Partial Functional Linear Models with ARCH Errors

Yafei Wang, Tianfa Xie, Zhongzhan Zhang

College of Applied Sciences, Beijing University of Technology, Beijing, China
Email: zzhang@bjut.edu.cn

Abstract

In this paper, the estimation of the parameters in partial functional linear models with ARCH(p) errors is discussed. With employing the functional principle component, a hybrid estimating method is suggested. The asymptotic normality of the proposed estimators for both the linear parameter in the mean model and the parameter in the ARCH error model is obtained, and the convergence rate of the slope function estimate is established. Besides, some simulations and a real data analysis are conducted for illustration, and it is shown that the proposed method performs well with a finite sample.

Keywords

Asymptotic Normality, ARCH(p) Errors, Functional Principal Components, Convergence Rate, Least Absolute Deviation

1. Introduction

In order to combine the flexibility of linear regression models with the recent methodology for the functional linear regression models, partial functional linear models, which was introduced by [1], is considered as follows:

$$Y = \boldsymbol{\beta}'\boldsymbol{z} + \int_{\mathcal{T}} \gamma(t) X(t) \mathrm{d}t + \varepsilon, \qquad (1.1)$$

where Y is a real-valued response random variable, $\boldsymbol{z}$ is a d-dimensional vector of random variables with zero means and finite second moments, and $X(t)$ is an explanatory functional variable defined on $\mathcal{T}$ with zero mean and finite second moments (*i.e.* $\mathrm{E}|X(t)|^2 < \infty$, for all $t \in \mathcal{T}$), $\boldsymbol{\beta}$ is a d-dimensional vector of unknown parameters, $\gamma(t)$ is a square integrable function on $\mathcal{T}$, ε is a random error and is independent of $\boldsymbol{z}$ and X. For simplicity, without loss of generality, it is assumed that $\mathcal{T} = [0,1]$ in the remainder of this paper. All the random variables are defined on the same probability space $(\Omega, \mathcal{A}, P)$.

Model (1.1) has been studied by many authors from different points. From the view of estimation of model (1.1), for example, reference [2] studied the estimate of the model (1.1) using the nonparametric kernel regression methods and showed the proposed estimators are asymptotically normal as well as the estimator of the slope function $\gamma(t)$ is consistent in supremum norm. Reference [3] considered the least square estimator of model (1.1) using the Karhunen-Loève (K-L) expansion to approximate the slope function, established asymptotic properties of the resulting estimation. Based on Tikhonov regularization, [4] introduced the functional ridge regression estimation procedure, and showed asymptotic normality of the estimated infinite dimensional regression coefficients as well as the convergence rate of the estimated slope function. Using the technique of polynomial splines, [5] considered the estimation of model (1.1) by minimizing the square of residuals, and furtherly considered the asymptotic property of the estimators. Recently, to get the robust estimator of coefficients of (1.1), the model has been also considered in the frame of qunatile regression ([6] [7]). Some authors also considered the model (1.1) from the view of hypothesis test, such as, [8] construct pivot by the square of residuals under the null and alternative hypothesis, to test whether the linearity term of (1.1) exists or not. Moreover, the generalized form of model (1.1), like semiparametric partially linear regression model for functional data and functional partial linear single-index model, has been respectively considered by [9] [10].

However, all the works have a common assumption that the responses are observed independently. As is well known, uncertainty such as volatility uncertainty is a common phenomenon in modern economic and financial theory. Therefore, the assumption of independence of the response observations is not valid in the real data analysis. Motivated by the fact mentioned above, we may want to reconsider the model (1.1) so that it can reflect the volatility of the data. Fortunately, conditional heteroscedasticity can reflect the size of volatility appropriately. One of the most popular models which can show the heteroscedasticity in econometrics is the autoregressive conditional heteroscedasticity (ARCH) model which was introduced by [11] and have had an enormous impact on the modeling of financial data. More importantly, many authors have studied the ARCH models to make it more perfect in theory. For example, reference [12] considered the existence of the strictly stationary and ergodic solution and high moment of the ARCH model; [13] studied strong law of large numbers of the absolute value sequence from ARCH.

If we have n observations $\{(z_1, X_1, Y_1), \cdots, (z_n, X_n, Y_n)\}$, model (1.1) can be written as

$$Y_i = \boldsymbol{\beta}' z_i + \int_0^1 \gamma(t) X_i(t) \mathrm{d}t + \varepsilon_i. \tag{1.2}$$

The ARCH(p) model for $\{\varepsilon_i\}$ is defined by the following equations:

$$\begin{cases} \varepsilon_i = e_i h_i^{1/2}, \\ h_i = \alpha_0 + \alpha_1 \varepsilon_{i-1}^2 + \alpha_2 \varepsilon_{i-2}^2 + \cdots + \alpha_p \varepsilon_{i-p}^2, \end{cases} \tag{1.3}$$

where $\alpha_0 > 0, \alpha_i \geq 0, i = 1, \cdots, p$. Besides, $\{e_i : i \geq 1\}$ is an independent and identically distributed (i.i.d.) random sequence and independent of $\{\varepsilon_t : t < i\}$ with $Ee_i = 0$ and $Ee_i^2 = 1$. For sake of establishing the asymptotic properties of the joint model (1.2) and (1.3), in this paper, we assume that the distribution functions $\{F_i\}$ of $\{\varepsilon_i^2\}$ are absolutely continuous with continuous densities f_i, which is uniformly bounded away from 0 and ∞ at the 1/2-th quantile points $\xi_i, i = 1, \cdots, n$. Moreover, similar to [3], assume that $(z_1, X_1), \cdots, (z_n, X_n)$ in (1.2) are i.i.d..

The ordinary regression models with ARCH errors have been considered by many authors. For example, [14] considered a *p*-th order autoregression process with ARCH errors; [15] studied the estimation of the partly linear regression models with ARCH(p) errors. Under some regularity conditions, we study the estimation of the unknown parameters in the joint model of (1.2) and (1.3), and propose a hybrid estimation method with combining the functional principle component analysis in the mean model with the least absolute deviation for the error model. The asymptotic normality of the real-valued parameter estimators is established, and the convergence rate of the slope function estimator is obtained.

The rest of the paper is organized as follows. Section 2 gives the estimation of parameters for the partial functional linear regression models as well as ARCH(p) errors. Asymptotic theory of the proposed estimators is given in Section 3. In Section 4, we carry out a simulation study to illustrate the finite sample performance, and a real data analysis is conducted in Section 5. Some preliminary lemmas and the proofs of the theorems are presented in **Appendix**.

2. Estimation

Firstly, we shall study how to produce the estimators $\hat{\boldsymbol{\beta}}, \hat{\gamma}$ of $\boldsymbol{\beta}, \gamma$ in this section. Let $\langle \cdot, \cdot \rangle$ and $\|\cdot\|$ denote inner product and norm on $L^2[0,1]$ respectively. Denote the covariance function of process X by C_X which is continuous on $\mathcal{T} \times \mathcal{T}$. Then we have the following expansion

$$C_X(s,t) = \sum_{j=1}^{\infty} \lambda_j \rho_j(s) \rho_j(t)$$

by Mercer's theorem ([16]) with nonnegative eigenvalues $\lambda_1, \lambda_2, \cdots$ and continuous orthonormal eigenfunctions $\rho_1, \rho_2, \cdots$ of the covariance operator. For convenience, we assume $\lambda_1 > \lambda_2 > \cdots > 0$ throughout this paper. Therefore, by K-L expansion, one has

$$X_i(t) = \sum_{j=1}^{\infty} U_{ij} \rho_j(t)$$

and

$$\gamma(t) = \sum_{j=1}^{\infty} \gamma_j \rho_j(t),$$

where $U_{ij} = \langle X_i, \rho_j \rangle$ are uncorrelated random variables with $\mathrm{E}[U_{ij}] = 0$ and $\mathrm{E}[U_{ij}^2] = \lambda_j$, and $\gamma_j = \langle \gamma, \rho_j \rangle$. Then (1.2) is equivalent to

$$Y_i = \boldsymbol{\beta}' z_i + \sum_{j=1}^{\infty} \gamma_j U_{ij} + \varepsilon_i, \ i = 1, \cdots, n. \tag{2.1}$$

