

Article

Not peer-reviewed version

---

# Varying-Coefficient Additive Models with Density Responses and Functional Auto-Regressive Error Process

---

Zixuan Han , [Tao Li](#)\*, Jinhong You , [Narayanaswamy Balakrishnan](#)

Posted Date: 14 July 2025

doi: 10.20944/preprints202507.1075.v1

Keywords: varying-coefficient; density response; functional auto-regressive error process; log quantile density transformation



Preprints.org is a free multidisciplinary platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This open access article is published under a Creative Commons CC BY 4.0 license, which permit the free download, distribution, and reuse, provided that the author and preprint are cited in any reuse.

Disclaimer/Publisher's Note: The statements, opinions, and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions, or products referred to in the content.

Article

# Varying-Coefficient Additive Models with Density Responses and Functional Auto-Regressive Error Process

Zixuan Han <sup>1</sup>, Tao Li <sup>2,\*</sup> , Jinhong You <sup>2</sup> and N. Balakrishnan <sup>3</sup>

<sup>1</sup> Division of Public Health Sciences, Fred Hutchinson Cancer Center, Seattle, USA

<sup>2</sup> School of Statistics and Data Science, Shanghai University of Finance and Economics, Shanghai, China

<sup>3</sup> Department of Mathematics and Statistics, McMaster University, Hamilton, Ontario, Canada

\* Correspondence: li.tao@mail.shufe.edu.cn

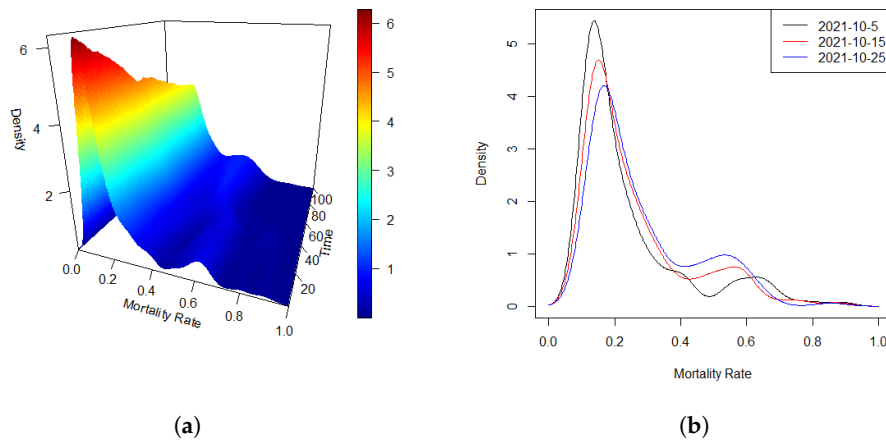
## Abstract

In many practical applications, data collected over time exhibit auto-correlation, which, if not addressed, can lead to incorrect statistical inferences. To address this, we propose a varying-coefficient additive model with density responses, incorporating a functional autoregressive (FAR) error process to account for serial dependency. We present a three-step spline-based estimation procedure for the varying-coefficient components after mapping densities into a linear space using the log-quantile density transformation. First, a B-spline series approximation provides initial estimates of the bi-variate varying-coefficient functions. Second, spline estimation of the error process is obtained from the residuals. Lastly, improved estimates of the additive components are obtained by removing the estimated error process. Theoretical results, including convergence rates and asymptotic properties, are provided, and the practical performance of the method is demonstrated through simulations and real data analyses.

**Keywords:** varying-coefficient; density response; functional auto-regressive error process; log quantile density transformation

## 1. Introduction

Density, or more generally, distributional data appear increasingly often in various real-world research domains. Specific examples include the distributions of cross-sectional or intraday stock returns [9,15], mortality densities [13], and distributions of intrahub connectivity in neuroimaging [11,14]. In many applications, the density curves are observed consecutively over time, which we call in the article a density time series. A motivating example is shown in Figure 1, (a) shows density time series of global mortality rate (‰) over an interval of 100 days from January 22, 2020 to April 15, 2021, while Figure (b) is an alternative look with plotting densities at three different days. In this article, we consider a regression model where density time series, served as response, is coupled with scalar predictors.



**Figure 1.** Densities of global mortality rate (%) of COVID-19 over an interval of 100 days. (a): Three-dimensional trend of density time series during the whole period; (b) Density curves at three different selected days.

Unlike conventional functional data, density data do not form a linear space due to inherent constraints, such as nonnegativity and the requirement to integrate to one. These characteristics present significant challenges when attempting to directly apply functional data analysis techniques to random densities. Several studies have addressed this issue and proposed various solution methods, which can be broadly categorized into two types. One natural approach to overcoming the nonlinear limitations of density functions is to map them into a Hilbert space using a suitable transformation method. In this direction, [12] proposed two continuous and invertible transformations, which are log hazard transformation and log quantile density (LQD) transformation, to map probability densities to an unrestricted space of square integrable functions. [8] map densities by LQD transformation as in [12] to unrestricted square integrable functions, model and fit the responses with an additive functional-to-scalar regression. In order to forecast density functions derived from cross-sectional and intraday returns, [9] proposed two approaches based on compositional data analysis and modified log quantile transformation, combining the FPC representation and exponential smoothing forecasting method. The second way involves choosing appropriate metrics and adopting a geometric approach. For instance, [17] adopted the infinite-dimension version of the Aitchison geometry to construct a density-on-scalar linear regression model in Bayes–Hilbert spaces, while [13] discussed Fréchet regression in a general metric space with the Wasserstein metric. [5] utilized the geometry of tangent bundles of the Wasserstein space with the Wasserstein metric to propose distribution-on-distribution regression. They also studied an extension to autoregressive models of order one for distribution-valued time series. With different methodology, [19] also discussed order- $p$  Wasserstein autoregressive model.

Denote  $\mathcal{F}$  as the space of density functions  $f$  defined on a common support  $\mathcal{U}$ . Without of generality, we assume that  $\mathcal{U} = [0, 1]$ . Given the transformation  $\Psi : \mathcal{F} \rightarrow L_2$ , for any given  $\mathbf{X} \in \mathbb{R}^d$ , the conditional Fréchet mean of the random density  $f$  is defined as

$$\mu(\cdot | \mathbf{X}) = \arg \min_{d \in \mathcal{F}} E \left( \|\Psi(f) - \Psi(d)\|_2^2 | \mathbf{X} \right),$$

where the expectation  $E$  represents the joint distribution of  $(\mathbf{X}, f)$ .

This is equivalent to the following equation

$$\Psi(\mu(\cdot | \mathbf{X}))(u) = E \left( \Psi(f)(u) | \mathbf{X} \right), \quad 0 \leq u \leq 1,$$

leading to the fact that

$$\mu(s|X) = \Psi^{-1}\left(E(\Psi(f)(u)|X)\right)(s), \quad 0 \leq s \leq 1.$$

Data that we consider in this article are the density time series  $d_t$  coupled with scalar predictors  $(X_t, Z_t)$ . To deal with the density functions, we adopt the LQD transformation  $\Psi : \mathcal{F} \rightarrow L_2$ , where  $\mathcal{F}$  is the class of density function  $d$  satisfying  $\int_{\mathcal{R}} u^2 d(u) du < \infty$ . For each  $d_t \in \mathcal{F}$ , let  $Q_t(u)$  be the quantile function corresponding to the cumulative distribution function  $F_t(y)$  of  $d_t(y)$  with support on  $[0,1]$ , and  $q_t(u)$  be the quantile density function, i.e.,  $q_t(u) = Q'_t(u) = \frac{d}{du} F_t^{-1}(u)$  for  $u \in [0,1]$ , then the LQD transformation of  $d_t$  is defined as

$$\Psi(d_t)(u) = \log\left(\frac{d}{du} F_t^{-1}(u)\right), \quad u \in [0,1].$$

In this article, a varying-coefficient additive model with functional auto-regressive error process is proposed to estimate  $E(\Psi(d_t)|X_t, Z_t)$ . Then  $d_t$  can be written as

$$d_t = \Psi^{-1}(E(\Psi(d_t)|X_t, Z_t)) + \delta_{t1},$$

where  $\delta_{t1}$  is the error from regression.

Except for  $\delta_{t1}$ , there is usually another source of error, which comes from the estimation of  $d_t$ , i.e.,  $\hat{d}_t = d_t + \delta_{t2}$ . That is because in most practical applications, instead of the density  $d_t$  itself, only  $n_t$  random samples generated from each  $d_t$  can be observed. Namely,  $Y_{t1}, \dots, Y_{tn_t} \sim d_t$ . In this article, we assume that  $n_t = n$ . Following [12],  $d_t$  is estimated by the modified kernel smoothing method, i.e.,

$$\hat{d}_t(y) = \sum_{i=1}^n \mathcal{K}\left(\frac{y - Y_{ti}}{h}\right) w(y, h) \bigg/ \sum_{i=1}^n \int_0^1 \mathcal{K}\left(\frac{s - Y_{ti}}{h}\right) w(s, h) ds,$$

where the weight function  $w(y, h)$  is defined so as to rectify the boundary effects as follows:

$$w(y, h) = \left(\int_{-y/h}^1 \mathcal{K}(u) du\right)^{-1} I_{y \in [0, h]} + \left(\int_{-1}^{(1-y)/h} \mathcal{K}(u) du\right)^{-1} I_{y \in (1-h, 1]} + I_{y \in [h, 1-h]},$$

in which  $\mathcal{K}$  is the kernel with bandwidth  $h < 1/2$  and bounded variation that is symmetric about 0 such that  $\int_0^1 \mathcal{K}(u) du > 0$ ,  $\int_{\mathcal{R}} |u| \mathcal{K}(u) du$ ,  $\int_{\mathcal{R}} \mathcal{K}^2(u) du$ , and  $\int_{\mathcal{R}} |u| \mathcal{K}^2(u) du$  are finite. Thus, when we fit the regression model with  $\hat{d}_t$  as the observation of  $d_t$ , we have

$$\hat{d}_t = \Psi^{-1}(E(\Psi(d_t)|X_t, Z_t)) + \delta_{t1} + \delta_{t2}.$$

The contribution of this article lies in integrating density time series modeling with a functional autoregressive error process. A common assumption in regression modeling is the independence of random errors. However, data collected over time often exhibit serial dependence, rendering this assumption potentially inappropriate. By incorporating autoregressive structures into the error process, this work accounts for the temporal correlation inherent in density time series, thereby improving the model's flexibility and accuracy. For example, daily price curves or liquidity demand-supply curves in economic and financial applications or climatological curves describing the natural phenomenon [1,3,4,6] all change over time. In such cases, the daily fluctuations in the collected data are heavily influenced by previous values, potentially resulting in what is known as sequence dependence. Numerous empirical studies have demonstrated that neglecting the temporal dependencies between observations can lead to biased statistical results and flawed inferences. Consequently, it is crucial to incorporate autoregressive error structures that explicitly account for the autocorrelation present within the data. This approach ensures a more accurate representation of the underlying temporal dynamics in regression modeling. While functional autoregressive (FAR) models have been extensively studied in the literature [1–4,6,10,18], there is a notable gap in the literature regarding the use of FAR

processes to model the random error component in functional regression models. In this article, to improve the accuracy of density time series predictions, we integrate the FAR error structure within a varying coefficient additive regression model. This novel approach allows for better handling of temporal dependencies in the error process, ultimately enhancing the model's predictive performance.

The remainder of this article is organized as follows. Section 2 outlines the methodology for constructing a varying coefficient additive model with a density response, incorporating a functional autoregressive (FAR) error process. In Section 3, we introduce a three-step procedure for estimating the bivariate varying-coefficient components within the model. Section 4 provides the theoretical results and corresponding inferences derived from the model. Monte Carlo simulations are conducted in Section 5 to demonstrate the efficiency and robustness of the proposed approach. In Section 6, we showcase the practical applicability of the model through analyses of COVID-19 data and U.S. income data. Finally, a discussion of the findings is provided in Section 7, with detailed proofs of the theoretical results included in the Appendix.

## 2. Model Setup

In this article, we focus on the modeling of density response. Due to the constraints of density function which is nonnegative and integrated to 1, we consider the functions after LQD transformation.

In the article, the main goal is to estimate  $E(\Psi(d_t)(u)|x)$  based on the transfer of density function as

$$E(\Psi(d_t)(u)|x) = \sum_{m=1}^k z_{t,m} g_m(u, x_{t,m}), \quad 0 \leq u \leq 1, \quad (1)$$

leading to our proposed varying-coefficient additive models with density responses and functional auto-regressive  $FAR(p)$  error process (DVCA-FAR):

$$f_t(u) = \Psi(d_t)(u) = \sum_{m=1}^k z_{t,m} g_m(u, x_{t,m}) + \varepsilon_t(u), \quad 0 \leq u \leq 1, \quad 1 \leq t \leq T, \quad (2)$$

where  $\varepsilon_t(u)$  is the  $FAR(p)$  process

$$\varepsilon_t(u) = \int \gamma_1(s, u) \varepsilon_{t-1}(s) ds + \cdots + \int \gamma_p(s, u) \varepsilon_{t-p}(s) ds + e_t(u). \quad (3)$$

In the above model, the random density  $d_t(\cdot) \in \mathcal{F}$  serves as a response, and  $\Psi : \mathcal{F} \rightarrow L_2$  is the LQD transformation. Each density is coupled with the  $k$ -dimensional covariates  $x_t = (x_{t,1}, \dots, x_{t,k})^\tau$  and  $z_t = (z_{t,1}, \dots, z_{t,k})^\tau$  with supports  $S_x$  and  $S_z$  respectively. Without loss of generality, we take  $S_x = S_z = [0, 1]$ . In this article, covariate  $x_t$  can be  $z_t$  or  $t/T$ .

The bivariate functions  $g_m(\cdot, x_m)$  quantify the effect of  $z$ ,  $\gamma_l(\cdot, \cdot)$  are smooth functions satisfying  $\iint \gamma_l^2(s, u) du ds < \infty$ . Moreover, the error  $e_t(u)$  is an independent and identically distributed random process satisfying  $E(e_t(u)) = 0$  and  $Cov(e_t(u), e_t(s)) = \sigma_t^2(u, s)$ .

With the estimated density  $\hat{d}$ , let  $\hat{f}_t = \Psi(\hat{d}_t)$ . Then DVCA-FAR model can be written as

$$\hat{f}_t(u) = \sum_{m=1}^k z_{t,m} g_m(u, x_{t,m}) + \varepsilon_t(u) + \varepsilon_{f_t}(u), \quad 0 \leq u \leq 1, \quad (4)$$

where  $\varepsilon_{f_t}(u)$  is random error generated from the transformation of estimated density.

## 3. Three-Step Estimation Methodology

We propose a three-step estimation procedure. First, the B-spline smoothing method is employed to obtain the initial estimate of the bivariate varying coefficients, disregarding the error structure. Second, using the initial estimate and the response, we estimate the error process. The functional autoregressive (FAR) process is then estimated using the sequential test proposed by [10]. Finally, after removing the estimated FAR error, the optimal estimate is derived using the spline method.



