



# Convolutional autoregressive models for functional time series



Xialu Liu<sup>a</sup>, Han Xiao<sup>b</sup>, Rong Chen<sup>b,\*</sup>

<sup>a</sup> Management Information Systems Department, San Diego State University, San Diego, CA 92182, United States

<sup>b</sup> Department of Statistics, Rutgers University, Piscataway, NJ 08854, United States

## ARTICLE INFO

### Article history:

Available online 9 June 2016

### Keywords:

Functional time series  
Nonparametric methods  
Spline methods  
Sieve estimation

## ABSTRACT

Functional data analysis has become an increasingly popular class of problems in statistical research. However, functional data observed over time with serial dependence remains a less studied area. Motivated by Bosq (2000), who first introduced the functional autoregressive models, we propose a convolutional functional autoregressive model, where the function at time  $t$  is a result of the sum of convolutions of the past functions and a set of convolution functions, plus a noise process, mimicking the vector autoregressive process. It provides an intuitive and direct interpretation of the dynamics of a stochastic process. Instead of principal component analysis commonly used in functional data analysis, we adopt a sieve estimation procedure based on B-spline approximation of the convolution functions. We establish convergence rate of the proposed estimator, and investigate its theoretical properties. The model building, model validation, and prediction procedures are also developed. Both simulated and real data examples are presented.

© 2016 Elsevier B.V. All rights reserved.

## 1. Introduction

Functional data analysis has received much attention over the last few decades, and has been widely applied in many fields, including medical science (Houghton et al., 1980; Gasser et al., 1984; Ratcliffe et al., 2002a,b), behavioral science (Kesselman and Kesselman, 1993), and economics (Roberts, 1995; Diebold and Li, 2006). Nonparametric methods, such as spline methods (Silverman, 1984; Brumback and Rice, 1998; Zhou et al., 1998; Cai et al., 2000) and kernel smoothing (Nadaraya, 1964; Watson, 1964; Fan and Gijbels, 1996), were often implemented to analyze functional data. Unsupervised learning methods, such as principal component analysis (James et al., 2000) and clustering analysis (James and Sugar, 2003) were extended for functional data as well. Books by Ramsay and Silverman (2005), Ferraty and Vieu (2006), and Horváth and Kokoszka (2012) provide comprehensive introductions on various aspects of functional data analysis.

Often, a variety of functional data is observed over time and has serial dependence. For example, in financial industry, the implied volatility of an option as a function of moneyness changes over time. In insurance industry, age-specific mortality rate as a function of age changes over time. In banking industry, term

structure of interest rates (yield as a function of time to maturity of a bond) changes over time. In meteorology, daily records of temperature, precipitation and cloud cover for a region, viewed as three related functional surfaces, change over time.

Time series analysis, designed to explore the underlying dynamics of data, is well studied and understood, with modern development in nonlinear (Tong and Lim, 1980; Chan, 1993), nonparametric (Chen and Tsay, 1993a,b; Härdle et al., 1997; Xia and Li, 1999; Cai et al., 2000; Fan and Yao, 2003), multivariate (Tiao and Tsay, 1983, 1989; Lütkepohl, 2005) and spatial-temporal modeling (Handcock and Wallis, 1994; Cressie and Huang, 1999; Gneiting, 2002). Functional data with serial dependence poses new challenges, and requires new methodology in time series analysis.

Bosq (2000) first introduced functional autoregressive (FAR) models of order  $p$ ,

$$X_t = \Delta_1 X_{t-1} + \cdots + \Delta_p X_{t-p} + \varepsilon_t,$$

where  $X = (X_t, t \in \mathbb{Z})$  and  $\varepsilon = (\varepsilon_t, t \in \mathbb{Z})$  are a sequence of random functions and a functional white noise process, respectively, and  $\Delta_i$  is a linear operator in Hilbert functional space  $\mathbf{H}$ . Only under some special cases, these linear operators can be estimated by performing functional principal component analysis on the sample autocovariance operators. The consistency of such estimators has been proved (Bosq, 2000; Hörmann et al., 2013). All the theoretical and empirical results in the literature were developed based on the models and methods in

\* Corresponding author.