To estimate the parameters in (1.2), following [3]'s idea, we approximate the second term in (2.1) by finite sum

$$Y_i \doteq \boldsymbol{\beta}' z_i + \sum_{j=1}^{m} \gamma_j U_{ij} + \varepsilon_i, \ i = 1, \cdots, n, \tag{2.2}$$

where $m \to \infty$ as $n \to \infty$. Furthermore, we employ the empirical version of C_X

$$\hat{C}_X(s, t) = \frac{1}{n} \sum_{i=1}^{n} X_i(s) X_i(t) = \sum_{j=1}^{\infty} \hat{\lambda}_j \hat{\rho}_j(s) \hat{\rho}_j(t)$$

with $(\hat{\lambda}_j, \hat{\rho}_j)$ being the pairs of eigenvalues and eigenfunctions of covariance operator related to $\hat{C}_X$ and $\hat{\lambda}_1 \geq \hat{\lambda}_2 \geq \cdots \geq 0$, and substitute U_{ij} in (2.2) with $\hat{U}_{ij} = \langle X_i, \hat{\rho}_j \rangle$. To get an elegant matrix form for model (2.2), denote $\boldsymbol{Y} = (Y_1, \cdots, Y_n)'$, $\boldsymbol{Z} = (z_1, \cdots, z_n)'$, $\boldsymbol{U}_m = (\hat{U}_{ij})_{\substack{i=1,\cdots,n \\ j=1,\cdots,m}}$, $\tilde{\boldsymbol{\gamma}} = (\gamma_1, \cdots, \gamma_m)'$ and $\boldsymbol{\varepsilon} = (\varepsilon_1, \cdots, \varepsilon_n)'$. Then (2.2) can be rewritten as

$$\boldsymbol{Y} \doteq \boldsymbol{Z}\boldsymbol{\beta} + \boldsymbol{U}_m \tilde{\boldsymbol{\gamma}} + \boldsymbol{\varepsilon},$$

and the least square estimator $\hat{\boldsymbol{\beta}}$ and $\hat{\tilde{\boldsymbol{\gamma}}}$ are given by

$$\left(\hat{\boldsymbol{\beta}}', \hat{\tilde{\boldsymbol{\gamma}}}' \right)' = \arg\min \left(\boldsymbol{Y} - \boldsymbol{Z}\boldsymbol{\beta} - \boldsymbol{U}_m \tilde{\boldsymbol{\gamma}} \right)' \left(\boldsymbol{Y} - \boldsymbol{Z}\boldsymbol{\beta} - \boldsymbol{U}_m \tilde{\boldsymbol{\gamma}} \right).$$

By simple calculation, we have

$$\hat{\boldsymbol{\beta}} = \left(\boldsymbol{Z}'(I - \boldsymbol{V}_m) \boldsymbol{Z} \right)^{-1} \boldsymbol{Z}'(I - \boldsymbol{V}_m) Y$$

with $\boldsymbol{V}_m = \boldsymbol{U}_m \left(\boldsymbol{U}_m' \boldsymbol{U}_m \right)^{-1} \boldsymbol{U}_m'$ and

$$\hat{\tilde{\boldsymbol{\gamma}}} = \left(\boldsymbol{U}_m' \boldsymbol{U}_m \right)^{-1} \boldsymbol{U}_m' \left(\boldsymbol{Y} - \boldsymbol{Z}\hat{\boldsymbol{\beta}} \right)$$

provided that $\left(\boldsymbol{Z}'(I - \boldsymbol{V}_m) \boldsymbol{Z} \right)^{-1}$ exists (this is true with probability tending to 1, see Lemma 1 in [8]). The estimator $\hat{\gamma}$ of γ can be given as

$$\hat{\gamma}(\cdot) = \sum_{j=1}^{m} \hat{\gamma}_j \hat{\rho}_j(\cdot),$$

where $\hat{\gamma}_j$ is the jth element of $\hat{\tilde{\boldsymbol{\gamma}}}$.

To get asymptotic properties of $\hat{\boldsymbol{\beta}}$, let $\hat{C}_z = n^{-1} \sum_{i=1}^{n} z_i z_i'$, $\hat{C}_{zY} = n^{-1} \sum_{i=1}^{n} z_i Y_i$, $\hat{C}_{zX}(t) = n^{-1} \sum_{i=1}^{n} z_i X_i(t)$, $\hat{C}_{Xz}(t) = \left(\hat{C}_{zX}(t) \right)'$ and $\hat{C}_{YX}(t) = n^{-1} \sum_{i=1}^{n} Y_i X_i(t)$. Then $\hat{\boldsymbol{\beta}}$ is equal to

$$\hat{\boldsymbol{\beta}} = \left(\hat{C}_z - \sum_{j=1}^{m} \frac{\langle \hat{C}_{zX}, \hat{\rho}_j \rangle \langle \hat{C}_{Xz}, \hat{\rho}_j \rangle}{\hat{\lambda}_j} \right)^{-1} \left(\hat{C}_{zY} - \sum_{j=1}^{m} \frac{\langle \hat{C}_{zX}, \hat{\rho}_j \rangle \langle \hat{C}_{YX}, \hat{\rho}_j \rangle}{\hat{\lambda}_j} \right), \tag{2.3}$$

with $\langle \hat{C}_{zX}, \hat{\rho}_j \rangle = n^{-1} \sum_{i=1}^{n} z_i \langle X_i, \hat{\rho}_j \rangle$ and $\langle \hat{C}_{YX}, \hat{\rho}_j \rangle = n^{-1} \sum_{i=1}^{n} Y_i \langle X_i, \hat{\rho}_j \rangle$. Simi-

larly, $\hat{\gamma}_j$ can be represented as $\hat{\gamma}_j = \left\langle \hat{C}_{YX} - \hat{\boldsymbol{\beta}}' \hat{C}_{zX}, \hat{\rho}_j \right\rangle \big/ \hat{\lambda}_j$, $j = 1, \cdots, m$.

So far, we have already obtained the estimator $\hat{\boldsymbol{\beta}}$ and $\hat{\gamma}$, now we turn to consider the estimation of $\boldsymbol{\alpha} = \left(\alpha_0, \alpha_1, \cdots, \alpha_p \right)'$. Denote by

$$\hat{\varepsilon}_i = Y_i - \hat{\boldsymbol{\beta}}' \boldsymbol{z}_i - \sum_{j=1}^m \hat{\gamma}_j \hat{U}_{ij}, i = 1, \cdots, n$$

the residuals. For ARCH(p) models, in view of the higher peak and heavy tail phenomenon, unlike Sastry's idea that regress $\hat{\varepsilon}_i^2$ on a column of ones and $\hat{\varepsilon}_{i-1}^2$ by minimizing the sum of the square of residuals, after getting the residuals $\hat{\varepsilon}_i, i = 1, \cdots, n$ to get the parameter's estimate of ARCH (1) sequence ([14]), minimizing the sum of the absolute residuals is used in this paper to get an estimator of $\boldsymbol{\alpha} = \left(\alpha_0, \alpha_1, \cdots, \alpha_p \right)'$. That is to say

$$\hat{\boldsymbol{\alpha}} = \arg\min_{\alpha \in R^{p+1}} \sum_{i=p+1}^n \left| \hat{\varepsilon}_i^2 - \alpha_0 - \alpha_1 \hat{\varepsilon}_{i-1}^2 - \cdots - \alpha_p \hat{\varepsilon}_{i-p}^2 \right|, \tag{2.4}$$

where $\hat{\boldsymbol{\alpha}} = \left(\hat{\alpha}_0, \hat{\alpha}_1, \cdots, \hat{\alpha}_p \right)'$.