### 3.1. Initial Estimation of Bivariate Varying-Coefficient Function

First, without considering error structure, we adopt the spline method to obtain the initial estimations of  $g_m(u, x_m)$ ,  $m = 1, \dots, k$ , regardless of the error process, by the spline series approximation method.

Let  $\{B_0(u), \dots, B_{N_0}(u)\}$  be the set of B-spline basis functions for  $u$  of order  $q$  with  $L_0$  interior knots, where  $N_0 + 1 = L_0 + q$ . Similarly, for each  $m = 1, \dots, k$ , let  $\{B_{0,m}(x_m), \dots, B_{N_m,m}(x_m)\}$  be the set of B-spline basis functions of order  $q$  for  $x_m$  ( $m = 1, \dots, k$ ) with  $L_m$  interior knots, where  $N_m + 1 = L_m + q$ . Denote  $b_{j,m}^*(x_m)$  as the normalization of  $B_{j,m}(x_m)$  and  $b_r(u) = N_0^{1/2} B_r(u)$  as the scaled version of  $B_r(u)$ .

Using these basis functions, we define the tensor product of the B-spline basis as

$$b_{r,j,m}(u, x_m) = b_r(u) b_{j,m}^*(x_m), \quad 1 \leq r \leq N_0, \quad 1 \leq j \leq N_m, \quad 1 \leq m \leq k.$$

The spline approximation of  $g_m(u, x_m)$  is given by

$$g_m(u, x_m) \approx \sum_{r=1}^{N_0} \sum_{j=1}^{N_m} \lambda_{r,j,m} b_{r,j,m}(u, x_m), \quad 1 \leq m \leq k.$$

With the least square method, the estimation of  $g_m(u, x_m)$  is

$$\tilde{g}_m(u, x_m) = \sum_{r=1}^{N_0} \sum_{j=1}^{N_m} \tilde{\lambda}_{r,j,m} b_{r,j,m}(u, x_m), \quad 1 \leq m \leq k, \quad (5)$$

where  $\tilde{\lambda} = (\tilde{\lambda}_{1,1,1}, \dots, \tilde{\lambda}_{N_0, N_k, k})^\tau$  is a  $(N_0 \sum_{m=1}^k N_m)$ -dimensional vector satisfying

$$\tilde{\lambda} = \arg \min_{\tilde{\lambda}} \sum_{t=1}^T \sum_{i=1}^n \left[ \hat{f}_t(u_i) - \sum_{m=1}^k z_{t,m} \sum_{r=1}^{N_0} \sum_{j=1}^{N_m} \lambda_{r,j,m} b_{r,j,m}(u_i, x_{t,m}) \right]^2. \quad (6)$$

Theorem 1 shows that the initial estimation  $\tilde{g}_m(u, x_m)$  is uniformly consistent.

### 3.2. Estimation of FAR Error Process

With the initial estimation  $\tilde{g}_m(u, x_m)$ , we now estimate the FAR error process. To this end, let

$$\tilde{\varepsilon}_t(u) = f_t(u) - \sum_{m=1}^k z_{t,m} \tilde{g}_m(u, x_{t,m}), \quad 1 \leq t \leq T. \quad (7)$$

Denote the additive function  $\rho_t(u) = \sum_{l=1}^p \int \gamma_l(s, u) \varepsilon_{t-l}(s) ds$ , thus the FAR( $p$ ) error process (3) can be written as  $\varepsilon_t(u) = \rho_t(u) + e_t(u)$ .

Let  $\{B_0(u), B_2(u), \dots, B_N(u)\}$  be the set of B-spline basis functions of order  $q$  with  $L$  interior knots, where  $N + 1 = L + q$ . Define the tensor product of the B-spline basis as

$$b_{r,j}(u, s) = B_r(u) B_j(s), \quad 1 \leq r, j \leq N.$$

Then, the spline approximations of the functions  $\gamma_l(\cdot, \cdot)$  is given by

$$\gamma_l(s, u) = \sum_{r=1}^N \sum_{j=1}^N \mu_{r,j,l} b_{r,j}(u, s), \quad 1 \leq l \leq p.$$

The estimation of the  $pN^2$ -dimensional vector  $\mu = (\mu_{1,1,1}, \dots, \mu_{N,N,p})^\tau$  can be obtained by minimizing the following quadratic loss function:

$$\hat{\mu} = \arg \min_{\mu} \sum_{t=p+1}^T \sum_{i=1}^n \left[ \tilde{\varepsilon}_t(u_i) - \sum_{l=1}^p \sum_{r=1}^N \sum_{j=1}^N \mu_{r,j,l} \int b_{r,j}(u_i, s) \tilde{\varepsilon}_{t-l}(s) ds \right]^2.$$

Then, the corresponding estimation of  $\gamma_l(\cdot, \cdot)$  and the additive function  $\rho_t(u)$  are given by

$$\hat{\gamma}_l(s, u) = \sum_{r=1}^N \sum_{j=1}^N \hat{\mu}_{r,j,l} b_{r,j}(u, s), \quad 1 \leq l \leq p,$$

and

$$\hat{\rho}_t(u) = \sum_{l=1}^p \sum_{r=1}^N \sum_{j=1}^N \hat{\mu}_{r,j,l} \int b_{r,j}(u, s) \hat{\varepsilon}_{t-l}(s) ds, \quad 0 \leq u \leq 1, \quad p+1 \leq t \leq T, \quad (8)$$

respectively.

In practice, the order  $p$  of error process is unknown. To solve this problem, we utilize a sequential test by [10] to identify the order  $p$ . The detailed identification procedure is given in Section 3.4.2.

### 3.3. Improved Estimation of Bivariate Varying-Coefficient Function

By removing the estimated error process (8) from the response function  $f_t(u)$  and repeating the same procedure as in Section 3.1, we obtain an improved estimation of  $g_m(u, x_m)$ .

To do so, we first denote

$$f_t^c(u) = f_t(u) - \sum_{l=1}^p \int \gamma_l(s, u) \varepsilon_{t-l}(s) ds, \quad 0 \leq u \leq 1, \quad p+1 \leq t \leq T,$$

and the estimation of  $f_t^c(u)$  as

$$\hat{f}_t^c(u) = f_t(u) - \sum_{l=1}^p \int \hat{\gamma}_l(s, u) \hat{\varepsilon}_{t-l}(s) ds, \quad 0 \leq u \leq 1, \quad p+1 \leq t \leq T.$$

On the other side,

$$f_t^c(u) = \sum_{m=1}^k z_{t,m} g_m(u, x_{t,m}) + e_t(u), \quad 0 \leq u \leq 1.$$

Therefore, following the same procedure as the first step, the improved spline approximation estimations are given by

$$\hat{g}_m(u, x_m) = \sum_{r=1}^{N_0} \sum_{j=1}^{N_m} \hat{\lambda}_{r,j,m} b_{r,j,m}(u, x_m), \quad 1 \leq m \leq k, \quad (9)$$

where  $\hat{\lambda} = (\hat{\lambda}_{1,1,1}, \dots, \hat{\lambda}_{N_0, N_k, k})^T$  is a  $(N_0 \sum_{m=1}^k N_m)$ -dimensional vector satisfying

$$\hat{\lambda} = \arg \min_{\lambda} \sum_{t=1}^T \sum_{i=1}^n \left[ \hat{f}_t^c(u_i) - \sum_{m=1}^k z_{t,m} \sum_{r=1}^{N_0} \sum_{j=1}^{N_m} \lambda_{r,j,m} b_{r,j,m}(u_i, x_{t,m}) \right]^2. \quad (10)$$

Theorem 2 and 3 show the uniform convergence and asymptotic normality of the improved estimation  $\hat{g}_m(u, x_m)$ . Besides, simulation studies in Section 5 show that the improved estimation is more efficient than the initial estimation  $\tilde{g}_m(u, x_m)$ .

### 3.4. Implementation

#### 3.4.1. Selection of Bandwidth

In empirical applications, prior to modeling, it is necessary to estimate the density, which involves selecting the appropriate bandwidth for the modified kernel estimation. In this section, we employ the leave-one-out cross-validation method to determine the optimal bandwidth. Specifically, the bandwidth  $h$  is chosen by minimizing the following mean squared error (MSE):

$$CV(h) = \frac{1}{nT} \sum_{t=1}^T \sum_{i=1}^n [d_t(y_{ti}) - \hat{d}_t(y_{ti})]^2,$$

where for each  $i = 1, \dots, n$ ,  $\hat{d}_t(y_{ti})$  is the estimate of  $d_t(y_{ti})$  with bandwidth  $h$  obtained by using observations from the  $t$ -th section other than the  $i$ -th one.

#### 3.4.2. Identifying the Order of the FAR Process

In this section, we utilize the order determination procedure proposed by [10] to identify the order  $p$  of FAR error process. The main idea is to represent the FAR process as a fully functional linear model with dependent regressors and construct a sequential test procedure.

Given a sequence of hypotheses:

$$H_{0,p} : \{\varepsilon_t\} \text{ are FAR}(p) \text{ vs } H_{a,p+1} : \{\varepsilon_t\} \text{ are FAR}(p+1), \quad p = 0, 1, 2, \dots,$$

Here,  $\text{FAR}(0)$  means independent and identically distributed process. The sequential test begins with  $p = 0$  and terminated if  $H_{0,p}$  is not rejected, the order of the process is then identified as  $p$ . See [10] for more relevant details and further explanation.

To construct the test statistic, define  $\eta_j(s) = \sum_{l=1}^p \tilde{\varepsilon}_{j-l}(sp - (l-1))I_l(s)$ ,  $\varphi(s, u) = p \sum_{l=1}^p \gamma_l(sp - (l-1), u)I_l(s)$ , where  $I_l$  as the indicator function of the interval  $[(l-1)/p, l/p]$ . Denote the orthonormal basis of  $L^2$  as  $\{\hat{x}_j, 1 \leq j \leq T\}$  from the eigenfunctions of  $\hat{C}_\eta(s, u) = \frac{1}{T} \sum_{j=1}^T (\eta_j(s) - \bar{\eta}(s))(\eta_j(u) - \bar{\eta}(u))$ , with the corresponding decreasingly ordered eigenvalues  $\hat{\lambda}_j$ , where  $\bar{\eta}$  is the mean function of  $\eta_j$ . In the proposed method, only the first  $q_\eta$  eigenfunction/eigenvalue pairs are used. Moreover, define  $\{\hat{y}_j, 1 \leq j \leq T\}$  and  $q_\pi$  analogously to  $\{\hat{x}_j\}$  and  $q_\eta$  for the response functions denoted as  $\pi_j$ .

For the product space  $L^2([0, 1] \times [0, 1])$ , denote  $\boldsymbol{\eta}(j, k) = \langle \eta_j, \hat{x}_k \rangle$ ,  $\boldsymbol{\pi}(j, m) = \langle \pi_j, \hat{y}_m \rangle$ ,  $\boldsymbol{\psi}(k, m) = \langle \varphi, \hat{x}_k \otimes \hat{y}_m \rangle$ . Let  $\boldsymbol{\pi} = [\boldsymbol{\pi}(j, m)]_{T \times q_\pi}$ ,  $\boldsymbol{\eta} = [\boldsymbol{\eta}(j, k)]_{T \times q_\eta}$ , and  $\boldsymbol{\psi} = [\boldsymbol{\psi}(k, m)]_{q_\eta \times q_\pi}$ ,  $j = 1, \dots, T; k = 1, \dots, q_\eta; m = 1, \dots, q_\pi$ .

Construct the matrix  $\hat{\mathbf{A}}$  in which the entries are  $\hat{\mathbf{A}}(k, k') = \langle \hat{x}_{k,p}, \hat{x}_{k',p} \rangle$ , where  $\hat{x}_{k,p}(s) = \hat{x}_k(\frac{s+p-1}{p})$ ,  $0 \leq s \leq 1$ . Define the orthonormal eigenvectors  $\hat{\boldsymbol{\beta}}_k$  with corresponding ordered eigenvalues  $\hat{\zeta}_1 \geq \dots \geq \hat{\zeta}_{q_\eta}$  as  $\hat{\mathbf{A}}\hat{\boldsymbol{\beta}}_k = \hat{\zeta}_k\hat{\boldsymbol{\beta}}_k$ ,  $1 \leq k \leq q_\eta$ , and denote  $\hat{\mathbf{B}} = [\hat{\boldsymbol{\beta}}_1, \dots, \hat{\boldsymbol{\beta}}_{q^*}]$ , where  $q^* = \max\{k \in \{1, \dots, q_\eta\} : \|\hat{z}_{k,p}\|^2 \geq 0.9p\}$  and  $\hat{z}_{k,p}(s) = \sum_{i=1}^{q_\eta} \hat{\beta}_{k,i}\hat{x}_{i,p}(s)$ .

Following [10], the test statistic is constructed as

$$\hat{\tau}_p = \frac{1}{T} \left( \text{vec}[\hat{\mathbf{B}}^\tau \hat{\boldsymbol{\psi}}] \right)^\tau \left[ (I_{q_\varepsilon} \otimes \hat{\mathbf{B}}^\tau)(\hat{\mathbf{C}} \otimes \hat{\mathbf{A}})(I_{q_\varepsilon} \otimes \hat{\mathbf{B}}) \right]^{-1} \text{vec}[\hat{\mathbf{B}}^\tau \hat{\boldsymbol{\psi}}], \quad (11)$$

where  $\hat{\mathbf{A}} = \text{diag}(\hat{\lambda}_1, \dots, \hat{\lambda}_{q_\eta})$ ,  $\hat{\mathbf{C}} = \frac{1}{T}(\boldsymbol{\pi} - \boldsymbol{\eta}\hat{\boldsymbol{\psi}})^\tau(\boldsymbol{\pi} - \boldsymbol{\eta}\hat{\boldsymbol{\psi}})$ . The test statistic  $\hat{\tau}_{p+1}$  has an approximately chi-squared distribution with degree  $q_\pi q^*$ .

## 4. Theoretical Results

In this section, we discuss the asymptotic properties of both initial and improved estimation of  $g_m(u, x_m)$ . Moreover, the consistency of estimation of order  $p$  is derived. All proofs are given in Appendix.



Throughout the remainder of this article, for any fixed interval  $[a, b]$ , we denote the space of  $l$ -th order smooth functions as  $C^{(l)}[a, b] = \{g|g^{(l)} \in [a, b]\}$  and the class of Lipschitz-continuous functions for some fixed constant  $C > 0$  as  $Lip([a, b], C) = \{g| |g(x) - g(x')| \leq |x - x'|, \forall x, x' \in [a, b]\}$ . Let  $S_{x_m}$  and  $S_{z_m}$  denote the supports of  $x_m$  and  $z_m$ , respectively. Then, the supports of  $x$  and  $z$  are  $S_x = \prod_{m=1}^k S_{x_m}$  and  $S_z = \prod_{m=1}^k S_{z_m}$ , respectively. The necessary assumptions for the asymptotic results are as follows.