E-mail addresses: [xialu.liu@mail.sdsu.edu](mailto:xialu.liu@mail.sdsu.edu) (X. Liu), [hxiao@stat.rutgers.edu](mailto:hxiao@stat.rutgers.edu) (H. Xiao), [rongchen@stat.rutgers.edu](mailto:rongchen@stat.rutgers.edu) (R. Chen).

Bosq (2000), including Hörmann and Kokoszka (2010), Horváth et al. (2010), Aue et al. (2012), Horváth and Kokoszka (2012), Horváth et al. (2012), Berkes et al. (2013), and Hörmann et al. (2013).

In this article, we develop a new class of functional time series models called the convolutional functional autoregressive (CFAR) models, along with its associated estimation procedure using splines and sieve methods. As a special case of the FAR model, our model provides an intuitive and direct interpretation of the dynamics of a stochastic process. It assumes that the function at time  $t$  is a result of the sum of convolutions of the past functions and convolution functions plus a noise process, mimicking the autoregressive process commonly used in scalar time series. It is also an extension of the vector autoregressive process. For the FAR model, Bosq (2000) proposed a Yule–Walker type estimator of the autocorrelation operator, obtained using functional principal component analysis. In contrast, our method fully exploits the advantage of the convolution structure and the assumption that the impact of the past on the present is smooth. Both simulated and real examples show that the sieve estimator outperforms in estimation and prediction. The paper makes contributions to the literature in three aspects. First, we propose a new class of functional time series model, and introduce the sieve estimation of the autoregressive operators. Second, we establish the central limit theorems and convergence rates for the convolution function estimators. For the FAR model, Bosq (2000) only considered consistency, and Mas (2002) obtained a partial result on the weak convergence of the autoregressive operator. Third, we develop model building and model validation procedures for CFAR models, while the study of FAR models is less complete due to lack of specific model assumptions.

The rest of the paper is organized as follows. In Sections 2 and 3, the CFAR model and the associated statistical inference procedures are introduced. The asymptotic theories are developed in Section 4. Simulation results are presented in Section 5 and a real example is analyzed in Section 6. All proofs are collected in the Appendix.

We first introduce some notations. For a vector  $\mu$ ,  $(\mu)_i$  denotes its  $i$ th entry. For a matrix  $\mathbf{A}$ ,  $(\mathbf{A})_{ij}$  denotes its  $(i, j)$ th entry. Without loss of generality, we only consider time series on the function space  $L_2([0, 1], \mathcal{B}_{[0,1]}, \lambda)$ , abbreviated as  $L_2[0, 1]$ , where  $\mathcal{B}_{[0,1]}$  is the Borel  $\sigma$ -field, and  $\lambda$  is the Lebesgue measure. For a function  $f \in L_2[0, 1]$ ,  $\|f\| := \|f\|_2$  denotes its  $L_2$  norm. If  $f$  is also continuous, we use  $\|f\|_\infty$  to denote its maximum norm. We consider the following classes of smooth functions:

$$\text{Lip}^h[-1, 1] = \{f \in [-1, 1] : |f(x + \delta) - f(x)| \leq M\delta^h,$$

$$M < \infty, 0 < h \leq 1\},$$

$$\text{Lip}_2^r[-1, 1] = \{f \in C^r[-1, 1] : f^{(r)} \in \text{Lip}^h[-1, 1],$$

$$\zeta > 1, r = \lfloor \zeta \rfloor, h = \zeta - r\},$$

If  $f \in \text{Lip}_2^\zeta[-1, 1]$ , then  $\zeta$  is called moduli of smoothness of  $f(\cdot)$ .

## 2. Convolutional functional autoregressive models

A sequence of random functions  $X = (X_t, t \in \mathbb{Z})$  in  $L_2[0, 1]$  is called a convolutional functional autoregressive model of order  $p$ , denoted by CFAR( $p$ ), if

$$X_t(s) = \sum_{i=1}^p \int_0^1 \phi_i(s-u)X_{t-i}(u) du + \varepsilon_t(s), \quad s \in [0, 1], \quad (1)$$

where  $\phi_i \in L_2[-1, 1]$  for  $i = 1, \dots, p$ , are called convolution functions, and  $\varepsilon_t$  are i.i.d. Ornstein–Uhlenbeck (O–U) processes defined on  $[0, 1]$ , following the stochastic differential equation,  $d\varepsilon_t(s) = -\rho\varepsilon_t(s)ds + \sigma dW_s$ ,  $\rho > 0$ , and  $W_s$  being a Wiener process.