3. Asymptotic Properties

We first state the assumptions under which the asymptotic properties are proved, then present the theorems. Let $\rho(D)$ and $D^{\otimes m}$ denote the spectral radius and Kronecker product of matrix D respectively, and $\mathcal{F}_t = \sigma \{ \varepsilon_i : i \leq t \}$ in the following.

It is easy to see that $\mathrm{E}[\varepsilon_t \mid \mathcal{F}_{t-1}] = 0$, $\mathrm{E}[\varepsilon_t^2 \mid \mathcal{F}_{t-1}] = h_t$, namely, the ARCH(p) process forms a martingale difference sequence with $\mathrm{E}[\varepsilon_t^2] = \dfrac{\alpha_0}{1 - \alpha_1 - \cdots - \alpha_p}$.

In order to attain the stationary solution and guarantee the existence of high moment of $\{\varepsilon_t\}$, we suppose that

$$0 < \alpha_0 < \infty, \ \sum_{j=1}^p \alpha_j < 1, \ \rho(\Sigma_r) < 1 \tag{3.1}$$

for some integer $r \geq 1$, where $\Sigma_r = \mathrm{E}\left(D_t^{\otimes r} \right)$,

$$D_t = \begin{pmatrix} \alpha_1 e_t^2 & \cdots & \alpha_{p-1} e_t^2 & \alpha_p e_t^2 \\ 1 & \cdots & 0 & 0 \\ \vdots & \ddots & \vdots & \vdots \\ 0 & \cdots & 1 & 0 \end{pmatrix}.$$

Then, as [12] and [14] proved, there exists a strictly stationary solution for the *p*-th order ARCH process given by

$$\varepsilon_t^2 = \alpha_0 \left[e_t^2 + \sum_{j=1}^\infty e_{t-j}^2 \boldsymbol{\delta}_1' \left(\prod_{i=0}^{j-1} D_{t-i} \right) \boldsymbol{\delta}_1 \right] \tag{3.2}$$

with $\boldsymbol{\delta}_1 = (1, 0, \cdots, 0)'$.

In the following, let C denote positive constant which may change from line to line. It is assumed that the random function X satisfies

$$\mathrm{E}\|X\|^4 < \infty \tag{3.3}$$

and for each j

$$\mathrm{E}\left[U_j^4\right] \le C\lambda_j^2 \tag{3.4}$$

for some constant C. For the eigenvalues of C_X, assume that there exist C and $a > 1$ such that

$$C^{-1}j^{-a} \le \lambda_j \le Cj^{-a}, \; \lambda_j - \lambda_{j-1} \ge Cj^{-a-1}, \; j \ge 1 \tag{3.5}$$

to prevent the spacings among eigenvalues being too small. In order to guarantee that the regression weight function γ is smoother than the sample path X, for the Fourier coefficients γ_j, we suppose that

$$\left|\gamma_j\right| \le Cj^{-b} \tag{3.6}$$

for some constant C and $b > a/2 + 1$. On the tuning parameter, we assume that

$$m \sim n^{1/(a+2b)}, \tag{3.7}$$

where $a_n \sim b_n$ means there exist constants $0 < L < M < \infty$ s.t. $L \le \dfrac{a_n}{b_n} \le M$ for all n. Besides, we also assume that

$$E\|z\|_{R^d}^4 < \infty \tag{3.8}$$

for the random vector z with $\|z\|_{R^d} = (z'z)^{\frac{1}{2}}$ and $C_{z_kX}(\cdot) = \mathrm{Cov}(z_k, X(\cdot))$ satisfies

$$\left|\left\langle C_{z_kX}, \rho_j \right\rangle\right| \le Cj^{-(a+b)} \tag{3.9}$$

for each $k = 1, 2, \cdots, d$ and $j \ge 1$.

Let $\eta_{ik} = z_{ik} - \left\langle \chi_k, X_i \right\rangle$, where $\chi_k = \sum_{j=1}^{\infty}\left(\left\langle C_{z_kX}, \rho_j \right\rangle \big/ \lambda_j\right)\rho_j$. Then, $\eta_{1k}, \cdots, \eta_{nk}$ are i.i.d. random variables. We suppose that

$$\mathrm{E}\left[\eta_{1k} \mid X_1, \cdots, X_n\right] = 0, \mathrm{E}\left[\eta_{1k}^2 \mid X_1, \cdots, X_n\right] = B_{kk},$$

where B_{kk} is the kth diagonal element of

$$\boldsymbol{B} = \mathrm{E}\left[\boldsymbol{\eta}_1\boldsymbol{\eta}_1'\right] = C_z - \sum_{j=1}^{\infty} \frac{\left\langle C_{zX}, \rho_j \right\rangle \left\langle C_{Xz}, \rho_j \right\rangle}{\lambda_j}, \tag{3.10}$$

which is assumed positive definite, and $\boldsymbol{\eta}_i = \left(\eta_{i1}, \cdots, \eta_{id}\right)'$.

With the assumptions that mentioned above, we have the following results.

Theorem 1. If the assumptions (3.1) with $r = 2$, (3.3)-(3.10) hold, we have

$$n^{1/2}\left(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}\right) \xrightarrow{d} N\left(0, \frac{\alpha_0}{1 - \alpha_1 - \cdots - \alpha_p}\boldsymbol{B}^{-1}\right)$$

as $n \to \infty$, where "$\xrightarrow{d}$" denotes convergence in distribution.

Theorem 2. Under the assumptions (3.1) with $r = 1$, (3.3)-(3.10), one has

$$\|\hat{\gamma} - \gamma\|^2 = O_p\left(n^{-(2b-1)/(a+2b)}\right).$$

Theorem 3. Under the conditions of $\{\varepsilon_i\}$ and the assumptions (3.1) with

$r = 2$, (3.3)-(3.10), we have

$$n^{1/2}(\hat{\alpha} - \alpha) \xrightarrow{d} N\left(0, \frac{1}{4}D_1^{-1}PD_1^{-1}\right)$$

as $n \to \infty$, where

$$P = \mathrm{E}\begin{pmatrix} 1 & \varepsilon_p^2 & \varepsilon_{p-1}^2 & \cdots & \varepsilon_1^2 \\ \varepsilon_p^2 & \varepsilon_p^4 & \varepsilon_p^2\varepsilon_{p-1}^2 & \cdots & \varepsilon_p^2\varepsilon_1^2 \\ \vdots & \vdots & \vdots & & \vdots \\ \varepsilon_1^2 & \varepsilon_1^2\varepsilon_p^2 & \varepsilon_1^2\varepsilon_{p-1}^2 & \cdots & \varepsilon_1^4 \end{pmatrix},$$

$$D_1 = \lim_{n\to\infty} n^{-1}\sum_{i=p+1}^{n} f_i(\xi_i)v_i v_i',$$

with $v_i = \left(1, \varepsilon_{p+i-1}^2, \varepsilon_{p+i-2}^2, \cdots, \varepsilon_i^2\right)'$.

Remark 1. Compared with [3], we can find that the estimator of the regression coefficient vector also has the convergence rate $\sqrt{n}$ and is asymptotically normal under the ARCH(p) errors.

Remark 2. To implement the proposed method, we need to know how to choose the cut-off point *m*. Theoretically, if m is too large, the number of parameters in model (2.2) is also too large and the estimate of the slope function γ may goes terrible by the properties of Functional Principal Component Analysis (FPCA); if *m* is taken as a small value, the approximation of model (2.2) to model (2.1) may not be enough. This is the role that condition (3.7) plays. There are well-established methods for choosing such tuning parameter m, such as Generalized Cross-Validation (GCV), AIC, BIC and FPCA. As we all know, the first three criteria are data-driven and the FPCA is based on the ratio of variance explained by the first *m* eigenvalues to the total variation of *X*. In section 4, GCV and FPCA are respectively considered.