(A1) For any  $d \in \mathcal{F}$ ,  $d$  is differentiable and there exists a constant  $M > 1$  such that  $\|d\|_\infty, \|1/d\|_\infty$ , and  $\|d'\|_\infty$  are all bounded by  $M$ .

(A2) (a) The kernel density  $\mathcal{K}$  is Lipschitz-continuous, bounded, and symmetric about 0. Furthermore,  $\mathcal{K} \in Lip([-1, 1], L_k)$  for some constant  $L_k > 0$ . (b) The kernel density  $\mathcal{K}$  is such that  $\int_0^1 \mathcal{K}(u) du > 0$ ,  $\int_{\mathcal{R}} |u| \mathcal{K}(u) du$ ,  $\int_{\mathcal{R}} \mathcal{K}^2(u) du$ , and  $\int_{\mathcal{R}} |u| \mathcal{K}^2(u) du$  are finite.

(A3) The covariates  $x_{t,m}, z_{t,m}$ ,  $1 \leq m \leq k$ , and the errors  $\varepsilon_t(u)$  satisfy the following moment conditions: for some  $s > 2$ ,

$$\max_{1 \leq t \leq T} \max_{1 \leq m \leq k} E(|x_{t,m}|^{2s}) < \infty, \max_{1 \leq t \leq T} \max_{1 \leq m \leq k} E(|z_{t,m}|^{2s}) < \infty, \max_{1 \leq t \leq T} \sup_u E(|\varepsilon_t(u)|^{2s}) < \infty.$$

For each  $t = 1, \dots, T$ , the covariance function  $Cov(\varepsilon_t(s), \varepsilon_t(v)) = \Sigma_t(s, v)$  has finite nondecreasing eigenvalues  $\lambda_1 \leq \dots \leq \lambda_{\max}$  satisfying  $\sum_j \lambda_j < \infty$ .

(A4) The additive component functions  $g_m(u, x_m)$ ,  $1 \leq m \leq k$ , are continuous on  $[0, 1] \times [a_m, b_m]$  and twice continuously partially differentiable with respect to  $u$  and  $x_m$ , where  $[a_m, b_m]$  is a compact subset of  $S_{x_m}$ .

(A5)  $N_0 \sim (nT)^{1/6} \log nT$ ,  $N_m \sim (nT)^{1/6} \log nT$ ,  $1 \leq m \leq k$ ,  $h \sim n^{-1/3}$ , as  $n, T \rightarrow \infty$ .

**Remark 1.** Assumption (A1) is basic and essential for deriving the consistency of densities after transformation. The conditions in (A2) on the kernel function  $\mathcal{K}(\cdot)$  are mild and can be satisfied by commonly used kernel functions such as the uniform and Epanechnikov kernels. The moment conditions in (A3) are crucial for deriving the uniform convergence and other asymptotic properties based on the spline function. The smoothness conditions for the component functions in (A4) are greatly relaxed. The conditions in (A5) are commonly applied in spline smoothing to ensure optimal convergence rates.

We first derive the uniform consistency of the initial estimates of bivariate functions  $g_m(u, x_m)$ , as stated in Theorem 1.

**Theorem 1.** *Theorem 1. Assume that assumptions (A1)–(A5) hold, and that  $\tilde{g}_m(u, x_m)$  are the initial estimates of  $g_m(u, x_m)$  defined by (5),  $m = 1, \dots, k$ . Then, as  $n \rightarrow \infty$  and  $T \rightarrow \infty$ , it holds that*

$$\sup_{u, x_m \in [0, 1]} |\tilde{g}_m(u, x_m) - g_m(u, x_m)| = O_p\left((nT)^{-1/3} \log(nT) + n^{-1/3}\right).$$

Theorem 2 characterizes the uniform convergence of the improved estimation of  $g_m(u, x_m)$ , and Theorem 3 describes the asymptotic properties of both the initial and improved estimations.

**Theorem 2.** *Theorem 2. Assume that assumptions (A1)–(A5) hold,  $\hat{g}_m(u, x_m)$  are the improved estimates of  $g_m(u, x_m)$  defined by (9),  $m = 1, \dots, k$ , and that the order of the functional error process  $p$  is known. Then, as  $n \rightarrow \infty$  and  $T \rightarrow \infty$ , the following holds:*

$$\sup_{u, x_m \in [0, 1]} |\hat{g}_m(u, x_m) - g_m(u, x_m)| = O_p\left((nT)^{-1/3} (\log(nT))^{-2} + n^{-1/3}\right).$$

To present the asymptotic normality of the estimations, some notations are given as follows. Denote  $\mathbf{b}(u, x_{t,m}) = (b_{1,1,m}(u, x_{t,m}), \dots, b_{N_0, N_m, m}(u, x_{t,m}))^\tau$ ,  $\mathbf{b}_z(u, x_{t,m}) = z_{t,m} \mathbf{b}(u, x_{t,m})$ ,  $\mathbf{B} \mathbf{z}_{t,m} = (\mathbf{b}_z(u_{1,1,m}, x_{t,m}), \dots, \mathbf{b}_z(u_{N_0, N_m, m}, x_{t,m}))^\tau_{n \times N_0 N_m}$ ,  $\mathbf{B}_m = (\mathbf{B} \mathbf{z}_{1,m}^\tau, \dots, \mathbf{B} \mathbf{z}_{T,m}^\tau)^\tau$ , and  $\mathbf{B} = (\mathbf{B}_1, \dots, \mathbf{B}_k)$ .

Let  $B_* = B/\sqrt{nT}$ , and  $A_m = (0, \dots, I, \dots, 0)$  is an  $1 \times k$  block matrix, with the  $m$ -th block being  $N_0 N_m \times N_0 N_m$  identity matrix and the  $j$ -th ( $j \neq m$ ) block being  $N_0 N_m \times N_0 N_j$  zero matrix.

**Theorem 3.** *Theorem 3. Assume that assumptions (A1)–(A5) hold,  $\tilde{g}_m(u, x_m)$  and  $\hat{g}_m(u, x_m)$  are the initial and improved estimates of  $g_m(u, x_m)$  defined by (5) and (9),  $m = 1, \dots, k$ , respectively. Then, as  $n \gg T \rightarrow \infty$ , the following hold for all  $u \in (0, 1)$  and  $x_m \in [0, 1]$ :*

(i) *The initial estimate  $\tilde{g}_m(u, x_m)$  is asymptotically normally distributed, i.e.,*

$$\sqrt{nT}(\mathbf{C}_m \boldsymbol{\Sigma}_\varepsilon \mathbf{C}_m^\tau)^{-1}(\tilde{g}_m(u, x_m) - g_m(u, x_m)) \xrightarrow{D} N(0, 1),$$

where  $\mathbf{C}_m = \mathbf{b}^\tau(u, x_m)E(\mathbf{A}_m(\mathbf{B}_*^\tau \mathbf{B}_*)^{-1} \mathbf{B}_*^\tau)$ , the covariance matrix  $\boldsymbol{\Sigma}_\varepsilon = (\boldsymbol{\Sigma}_{t,s})_{1 \leq t, s \leq T}$ , with  $\boldsymbol{\Sigma}_{t,s} = \text{Cov}(\varepsilon_t, \varepsilon_s)$ .

(ii) *The improved estimate  $\hat{g}_m(u, x_m)$  is asymptotically normally distributed, i.e.,*

$$\sqrt{nT}(\mathbf{C}_m \boldsymbol{\Xi}_\varepsilon \mathbf{C}_m^\tau)^{-1}(\hat{g}_m(u, x_m) - g_m(u, x_m)) \xrightarrow{D} N(0, 1),$$

where the covariance matrix  $\boldsymbol{\Xi}_\varepsilon = \text{diag}(\boldsymbol{\Xi}_{t,t})_{1 \leq t \leq T}$ , with  $\boldsymbol{\Xi}_{t,t}(u, s) = \sigma_t^2(u, s)$ .

## 5. Numerical Study

In this section, we conduct two simulation studies to demonstrate the performance of the proposed identification and estimation procedure for the additive model.

### 5.1. Case 1

With the assumption that the auto-regressive error process is known, this case is conducted to demonstrate the performance of the estimation procedure with finite  $n$  and  $T$ . We consider the following DVCA-FAR(1) model:

$$f_t(u) = z_{t,1}g_1(u, x_{t,1}) + z_{t,2}g_2(u, x_{t,2}) + \varepsilon_t(u), \quad 0 \leq u \leq 1, \quad (12)$$

and the error function  $\varepsilon_t(u)$  is

$$\varepsilon_t(u) = \int \gamma_1(s, u)\varepsilon_{t-1}(s)ds + e_t(u), \quad 2 \leq t \leq T.$$

Let the bivariate varying-coefficient functions be

$$g_1(u, x_{t,1}) = \sin(2\pi u)(2x_{t,1} - 1), \quad g_2(u, x_{t,2}) = \sin(2\pi u) \sin(2\pi x_{t,2}),$$

and the coefficient functions be

$$\gamma_1(s, u) = 0.2us, \quad e_t(u) = 0.2\eta_{t,1} \sin(\pi u) + \eta_{t,2} \sin(2\pi u),$$

where  $\eta_{t,1} \sim N(0, 0.1^2)$ ,  $\eta_{t,2} \sim N(0, 0.05^2)$ , and  $\eta_{t,1}$  are independent of  $\eta_{t,2}$  for  $u \in [0, 1]$ .

The covariates  $z_{t,1}, z_{t,2}$  are generated by  $N(0, 1)$  and  $N(0, 0.5^2)$ , respectively, while  $x_{t,1}, x_{t,2}$  are generated by  $(x_{t,1}, x_{t,2})^\tau = (\Phi(v_{t,1}), \Phi(v_{t,2}))^\tau$ ,  $1 \leq t \leq T$ , where  $\Phi$  is the cumulative distribution function of the standard normal distribution and  $v_{t,1}, v_{t,2}$  are mutually independent variables of the standard normal distribution.

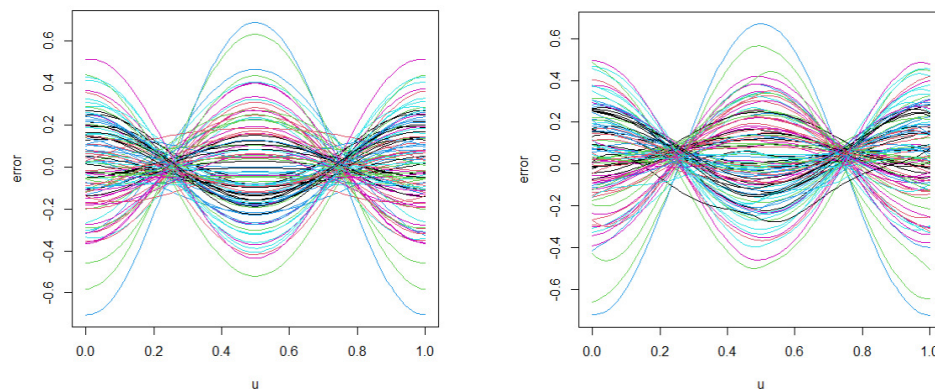
To generate the response densities, for each given  $\mathbf{Z} = \mathbf{z}$  and  $\mathbf{X} = \mathbf{x}$ , let  $\alpha(u, \mathbf{x}, \mathbf{z})$  be the additive surface given by  $\alpha(u, \mathbf{x}, \mathbf{z}) = \sum_{m=1}^p z_m g_m(u, x_m)$ . The conditional quantile function  $Q(\cdot | \mathbf{x}, \mathbf{z})$  with the error process  $\varepsilon(u)$  satisfy  $Q(u | \mathbf{x}, \mathbf{z}) = F^{-1}(u | \mathbf{x}, \mathbf{z}) = \theta(\mathbf{x}, \mathbf{z})^{-1} \int_0^u \exp\{\alpha(v, \mathbf{x}, \mathbf{z}) + \varepsilon(v)\} dv$ , where  $\theta(\mathbf{x}, \mathbf{z}) = \int_0^1 \exp\{\alpha(v, \mathbf{x}, \mathbf{z}) + \varepsilon(v)\} dv$ .

Implementing the conditional quantile function to  $\{U_{t,1} \leq \dots \leq U_{t,n_t}\} \sim U(0, 1)$ , which are independent of  $\mathbf{X}_t$  and  $\mathbf{Z}_t$ , we can get the random samples  $\mathbf{Y}_t = \{Y_{t,j} = Q(U_{t,j} | \mathbf{X}_t, \mathbf{Z}_t) : 1 \leq j \leq n_t\}$

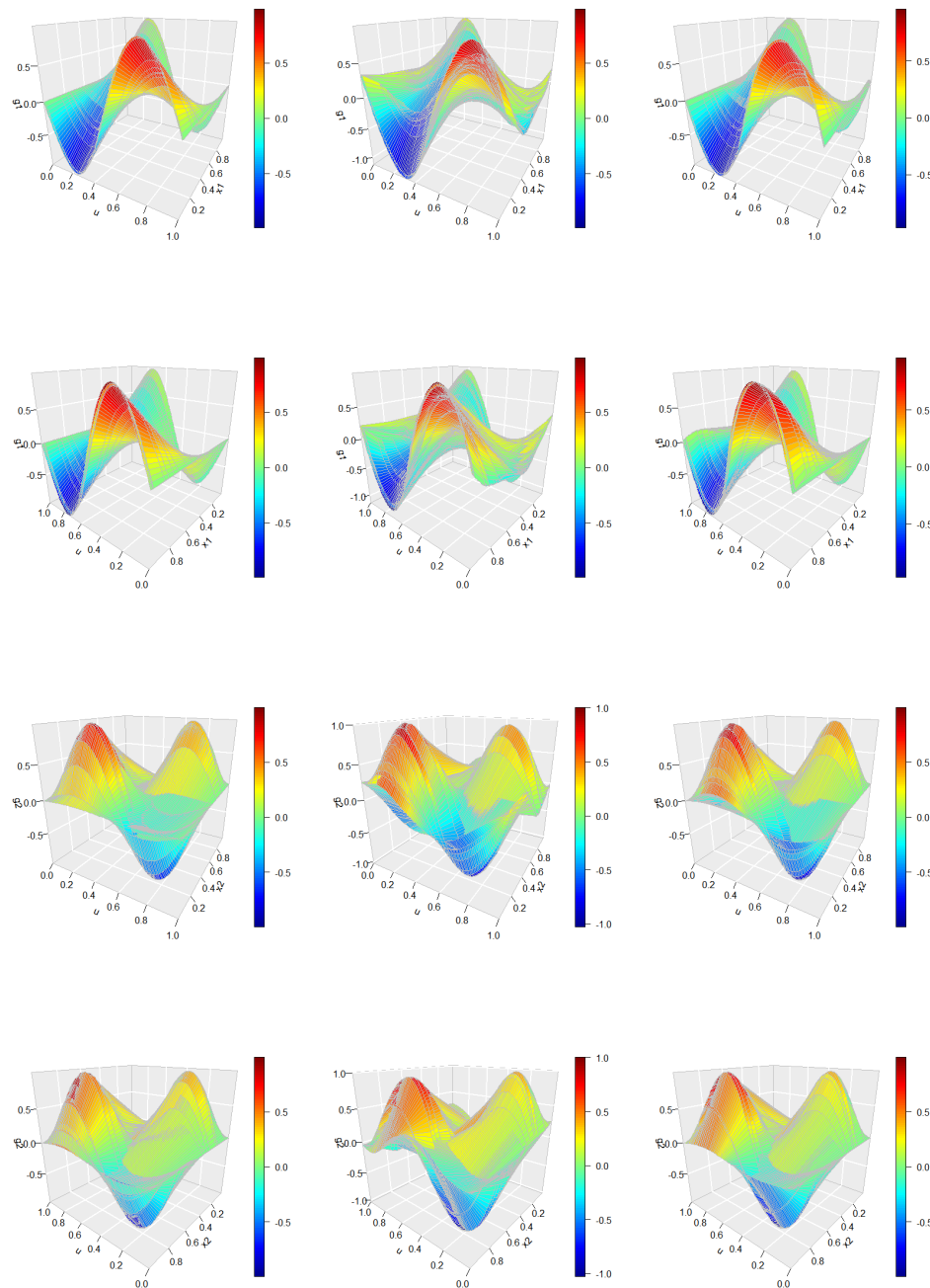
for each  $1 \leq t \leq T$ , so that  $Y_{t,1}, \dots, Y_{t,n_t} \sim d_t$ , where  $d_t$  is the random response density. Denoting  $f_t(u) = \Psi(d_t(u))$ , we obtain the transformed density expressed in model (2). Without loss of generality, we assume that  $n_t = n$  independent and identically distributed observations are available for each response distribution.