**Remark 1.** Following Bosq (2000), a natural generalization of vector autoregressive process of order  $p$  on the function space is

$$X_t(s) = \sum_{i=1}^p \int_0^1 \phi_i(s, u)X_{t-i}(u) du + \varepsilon_t(s), \quad s \in [0, 1].$$

From a pointwise view, the function at time  $t$  and point  $s$  is a weighted sum of  $p$  past functions plus noise. Our CFAR( $p$ ) model can be viewed as a special case when  $\psi(s, u) = \phi(s - u)$ , where  $\phi(\cdot)$  is a smooth function. The autoregressive operator of our model thus has the Toeplitz structure. Under this model, the conditional mean of  $X_t(s)$  is obtained as a kernel type average of  $X_{t-1}(\cdot)$  around the same argument  $s$ . In functional data analysis, it is often the case that  $X_t(s)$  has a stronger relationship with  $X_{t-1}(u)$  for the  $u$  close to  $s$ , than those  $u$  far away from  $s$ . Our model is able to exploit this type of dependence among the data. The real data example on volatility smiles shows that our model has better prediction performance, as compared with the more general model.

**Remark 2.** For functional data, it is common to assume that it is continuous for both practical and technical reasons; see Ramsay and Silverman (2005). For this reason, we choose the noise process with continuous sample paths. To reduce model complexity, we also require spatial dependence to be stationary. Due to these considerations, we assume that the error processes  $\varepsilon_t$  follow the O–U process, which is Gaussian with the following covariance structure:

$$\varepsilon_t(s_1) \sim N\left(0, \frac{\sigma^2}{2\rho}\right), \quad \text{Corr}(\varepsilon_t(s_1), \varepsilon_t(s_2)) = e^{-\rho|s_1-s_2|},$$

$$\forall s_1, s_2 \in [0, 1].$$

To account for spatial heteroscedasticity, we can include a variance function of the noise process in the model.

$$X_t(s) = \sum_{i=1}^p \int_0^1 \phi_i(s-u)X_{t-i}(u) du + w(s)\varepsilon_t(s), \quad s \in [0, 1], \quad (2)$$

where  $w(s)$  is a heteroscedasticity function for the noise process. We note that O–U process is just one of the many possible choices here. Other noise process, including various Gaussian processes, can be used here, though we require a parametric family for our estimation procedure. Our asymptotic results are derived under O–U process but can be extended to other noise processes.

If all the convolution functions  $\{\phi_i(\cdot), i = 1, \dots, p\}$  are continuous,  $X_t$  is also continuous, but not differentiable. The skeleton of  $X_t(\cdot)$ , excluding the noise process, defined as

$$f_t(s) = \sum_{i=1}^p \int_0^1 \phi_i(s-u)X_{t-i}(u) du$$

is differentiable.

In model (1), convolution functions  $\{\phi_i(\cdot), i = 1, \dots, p\}$  allow various sample paths of the  $X_t(\cdot)$  process, Fig. 1 shows two simulated examples. The top panel uses  $\phi(s) = 1, s \in [-1, 1]$ , and  $X_0(\cdot) = 0$  and the bottom one uses  $\phi(s) = I(s > 0), s \in [-1, 1]$ , and  $X_0(\cdot) = 10$ . Both use  $\rho = 5, \sigma^2 = 10$ . The solid lines and dashed lines are  $X_t(\cdot)$  and  $f_t(\cdot)$ , respectively,  $t = 1, 2, 3$  and 100. In the top panel, since  $\phi(\cdot)$  is a constant function,  $f_t(\cdot)$  is simply the average of  $X_t(\cdot)$  hence a constant function. In the bottom panel,  $\phi(\cdot)$  is an indicator function on  $(0, 1]$ , so the skeleton of  $f_t(s)$  would be a partial integration of  $X_{t-1}(\cdot)$  on the left of  $s$  in  $[0, s]$ . At  $s = 0$ ,  $X_t(0)$  contains no information of  $X_{t-1}$ , but only noise; as  $s$  increases, the weight  $\phi(s - \cdot)$  increases and information carried by  $X_t(s)$  on  $X_{t-1}(\cdot)$  increases as well; at  $s = 1$ ,  $X_t(1)$  is the integration of the function  $X_{t-1}(\cdot)$  in the entire range of  $[0, 1]$  plus noise. It is worth noting that the process in the top panel is nonstationary. We start