Remark 3. In order to make inference for α, the estimation of the asymptotic variance, mainly involving the estimation of P and $f_i(\xi_i)$, is needed to be given. Based on (A.8) in the **Appendix**, it is reasonable to use $n^{-1}\sum_{i=p+1}^{n}\hat{v}_i\hat{v}_i'$ as the estimate of P with $\hat{v}_i = \left(1, \hat{\varepsilon}_{p+i-1}^2, \hat{\varepsilon}_{p+i-2}^2, \cdots, \hat{\varepsilon}_i^2\right)'$. For $f_i(\xi_i)$, the sparsity estimation methods or the kernel density estimation ideas, suggested by [17] and [18] respectively, can be used for this paper.

4. Simulation Studies

In this section, simulations are carried out to show the finite sample performance of the proposed method. The data is generated from the model (1.1) in the case where z_{i1} and z_{i2} are standard normal,

$$X(t) = \sum_{j=1}^{200} U_j \rho_j(t), \ t \in [0,1],$$

where the U_js are distributed as independent normal with mean 0 and variance $\lambda_j = \left((j-0.5)\pi\right)^{-2}$ respectively, $\rho_j(t) = \sqrt{2}\sin\left((j-0.5)\pi t\right)$ and

$$Y_i = z_{i1}\beta_1 + z_{i2}\beta_2 + \int_0^1 \gamma(t)X_i(t)\mathrm{d}t + \varepsilon_i$$

with $\boldsymbol{\beta} = (2, -1)'$, $\gamma(t) = \sqrt{2}\sin(\pi t/2) + 3\sqrt{2}\sin(3\pi t/2)$. For the random error, we take the following form: $\varepsilon_i = e_i h_i^{1/2}$, $h_i = \alpha_0 + \alpha_1 \varepsilon_{i-1}^2 + \alpha_2 \varepsilon_{i-2}^2$, $e_i \overset{i.i.d.}{\sim} t(5)$, where α_0 takes value 0.1, α_1 takes value from 0.1, 0.3 and α_2 takes value from 0.3, 0.1 correspondingly. Note that $t(5)$ has finite 4$^{\text{th}}$ order moment, and the condition (3.1) is satisfied by $\boldsymbol{\alpha} = (0.1, 0.1, 0.3)'$ and $\boldsymbol{\alpha} = (0.1, 0.3, 0.1)'$ with $r = 1$, where it may be shown that both $\hat{\boldsymbol{\beta}}$ and $\hat{\boldsymbol{\alpha}}$ given by (2.3) and (2.4) are consistent. For $\boldsymbol{\alpha} = (0.1, 0.3, 0.3)'$ with $r = 1$, $\rho(\Sigma_r) = 1$. That is, it is on the boundary of the condition region.

We also consider the situation that α_1 and α_2 take values 0 to compare with the independent structure. For each $\boldsymbol{\alpha}$, we simulate 1000 random samples, each with sample size $n = 100, 300, 500$ respectively. For the determination of m by FPCA, $m = \min\left\{k : \sum_{i=1}^{k}\hat{\lambda}_i \Big/ \sum_{i=1}^{n}\hat{\lambda}_i \geq 0.85\right\}$ is used. The accuracy of the slope function estimate is checked by the mean integrated square error (MISE) which is defined as

$$\text{MISE} = \frac{1}{1000}\sum_{i=1}^{1000}\left[\frac{1}{N}\sum_{s=1}^{N}\left(\hat{\gamma}_i(t_s) - \gamma(t_s)\right)^2\right],$$

where $\hat{\gamma}_i(\cdot)$ is the estimate of the slope function $\gamma(\cdot)$ obtained from the *i*-th replication, and $t_s, s = 1, \cdots, N$ are the equally spaced grid points at which the function $\hat{\gamma}_i(t)$ is evaluated. In our implementation, $N = 100$ is used. In this section, the results of the estimators of $\boldsymbol{\alpha}$ using the Least Square (LS) method is also carried out to compare with the Least Absolute Deviation (LAD) method which is proposed by this paper. The results are summarized into **Tables 1-3** and the shape of the true function γ and the estimated function $\hat{\gamma}$, based on the average of 1000 replications with $\boldsymbol{\alpha} = (0.1, 0.1, 0.3)'$ are depicted in **Figure 1**.

Table 1. MSE and MISE under LAD (GCV).

n	$(\alpha_1\ \alpha_2)$	β_1	β_2	γ	α_0	α_1	α_2
100	(0.0 0.0)	0.0021	0.0022	0.4414	0.0029	0.0021	0.0019
	(0.1 0.3)	0.0056	0.0060	0.9230	0.0024	0.0093	0.0458
	(0.3 0.1)	0.0053	0.0068	0.9593	0.0028	0.0466	0.0086
	(0.3 0.3)	0.0163	0.0178	1.7093	0.0084	0.0477	0.0482
300	(0.0 0.0)	0.0006	0.0006	0.1400	0.0026	0.0006	0.0004
	(0.1 0.3)	0.0017	0.0019	0.3676	0.0021	0.0052	0.0364
	(0.3 0.1)	0.0017	0.0017	0.3203	0.0021	0.0353	0.0060
	(0.3 0.3)	0.0089	0.0089	1.0242	0.0076	0.0381	0.0402
500	(0.0 0.0)	0.0003	0.0003	0.0865	0.0025	0.0002	0.0002
	(0.1 0.3)	0.0009	0.0009	0.1942	0.0022	0.0041	0.0320
	(0.3 0.1)	0.0010	0.0009	0.1924	0.0021	0.0310	0.0050
	(0.3 0.3)	0.0039	0.0045	0.6985	0.0024	0.0352	0.0386

Table 2. MSE and MISE under LAD (FPCA).

n	$(\alpha_1\ \alpha_2)$	β_1	β_2	γ	α_0	α_1	α_2
100	(0.1 0.3)	0.0056	0.0059	0.6424	0.0028	0.0096	0.0464
	(0.3 0.1)	0.0054	0.0066	0.6453	0.0031	0.0465	0.0086
	(0.3 0.3)	0.0164	0.0176	0.8812	0.0094	0.0480	0.0491
300	(0.1 0.3)	0.0017	0.0019	0.0910	0.0022	0.0053	0.0362
	(0.3 0.1)	0.0017	0.0017	0.0911	0.0022	0.0351	0.0061
	(0.3 0.3)	0.0059	0.0084	0.2079	0.0021	0.0382	0.0416
500	(0.1 0.3)	0.0009	0.0009	0.0581	0.0022	0.0041	0.0318
	(0.3 0.1)	0.0010	0.0009	0.0573	0.0021	0.0311	0.0050
	(0.3 0.3)	0.0030	0.0030	0.1284	0.0020	0.0357	0.0399

Table 3. MSE and MISE under LS (GCV).

n	$(\alpha_1\ \alpha_2)$	β_1	β_2	γ	α_0	α_1	α_2
100	(0.1 0.3)	0.0056	0.0060	0.9230	0.1559	0.0203	0.0411
	(0.3 0.1)	0.0053	0.0068	0.9593	0.1286	0.0412	0.0252
	(0.3 0.3)	0.0163	0.0178	1.7093	4.1482	0.0430	0.0541
300	(0.1 0.3)	0.0017	0.0019	0.3676	0.1011	0.0122	0.0258
	(0.3 0.1)	0.0017	0.0017	0.3203	0.0619	0.0271	0.0206
	(0.3 0.3)	0.0089	0.0089	1.0242	25.1315	0.0318	0.0690
500	(0.1 0.3)	0.0009	0.0009	0.1942	0.1215	0.0123	0.0230
	(0.3 0.1)	0.0010	0.0009	0.1924	0.1684	0.0259	0.0251
	(0.3 0.3)	0.0039	0.0045	0.6985	18.6453	0.0305	0.0525