With  $T = 100, n = 100$ , the estimation is conducted with 200 Monte Carlo runs. Figure 2 displays the true curves of the FAR error process  $\varepsilon(u)$  in panel (a) and its corresponding spline-based estimations in panel (b). Figure 3 presents a comprehensive view of the true surface of  $g_m(u, x_m)$  alongside the average estimations obtained from 200 Monte Carlo simulations. Specifically, the left panel displays the true densities, the middle panel shows the initial spline estimates derived without accounting for the FAR(1) error process, and the right panel provides the improved estimations achieved through the same method after removing the estimated error process. To better illustrate the performance, the bivariate function estimations are presented from two different relative perspectives, offering a clearer understanding of the improvements made through the error correction process.

Figure 2 indicates that the estimation of  $\varepsilon(u)$  yields highly accurate results, which is further corroborated by the findings presented in Figure 3. Specifically, the right panel of Figure 3 illustrates that the improved estimation, after accounting for the FAR error process, significantly outperforms the initial estimation shown in the middle panel. This improvement highlights the effectiveness of the proposed methodology in refining the bivariate function estimates by appropriately addressing the error structure.



**Figure 2.** Average estimations of FAR(1) error process  $\varepsilon(u)$  obtained from 200 Monte Carlo runs with sample size  $T = 100$  and  $n = 100$  observations. Picture (a): true curves, (b): spline estimations.



**Figure 3.** Average estimations of  $g_m(u, x_m)$ ,  $m = 1, 2$ . Left panels: true densities, middle panels: initial estimations, right panels: improved estimations. Upper two panels display  $g_1(u, x_1)$  from two angles, lower panels display  $g_2(u, x_2)$ .

For further comparison, we conducted simulations with sample sizes of  $T = 50, 100$  and  $n = 50, 100$  observations. We use the root mean squared errors (RMSEs) to measure the accuracy of the estimations, including the initial and improved estimations of  $g_m(u, x_m)$ . The RMSE is defined as

$$RMSE(\hat{g}_m) = \frac{1}{T} \sum_{t=1}^T \left\{ \frac{1}{n} \sum_{i=1}^n \|\hat{g}_m(u_i, x_{t,m}) - g_m(u_i, x_{t,m})\|_2^2 \right\}^{\frac{1}{2}}.$$

where  $\hat{g}_m$  is  $\tilde{g}_m$  or  $\hat{g}_m$ .

Based on the results from 200 Monte Carlo simulations, Table 1 presents the average root mean square errors (RMSEs) along with their standard deviations for both the initial and improved estimations of  $g_m(u, x_m)$ . The findings reveal that the RMSEs of the bivariate functions decrease as both the sample size  $T$  and the number of observations  $n$  increase. Notably, the RMSEs associated with the improved estimation are consistently smaller than those for the initial estimation. This is to be expected, as the initial estimates were derived without accounting for the error process, which, as demonstrated, has a significant impact on the accuracy of the results. By incorporating the error process and removing its estimated effects in the improved estimation, the model yields more accurate and refined results.

**Table 1.** Average RMSEs of both initial and improved estimations of  $g_m(u, x_m)$ .

Average RMSEs of Bivariate Varying-Coefficient Additive Functions					
Sample Size		$g_1(u, x_1)$		$g_2(u, x_2)$	
$T$	$n$	Initial	Improved	Initial	Improved
50	50	0.2247	0.1848	0.2139	0.1785
	100	0.1759	0.1325	0.1844	0.1521
100	50	0.1826	0.1471	0.1732	0.1354
	100	0.1431	0.1164	0.1319	0.1057

To provide a clearer explanation of the theoretical validity differences between the two estimations, we calculate their biases and standard deviations and compare them. Taking the first example in the numerical simulation section as an example.

**Table 2.** Average Standard Deviation (SD) and Bias of both initial and improved estimations of  $g_m(u, x_m)$ .

Average SD and Bias of Bivariate Varying-Coefficient Additive Functions									
Sample Size		$g_1(u, x_1)$				$g_2(u, x_2)$			
$T$	$n$	Initial		Improved		Initial		Improved	
		SD	Bias	SD	Bias	SD	Bias	SD	Bias
50	50	0.205	0.147	0.168	0.104	0.219	0.137	0.183	0.117
	100	0.179	0.122	0.142	0.093	0.196	0.128	0.164	0.095
100	50	0.174	0.136	0.151	0.082	0.187	0.131	0.158	0.086
	100	0.133	0.099	0.112	0.057	0.153	0.111	0.129	0.061

Table 2 presents the average bias and standard deviation (SD) of both the initial and improved estimations of  $g_m(u, x_m)$ . The results clearly indicate that, across all settings of the model sample size, the standard deviation of the improved estimates of the bivariate additive functions is substantially smaller than that of the initial estimates, with the bias also being correspondingly reduced. Moreover, as the sample size increases, both the standard deviation and bias of the estimates decrease, further reinforcing the reliability of the improved method. This finding numerically substantiates the claim that the improved estimation method results in a smaller asymptotic variance-covariance matrix compared to the initial estimation, thereby enhancing the precision and robustness of the estimates.

5.2. Case 2

Case 2 is conducted to demonstrate the efficiency of identifying the auto-regressive order of the functional error process. The densities are also generated from model (12), but with the FAR(2) error function, where  $\gamma_2(s, u) = \frac{1}{4}us^2$ . All other settings are the same as for Case 1.

Table 3 presents the empirical power of the testing algorithm used to determine the order of the FAR error process under various settings and significance levels. The results clearly demonstrate that the power of the test increases as the sample size  $T$  and the number of observations  $n$  grow. Specifically, the power approaches 1 as both  $T$  and  $n$  increase to 100 when testing the null hypothesis of an independent and identically distributed (i.i.d.) sample. This suggests that the test becomes increasingly reliable with larger sample sizes. However, the power is slightly lower when testing the



null hypothesis of order 1 against order 2, which is expected given the complexity of distinguishing between these two orders. Furthermore, the size of the test remains low when testing order 2 against a higher-order process, confirming the accuracy and feasibility of the testing algorithm for determining the appropriate order of the functional error process. These findings validate the effectiveness of the proposed testing algorithm in practical applications, ensuring its robustness and precision in various settings.

Furthermore, to assess the efficiency of the auto-regressive order  $p$  on the overall estimation results, Table 4 presents the average RMSEs for the bivariate varying-coefficient functions. The observed pattern closely mirrors the results from Case 1, further reinforcing the effectiveness and reliability of the proposed model’s identification and estimation procedures. These findings provide strong empirical evidence that the auto-regressive order plays a crucial role in enhancing the accuracy of the estimation process. The consistency of the RMSE results across different settings underscores the robustness of the model, confirming its ability to effectively account for the error structure and yield precise estimates of the bivariate varying-coefficient functions.

**Table 3.** Empirical power of testing algorithm to determine the order of FAR error process under different significance levels.

Null Hypothesis		$p = 0$		$p \leq 1$		$p \leq 2$	
Alternative Hypothesis		$p \geq 1$		$p \geq 2$		$p \geq 3$	
Sample Size		Significance Level		Significance Level		Significance Level	
$T$	$n$	0.05	0.1	0.05	0.1	0.05	0.1
50	50	0.893	0.962	0.787	0.846	0.082	0.134
	100	0.931	0.985	0.824	0.893	0.073	0.125
100	50	0.942	0.972	0.821	0.881	0.071	0.121
	100	0.985	1.000	0.889	0.935	0.064	0.113

**Table 4.** Average RMSEs of both initial and improved estimations of  $g_m(u, x_m)$ .

Average RMSEs of Bivariate Varying-Coefficient Additive Functions					
Sample Size		$g_1(u, x_1)$		$g_2(u, x_2)$	
$T$	$n$	Initial	Improved	Initial	Improved
50	50	0.2739	0.2438	0.2691	0.2235
	100	0.2264	0.1852	0.2157	0.1809
100	50	0.2136	0.1817	0.2232	0.1761
	100	0.1729	0.1263	0.1816	0.1224

6. Real Data Analysis

In this section, we demonstrate the feasibility and efficiency of the proposed estimation procedure through the analysis of two real-world datasets. By applying the developed methodology to empirical data, we aim to showcase the practical utility of the model in capturing the underlying patterns and dependencies inherent in the datasets. The analysis serves not only to validate the effectiveness of the proposed estimation approach but also to highlight its applicability across different domains, offering insights into the model’s versatility and robustness in handling complex, time-dependent, and non-Euclidean data structures.

6.1. COVID-19 Data

On March 11, 2020, the World Health Organization (WHO) declared COVID-19, an infectious disease caused by the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), a global pandemic. The rapid and widespread transmission of the virus led to unprecedented global health challenges, with countries around the world instituting lockdowns and other measures to curb the spread of the disease. As of August 15, 2021, official statistics from the WHO reported a staggering 221,885,822 confirmed cases and 4,583,539 deaths across nearly all countries, reflecting the profound and far-reaching



impact of the pandemic. Given the magnitude of this crisis, it is essential for international health organizations and research institutions to closely monitor the evolving global trends of COVID-19. Such monitoring facilitates timely and accurate analysis, enabling effective public health responses and the development of strategies for medical treatment, prevention, and control of future outbreaks. Understanding the dynamics of the epidemic through data-driven models is thus critical for informing policy decisions and improving global health outcomes in the face of such a devastating crisis.

To demonstrate this proposition, we select the mortality rate as an appropriate metric for measuring the global trend of the COVID-19 pandemic. The mortality rate is defined as the ratio of cumulative deaths per day to the total population of each country, serving as a key indicator of the disease's lethality and spread. Notably, the data required to calculate the mortality rate are inherently temporally dependent, as they rely on the data from the previous day. This results in the mortality rate, and thus the global trend of the epidemic, exhibiting temporal auto-correlation.

The data on COVID-19-related deaths, which are critical for our analysis, can be accessed from the Johns Hopkins University repository. This repository includes a dynamic tracking map that provides a comprehensive view of the global trends related to the pandemic. The dataset, which is publicly available at <https://www.jhu.edu/>, covers the period from January 22, 2020, to April 15, 2021. Additionally, the most recent total population data for each country, necessary for calculating the mortality rate, can be obtained from the World Bank's online platform, accessible at <https://data.worldbank.org/indicator>. These publicly available datasets serve as a rich resource for tracking the progression of the pandemic and for conducting rigorous statistical and epidemiological analyses aimed at understanding the disease dynamics across different regions.

Due to the varying outbreak times across different countries and regions, we define the origin of the time scale as the point at which the cumulative number of confirmed COVID-19 cases reached 100 in all countries. For this analysis, we focus on a dataset containing daily cumulative deaths from 189 countries, considering a 100-day period following this reference time. At each time point  $t$ , we estimate the density function of the mortality rate, denoted as  $\hat{d}_t(y)$ , using the observations from the 189 countries. Figure 1 (a) presents the estimated densities of the global mortality rate (%) over the 100-day interval, with data records from up to 189 countries for each time point. Figure 1 (b) offers an alternative perspective by displaying the estimated densities for three selected days. From the figures, it is evident that the mortality rate densities are well-defined across the observed period. Moreover, a temporal dependency among the distributions is clearly observable, which suggests the presence of an auto-regressive process in the data, potentially supporting the hypothesis of a FAR error structure.

The primary goal of this analysis, based on the COVID-19 data, is to identify the FAR process underlying the mortality rate and estimate its component functions. For the sake of simplicity, we begin by considering a special case in which the covariate  $z$  is constant ( $=1$ ) and  $x$  represents a time scale, denoted as  $t/T$ , namely, we consider the following model:

$$\hat{f}_t(u) = \Psi(\hat{d}_t)(u) = g_1(u, x_{t,1}) + \varepsilon_t(u), \quad 1 \leq t \leq 100,$$

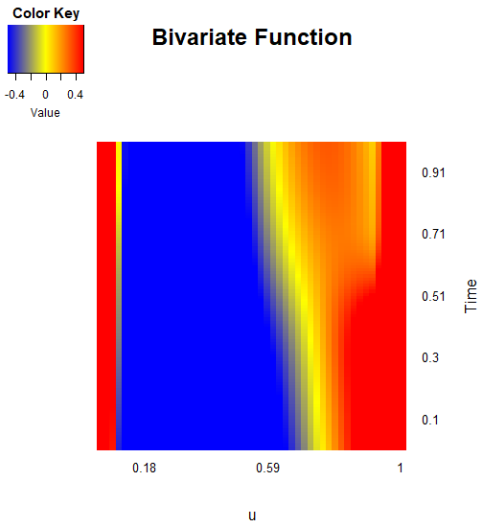
where  $\varepsilon_t(u) = \sum_{l=1}^p \int \gamma_l(u, s) \varepsilon_{t-l}(s) ds + e_t(u)$  and  $x_{t,1}$  denotes the time scale  $t/T$ .

