e_i - b_{0k}) \\ =& \frac{1}{n}\sum_{k=1}^{K}\sum_{i=1}^{n} w_k h_{\tau_k}(b_{0k})\left[\boldsymbol{Z}_i\boldsymbol{Z}_i^{\mathrm{T}}(\hat{\boldsymbol{\alpha}}-\boldsymbol{\alpha}_0)+\boldsymbol{Z}_i\hat{\boldsymbol{U}}_i^{\mathrm{T}}(\hat{\boldsymbol{\gamma}}-\boldsymbol{\gamma}_0)\right](1+o_p(1)).\end{aligned} \tag{A.11}$$

Similarly, Eq $(A.10)$ can be changed into

$$\begin{aligned}&\frac{1}{n}\sum_{k=1}^{K} \omega_k \sum_{i=1}^{n} \hat{\boldsymbol{U}}_i \psi_{\tau_k}(e_i - b_{0k}) \\ =& \frac{1}{n}\sum_{k=1}^{K}\sum_{i=1}^{n} \omega_k h_{\tau_k}(b_{0k})\left[\hat{\boldsymbol{U}}_i\boldsymbol{Z}_i^{T}(\hat{\boldsymbol{\alpha}}-\boldsymbol{\alpha}_0)+\hat{\boldsymbol{U}}_i\hat{\boldsymbol{U}}_i^{T}(\hat{\boldsymbol{\gamma}}-\boldsymbol{\gamma}_0)\right](1+o_p(1)).\end{aligned}$$

Now let $\Phi_n = \frac{1}{n}\sum_{i=1}^n \hat{\boldsymbol{U}}_i\hat{\boldsymbol{U}}_i^{\mathrm{T}}$, $\Gamma_n = \frac{1}{n}\sum_{i=1}^n \hat{\boldsymbol{U}}_i\boldsymbol{Z}_i^{\mathrm{T}}$, $\Gamma = E(\boldsymbol{U}_i\boldsymbol{Z}_i^{\mathrm{T}})$, $\Sigma_n = \frac{1}{n}\sum_{i=1}^n \widetilde{\boldsymbol{Z}}_i\widetilde{\boldsymbol{Z}}_i^{\mathrm{T}}$, $\widetilde{\boldsymbol{Z}}_i = \boldsymbol{Z}_i - \Gamma_n^{\mathrm{T}}\Phi_n^{-1}\hat{\boldsymbol{U}}_i$, $\Upsilon_{nk} = \frac{1}{n}\sum_{i=1}^n \hat{\boldsymbol{U}}_i\psi_{\tau_k}(e_i - b_{0k})$. Then, above equation allows $\sum_{k=1}^K w_k h_{\tau_k}(b_{0k})(\hat{\boldsymbol{\gamma}}-\boldsymbol{\gamma}_0) = \sum_{k=1}^K w_k \left(\Phi_n + o_p(1)\right)^{-1}[\Upsilon_{nk} + h_{\tau_k}(b_{0k})\Gamma_n(\boldsymbol{\alpha}_0 - \hat{\boldsymbol{\alpha}})]$. Substituting it into $(A.11)$, we obtain

$$\begin{aligned}&\frac{1}{n}\sum_{k=1}^{K} w_k \sum_{i=1}^{n} h_{\tau_k}(b_{0k})\boldsymbol{Z}_i\left[\boldsymbol{Z}_i - \Gamma_n^{\mathrm{T}}\Phi_n^{-1}\hat{\boldsymbol{U}}_i\right]^{\mathrm{T}}(\hat{\boldsymbol{\alpha}}-\boldsymbol{\alpha}_0)\left(1+o_p(1)\right) \\ =&\frac{1}{n}\sum_{k=1}^{K} w_k \sum_{i=1}^{n} \boldsymbol{Z}_i\left[\psi_{\tau_k}(e_i - b_{0k}) - \hat{\boldsymbol{U}}_i^{\mathrm{T}}(\Phi_n)^{-1}\Upsilon_{nk}\right].\end{aligned}$$