Table 4. MSE and MISE with $N(0,1)$ produced errors under GCV ($\boldsymbol{\alpha}=(0.1,0.3,0.3)'$).

n		β_1	β_2	γ	α_0	α_1	α_2
100	LAD	0.0030	0.0028	0.6190	0.0027	0.0455	0.0473
	LS	0.0030	0.0028	0.6190	0.0049	0.0379	0.0472
300	LAD	0.0009	0.0009	0.1982	0.0028	0.0373	0.0387
	LS	0.0009	0.0009	0.1982	0.0023	0.0201	0.0261
500	LAD	0.0006	0.0005	0.1088	0.0029	0.0341	0.0345
	LS	0.0006	0.0005	0.1088	0.0022	0.0171	0.0220

We also would like to know that how will the LAD method behave when the error of the ARCH sequence is not heavy tailed, such as $e \sim N(0,1)$. The simulation results are summarized into Table 4 with $\boldsymbol{\alpha}=(0.1,0.3,0.3)'$. In this case, (3.1) is satisfied with both $r=1$ and $r=2$.

We can derive the following conclusions from Tables 1-4.

1) From Table 1 and Table 2, with the increasing of sample size *n*, it can be

seen MSEs, MISE decrease in all scenarios that we considered about error. This reflects the proposed estimators fit better to the real values as the sample size increases and thus is promising.

2) For every fixed sample size n, it can be seen the larger value of coefficients α, the larger the corresponding MSE for the different coefficients form of errors. For example, when $n = 100$, α take values $(0.1, 0.1, 0.3)$ and $(0.1, 0.3, 0.1)$ respectively, the MSE of $\alpha_1 = 0.3$ is larger than the MSE of $\alpha_1 = 0.1$ and so is α_2. Moreover, the MSE of $\hat{\alpha}$ and MISE of $\hat{\gamma}$ become large when the coefficients α_1, α_2 take relative large values simultaneously, such as $(\alpha_1, \alpha_2) = (0.3, 0.3)$ in **Table 1**. This is due to the stronger volatility for larger $\alpha_j, j = 1, 2$.

3) From **Table 1**, for every fixed sample size n, when α_1, α_2 take values 0, which is the case considered by [3], the MSE and MISE of $\hat{\beta}$ are smaller than those with ARCH errors. This shows that the dependence of errors makes the estimators more varying. However, with the increasing of sample size, the later ones decrease and could reach the former quantities.

4) The MSE of the coefficients $\hat{\alpha}$, in **Table 3**, using the LS method for $t(5)$ produced errors get larger values, compared to the results of estimator given by (2.3). Specifically, unlike the results in **Table 1** and **Table 2**, the results of α_0 for the boundary case $(\alpha_1, \alpha_2) = (0.3, 0.3)$ are unstable, illustrating the reasonableness of the proposed method.

5) **Table 4** shows that the LAD method could perform as well as the LS method even for non-heavy tailed distribution of the error.

6) As **Table 1** and **Table 2** show, the difference of the estimators of β between the two selection methods of m is very small.

Based on simulation results from **Table 1** and **Table 2**, it seems that the estimator of γ corresponding to FPCA is better in view of MISE. As we know, when using FPCA to choose m, a threshold value for the ratio is needed. We reset the threshold value as 0.80 rather than 0.85 for the case $n = 100$ and $\alpha = (0.1, 0.1, 0.3)'$, the MISE will become 6.6373 which is bigger than 0.9230 given in **Table 1**. As far as we know, there is no theoretical research on how the threshold value should be set to get a compromise between goodness of fit and the precision of the estimated slope function.

From **Figure 1**, it can be seen that the estimated function can fit the true function approximately no matter which method is used to choose the tuning parameter m, which demonstrate that the proposed method works well.

From the above observation, we see that the estimator (2.4) performs well, even under the boundary condition. It may be theoretically interesting to know the performance of the estimator in this case, but it is beyond our focus here.

5. Real Data Analysis

In this section, we apply the proposed method to deal with a real dataset. The data consist of monthly electricity consumption, denote by C, consumed by
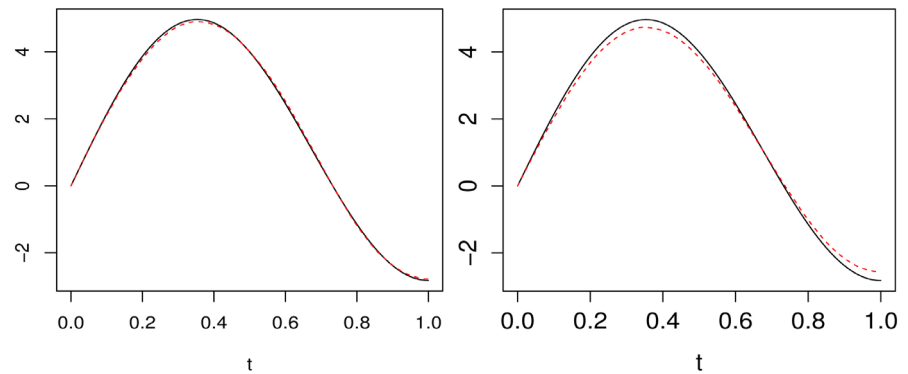
Figure 1. The true function γ (solid line) and the estimated function $\hat{\gamma}$ (dashed line) using GCV (left) and FPCA (right) with $n = 100$.

commercial sectors from January 1972 to January 2005 (397 months) and their annual average retail price P (33 years). A main goal of this study is to consider the effect of dependence structure of the error on the asymptotic variance of $\hat{\beta}$, when using the price and consumption to predict the consumption 6 months later.

According to the stationary test of the electricity consumption data, the heteroscedasticity and linear trend can be found and then may be eliminated by differencing the ln data. Corresponding to the general notation introduced in model (1.1), let