Based on the initial spline estimation of  $g_1(u, x_1)$ , the testing algorithm is employed to determine the order of the FAR process. Table 5 presents the test results, specifically the  $p$ -values under different hypotheses. The table reveals that the observed  $p$ -values indicate significant evidence of auto-correlation in the data. This suggests that the underlying process can indeed be effectively modeled as a first-order functional auto-regressive process, denoted as  $FAR(1)$ . Such findings provide strong empirical support for the presence of temporal dependencies in the COVID-19 mortality rate, further justifying the application of a FAR error structure in modeling the dynamics of the epidemic.

**Table 5.** *P*-values of the testing algorithm for identifying the order of functional error process based on the mortality rate data of COVID-19.

Null Hypothesis	$p = 0$	$p \leq 1$
Alternative Hypothesis	$p \geq 1$	$p \geq 2$
P-value	0.000	0.194

Figure 4 presents the heat map of the estimated bivariate function  $g_1(u, x_1)$  after accounting for the functional error process and identifying the auto-regressive order. The heat map reveals a relatively stable temporal pattern, with the function initially reaching a minimum value at lower values of  $u$ , gradually increasing, and eventually reaching a maximum at later time points. This pattern highlights the underlying dynamics of the COVID-19 mortality rate across successive days. The observed correlation between the data from consecutive days supports the notion that the global mortality rate exhibits significant temporal dependencies. This is consistent with the design of the mortality rate measure, which is derived from previous daily data, reflecting the evolving trend of the pandemic on a global scale. The results further validate the presence of auto-correlation in the mortality rate data and underscore the necessity of incorporating a functional auto-regressive process to capture these dependencies effectively.



**Figure 4.** Heat map of bivariate function  $g_1(u, x_1)$  in the model based on the mortality rate (%) data of COVID-19.

6.2. USA Income Data

Personal income statistics play a crucial role in enabling governments to understand the dynamics between national income, spending, and saving, while also serving as an important tool for evaluating and comparing the economic well-being across different regions or nations. In this context, we focus on the density time series of per capita personal income, which is defined as the total personal income of an area divided by its population. This measure provides a more granular perspective on the economic conditions within a region, reflecting the distribution and trends in income on a per-person basis over time. By analyzing such time series, policymakers and researchers can gain insights into the long-term economic trajectory of a region, assess disparities in income distribution, and make informed decisions regarding fiscal policies, social welfare programs, and economic development strategies.

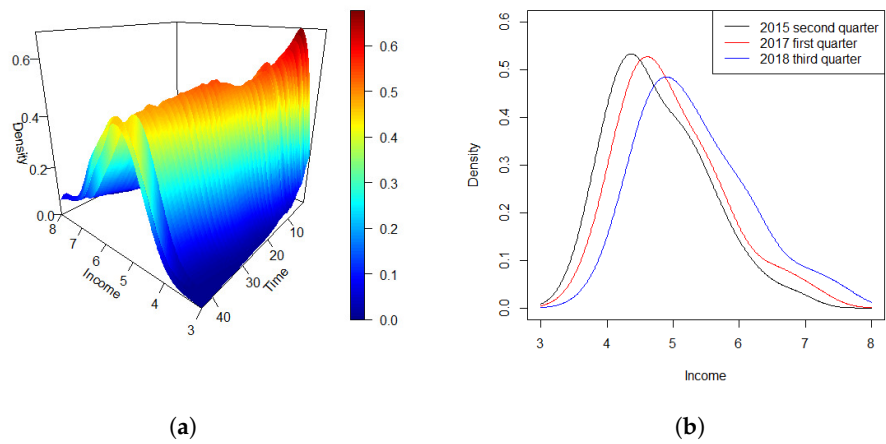
Income data for the USA are publicly available at the official website of the United States Bureau of Economic Analysis (<http://www.bea.gov/>). We consider the quarterly per capita personal income of 50 states in the USA from the first quarter of 2010 to the fourth quarter of 2020, namely,  $t = 1, \dots, 44$ . For each  $t$ , we obtain the density function of per capita income,  $\hat{d}_t(y)$ , each with 50 observations. As the quarterly personal income is an economic measure based on national conditions, we choose two related covariates, ‘GDP’ (quarterly gross domestic product of the USA) and ‘Population’ (quarterly total population of the USA), which can also be obtained from <http://www.bea.gov/>.

The income curve, traditionally studied in economics as panel data, primarily reflects the relationship between consumers’ equilibrium points. As individuals’ income levels fluctuate, the connections between these equilibrium points form a trajectory, symbolizing not only an increase in income but also a corresponding rise in consumer satisfaction. This approach emphasizes the dynamic nature of income growth and its impact on individual well-being, providing valuable insights into consumer behavior over time.

In contrast, the income density curve, treated as functional data, is used to describe the distribution of income within a specific region or demographic group. It offers a graphical representation of the characteristics and trends in income across various income intervals, providing a more holistic view of the socio-economic landscape. By analyzing the income density curve, the degree of income inequality within a population can be effectively observed, revealing important patterns of wealth distribution. This type of curve is crucial for economic research as it facilitates a deeper understanding of consumption behavior, socio-economic conditions, and the formulation of societal policies.

Moreover, income density curves are instrumental in economic forecasting and analysis. By examining trends in income distribution, economists can combine insights into consumer preferences and consumption habits at different income levels. This integration enables predictions about future economic conditions and shifts in consumption patterns, making the income density curve a key tool for anticipating changes in both microeconomic and macroeconomic contexts. In this way, the income density curve plays an essential role in informing policy decisions, economic strategies, and the broader understanding of economic well-being.

Figure 5 (a) illustrates the density time series of quarterly personal income over a period of 44 quarters. The density curves reveal that, over the past decade, the overall distribution of per capita income across various states in the United States has exhibited a consistent pattern. Specifically, there are relatively few individuals in the high-income and middle-to-high-income brackets, a moderate number in the middle-income category, and a larger proportion in the middle-to-low-income segments.



**Figure 5.** Densities of national quarterly personal income in the USA over an interval of 44 quarters. (a): Three-dimensional trend of density time series during the whole period; (b): Density curves at three different selected quarters .

To further elucidate the temporal evolution of income distribution, Figure 5 (b) presents the density curves for three distinct time points: the second quarter of 2015, the first quarter of 2017, and the third quarter of 2018. A clear trend emerges, showing a gradual shift towards higher income values over time, accompanied by a corresponding decrease in the peak of the density curve. This shift is unsurprising given the broader economic and technological advancements in modern society. As the economy progresses, the proportion of low-income individuals in the United States has been steadily declining, while the number of middle-to-high-income individuals has been rising. Consequently,

the distribution of income is becoming more balanced, with a growing proportion of the population occupying the middle to high-income brackets. This pattern reflects the broader trends of economic development and income redistribution that have taken place in recent years.

We consider the following DVCA-FAR model:

$$\hat{f}_t(u) = \Psi(\hat{d}_t)(u) = g_0(u, x_{t,0}) + z_{t,1}g_1(u, x_{t,1}) + z_{t,2}g_2(u, x_{t,2}) + \varepsilon_t(u), \; 1 \leq t \leq 44,$$

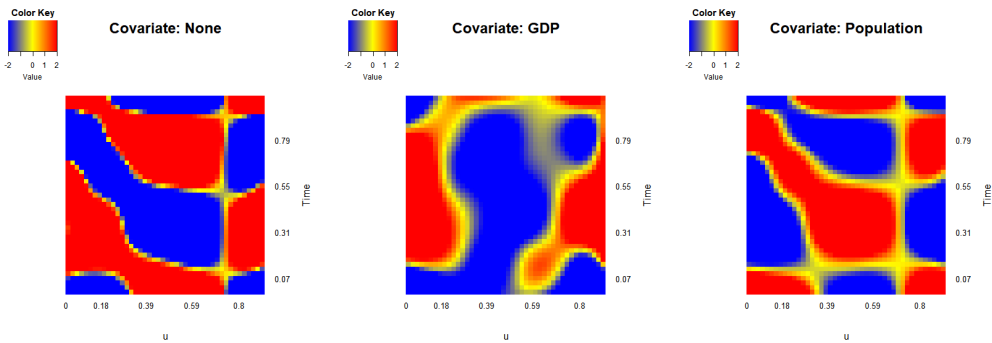
where  $\varepsilon_t(u) = \sum_{l=1}^p \int \gamma_l(u, s)\varepsilon_{t-l}(s)ds + e_t(u)$ .  $z_{t,1}$  denotes the quarterly gross domestic product of the USA,  $z_{t,2}$  denotes the quarterly total population of the USA, and  $x_{t,0}, x_{t,1}, x_{t,2}$  denote the time scale  $t/T$ .

Similar to the previous estimation procedure, the testing algorithm is applied to determine the optimal order based on the initial spline estimations. Table 6 presents the  $p$ -values of the test under various hypotheses, which indicate that the  $FAR(2)$  model is the most appropriate for modeling the error process in this context. This result suggests that a second-order functional autoregressive process effectively captures the autocorrelated structure of the error terms in the income data.

**Table 6.**  $P$ -values of the testing algorithm for identifying the order of the functional error process based on USA income data.

Null Hypothesis	$p = 0$	$p \leq 1$	$p \leq 2$
Alternative Hypothesis	$p \geq 1$	$p \geq 2$	$p \geq 3$
P-value	0.000	0.000	0.436

Using the three-step estimation procedure, estimations of the bivariate varying-coefficient functions can be obtained. Figure 6 displays the heat maps of the three bivariate functions, where  $g_0(u, x_0)$  represents the common effect of the data over time,  $g_1(u, x_1)$  and  $g_2(u, x_2)$  reflect the impact of quarterly GDP and quarterly total population on data respectively. The heat map of  $g_0$  reveals a clear pattern, where high and low values alternate over time, indicating that individuals with both higher and lower per capita income experience similar effects, in contrast to those in the middle-income bracket, who show an opposing trend. For  $g_1$ , the figure demonstrates a mode that is consistent for both small and large values of  $u$ , but varies over time. Initially, the function reaches a maximum, then decreases to a minimum before increasing again towards the end of the time scale. In contrast, the effect of population, as shown in the heat map of  $g_2$ , exhibits a trend that is opposite to that of the common effect. The population impact is relatively consistent across both higher and lower per capita income groups. Taken together, these results suggest a significant dependency of the quarterly personal income distributions in the United States on previous statistics, with the dynamics of income closely tied to both macroeconomic factors (such as GDP) and demographic factors (such as population).



**Figure 6.** Heat maps of bivariate varying-coefficient functions  $g_m(u, x_m), m = 0, 1, 2$ , in the USA income data model.

## 7. Discussion

Data collected from sequential time points often exhibit autocorrelation, which must be addressed in the modeling process. Additionally, the analysis of non-Euclidean data has become increasingly common. To tackle these challenges, we propose a varying-coefficient additive model with density responses, incorporating a Functional Auto-Regressive (FAR) error process. Due to the complexity and constraints of the data, we first map density functions into a linear space using the transformation method by [12]. We then develop a three-step estimation procedure for the varying-coefficient components. First, B-spline series approximation is used to obtain initial estimates of the bivariate varying-coefficient components without considering the functional errors. Next, the FAR error process order is determined using the test statistic from [10] based on the residuals from the initial estimation. Finally, the FAR error process is removed, and improved spline estimators are constructed for the varying-coefficient components. Theoretical results, including convergence rates and asymptotic properties, are derived for both the initial and improved estimations. The performance of the proposed method is demonstrated through simulations and two real datasets, showing the importance of accounting for autocorrelation and validating the efficiency of the proposed approach.

Future research can explore a variety of related problems. In this study, we employed the varying-coefficient additive model to establish the relationship between density function responses and scalar predictors. With the advent of large and complex datasets, functional predictors have become increasingly valuable in analyzing practical data applications. Furthermore, to account for sequence dependence, we utilized the FAR process to model the correlation structures in the data. However, more intricate models representing autocorrelation could also be explored to further refine the representation of sequence dependence in functional errors. Future work will focus on extending these approaches and conducting additional studies in these areas.

## Appendix A

In this appendix, we provide detailed proofs of the theoretical results.

**Proof of Theorem 1.** As described in the article, the proposed varying-coefficient additive model with functional error process can be written as

$$f_t(u) = \sum_{m=1}^k z_{t,m} g_m(u, x_{t,m}) + \varepsilon_t(u), \quad 0 \leq u \leq 1.$$

Denote  $\mathbf{b}(u, x_m) = (b_{1,1,m}(u, x_m), \dots, b_{N_0, N_m, m}(u, x_m))^T$ ,  $\boldsymbol{\lambda}_m = (\lambda_{1,1,m}, \dots, \lambda_{N_0, N_m, m})^T$ ,  $\tilde{\boldsymbol{\lambda}}_m = (\tilde{\lambda}_{1,1,m}, \dots, \tilde{\lambda}_{N_0, N_m, m})^T$ ,  $\boldsymbol{\lambda} = (\lambda_1, \dots, \lambda_k)^T$ , and  $\tilde{\boldsymbol{\lambda}} = (\tilde{\lambda}_1, \dots, \tilde{\lambda}_k)^T$ ,  $m = 1, \dots, k$ ,  $t = 1, \dots, T$ .

By using the spline approximation method,  $g_m(u, x_m)$  can be written in the matrix form as

$$g_m(u, x_m) \approx \sum_{r=1}^{N_0} \sum_{j=1}^{N_m} \lambda_{r,j,m} b_{r,j,m}(u, x_m) = \mathbf{b}^T(u, x_m) \boldsymbol{\lambda}_m.$$

Ignoring the FAR error term, the initial estimator of bivariate varying-coefficient functions  $g_m(u, x_m)$  can be written as

$$\tilde{g}_m(u, x_m) = \sum_{r=1}^{N_0} \sum_{j=1}^{N_m} \tilde{\lambda}_{r,j,m} b_{r,j,m}(u, x_m) = \mathbf{b}^T(u, x_m) \tilde{\boldsymbol{\lambda}}_m, \quad 1 \leq m \leq k,$$

where  $\tilde{\boldsymbol{\lambda}} = (\tilde{\lambda}_{1,1,1}, \dots, \tilde{\lambda}_{N_0, N_k, k})^T$  is a  $(N_0 \sum_{m=1}^k N_m)$ -dimensional vector defined by

$$\tilde{\boldsymbol{\lambda}} = \arg \min_{\boldsymbol{\lambda}} \sum_{t=1}^T \sum_{i=1}^n \left[ f_t(u_i) - \sum_{m=1}^k z_{t,m} \sum_{r=1}^{N_0} \sum_{j=1}^{N_m} \lambda_{r,j,m} b_{r,j,m}(u_i, x_{t,m}) \right]^2.$$

We first give some notations. Denote  $B = (B_1, \dots, B_k)$ ,  $B_m = (Bz_{1,m}^\tau, \dots, Bz_{T,m}^\tau)^\tau$ ,  $Bz_{t,m} = (b_z(u_1, x_{t,m}), \dots, b_z(u_n, x_{t,m}))^\tau$ ,  $b_z(u, x_{t,m}) = z_{t,m}b(u, x_{t,m})$ ,  $f_t = (f_t(u_1), \dots, f_t(u_n))^\tau$ ,  $f = (f_1^\tau, \dots, f_T^\tau)^\tau$ ,  $\varepsilon_t = (\varepsilon_t(u_1), \dots, \varepsilon_t(u_n))^\tau$ ,  $\varepsilon = (\varepsilon_1^\tau, \dots, \varepsilon_T^\tau)^\tau$ .