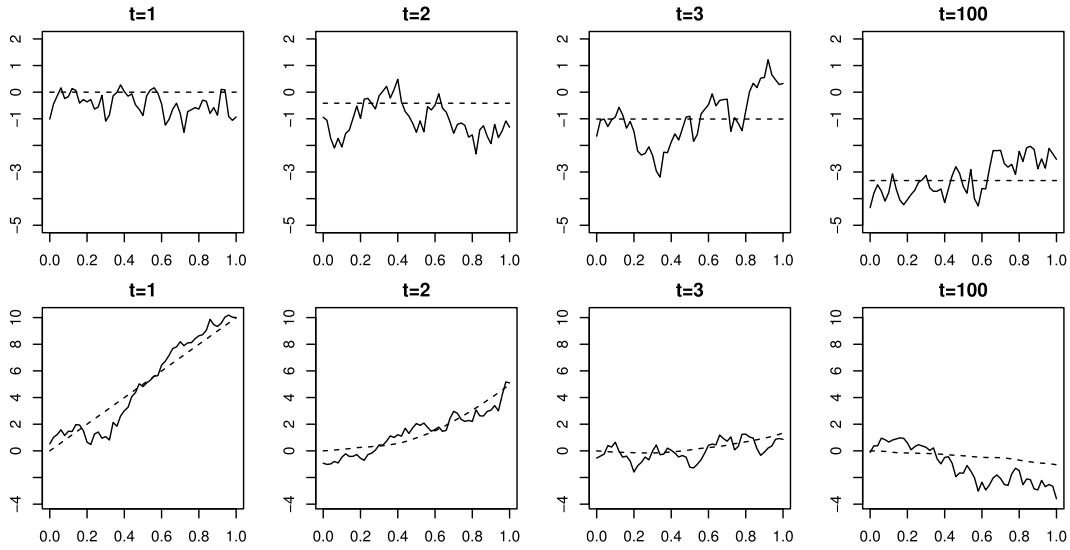


Fig. 1. Plots of  $X_t(\cdot)$  and  $f_t(\cdot)$ ,  $t = 1, 2, 3, 100$  under the two models in Section 2.

at  $X_0 = 0$ , but  $X_t(\cdot)$  becomes explosive as time increases. The process shown in the bottom panel is stationary. Although we start at a large value,  $X_0(s) = 10$  for case (ii), the process becomes close to 0 when  $t = 100$ .

**Definition 1.** A sequence of random functions  $X = (X_t, t \in \mathbb{Z})$  in  $L_2[0, 1]$  is **(weakly) stationary**, if the mean and covariance functions do not vary with time, i.e., for  $\forall h \in \mathbb{Z}, s_1, s_2 \in [0, 1]$ ,

$$E(X_t(s_1)) = \mu(s_1), \quad \text{and}$$

$$\text{Cov}(X_{t+h}(s_1), X_{m+h}(s_2)) = \text{Cov}(X_t(s_1), X_m(s_2)).$$

**Theorem 1** presents a sufficient condition for the stationarity of CFAR(1) models.

**Theorem 1.** The CFAR(1) process  $X = (X_t, t \in \mathbb{Z})$ , defined in (1) is (weakly) stationary, if

$$\kappa = \sup_{0 \leq s \leq 1} \left( \int_0^1 \phi^2(s-u) du \right)^{1/2} < 1. \tag{3}$$

Here weak stationarity is actually equivalent to strong stationarity, since we assume that the noise process is Gaussian. It is easy to see that  $\|\phi\|_2^2 < 1$  is also a sufficient condition for stationarity. However, the condition in **Theorem 1** is a weaker one, since for any  $s, \int_0^1 \phi^2(s-u) du \leq \int_{-1}^1 \phi^2(u) du$ . Note that  $\int_0^1 \phi^2(s-u) du$  is the sum of squares of the convolution weight of  $X_{t-1}(\cdot)$  to obtain  $f_t(s)$ .