$$D_j = \ln C_j - \ln C_{j-1}, \ j = 1, 2, \cdots, 397,$$

$$X_i = \left\{ D_{12(i-1)+t}, t \in [1,12] \right\}, \ i = 1, 2, \cdots, 32.$$

The response variable is

$$Y_i = D_{12i+6}, \ i = 1, 2, \cdots, 32,$$

and the additional real variable is defined by

$$z_i = P_i, \ i = 1, 2, \cdots, 32.$$

Regress Y on Z and X with m chosen by the FPCA with threshold 0.85, then the residuals are obtained. Although it seems reasonable to treat the residuals as white noise sequence, the characteristics of volatility clustering may exists according to **Figure 2**. For this sequence, a further analysis is conducted numerically, and the significant level takes value 0.05 in the following test. The stationary is tested firstly using the function "adf.test()" in R packages with the p value 0.019. Whether the sequence is uncorrelated or not is considered using the Box-Ljung statistics, accepting the null hypothesis with the p value 0.08. However, for this sequence, the skewness value 0.03 and kurtosis value 2.48 shows it has high peak and heavy tailed features. Box-Ljung test is adopted again for the square of the residuals sequence, demonstrating the existence of the ARCH structure with p value 0.03. As **Figure 3** show, it is appropriate to use ARCH(3) to fit the square residuals sequence. By calculation, under the existence of volatility clustering for the errors, the asymptotic standard variation
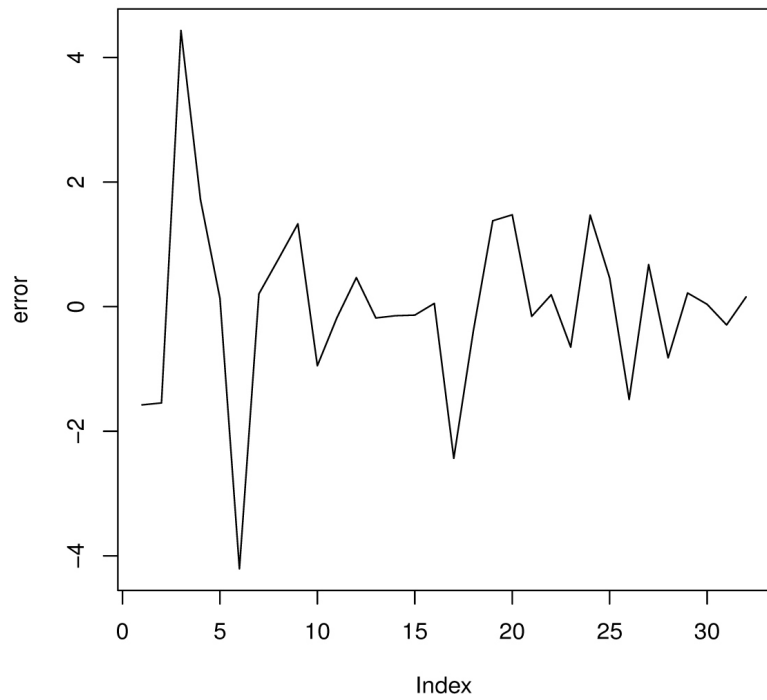
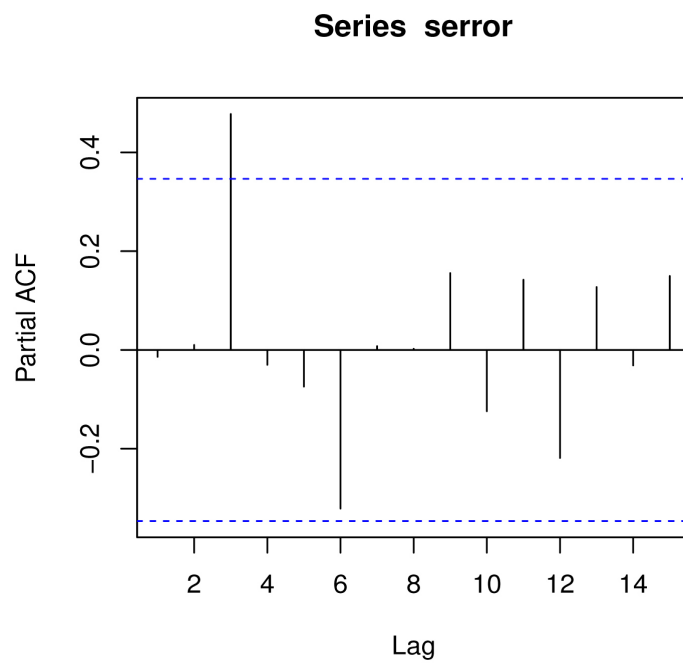**Figure 2.** The graph of error sequence.

Series serror

Figure 3. The pacf of the errors square.

of $\hat{\beta}$ is 0.01, which is reduced by 94% comparing with the value 0.18, which is given under ignoring concrete form of the error, showing it is promising to consider the ARCH structure.

6. Discussion

In this paper, the estimation of partial functional linear models with ARCH(p)

errors using the LS method, as well as the parameters of ARCH(p) sequence using the LAD method are respectively considered. Considering that the dimensionality of the slope function is infinite, for this paper, the key point we have given consists in transforming the partial functional linear models with ARCH errors into the corresponding linear regression models by the K-L expansion and the idea of FPCA. The linear relationship between z and X is essentially assumed (see Remark 1 in [8]). In the future study, under the errors' dependent structure, we will further consider the estimation of the model (1.1) using the kernel method noticing that the relationship between z and X may be relaxed. Since the heteroscedasticity in economics is a common phenomenon, the theory study of the model is practically useful and worthy to be explored. Furthermore, based on the fact the consistency of $\hat{\alpha}$ and $\hat{\beta}$ can be respectively obtained from Theorem 3 and the proof of Theorem 1, the inference to the models could be made precisely within this paper by the asymptotic normality of $\hat{\beta}$.

Acknowledgements

References

[1] Aneiros-Pérez, G. and Vieu, P. (2006) Semi-Functional Partial Linear Regression. *Statistics & Probability Letters*, **76**, 1102-1110. https://doi.org/10.1016/j.spl.2005.12.007

[2] Aneiros-Pérez, G. and Vieu, P. (2008) Nonparametric Time Series Prediction: A Semi-Functional Partial Linear Modeling. *Journal of Multivariate Analysis*, **99**, 834-857. https://doi.org/10.1016/j.jmva.2007.04.010

[3] Shin, H. (2009) Partial Functional Linear Regression. *Journal of Statistical Planning and Inference*, **139**, 3405-3418. https://doi.org/10.1016/j.jspi.2009.03.001

[4] Shin, H. and Lee, M. (2012) On Prediction Rate in Partial Functional Linear Regression. *Journal of Multivariate Analysis*, **103**, 93-106. https://doi.org/10.1016/j.jmva.2011.06.011

[5] Zhou, J.J., Chen, Z. and Peng, Q.Y. (2016) Polynomial Spline Estimation for Partial Functional Linear Regression Models. *Computational Statistics*, **31**, 1107-1129. https://doi.org/10.1007/s00180-015-0636-0

[6] Lv, Y., Du, J. and Sun, Z.M. (2014) Functional Partially Linear Quantile Regression Model. *Metrika*, **77**, 317-332.

[7] Zhou, J., Du, J. and Sun, Z.M. (2016) M-Estimation for Partially Functional Linear Regression Model Based on Splines. *Communications in Statistics-Theory and Methods*, **45**, 6436-6466. https://doi.org/10.1080/03610926.2014.921309

[8] Yu, P., Zhang, Z.Z. and Du, J. (2016) A Test of Linearity in Partial Functional Linear Regression. *Metrika*, **79**, 953-969. https://doi.org/10.1007/s00184-016-0584-x

[9] Zhang, T. and Wang, Q.H. (2012) Semiparametric Partially Linear Regression Models for Functional Data. *Journal of Statistical Planning and Inference*, **142**,

2518-2529. https://doi.org/10.1016/j.jspi.2012.03.004

[10] Wang, G.C., Feng, X.N. and Chen, M. (2015) Functional Partial Linear Single-Index Model. *Scandinavian Journal of Statistics*, **43**, 261-274.

[11] Engle, R.F. (1982) Autoregressive Conditional Heteroscedasticity with Estimates of United Kingdom Inflation. *Econometrica*, **50**, 987-1007.
https://doi.org/10.2307/1912773

[12] Chen, M. and An, H.Z. (1995) The Strictly Stationary Ergodicity and High Moment of ARCH(p) Models. *Chinese Science Bulletin*, **40**, 2118-2123.

[13] Zhang, Z.Q., Feng, J.Y. and Zhang, R.Q. (2007) Strong Law of Large Numbers of the Absolute Value Sequences from ARCH. *Journal of Yanbei Normal University*, **23**, 9-11.

[14] Sastry, G.P. (1988) Estimation of Autoregressive Models with ARCH Errors. *The Indian Journal of Statistics Series B*, **50**, 119-138.

[15] Lu, Z.D. and Gijbels, I. (2001) Asymptotics for Partly Linear Regression with Dependent Samples and ARCH Errors: Consistency with Rates. *Science in China Series A*: *Mathematics*, 44, 168-183. https://doi.org/10.1007/BF02874419

[16] Riesz, F. and Sz-Nagy, B. (1955) Functional Analysis. Dover Publications, New York.