Let  $\hat{f}_t(u) = \Psi(\hat{d}_t)(u)$  be the estimate of  $f_t(u)$  based on observations  $\{Y_t\}$ . Denote  $\hat{f} = (\hat{f}_1^\tau, \dots, \hat{f}_T^\tau)^\tau$ , where  $\hat{f}_t = (\hat{f}_t(u_1), \dots, \hat{f}_t(u_n))^\tau$ . We then have  $\hat{f} = f + \varepsilon_f$ , where  $\varepsilon_f = (\varepsilon_{f_1}^\tau, \dots, \varepsilon_{f_T}^\tau)^\tau$ , and  $\varepsilon_{f_t} = \Psi(\hat{d}_t) - \Psi(d_t)$  is the error generated from the LQD transformation of the density estimation  $\hat{d}_t$ . For each  $d_t$ , we assume the error from transfer and kernel smoothing is identically and independent distributed. The estimation of  $\lambda$  is given by  $\tilde{\lambda} = (B^\tau B)^{-1} B^\tau \hat{f}$ .

For simplicity, denote  $g_z(u, x_t) = \sum_{m=1}^k z_{t,m} g_m(u, x_{t,m})$ ,  $g_t = (g_z(u_1, x_t), \dots, g_z(u_n, x_t))^\tau$ , and  $g = (g_1^\tau, \dots, g_T^\tau)^\tau$ . Let  $A_m = (0, \dots, I, \dots, 0)$  be an  $N_0 N_m \times N_0 \sum_{i=1}^k N_i$  block matrix, with each block be an  $N_0 N_m \times N_0 N_i$  matrix,  $i = 1, \dots, k$ , and the  $m$ -th position an identity matrix.

To prove the consistency of  $\tilde{g}_m(u, x_m)$ , we first decompose  $\tilde{g}_m(u, x_m) - g_m(u, x_m)$  into three parts, which is as follows.

$$\begin{aligned} & \tilde{g}_m(u, x_m) - g_m(u, x_m) \\ &= b^\tau(u, x_m) A_m (B^\tau B)^{-1} B^\tau \hat{f} - g_m(u, x_m) \\ &= b^\tau(u, x_m) A_m (B^\tau B)^{-1} B^\tau f + b^\tau(u, x_m) A_m (B^\tau B)^{-1} B^\tau \varepsilon_f - g_m(u, x_m) \\ &= b^\tau(u, x_m) A_m (B^\tau B)^{-1} B^\tau g - g_m(u, x_m) \\ & \quad + b^\tau(u, x_m) A_m (B^\tau B)^{-1} B^\tau \varepsilon + b^\tau(u, x_m) A_m (B^\tau B)^{-1} B^\tau \varepsilon_f \\ &= g^B(u, x_m) + g^V(u, x_m) + g^e(u, x_m), \end{aligned}$$

where

$$\begin{aligned} g^B(u, x_m) &= b^\tau(u, x_m) A_m (B^\tau B)^{-1} B^\tau g - g_m(u, x_m), \\ g^V(u, x_m) &= b^\tau(u, x_m) A_m (B^\tau B)^{-1} B^\tau \varepsilon, \\ g^e(u, x_m) &= b^\tau(u, x_m) A_m (B^\tau B)^{-1} B^\tau \varepsilon_f. \end{aligned}$$

For  $g^B(u, x_m)$ , we have

$$\begin{aligned} & g^B(u, x_m) \\ &= b^\tau(u, x_m) A_m (B^\tau B)^{-1} B^\tau g - g_m(u, x_m) \\ &= b^\tau(u, x_m) A_m \left[ (B^\tau B)^{-1} B^\tau g - \lambda \right] + b^\tau(u, x_m) A_m \lambda - g_m(u, x_m) \\ &= b^\tau(u, x_m) A_m (B^\tau B)^{-1} B^\tau [g - B\lambda] + \left[ b^\tau(u, x_m) \lambda_m - g_m(u, x_m) \right] \\ &= b^\tau(u, x_m) A_m \left( \frac{1}{nT} B^\tau B \right)^{-1} B^\tau \left[ \frac{1}{nT} (g - B\lambda) \right] + \left[ b^\tau(u, x_m) \lambda_m - g_m(u, x_m) \right]. \end{aligned}$$

It indicates by [16] that the order of the traditional bivariate spline estimator is  $O_p(N_m^{-2})$ , i.e., there exists a constant  $C_0$ , such that

$$\sup_{u, x_m} |g_m(u, x_m) - b(u, x_m)^\tau \lambda_m| \leq C_0 N_m^{-2}.$$

For simplicity, we assume that there exists a constant  $N_1$  such that  $N_0 = N_m = N_1$ ,  $1 \leq m \leq k$ . As a result, we obtain that  $\sup_{u, x} \frac{1}{nT} |g - B\lambda| \leq C_1 N_1^{-2}$  with a constant  $C_1$ . Combined with the result that  $\|(\frac{1}{nT} B^\tau B)^{-1}\| = O_p(N_1^2)$ , which can be derived from DeVore & Lorentz (1993), therefore, we can get that

$$\sup_{u, x_m \in [0,1]} |g^B(u, x_m)| = O_p(N_1^{-2}).$$



For  $g^V(u, x_m) = \mathbf{b}^\tau(u, x_m) \mathbf{A}_m (\mathbf{B}^\tau \mathbf{B})^{-1} \mathbf{B}^\tau \varepsilon$ ,  $\varepsilon$  is a functional auto-regressive process, defined as  $\varepsilon_t(u) = \sum_{l=1}^p \int \gamma_l(s, u) \varepsilon_{t-l}(s) ds + e_t(u)$ . Based on the assumption that  $\{e_t\}_{t=1}^T$  is independent with  $x_t, z_t$ , satisfying  $E(e_t(u)|x_t, z_t) = 0$ , and the largest eigenvalues of the covariance function  $\Sigma_\varepsilon(u)$ ,  $\lambda_{max}$ , is finite. Therefore,

$$\begin{aligned} E(g^V(u, x_m)) &= E[E(g^V(u, x_m)|x, z)] = E[E(\mathbf{b}^\tau(u, x_m) \mathbf{A}_m (\mathbf{B}^\tau \mathbf{B})^{-1} \mathbf{B}^\tau \varepsilon|x, z)] \\ &= E[\mathbf{b}^\tau(u, x_m) \mathbf{A}_m (\mathbf{B}^\tau \mathbf{B})^{-1} \mathbf{B}^\tau E(\varepsilon|x, z)] = 0 \end{aligned}$$

and

$$\begin{aligned} E(g^V(u, x_m))^2 &= E[E(g^V(u, x_m))^2|x, z] \\ &= E\left[E\left((\mathbf{b}^\tau(u, x_m) \mathbf{A}_m (\mathbf{B}^\tau \mathbf{B})^{-1} \mathbf{B}^\tau \varepsilon)^2|x, z\right)\right] \\ &= E\left[E\left(\mathbf{b}^\tau(u, x_m) \mathbf{A}_m (\mathbf{B}^\tau \mathbf{B})^{-1} \mathbf{B}^\tau \varepsilon \varepsilon^\tau \mathbf{B} (\mathbf{B}^\tau \mathbf{B})^{-1} \mathbf{A}_m^\tau \mathbf{b}(u, x_m)|x, z\right)\right] \\ &= E\left[\mathbf{b}^\tau(u, x_m) \mathbf{A}_m (\mathbf{B}^\tau \mathbf{B})^{-1} \mathbf{B}^\tau E(\varepsilon \varepsilon^\tau|x, z) \mathbf{B} (\mathbf{B}^\tau \mathbf{B})^{-1} \mathbf{A}_m^\tau \mathbf{b}(u, x_m)\right] \\ &= \mathbf{b}^\tau(u, x_m) \mathbf{A}_m (\mathbf{B}^\tau \mathbf{B})^{-1} \mathbf{B}^\tau \Sigma_\varepsilon \mathbf{B} (\mathbf{B}^\tau \mathbf{B})^{-1} \mathbf{A}_m^\tau \mathbf{b}(u, x_m) \\ &\leq \frac{C\lambda_{max}}{nT} \mathbf{b}^\tau(u, x_m) \mathbf{A}_m \left(\frac{1}{nT} \mathbf{B}^\tau \mathbf{B}\right)^{-1} \mathbf{A}_m^\tau \mathbf{b}(u, x_m) \\ &\leq \frac{C\lambda_{max}}{nT} \mathbf{b}(u, x_m) \mathbf{b}^\tau(u, x_m) \left\| \left(\frac{1}{nT} \mathbf{B}^\tau \mathbf{B}\right)^{-1} \right\| \end{aligned}$$

Subsequently, combining the above result, we have

$$\sup_{u, x_m \in [0,1]} |g^V(u, x_m)| = O_p(\sqrt{N_1/nT}) = O_p\left(N_1^{1/2}/\sqrt{nT}\right).$$

For  $g^e(u, x_m) = \mathbf{b}^\tau(u, x_m) \mathbf{A}_m (\mathbf{B}^\tau \mathbf{B})^{-1} \mathbf{B}^\tau \varepsilon_f$ , note that  $f_t(u) = \Psi(d_t)(u)$  and the estimation obtained from observations is  $\hat{f}_t(u) = \Psi(\hat{d}_t(u))$ . From Petersen & Müller (2016), it can be obtained that  $\sup_{d_t \in \mathcal{F}} |\hat{d}_t - d_t| = O_p(h + (nh)^{-1/2})$ . Since  $\varepsilon_{f_t} = \hat{f}_t - f_t$  is the error from the transformation of  $\hat{d}_t(\cdot)$  and  $d_t(\cdot)$ , the consistency of LQD transformation indicates that  $\sup_{d_t} |\Psi(\hat{d}_t) - \Psi(d_t)| = O_p(h + (nh)^{-1/2})$ . Then, under the assumption of smoothness condition, we can get that

$$\sup_{u, x_m \in [0,1]} |g^e(u, x_m)| = O_p(h + (nh)^{-1/2}).$$

Therefore, as  $n, T \rightarrow \infty$ ,  $h \sim n^{-1/3}$ ,  $N_0, N_m \sim (nT)^{1/6} \log nT$ , thus  $N_1 \sim (nT)^{1/6} \log nT$ , it is easy to have

$$\begin{aligned} &\sup_{u, x_m \in [0,1]} |\tilde{g}_m(u, x_m) - g_m(u, x_m)| \\ &\leq \sup_{u, x_m \in [0,1]} |g^B(u, x_m)| + \sup_{u, x_m \in [0,1]} |g^V(u, x_m)| + \sup_{u, x_m \in [0,1]} |g^e(u, x_m)| \\ &= O_p(N_1^{-2}) + O_p(N_1/\sqrt{nT}) + O_p(h + (nh)^{-1/2}) \\ &= O_p\left((nT)^{-1/3}(\log nT) + n^{-1/3}\right). \end{aligned}$$

The proof of the theorem is completed.  $\square$

**Proof of Theorem 2.** The improved spline approximation of error process  $\varepsilon_t(u)$  is given by

$$\hat{\varepsilon}_t(u) = \sum_{l=1}^p \sum_{r=1}^N \sum_{j=1}^N \hat{\mu}_{r,j,l} \int b_{r,j}(u,s) \hat{\varepsilon}_{t-l}(s) ds, \quad 0 \leq u \leq 1,$$

where  $\hat{\mu} = (\hat{\mu}_{1,1,1}, \dots, \hat{\mu}_{N,N,p})^\tau$  is a  $pN^2$ -dimensional vector, minimizing the following problem, i.e.,

$$\hat{\mu} = \arg \min_{\mu} \sum_{t=p+1}^T \sum_{i=1}^n \left[ \hat{\varepsilon}_t(u_i) - \sum_{l=1}^p \sum_{r=1}^N \sum_{j=1}^N \mu_{r,j,l} \int b_{r,j}(u_i,s) \hat{\varepsilon}_{t-l}(s) ds \right]^2.$$

Let  $f_t^c(u) = f_t(u) - \varepsilon_t(u)$ , and  $\hat{f}_t^c(u) = f_t(u) - \hat{\varepsilon}_t(u)$ , with the spline algorithm again, the improved estimate of  $g_m(u, x_m)$  is given by

$$\hat{g}_m(u, x_m) = \sum_{r=1}^{N_0} \sum_{j=1}^{N_m} \hat{\lambda}_{r,j,m} b_{r,j,m}(u, x_m), \quad 1 \leq m \leq k,$$

where  $\hat{\lambda} = (\hat{\lambda}_{1,1,1}, \dots, \hat{\lambda}_{N_0, N_k, k})^T$  is a  $(N_0 \sum_{m=1}^k N_m)$ -dimensional vector satisfying

$$\hat{\lambda} = \arg \min_{\lambda} \sum_{t=1}^T \sum_{i=1}^n \left[ \hat{f}_t^c(u_i) - \sum_{m=1}^k z_{t,m} \sum_{r=1}^{N_0} \sum_{j=1}^{N_m} \lambda_{r,j,m} b_{r,j,m}(u_i, x_{t,m}) \right]^2.$$

Denote  $f_t^c = (f_t^c(u_1), \dots, f_t^c(u_n))^\tau$  and  $f^c = (f_1^c, \dots, f_T^c)^\tau$ ,  $\hat{f}_t^c = (\hat{f}_t^c(u_1), \dots, \hat{f}_t^c(u_n))^\tau$  and  $\hat{f}^c = (\hat{f}_1^c, \dots, \hat{f}_T^c)^\tau$ , meanwhile  $e_t = (e_t(u_1), \dots, e_t(u_n))^\tau$  and  $e = (e_1, \dots, e_T)^\tau$ .

Based on the estimation of  $\varepsilon$ , the estimation of  $\lambda$  is given by  $\hat{\lambda} = (B^\tau B)^{-1} B^\tau \hat{f}^c$ , where  $\hat{f}^c = f - \hat{\varepsilon}$ . Since random error exists in the density estimation process, namely,  $\hat{f} = f + \varepsilon_f$ , where  $\varepsilon_f$  is defined similarly as  $\varepsilon$ , therefore, denote  $\tilde{f}^c = \hat{f} - \hat{\varepsilon}$ , the estimate of  $\lambda$  based on the observations is given by  $\hat{\lambda} = (B^\tau B)^{-1} B^\tau \tilde{f}^c$ .