The preceding condition can be generalized to provide a sufficient condition for the stationarity of CFAR( $p$ ) models.

**Theorem 2.** The CFAR( $p$ ) process  $X = (X_t, t \in \mathbb{Z})$  defined in (1) is (weakly) stationary, if all the roots of the characteristic function

$$1 - \kappa_1 z - \kappa_2 z^2 - \dots - \kappa_p z^p = 0, \tag{4}$$

are outside the unit circle, where  $\kappa_i = \sup_{0 \leq s \leq 1} \sqrt{\int_0^1 \phi_i^2(s-u) du}$  for  $i = 1, \dots, p$ .

Condition in **Theorem 2** is similar to the sufficient condition for scalar AR( $p$ ) models to be stationary, replacing the AR coefficients by the maximum of the norms of the weight functions  $\phi_i(s - \cdot)$ ,  $s \in [0, 1]$ .

**Corollary 1.** Assume the CFAR(1) process  $X = (X_t, t \in \mathbb{Z})$  satisfies (3). Define a sequence of functions on  $[0, 1]^2$ :

$$\Psi_1(s, u) = \phi(s-u), \quad \Psi_\ell(s, u) = \int_0^1 \Psi_{\ell-1}(s, v) \phi(v-u) dv,$$

for  $\ell \geq 2$ . (5)

Then  $X_t$  has the following functional MA ( $\infty$ ) representation

$$X_t(s) = \varepsilon_t(s) + \sum_{\ell=1}^{\infty} \int_0^1 \Psi_\ell(s, u) \varepsilon_{t-\ell}(u) du. \tag{6}$$

**Corollary 2.** The CFAR( $p$ ) process  $X = (X_t, t \in \mathbb{Z})$  satisfies (4). Define a sequence of functions on  $[0, 1]^2$ :

$$\Psi_1(s, u) = \phi_1(s-u),$$

$$\Psi_\ell(s, u) = \sum_{i=1}^{\ell-1} \int_0^1 \phi_i(s-v) \Psi_{\ell-i}(v, u) dv + \phi_\ell(s-u),$$

$2 \leq \ell \leq p,$

$$\Psi_\ell(s, u) = \sum_{i=1}^p \int_0^1 \phi_i(s, v) \Psi_{\ell-i}(v, u) dv, \quad \ell > p.$$

Then  $X_t$  has the following representation

$$X_t(s) = \varepsilon_t(s) + \sum_{\ell=1}^{\infty} \int_0^1 \Psi_\ell(s, u) \varepsilon_{t-\ell}(u) du. \tag{7}$$

**Corollaries 1 and 2** are derived from **Theorem 3.1** and **Theorem 5.1** of **Bosq (2000)** directly. This is similar to that of the stationary scalar AR( $p$ ) case when there is no intercept in the model. To include nonzero mean function  $\mu(\cdot)$ , we can use  $X_t(s) - \mu(s) = \sum_{i=1}^p \int_0^1 \phi_i(s-u) (X_{t-i}(u) - \mu(u)) du$ .

### 3. Estimation, prediction and order determination

For simplicity, we assume that each  $X_t(\cdot)$  is observed at discrete points,  $s = n/N, n = 0, \dots, N$ , for time  $t = 1, \dots, T$ .

#### 3.1. Estimation

Since the convolution operator guarantees continuous path of  $X_t(\cdot)$ , we approximate it by linearly interpolating the observations.

For any  $s \in [-1, 1]$ , if  $s_{n-1} \leq s < s_n$ , let

$$\tilde{X}_t(s) = \frac{(s_n - s)X_t(s_{n-1}) + (s - s_{n-1})X_t(s_n)}{1/N}. \tag{8}$$

Linear interpolation is used to approximate  $\int \phi_i(s - u)X_{t-i}(u)du$  since  $X_t(\cdot)$  is continuous but not differentiable.