[17] Hendrick, W. and Koenker, R. (1991) Hierarchical Spline Models for Conditional Quantiles and the Demand for Electricity. *Journal of the American Statistical Association*, **87**, 58-68. https://doi.org/10.1080/01621459.1992.10475175

[18] Pollard, D. (1991) Asymptotics for Least Absolute Deviation Regression Estimators. *Econometrics Theory*, **7**, 186-199. https://doi.org/10.1017/S0266466600004394

[19] Revesz, P. (1968) The Laws of Large Numbers. Academic Press, New York.

[20] Hall, P. and Heyde, C.C. (1980) Martingale Limit Theory and Its Application. Academic Press, New York.

[21] Pollard, D. (1984) Convergence of Stochastic Process. Springer-Verlag, New York.
https://doi.org/10.1007/978-1-4612-5254-2

[22] Bosq, D. (2000) Linear Processes in Function Spaces. Springer-Verlag, New York.
https://doi.org/10.1007/978-1-4612-1154-9

[23] Koenker, R. (2005) Quantile Regression. Cambridge University Press, Cambridge.
https://doi.org/10.1017/CBO9780511754098

Appendix. Proofs of the Theorems

We will state the proofs of the theorems given in Section 3. Firstly, some lemmas will be given.

Lemma A.1. ([12], Theorem 1) $\{\varepsilon_t\}$ is a strictly stationary solution of model (1.3) and $E\varepsilon_0^2 < \infty$ if and only if $\sum_{j=1}^{p} \alpha_j < 1$. Furthermore, this solution is unique and ergodic.

Lemma A.2. ([12], Theorem 3) Let

$$L^r = \left\{ x \Big| \|x\|_r = \left(E|x|^r \right)^{1/r} < \infty, x \text{ is random variable} \right\} \quad \text{and suppose (3.1) holds,}$$

$Ee_t^{4(r-1)} < \infty$, where $r \geq 1$ is an integer, then $\varepsilon_t^2 \in L^r$.

Lemma A.3. Consider $\{\varepsilon_i : i \geq 1\}$ forms an ARCH(p) process. Besides, (3.1) holds, then

$$n^{-1} \sum_{i=1}^{n} \varepsilon_i^{2r} \to E\left[\varepsilon_i^{2r} \right] \ a.s.$$

for the integer r in condition (3.1); furthermore, if $r \geq 2$, then

$$n^{-1} \sum_{i=1}^{n} \varepsilon_i^2 \varepsilon_{i-j}^2 \to E\left[\varepsilon_i^2 \varepsilon_{i-j}^2 \right] \ a.s..$$

Proof. From Lemma A.1 and the representation (3.2), it follows that $\{\varepsilon_t\}$ and $\{\varepsilon_t^2\}$ are strictly stationary ergodic sequences. Combining with Lemma A.2, the results follow immediately from the ergodic theorem ([19] [20]).□

Lemma A.4. If ε is independent of X and (3.1)-(3.2) hold, one has

$$\left\| n^{-1} \sum_{i=1}^{n} X_i \varepsilon_i \right\| = O_p\left(n^{-\frac{1}{2}} \right).$$

Proof. By simple calculation, the conclusion can be easily derived under the fact $E\left[\varepsilon_i^2 \right] = \alpha_0 / \left(1 - \alpha_1 - \cdots - \alpha_p \right)$.□

Proof of Theorem 1. Let $\hat{\Phi}_k(x) = \sum_{j=1}^{m} \left(\left\langle \hat{C}_{z_k X}, \hat{\rho}_j \right\rangle / \hat{\lambda}_j \right) \left\langle \hat{\rho}_j, x \right\rangle$ and

$\Phi_k(x) = \sum_{j=1}^{\infty} \left(\left\langle C_{z_k X}, \rho_j \right\rangle / \lambda_j \right) \left\langle \rho_j, x \right\rangle$ with $x \in L^2([0,1])$. Set $\|A\|_\infty = \max_i \sum_j |A_{ij}|$ and $\|A\| = \sum_{i=1}^{d} \sum_{j=1}^{d} |A_{ij}|$ for $A = (A_{ij}) \in R^{d \times d}$. Observe that

$$n^{1/2}\left(\hat{\beta} - \beta \right) = \hat{B}^{-1} n^{1/2} \left\{ \frac{1}{n} \sum_{i=1}^{n} \left(z_i - \sum_{j=1}^{m} \frac{\left\langle \hat{C}_{zX}, \hat{\rho}_j \right\rangle \left\langle X_i, \hat{\rho}_j \right\rangle}{\hat{\lambda}_j} \right) \left(\left\langle \gamma, X_i \right\rangle + \varepsilon_i \right) \right\}$$

$$= \hat{B}^{-1} n^{-1/2} \left\{ \sum_{i=1}^{n} \left(z_i - \sum_{j=1}^{m} \frac{\left\langle \hat{C}_{zX}, \hat{\rho}_j \right\rangle \left\langle X_i, \hat{\rho}_j \right\rangle}{\hat{\lambda}_j} \right) \left\langle \gamma, X_i \right\rangle \right.$$

$$+ \sum_{i=1}^{n} \left(\sum_{j=1}^{\infty} \frac{\left\langle C_{zX}, \rho_j \right\rangle \left\langle X_i, \rho_j \right\rangle}{\lambda_j} - \sum_{j=1}^{m} \frac{\left\langle \hat{C}_{zX}, \hat{\rho}_j \right\rangle \left\langle X_i, \hat{\rho}_j \right\rangle}{\hat{\lambda}_j} \right) \varepsilon_i$$

$$+ \left. \sum_{i=1}^{n} \left(z_i - \sum_{j=1}^{\infty} \frac{\left\langle C_{zX}, \rho_j \right\rangle \left\langle X_i, \rho_j \right\rangle}{\lambda_j} \right) \varepsilon_i \right\}$$

with $\hat{\boldsymbol{B}} = \hat{C}_z - \left\{ \hat{\Phi}_k \left(\hat{C}_{z_m X} \right) \right\}_{k,m=1,\cdots,d}$.

According to Lemma A.4, similar to [3], one has

$$\left\| \hat{\boldsymbol{B}} - \boldsymbol{B} \right\|_\infty = O_p \left(n^{-(2b-1)/(a+2b)} \right),$$

$$n^{-1/2} \sum_{i=1}^n \left(z_i - \sum_{j=1}^m \frac{\left\langle \hat{C}_{zX}, \hat{\rho}_j \right\rangle \left\langle X_i, \hat{\rho}_j \right\rangle}{\hat{\lambda}_j} \right) \left\langle \gamma, X_i \right\rangle = o_p(1),$$

$$n^{-1/2} \sum_{i=1}^n \left(\sum_{j=1}^\infty \frac{\left\langle C_{zX}, \rho_j \right\rangle \left\langle X_i, \rho_j \right\rangle}{\lambda_j} - \sum_{j=1}^m \frac{\left\langle \hat{C}_{zX}, \hat{\rho}_j \right\rangle \left\langle X_i, \hat{\rho}_j \right\rangle}{\hat{\lambda}_j} \right) \varepsilon_i = o_p(1).$$

Now, we consider the term

$$n^{-1/2} \sum_{i=1}^n \left(z_i - \sum_{j=1}^\infty \frac{\left\langle C_{zX}, \rho_j \right\rangle \left\langle X_i, \rho_j \right\rangle}{\lambda_j} \right) \varepsilon_i := n^{-1/2} \sum_{i=1}^n \boldsymbol{\eta}_i \varepsilon_i . \text{ We will show}$$

$$n^{-1/2} \sum_{i=1}^n \boldsymbol{\eta}_i \varepsilon_i \xrightarrow{\ d\ } N \left(0, \frac{\alpha_0}{1 - \alpha_1 - \cdots - \alpha_p} \boldsymbol{B} \right). \tag{A.1}$$