Similar as the proof of Theorem 1, we decompose  $\hat{g}_m(u, x_m) - g_m(u, x_m)$  as

$$\begin{aligned} & \hat{g}_m(u, x_m) - g_m(u, x_m) \\ &= b^\tau(u, x_m) A_m (B^\tau B)^{-1} B^\tau \tilde{f}^c - g_m(u, x_m) \\ &= b^\tau(u, x_m) A_m (B^\tau B)^{-1} B^\tau (f - \hat{\varepsilon}) + b^\tau(u, x_m) A_m (B^\tau B)^{-1} B^\tau \varepsilon_f - g_m(u, x_m) \\ &= b^\tau(u, x_m) A_m (B^\tau B)^{-1} B^\tau g - g_m(u, x_m) \\ & \quad + b^\tau(u, x_m) A_m (B^\tau B)^{-1} B^\tau (\varepsilon - \hat{\varepsilon}) + b^\tau(u, x_m) A_m (B^\tau B)^{-1} B^\tau \varepsilon_f \\ &= g^B(u, x_m) + g^V(u, x_m) + g^e(u, x_m), \end{aligned}$$

where

$$\begin{aligned} g^B(u, x_m) &= b^\tau(u, x_m) A_m (B^\tau B)^{-1} B^\tau g - g_m(u, x_m), \\ g^V(u, x_m) &= b^\tau(u, x_m) A_m (B^\tau B)^{-1} B^\tau (\varepsilon - \hat{\varepsilon}), \\ g^e(u, x_m) &= b^\tau(u, x_m) A_m (B^\tau B)^{-1} B^\tau \varepsilon_f. \end{aligned}$$

For  $g^B(u, x_m)$ ,

$$\begin{aligned} g^B(u, x_m) &= b^\tau(u, x_m) A_m (B^\tau B)^{-1} B^\tau g - g_m(u, x_m) \\ &= b^\tau(u, x_m) A_m (B^\tau B)^{-1} B^\tau [g - B\lambda] + \left[ b^\tau(u, x_m) \lambda_m - g_m(u, x_m) \right] \\ &= b^\tau(u, x_m) A_m \left( \frac{1}{nT} B^\tau B \right)^{-1} B^\tau \left[ \frac{1}{nT} (g - B\lambda) \right] + \left[ b^\tau(u, x_m) \lambda_m - g_m(u, x_m) \right] \end{aligned}$$

Similar to the discussion in the proof of Theorem 1, we can have

$$\sup_{u, x_m \in [0,1]} |g^B(u, x_m)| = O_p(N_1^{-2}).$$

For  $g^V(u, x_m) = \mathbf{b}^\tau(u, x_m) \mathbf{A}_m (\mathbf{B}^\tau \mathbf{B})^{-1} \mathbf{B}^\tau (\boldsymbol{\varepsilon} - \hat{\boldsymbol{\varepsilon}})$ , the functional error process can be written as

$$\varepsilon_t(u) = \sum_{l=1}^p \int \gamma_l(u, s) \varepsilon_{t-l}(s) ds + e_t(u) = \sum_{l=1}^p \sum_{r=1}^N \sum_{j=1}^N \mu_{r,j,l} \int b_{r,j}(u, s) \varepsilon_{t-l}(s) ds + e_t(u),$$

with the corresponding approximation

$$\hat{\varepsilon}_t(u) = \sum_{l=1}^p \sum_{r=1}^N \sum_{j=1}^N \hat{\mu}_{r,j,l} \int b_{r,j}(u, s) \tilde{\varepsilon}_{t-l}(s) ds, \quad 0 \leq u \leq 1.$$

Denote  $\mathbf{b}_t(u) = (\int b_{1,1}(u, s) \tilde{\varepsilon}_{t-1}(s) ds, \dots, \int b_{N,N}(u, s) \tilde{\varepsilon}_{t-p}(s) ds)^\tau$ ,  $\mathbf{b}_t = (\mathbf{b}_t(u_1), \dots, \mathbf{b}_t(u_n))^\tau$ ,  $\mathbf{B}_\varepsilon = (\mathbf{b}_p^\tau, \dots, \mathbf{b}_T^\tau)^\tau$ . Denote  $\boldsymbol{\varepsilon}_t = (\varepsilon_t(u_1), \dots, \varepsilon_t(u_n))^\tau$ ,  $\boldsymbol{\varepsilon} = (\boldsymbol{\varepsilon}_{p+1}^\tau, \dots, \boldsymbol{\varepsilon}_T^\tau)^\tau$ .

The model can be rewritten in the form of matrix as  $\boldsymbol{\varepsilon} \approx \mathbf{B}_\varepsilon \boldsymbol{\mu} + \mathbf{e}$ . Based on the initial estimation, the model is  $\tilde{\boldsymbol{\varepsilon}} \approx \mathbf{B}_\varepsilon \boldsymbol{\mu} + \tilde{\mathbf{e}}$ , where  $\tilde{\mathbf{e}}$  is defined similarly as  $\boldsymbol{\varepsilon}$ , and the estimation of  $\boldsymbol{\mu}$  is given by  $\hat{\boldsymbol{\mu}} = (\mathbf{B}_\varepsilon^\tau \mathbf{B}_\varepsilon)^{-1} \mathbf{B}_\varepsilon^\tau \tilde{\boldsymbol{\varepsilon}}$ .

Then,

$$\begin{aligned} \hat{\varepsilon}_t(u) - \varepsilon_t(u) &= \sum_{l=1}^p \sum_{r=1}^N \sum_{j=1}^N \hat{\mu}_{r,j,l} \int b_{r,j}(u, s) \tilde{\varepsilon}_{t-l}(s) ds - \varepsilon_t(u) \\ &= \mathbf{b}_t^\tau(u) (\mathbf{B}_\varepsilon^\tau \mathbf{B}_\varepsilon)^{-1} \mathbf{B}_\varepsilon^\tau \tilde{\boldsymbol{\varepsilon}} - \varepsilon_t(u) \\ &= \mathbf{b}_t^\tau(u) (\mathbf{B}_\varepsilon^\tau \mathbf{B}_\varepsilon)^{-1} \mathbf{B}_\varepsilon^\tau (\tilde{\boldsymbol{\varepsilon}} - \mathbf{B}_\varepsilon \boldsymbol{\mu}) + (\mathbf{b}_t^\tau(u) \boldsymbol{\mu} - \varepsilon_t(u)) \end{aligned}$$

Since there exist a constant  $C_t$ , such that  $\sup_u |\varepsilon_t(u) - \mathbf{b}_t^\tau(u) \boldsymbol{\mu}| \leq C_t N^{-2}$ , and similarly,  $\sup_u |\tilde{\boldsymbol{\varepsilon}} - \mathbf{B}_\varepsilon \boldsymbol{\mu}| \leq C_p N^{-2}$  with another constant  $C_p$ . Then, under the assumption that sample size is infinite, it is clear that  $\sup_u |\tilde{\boldsymbol{\varepsilon}} - \boldsymbol{\varepsilon}| = O_p(N^{-2})$ . Therefore,

$$\sup_{u, x_m \in [0,1]} |g^V(u, x_m)| = \sup_{u, x_m \in [0,1]} |\mathbf{b}^\tau(u, x_m) \mathbf{A}_m (\mathbf{B}^\tau \mathbf{B})^{-1} \mathbf{B}^\tau (\boldsymbol{\varepsilon} - \hat{\boldsymbol{\varepsilon}})| = O_p(N^{-2}).$$

For  $g^e(u, x_m) = \mathbf{b}^\tau(u, x_m) \mathbf{A}_m (\mathbf{B}^\tau \mathbf{B})^{-1} \mathbf{B}^\tau \boldsymbol{\varepsilon}_f$ , under the smoothness condition of  $f_t$ , combining with the proof in Theorem 1, we can get that

$$\sup_{u, x_m \in [0,1]} |g^e(u, x_m)| = O_p(h + (nh)^{-1/2}).$$

Therefore, as  $n, T \rightarrow \infty$ ,  $h \sim n^{-1/3}$ ,  $N_0, N_m \sim (nT)^{1/6} \log nT$ , namely,  $N_1 \sim (nT)^{1/6} \log nT$ ,  $N \sim (nT)^{1/6} \log nT$ , it is easy to have

$$\begin{aligned} &\sup_{u, x_m \in [0,1]} |\hat{g}_m(u, x_m) - g_m(u, x_m)| \\ &\leq \sup_{u, x_m \in [0,1]} |g^B(u, x_m)| + \sup_{u, x_m \in [0,1]} |g^V(u, x_m)| + \sup_{u, x_m \in [0,1]} |g^e(u, x_m)| \\ &= O_p(N_1^{-2}) + O_p(N^{-2}) + O_p(N_1^{-2} + h + (nh)^{-1/2}) \\ &= O_p\left((nT)^{-1/3} (\log nT)^{-2} + n^{-1/3}\right). \end{aligned}$$

The proof of the theorem is completed.  $\square$

**Proof of Theorem 3.** (i) We first prove the asymptotic normality of the initial estimation  $\tilde{g}_m(u, x_m)$ . With (1)-(4), we rewrite  $\sqrt{nT}(\tilde{g}_m(u, x_m) - g_m(u, x_m))$  as

$$\begin{aligned} & \sqrt{nT}(\tilde{g}_m(u, x_m) - g_m(u, x_m)) \\ &= \sqrt{nT}[\mathbf{b}^\tau(u, x_m)\mathbf{A}_m(\mathbf{B}^\tau\mathbf{B})^{-1}\mathbf{B}^\tau\mathbf{g} - g_m(u, x_m)] \\ & \quad + \sqrt{nT}[\mathbf{b}^\tau(u, x_m)\mathbf{A}_m(\mathbf{B}^\tau\mathbf{B})^{-1}\mathbf{B}^\tau\boldsymbol{\varepsilon} + \mathbf{b}^\tau(u, x_m)\mathbf{A}_m(\mathbf{B}^\tau\mathbf{B})^{-1}\mathbf{B}^\tau\boldsymbol{\varepsilon}_f] \end{aligned}$$

Since the error process  $\boldsymbol{\varepsilon}_t$  is independent with the covariates  $\mathbf{x}_t, \mathbf{z}_t$ , we have  $E(\boldsymbol{\varepsilon}_t|\mathbf{x}_t, \mathbf{z}_t) = 0$ . Combined with the result of the Theorem 1, it is easy to have that

$$\begin{aligned} & E[\sqrt{nT}(\tilde{g}_m(u, x_m) - g_m(u, x_m))] \\ &= \sqrt{nT}E[\mathbf{b}^\tau(u, x_m)\mathbf{A}_m(\mathbf{B}^\tau\mathbf{B})^{-1}\mathbf{B}^\tau(\mathbf{g} - \mathbf{B}\boldsymbol{\lambda})] + \sqrt{nT}E[\mathbf{b}^\tau(u, x_m)\boldsymbol{\lambda}_m - g_m(u, x_m)] \\ & \quad + \sqrt{nT}E[\mathbf{b}^\tau(u, x_m)\mathbf{A}_m(\mathbf{B}^\tau\mathbf{B})^{-1}\mathbf{B}^\tau E(\boldsymbol{\varepsilon}|\mathbf{x}, \mathbf{z})] \\ & \quad + \sqrt{nT}E[\mathbf{b}^\tau(u, x_m)\mathbf{A}_m(\mathbf{B}^\tau\mathbf{B})^{-1}\mathbf{B}^\tau E(\boldsymbol{\varepsilon}_f|\mathbf{x}, \mathbf{z})] \\ &= 0 \end{aligned}$$

Meanwhile,

$$\begin{aligned} & \text{Var}[\sqrt{nT}(\tilde{g}_m(u, x_m) - g_m(u, x_m))] \\ &= nT\text{Var}[\mathbf{b}^\tau(u, x_m)\mathbf{A}_m(\mathbf{B}^\tau\mathbf{B})^{-1}\mathbf{B}^\tau\boldsymbol{\varepsilon} + \mathbf{b}^\tau(u, x_m)\mathbf{A}_m(\mathbf{B}^\tau\mathbf{B})^{-1}\mathbf{B}^\tau\boldsymbol{\varepsilon}_f] \end{aligned}$$

Based on the results proved in Theorem 1, we can get that

$$\begin{aligned} & nT\text{Var}[\mathbf{b}^\tau(u, x_m)\mathbf{A}_m(\mathbf{B}^\tau\mathbf{B})^{-1}\mathbf{B}^\tau\boldsymbol{\varepsilon}] \\ &= nTE\left[E\left(\mathbf{b}^\tau(u, x_m)\mathbf{A}_m(\mathbf{B}^\tau\mathbf{B})^{-1}\mathbf{B}^\tau\boldsymbol{\varepsilon}|\mathbf{x}, \mathbf{z}\right)^2\right] \\ &= nTE\left[E\left(\mathbf{b}^\tau(u, x_m)\mathbf{A}_m(\mathbf{B}^\tau\mathbf{B})^{-1}\mathbf{B}^\tau\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}^\tau\mathbf{B}(\mathbf{B}^\tau\mathbf{B})^{-1}\mathbf{A}_m^\tau\mathbf{b}(u, x_m)|\mathbf{x}, \mathbf{z}\right)\right] \end{aligned}$$

Denote  $\mathbf{D}_m = \mathbf{A}_m(\mathbf{B}_*^\tau\mathbf{B}_*)^{-1}\mathbf{B}_*^\tau$ , where  $\mathbf{B}_* = \mathbf{B}/\sqrt{nT}$ , then

$$nT\text{Var}[\mathbf{b}^\tau(u, x_m)\mathbf{A}_m(\mathbf{B}^\tau\mathbf{B})^{-1}\mathbf{B}^\tau\boldsymbol{\varepsilon}] = E\left[\mathbf{b}^\tau(u, x_m)E\left(\mathbf{D}_m\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}^\tau\mathbf{D}_m^\tau|\mathbf{x}, \mathbf{z}\right)\mathbf{b}^\tau(u, x_m)\right].$$

Denote  $\boldsymbol{\Sigma}_\varepsilon = E(\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}^\tau|\mathbf{x}, \mathbf{z})$ , since the error process has auto-correlation, the covariance matrix can be decomposed into two parts, namely,  $\boldsymbol{\Sigma}_\varepsilon = \boldsymbol{\Sigma}_1 + \boldsymbol{\Sigma}_2$ , where  $\boldsymbol{\Sigma}_1 = \text{diag}(\boldsymbol{\Sigma}_{t,t})_{1 \leq t \leq T}$  is a diagonal matrix with the  $t$ -th diagonal element as  $\boldsymbol{\Sigma}_{t,t} = \text{Cov}(\boldsymbol{\varepsilon}_t)$ , and  $\boldsymbol{\Sigma}_2$  contains the off-diagonal elements as  $\boldsymbol{\Sigma}_2 = (\boldsymbol{\Sigma}_{i,j})_{1 \leq t \neq s \leq T} = (\text{Cov}(\boldsymbol{\varepsilon}_t, \boldsymbol{\varepsilon}_s))_{1 \leq t \neq s \leq T}$ , representing the dependence between the auto-regressive errors.