We approximate the unknown convolution functions  $\phi_i(\cdot)$  using B-splines. Specifically,

$$\phi_i(\cdot) \approx \tilde{\phi}_{k,i}(\cdot) = \sum_{j=1}^k \tilde{\beta}_{k,i,j} B_{k,j}(\cdot), \quad \text{for } i = 1, \dots, p, \tag{9}$$

where  $\{B_{k,j}(\cdot), j = 1, \dots, k\}$  are uniform cubic B-spline basis functions with  $k$  degrees of freedom.

Plugging in the linear interpolation of  $X_{t-i}(\cdot)$  and B-spline approximation of  $\phi_i(\cdot)$  into (1), we have

$$\begin{aligned} \tilde{\varepsilon}_{t,n} &:= \varepsilon_t(s_n) \approx X_t(s_n) \\ &- \sum_{i=1}^p \sum_{j=1}^k \tilde{\beta}_{k,i,j} \int_0^1 B_{k,j}(s_n - u) \tilde{X}_{t-i}(u) du. \end{aligned} \tag{10}$$

Let  $\tilde{\varepsilon}_t = (\tilde{\varepsilon}_{t,0}, \dots, \tilde{\varepsilon}_{t,N})'$ . We can rewrite (10) as  $\tilde{\varepsilon}_t = X_t(\mathbf{s}) - \mathbf{M}_t \tilde{\beta}_k$ , where  $\mathbf{M}_t = (\mathbf{M}_{t,1}, \dots, \mathbf{M}_{t,p})$ ,  $\mathbf{M}_{t,i}$  is an  $(N + 1) \times k$  matrix with entries  $(\mathbf{M}_{t,i})_{nj} = \int_0^1 B_{k,j}(s_n - u) \tilde{X}_{t-i}(u) du$ , and  $\tilde{\beta}_k$  is a  $(pk) \times 1$  vector with entries  $\tilde{\beta}_k = (\tilde{\beta}'_{k,1}, \dots, \tilde{\beta}'_{k,p})'$ , and  $(\beta_{k,i})_j = \tilde{\beta}_{k,i,j}$ . Since  $B_{k,j}(\cdot)$  are fixed and known functions,  $\mathbf{M}_t$  is known, given the observations. Under the O-U process, for equally spaced  $s_n$ ,  $\{\varepsilon_t(s_n), 0 \leq n \leq N\}$  follows an AR(1) process with AR coefficient  $e^{-\rho/N}$  and covariance matrix  $\Sigma$ , where  $(\Sigma)_{ij} = e^{-\rho|i-j|/N} \sigma^2/2\rho$ . Therefore,  $\beta = \{\beta_{ij}, i = 1, \dots, p, j = 1, \dots, k\}$ ,  $\sigma^2$ , and  $\rho$  can be estimated by maximizing the approximated log-likelihood function,

$$\begin{aligned} Q_{k,T,N}(\beta, \sigma^2, \rho) &= -\frac{(N + 1)(T - p)}{2} \ln\left(\frac{\pi \sigma^2}{\rho}\right) \\ &- \frac{N(T - p)}{2} \ln(1 - e^{-2\rho/N}) - \frac{1}{2} \sum_{t=p+1}^T \mathbf{e}'_{\beta,t} \Sigma^{-1} \mathbf{e}_{\beta,t}, \end{aligned} \tag{11}$$

where  $\mathbf{e}_{\beta,t} = X_t(\mathbf{s}) - \mathbf{M}_t \beta$ . Let  $\hat{\beta}_k = (\hat{\beta}'_{k,1}, \dots, \hat{\beta}'_{k,p})'$  be the estimator of  $\tilde{\beta}_k$  by maximizing  $Q_{k,T,N}(\beta, \sigma^2, \rho)$ , where  $(\hat{\beta}_{k,i})_j = \hat{\beta}_{k,i,j}$ .