Noting that

$$\begin{aligned}&\frac{1}{n}\sum_{i=1}^{n}\sum_{k=1}^{K} w_k h_{\tau_k}(b_{0k})\Gamma_n^{\mathrm{T}}\Phi_n^{-1}\hat{\boldsymbol{U}}_i[\boldsymbol{Z}_i - \Gamma_n^{\mathrm{T}}\Phi_n^{-1}\hat{\boldsymbol{U}}_i]^{\mathrm{T}} = 0 \\ &\frac{1}{n}\sum_{i=1}^{n}\sum_{k=1}^{K} w_k \Gamma_n^{\mathrm{T}}\Phi_n^{-1}\hat{\boldsymbol{U}}_i\{\psi_{\tau_k}(e_i - b_{0k}) - \hat{\boldsymbol{U}}_i^{\mathrm{T}}\Phi_n^{-1}\Upsilon_{nk}\} = 0,\end{aligned}$$

then, it is easy to conclude that

$$\begin{aligned}&\left(\frac{1}{n}\sum_{k=1}^{K} w_k \sum_{i=1}^{n} h_{\tau_k}(b_{0k})\widetilde{\boldsymbol{Z}}_i\widetilde{\boldsymbol{Z}}_i^{\mathrm{T}}\right)\sqrt{n}(\hat{\boldsymbol{\alpha}}-\boldsymbol{\alpha}_0) \\ =&\frac{1}{\sqrt{n}}\sum_{k=1}^{K} w_k \sum_{i=1}^{n} \widetilde{\boldsymbol{Z}}_i\left[\psi_{\tau_k}(e_i - b_{0k}) - \hat{\boldsymbol{U}}_i^{\mathrm{T}}(\Phi_n)^{-1}\Upsilon_{nk}\right](1+o_p(1)).\end{aligned}$$

Note that

$$\frac{1}{\sqrt{n}}\sum_{k=1}^{K} w_k \sum_{i=1}^{n} \tilde{Z}_i\hat{U}_i^{\mathrm{T}} = 0.$$

Then we have

$$\left(\frac{1}{n}\sum_{k=1}^{K} w_k \sum_{i=1}^{n} h_{\tau_k}(b_{0k})\widetilde{\boldsymbol{Z}}_i\widetilde{\boldsymbol{Z}}_i^{\mathrm{T}}\right)\sqrt{n}\,(\hat{\boldsymbol{\alpha}}-\boldsymbol{\alpha}_0) = \frac{1}{\sqrt{n}}\sum_{k=1}^{K} w_k \sum_{i=1}^{n} \widetilde{\boldsymbol{Z}}_i \psi_{\tau_k}(e_i - b_{0k})(1+o_p(1)). \tag{A.12}$$

It is easy to see that $\Phi_n = \Lambda + o_p(1)$, $\Gamma_n = \Gamma + o_p(1)$. And based on Lemma 1 in [8] and the Condition C5, we can obtain $\frac{1}{n}\sum_{i=1}^{n}\widetilde{\boldsymbol{Z}}_i\widetilde{\boldsymbol{Z}}_i^{\mathrm{T}} \xrightarrow{p} \boldsymbol{\Sigma}$ $(n\to\infty)$. Through some simple calculations, one has

$$\begin{aligned} &\frac{1}{n}\sum_{k=1}^{K} w_k \sum_{i=1}^{n} h_{\tau_k}(b_{0k})\widetilde{\boldsymbol{Z}}_i\widetilde{\boldsymbol{Z}}_i^{\mathrm{T}} \xrightarrow{p} \sum_{k=1}^{K} w_k h_{\tau_k}(b_{0k})\boldsymbol{\Sigma},\\ &\mathrm{Var}\left(\sum_{k=1}^{K} w_k \psi_{\tau_k}(e_i - b_{0k})\right) = \boldsymbol{w}^{\mathrm{T}}\boldsymbol{V}\boldsymbol{w}, \end{aligned} \tag{A.13}$$