Let $\boldsymbol{P}_{j-1}(\cdot) = \mathrm{E}\left[\cdot \mid \mathcal{F}_{j-1} \right]$ and $\boldsymbol{\xi}_i = n^{-1/2} \boldsymbol{\eta}_i \varepsilon_i$, then $\{\boldsymbol{\xi}_i\}$ forms a martingale difference series due to the fact that $\boldsymbol{\xi}_i$ is $\mathcal{F}_{i-1}$-measurable and $\boldsymbol{P}_{i-1}(\boldsymbol{\xi}_i) = 0$. Let $\boldsymbol{u}_i$ denote the conditional variances of $\boldsymbol{\xi}_i$, then, for $i = 1, \cdots, n$,

$$\boldsymbol{u}_i = \boldsymbol{P}_{i-1}\left(\boldsymbol{\xi}_i \boldsymbol{\xi}_i' \right) = n^{-1} \boldsymbol{P}_{i-1}\left(\boldsymbol{\eta}_i \boldsymbol{\eta}_i' \varepsilon_i^2 \right) = n^{-1} \mathrm{E}\left(\boldsymbol{\eta}_i \boldsymbol{\eta}_i' \right) \boldsymbol{P}_{i-1}\left(\varepsilon_i^2 \right) = n^{-1} \boldsymbol{B} h_i.$$

Therefore,

$$\sum_i \boldsymbol{u}_i = n^{-1} \sum_i \boldsymbol{B} h_i \xrightarrow{\ p\ } \frac{\alpha_0}{1 - \alpha_1 - \cdots - \alpha_p} \boldsymbol{B},$$

according to the law of large numbers ([19]). Furthermore, for any $\delta > 0$,

$$\sum_j \boldsymbol{P}_{j-1}\left(\boldsymbol{\xi}_j \boldsymbol{\xi}_j' \left\{ \left\| \boldsymbol{\xi}_j \right\| > \delta \right\} \right)$$

$$= n^{-1} \sum_j \boldsymbol{P}_{j-1}\left(\boldsymbol{\eta}_j \boldsymbol{\eta}_j' \varepsilon_j^2 \left\{ \left\| \boldsymbol{\eta}_j \varepsilon_j \right\| > n^{1/2} \delta \right\} \right)$$

$$\leq n^{-1} \sum_j \boldsymbol{P}_{j-1}\left(\boldsymbol{\eta}_j \boldsymbol{\eta}_j' \varepsilon_j^2 \left[\left\{ \left\| \boldsymbol{\eta}_j \right\|^2 > n^{1/2} \delta \right\} \cup \left\{ \varepsilon_j^2 > n^{1/2} \delta \right\} \right] \right)$$

$$= n^{-1} \sum_j \boldsymbol{P}_{j-1}\left(\varepsilon_j^2 \right) E\left(\boldsymbol{\eta}_1 \boldsymbol{\eta}_1' \left\{ \left\| \boldsymbol{\eta}_1 \right\|^2 > n^{1/2} \delta \right\} \right)$$

$$+ n^{-1} \sum_j \boldsymbol{P}_{j-1}\left(\varepsilon_j^2 \left\{ \varepsilon_j^2 > n^{1/2} \delta \right\} \right) \mathrm{E}\left(\boldsymbol{\eta}_1 \boldsymbol{\eta}_1' \right).$$

For the first term, it converges to zero because $\left\| \boldsymbol{\eta}_j \boldsymbol{\eta}_j' \right\|$ is uniformly integrable. In view of the integrability of ε_i^2, the second term also converges to zero in probability. Using the martingale difference central limit theorem (CLT) ([21], we get (A.1) holds. Therefore, the conclusion of Theorem 1 holds.

Proof of Theorem 2. With Lemma A.4, the technics in the proof of Theorem 3.2 of [3] can be extended to the present model. So we omit it here.

Proof of Theorem 3. Firstly, we consider the following two equalities:

$$n^{-1} \sum_{i=p+1}^{n} \hat{\varepsilon}_i^2 = n^{-1} \sum_{i=p+1}^{n} \varepsilon_i^2 + o_p\left(n^{-\frac{1}{2}}\right), \tag{A.2}$$

$$n^{-1} \sum_{i=p+1}^{n} \hat{\varepsilon}_i^2 \hat{\varepsilon}_{i-j}^2 = n^{-1} \sum_{i=p+1}^{n} \varepsilon_i^2 \varepsilon_{i-j}^2 + o_p\left(n^{-\frac{1}{2}}\right). \tag{A.3}$$

From Theorem 1 and Theorem 2, we learn that

$$\sum_{j=1}^{m} \left(\hat{\gamma}_j - \gamma_j\right)^2 = \left\|\hat{\tilde{\gamma}} - \tilde{\gamma}\right\|^2 = O_p\left(n^{-(2b-1)/(a+2b)}\right), \tag{A.4}$$

$$\left\|\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}\right\| = O_p\left(n^{-1/2}\right). \tag{A.5}$$

Under the conditions (3.5)-(3.7) and $X \in L^2(\mathcal{T})$, one has

$$\sum_{j=m+1}^{\infty} \gamma_j^2 \left\langle X_i, \rho_j \right\rangle^2 \le C \sum_{j=m+1}^{\infty} j^{-2b} \left\langle X_i, \rho_j \right\rangle^2 = O_p\left(n^{-(2b-1)/(a+2b)}\right). \tag{A.6}$$

In addition, according to (3.3) and $\lambda_1 > \lambda_2 > \cdots > \lambda_m$, the relation

$$\limsup_{n \to \infty} n\mathrm{E}\left\|\hat{\rho}_j - \rho_j\right\|^2 < \infty \tag{A.7}$$

holds, see ([22], ch4). For the residual $\hat{\varepsilon}_i^2$, we have

$$\hat{\varepsilon}_i = Y_i - \hat{\boldsymbol{\beta}}' z_i - \sum_{j=1}^{m} \hat{\gamma}_j \hat{U}_{ij}$$

$$= \boldsymbol{\beta}' z_i + \sum_{j=1}^{\infty} \gamma_j U_{ij} + \varepsilon_i - \hat{\boldsymbol{\beta}}' z_i - \sum_{j=1}^{m} \hat{\gamma}_j \hat{U}_{ij}$$

$$= \varepsilon_i - \left(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}\right)' z_i - \sum_{j=1}^{m} \left(\hat{\gamma}_j - \gamma_j\right) \left\langle X_i, \hat{\rho}_j \right\rangle$$

$$- \sum_{j=1}^{m} \gamma_j \left\langle X_i, \hat{\rho}_j - \rho_j \right\rangle + \sum_{j=m+1}^{\infty} \gamma_j \left\langle X_i, \rho_j \right\rangle$$

by (2.1). Combining this equality with (A.4)-(A.7), (A.2) and (A.3) can be proved.

Now we turn to consider the asymptotic form of $\hat{\alpha}$. By Lemma A.3, we can conclude

$$\left(n^{-1} \hat{P}' \hat{P}\right)^{-1} \to P^{-1} \ a.s. \text{ as } n \to \infty, \tag{A.8}$$

where

$$\hat{P} = \begin{pmatrix} 1 & \hat{\varepsilon}_p^2 & \hat{\varepsilon}_{p-1}^2 & \cdots & \hat{\varepsilon}_1^2 \\ 1 & \hat{\varepsilon}_{p+1}^2 & \hat{\varepsilon}_p^2 & \cdots & \hat{\varepsilon}_2^2 \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & \hat{\varepsilon}_{n-1}^2 & \hat{\varepsilon}_{n-2}^2 & \cdots & \hat{\varepsilon}_{n-p}^2 \end{pmatrix}.$$

Combine (A.2), (A.3), (A.8) and the assumptions about the densities of $\left\{\varepsilon_i^2\right\}$, the results of Theorem 3 holds by the Theorem 4.1 of [23].