For another,

$$\begin{aligned} & nT\text{Var}[\mathbf{b}^\tau(u, x_m)\mathbf{A}_m(\mathbf{B}^\tau\mathbf{B})^{-1}\mathbf{B}^\tau\boldsymbol{\varepsilon}_f] \\ &= nTE\left[E\left(\mathbf{b}^\tau(u, x_m)\mathbf{A}_m(\mathbf{B}^\tau\mathbf{B})^{-1}\mathbf{B}^\tau\boldsymbol{\varepsilon}_f\boldsymbol{\varepsilon}_f^\tau\mathbf{B}(\mathbf{B}^\tau\mathbf{B})^{-1}\mathbf{A}_m^\tau\mathbf{b}(u, x_m)|\mathbf{x}, \mathbf{z}\right)\right] \\ &= nT\sigma_\varepsilon^2E\left[E\left(\mathbf{b}^\tau(u, x_m)\mathbf{A}_m(\mathbf{B}^\tau\mathbf{B})^{-1}\mathbf{A}_m^\tau\mathbf{b}(u, x_m)|\mathbf{x}, \mathbf{z}\right)\right], \end{aligned}$$

under the smoothness assumption, when sample size  $n, T \rightarrow \infty$ , the error generated from the density estimation tends to zero, thus the variance of this part tends to zero.

Therefore, combined with the proof that the property of the second moment in the Theorem 1 satisfies the Linderberg-Feller central limit theorem, then under the assumption that  $n \gg T \rightarrow \infty$ ,

$$\sqrt{nT}(\mathbf{C}_m \boldsymbol{\Sigma}_\varepsilon \mathbf{C}_m^\tau)^{-1}(\hat{g}_m(u, x_m) - g_m(u, x_m)) \xrightarrow{D} N(0, 1),$$

where  $\mathbf{C}_m = \mathbf{b}^\tau(u, x_m)E(\mathbf{D}_m) = \mathbf{b}^\tau(u, x_m)E(\mathbf{A}_m(\mathbf{B}_*^\tau \mathbf{B}_*)^{-1} \mathbf{B}_*^\tau)$ ,  $\boldsymbol{\Sigma}_\varepsilon = (\boldsymbol{\Sigma}_{t,s})_{1 \leq t,s \leq T}$  satisfying  $\boldsymbol{\Sigma}_{t,s} = \text{Cov}(\varepsilon_t, \varepsilon_s)$ .

(ii) After obtaining the spline estimation of error process as  $\hat{\varepsilon}$ , the improved estimation of bivariate varying-coefficient functions  $g_m(u, x_m)$  is estimated from the refined model as  $f_t^c(u) = f_t(u) - \varepsilon_t(u) = \sum_{m=1}^k z_{t,m} g_m(u, x_{t,m}) + e_t(u)$ .

As proved in the Theorem 3,

$$\begin{aligned} & \sqrt{nT}(\hat{g}_m(u, x_m) - g_m(u, x_m)) \\ &= \sqrt{nT}[\mathbf{b}^\tau(u, x_m) \mathbf{A}_m(\mathbf{B}^\tau \mathbf{B})^{-1} \mathbf{B}^\tau \tilde{f}^c - g_m(u, x_m)] \\ &= \sqrt{nT}[\mathbf{b}^\tau(u, x_m) \mathbf{A}_m(\mathbf{B}^\tau \mathbf{B})^{-1} \mathbf{B}^\tau (\mathbf{f} - \hat{\varepsilon}) + \mathbf{b}^\tau(u, x_m) \mathbf{A}_m(\mathbf{B}^\tau \mathbf{B})^{-1} \mathbf{B}^\tau \varepsilon_f - g_m(u, x_m)] \\ &= \sqrt{nT}[\mathbf{b}^\tau(u, x_m) \mathbf{A}_m(\mathbf{B}^\tau \mathbf{B})^{-1} \mathbf{B}^\tau \mathbf{g} - g_m(u, x_m)] \\ & \quad + \sqrt{nT}[\mathbf{b}^\tau(u, x_m) \mathbf{A}_m(\mathbf{B}^\tau \mathbf{B})^{-1} \mathbf{B}^\tau (\varepsilon - \hat{\varepsilon}) + \mathbf{b}^\tau(u, x_m) \mathbf{A}_m(\mathbf{B}^\tau \mathbf{B})^{-1} \mathbf{B}^\tau \varepsilon_f] \end{aligned}$$

Since  $\hat{\varepsilon}$  is the estimation of the error process  $\varepsilon$ , then based on the convergence results obtained in Theorem 3, it is clear to get that  $E(\sqrt{nT}[\hat{g}_m(u, x_m) - g_m(u, x_m)]) = 0$ .

Meanwhile,

$$\begin{aligned} & \text{Var}[\sqrt{nT}(\hat{g}_m(u, x_m) - g_m(u, x_m))] \\ &= nT \text{Var}[\mathbf{b}^\tau(u, x_m) \mathbf{A}_m(\mathbf{B}^\tau \mathbf{B})^{-1} \mathbf{B}^\tau (\mathbf{g} + \varepsilon - \hat{\varepsilon} - \varepsilon_f) - g_m(u, x_m)] \\ &= nT \text{Var}[\mathbf{b}^\tau(u, x_m) \mathbf{A}_m(\mathbf{B}^\tau \mathbf{B})^{-1} \mathbf{B}^\tau (\varepsilon - \hat{\varepsilon}) + \mathbf{b}^\tau(u, x_m) \mathbf{A}_m(\mathbf{B}^\tau \mathbf{B})^{-1} \mathbf{B}^\tau \varepsilon_f] \end{aligned}$$

Based on the results proved in Theorem 3 and the same notes, we can get that

$$\begin{aligned} & nT \text{Var}[\mathbf{b}^\tau(u, x_m) \mathbf{A}_m(\mathbf{B}^\tau \mathbf{B})^{-1} \mathbf{B}^\tau (\varepsilon - \hat{\varepsilon})] \\ &= nTE \left[ E \left( \mathbf{b}^\tau(u, x_m) \mathbf{A}_m(\mathbf{B}^\tau \mathbf{B})^{-1} \mathbf{B}^\tau (\varepsilon - \hat{\varepsilon}) | \mathbf{x}, \mathbf{z} \right)^2 \right] \\ &= nTE \left[ E \left( \mathbf{b}^\tau(u, x_m) \mathbf{A}_m(\mathbf{B}^\tau \mathbf{B})^{-1} \mathbf{B}^\tau (\varepsilon - \hat{\varepsilon}) (\varepsilon - \hat{\varepsilon})^\tau \mathbf{B} (\mathbf{B}^\tau \mathbf{B})^{-1} \mathbf{A}_m^\tau \mathbf{b}(u, x_m) | \mathbf{x}, \mathbf{z} \right) \right] \\ &= E \left[ \mathbf{b}^\tau(u, x_m) \mathbf{D}_m E((\varepsilon - \hat{\varepsilon})(\varepsilon - \hat{\varepsilon})^\tau | \mathbf{x}, \mathbf{z}) \mathbf{D}_m^\tau \mathbf{b}(u, x_m) \right] \end{aligned}$$

Denote  $\boldsymbol{\Xi}_\varepsilon = E((\varepsilon - \hat{\varepsilon})(\varepsilon - \hat{\varepsilon})^\tau | \mathbf{x}, \mathbf{z})$ , similarly, it can also be decomposed as two parts as  $\boldsymbol{\Xi}_\varepsilon = \boldsymbol{\Xi}_1 + \boldsymbol{\Xi}_2$ , where  $\boldsymbol{\Xi}_1 = \text{diag}(\boldsymbol{\Xi}_{t,t})_{1 \leq t \leq T}$  is a diagonal matrix with the  $t$ -th diagonal element as  $\boldsymbol{\Xi}_{t,t} = \text{Cov}(\varepsilon_t - \hat{\varepsilon}_t)$ , and  $\boldsymbol{\Xi}_2$  contains the off-diagonal elements as  $\boldsymbol{\Xi}_2 = (\boldsymbol{\Xi}_{t,s})_{1 \leq t \neq s \leq T} = (\text{Cov}(\varepsilon_t - \hat{\varepsilon}_t, \varepsilon_s - \hat{\varepsilon}_s))_{1 \leq t \neq s \leq T}$ .

Due to the convergence of  $\varepsilon - \hat{\varepsilon}$  proved before, the variance matrix becomes  $\boldsymbol{\Xi}_\varepsilon = \boldsymbol{\Xi}_1 = \text{Cov}(\mathbf{e}_t)$  as  $n, T \rightarrow \infty$ . Since  $\text{Cov}(e_t(u), e_t(s)) = \sigma_t^2(u, s)$ , then, the variance matrix can be written as  $\boldsymbol{\Xi}_\varepsilon = \text{Cov}(\mathbf{e}_t) = \text{diag}(\boldsymbol{\Xi}_{t,t})_{1 \leq t \leq T}$ , where  $\boldsymbol{\Xi}_{t,t}(u, s) = \sigma_t^2(u, s)$ .

Therefore,

$$\sqrt{nT}(\mathbf{C}_m \boldsymbol{\Xi}_\varepsilon \mathbf{C}_m^\tau)^{-1}(\hat{g}_m(u, x_m) - g_m(u, x_m)) \xrightarrow{D} N(0, 1),$$

where  $\mathbf{C}_m = \mathbf{b}^\tau(u, x_m)E(\mathbf{A}_m(\mathbf{B}_*^\tau \mathbf{B}_*)^{-1} \mathbf{B}_*^\tau)$ ,  $\boldsymbol{\Xi}_\varepsilon = \text{diag}(\boldsymbol{\Xi}_{t,t})_{1 \leq t \leq T}$  with  $\boldsymbol{\Xi}_{t,t}(u, s) = \sigma_t^2(u, s)$ .

The proof of the theorem is completed.  $\square$

**Author Contributions:** Conceptualization, Zixuan Han, TAO LI and Jinhong You; Data curation, Zixuan Han; Formal analysis, Zixuan Han; Funding acquisition, TAO LI and Jinhong You; Investigation, Zixuan Han; Methodology, Zixuan Han, TAO LI and Jinhong You; Project administration, TAO LI, Jinhong You and Narayanaswamy Balakrishnan; Resources, Zixuan Han; Software, Zixuan Han; Supervision, TAO LI, Jinhong You and Narayanaswamy Balakrishnan; Validation, Zixuan Han; Writing – original draft, Zixuan Han; Writing – review & editing, Zixuan Han, TAO LI, Jinhong You and Narayanaswamy Balakrishnan. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was funded by Dr. Li and Dr. You. Dr. Li's research is supported by grants from Humanities and Social Science Fund of Ministry of Education of China (No. 21YJA910001). Dr. You's research is supported by grants from the National Natural Science Foundation of China (NSFC) (No.11971291).

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** The original datasets employed in this study are publicly accessible from the official website of Johns Hopkins University at <https://www.jhu.edu/> and the World Bank's online platform at <https://data.worldbank.org/>.

**Acknowledgments:** Not applicable.

**Conflicts of Interest:** The authors declare no conflicts of interest.

## References

1. Berhoune, K.; Bensmain, N.. Sieves estimator of functional autoregressive process. *Statistics and Probability Letters* **2018**, *135*, 60–69.
2. Bosq, D.. *Linear processes in function spaces: theory and applications*. Springer Science & Business Media, New York, 2000.
3. Chen, Y.; Chua, W. S.; Hardle, W.. Forecasting limit order book liquidity supply-demand curves with functional autoregressive dynamics. *Quantitative Finance* **2019**, *19*(9), 1473–1489.
4. Chen, Y.; Li, B.. An adaptive functional autoregressive forecast model to predict electricity price curves. *Journal of Business and Economic Statistics* **2017**, *35*(3), 371–388.
5. Chen, Y.; Lin, Z.; Müller, H.. Wasserstein regression. *Journal of the American Statistical Association* **2023**, *118*(542), 869–882.
6. Daniel, R.; David, S.; David, R.. Functional autoregression for sparsely sampled data. *Journal of Business and Economic Statistics* **2019**, *37*(1), 97–109.
7. DeVore, R.; Lorentz, G.. *Constructive approximation*, Springer Science & Business media, 1993; volume 303.
8. Han, K.; Müller, H.; Park, B.. Additive functional regression for densities as responses. *Journal of the American Statistical Association* **2020**, *115*(530), 997–1010.
9. Kokoszka, P.; Miao, H.; Petersen, A.; Shang, H. L.. Forecasting of density functions with an application to cross-sectional and intraday returns. *International Journal of Forecasting* **2019**, *35*, 1304–1317.
10. Kokoszka, P.; Reimherr, M.. Determining the order of the functional autoregressive model. *Journal of Time Series Analysis* **2013**, *34*, 116–129.
11. Petersen, A.; Chen, C.; Müller, H.. Quantifying and visualizing intraregional connectivity in resting-state functional magnetic resonance imaging with correlation densities. *Brain Connectivity* **2019**, *9*(1), 37–47.
12. Petersen, A.; Müller, H.. Functional data analysis for density functions by transformation to a Hilbert space. *The Annals of Statistics* **2016**, *44*(1), 183–218.
13. Petersen, A.; Müller, H.. Fréchet regression for random objects with Euclidean predictors. *The Annals of Statistics* **2019**, *47*(2), 691–719.
14. Saha, A.; Banerjee, S.; Kurttek, S.; Narang, S.; Lee, J.; Rao, G.; Martinez, J.; Bharath, K.; Rao, A.; Baladandayuthapani, V.. DEMARCAT: Density-based magnetic resonance image clustering for assessing tumor heterogeneity in cancer. *NeuroImage: Clinical* **2016**, *12*, 132–143.
15. Sen, R.; Ma, C.. Forecasting density function: Application in finance. *Journal of Mathematical Finance* **2015**, *5*(5), 433–447.
16. Stone, C.. The use of polynomial splines and their tensor products in multivariate function estimation. *The Annals of Statistics* **1994**, *22*(1), 118–171.



17. Talská, R.; Menafoglio, A.; Machalová, J.; Hron, K.; Fiserová, E.. Compositional regression with functional response. *Computational Statistics & Data Analysis* **2018**, *123*(1), 66–85.
18. Xu, X.; Chen, Y.; Zhang, G.; Koch, T.. Modeling functional time series and mixed-type predictors with partially functional autoregressions. *Journal of Business and Economic Statistics* **2022**, *0*, 1–18.
19. Zhang, C.; Kokoszka, P.; Petersen, A.. Wasserstein autoregressive models for density time series. *Journal of Time Series Analysis* **2022**, *43*, 30–52.

**Disclaimer/Publisher's Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.