where $\boldsymbol{V} = (V_{k,l})_{1\le k,l\le K}$ with $V_{kl} = E[\psi_{\tau_k}(e_i - b_{0k})\psi_{\tau_l}(e_i - b_{0l})]$ and $\boldsymbol{w} = (w_1,\ldots,w_k)^{\mathrm{T}}$. Then, according to Eqs $(A.12)$ and $(A.13)$, the Slutsky's theorem, and the properties of multivariate normal distributions, we can obtain

$$\sqrt{n}(\hat{\boldsymbol{\alpha}}-\boldsymbol{\alpha}_0) \xrightarrow{d} N\left(0, \frac{\boldsymbol{w}^{\mathrm{T}}\boldsymbol{V}\boldsymbol{w}}{\left\{\sum_{k=1}^{K} w_k h_{\tau_k}(b_{0k})\right\}^2}\boldsymbol{\Sigma}^{-1}\right).$$

Lastly, we prove the third conclusion of Theorem 2. By the definition of MSPE_{WCAHR}, we have

$$\begin{aligned} &\mathrm{MSPE}_{WCAHR}\\ &\le 5\sum_{j=1}^{m}(\hat{\gamma}_j-\gamma_{0j})^2\lambda_j + 5C\left\|\sum_{j=1}^{m}\hat{\gamma}_j(v_j-\hat{v}_j)\right\|^2 + 5\sum_{j=m+1}^{\infty}\gamma_{0j}^2\lambda_j + 5C\|\hat{\boldsymbol{\alpha}}-\boldsymbol{\alpha}_0\|^2 + 5\left\{\sum_{k=1}^{K} w_k\left|\hat{b}_k - b_{0k}\right|\right\}^2\\ &\triangleq 5F_1 + 5CF_2 + 5F_3 + 5CF_4 + 5F_5. \end{aligned}$$

Firstly, according to the previous proof process, we have $F_1 = \|\Lambda^{1/2}(\hat{\boldsymbol{\gamma}}-\boldsymbol{\gamma}_0)\|^2 = O_p(\frac{m}{n})$. As for F_2, based on triangle inequality and C_R inequality,

$$\begin{aligned} F_2 &= \left\|\sum_{j=1}^{m}\hat{\gamma}_j(v_j-\hat{v}_j)\right\|^2 = \left\|\sum_{j=1}^{m}\gamma_{0j}(v_j-\hat{v}_j) + (\hat{\gamma}_j-\gamma_{0j})(v_j-\hat{v}_j)\right\|^2\\ &\le 2m\sum_{j=1}^{m}\gamma_{0j}^2\|v_j-\hat{v}_j\|^2 + 2\|\Lambda^{\frac{1}{2}}(\hat{\boldsymbol{\gamma}}-\boldsymbol{\gamma}_0)\|^2\sum_{j=1}^{m}\lambda_j^{-1}(v_j-\hat{v}_j)^2\\ &\triangleq 2F_{21} + 2F_{22}. \end{aligned}$$

By Eq (A.8), $F_{21} = m\sum_{j=1}^{m}\gamma_{0j}^2\|v_j-\hat{v}_j\|^2 = O_p\left(\frac{m}{n}\right)$ by . As for F_{22}, it is easy to know $\sum_{j=1}^{m}\lambda_j^{-1}(v_j-\hat{v}_j)^2 = o_p(1)$, then $F_{22} = \|\Lambda^{\frac{1}{2}}(\hat{\boldsymbol{\gamma}}-\boldsymbol{\gamma}_0)\|^2\sum_{j=1}^{m}\lambda_j^{-1}(v_j-\hat{v}_j)^2 = o_p\left(\frac{m}{n}\right)$. Next, by (A.8), we have $F_3 = \sum_{j=m+1}^{\infty} \gamma_{0j}^2 \lambda_j = O(m^{-a-2b+1})$. By (A.6), we know $F_4 = O_p\left(\frac{m}{n}\right)$ and $F_5 = O_p\left(\frac{m}{n}\right)$. Then, we have

$$\mathrm{MSPE}_{WCAHR} = O_p(n^{-\frac{a+2b-1}{a+2b}}).$$

The proof of Theorem 2 is complete.