After reparameterization, the objective function can be written as

$$\begin{aligned} Q_{k,T,N}(\beta, \varphi, \omega) &= -\frac{(N + 1)(T - p)}{2} \ln(2\pi\omega) \\ &- \frac{N(T - p)}{2} \ln(1 - \varphi^2) - \frac{e(\beta, \varphi)}{2\omega}, \end{aligned} \tag{12}$$

where  $\varphi = e^{-\rho/N}$ ,  $\omega = \sigma^2/2\rho$ ,  $\Sigma_0 = 2\rho\Sigma/\sigma^2$ ,  $e(\beta, \varphi) = \sum_{t=p+1}^T \mathbf{e}'_{\beta,t} \Sigma_0^{-1} \mathbf{e}_{\beta,t}$ , and  $\Sigma_0$  is the correlation matrix of  $\{\varepsilon_t(s_n), n = 0, \dots, N\}$ .

Given  $\beta$  and  $\varphi$ , the maximizer of  $w$  is

$$\hat{\omega} = \frac{e(\beta, \varphi)}{(N + 1)(T - p)}. \tag{13}$$

Hence

$$\begin{aligned} \max_{\beta, \varphi, \omega} Q_{k,T,N}(\beta, \varphi, \omega) &= \max_{\beta, \varphi} \left( -\frac{(N + 1)(T - p)}{2} \ln e(\beta, \varphi) \right. \\ &\quad \left. - \frac{N(T - p)}{2} \ln(1 - \varphi^2) + c \right) \\ &= \max_{\varphi} \left( -\frac{(N + 1)(T - p)}{2} \ln e(\beta(\varphi), \varphi) \right. \\ &\quad \left. - \frac{N(T - p)}{2} \ln(1 - \varphi^2) + c \right), \end{aligned} \tag{14}$$

where  $c$  is a constant and

$$\begin{aligned} \hat{\beta}(\varphi) &= \arg \min_{\beta} e(\beta, \varphi) \\ &= \left( \sum_{t=p+1}^T \mathbf{M}'_t \Sigma^{-1} \mathbf{M}_t \right)^{-1} \left( \sum_{t=p+1}^T \mathbf{M}'_t \Sigma^{-1} X_t(\mathbf{s}) \right). \end{aligned} \tag{15}$$

Then this problem becomes a one-parameter optimization problem, and the estimator of  $\varphi$  can be easily obtained by maximizing (14). Together with (13) and (15), we have the estimators for  $\sigma^2$ ,  $\rho$ ,  $\beta$ . Consequently, the convolution function  $\phi_i(\cdot)$  can be estimated by  $\hat{\phi}_{k,i}(\cdot) = \sum_{j=1}^k \hat{\beta}_{k,i,j} B_{k,j}(\cdot)$ .

**Remark 3.** With the linear interpolation, our method can deal with irregularly spaced  $\{s_n, n = 0, \dots, N\}$  as well. It can also be extended to cover the situation that at different times,  $X_t$  are observed at different spatial locations, as long as certain uniformity conditions are satisfied. For ease of theoretical analysis, we assume equally spaced  $\{s_n = n/N\}$  at every time  $t$ .

**Remark 4.** Comparing with the method proposed by Bosq (2000) that combines Yule-Walker estimation and functional principal component analysis, our method exploits the assumption that the convolution functions are smooth, and takes advantages of the convolution structure. Both simulated and empirical examples show that our method provides more accuracy predictions when model is correctly specified.

### 3.2. Fitted values

Given estimated  $\phi_i(\cdot)$ , the continuous process  $X_t(\cdot)$  can be interpolated more precisely than the simple linear interpolation. By taking advantage of information from the observed  $\{X_t(s_n), n = 0, \dots, N\}$  and the Markovian property of the O-U process, we obtain a better approximation of  $X_t(\cdot)$ . Specifically, for a fixed  $s \in (s_n, s_{n+1})$ , let

$$\tilde{f}_t(s) = \sum_{i=1}^p \int_0^1 \hat{\phi}_{k,i}(s - u) \tilde{X}_{t-i}(u) ds, \tag{16}$$

and let

$$\begin{aligned} \tilde{X}_t^*(s) &= \tilde{f}_t(s) + (e^{-\hat{\rho}(s-s_n)} e^{-\hat{\rho}(s_{n+1}-s)}) \hat{\Sigma}_0^{-1} \\ &\quad \times \begin{pmatrix} X_t(s_n) - \tilde{f}_t(s_n) \\ X_t(s_{n+1}) - \tilde{f}_t(s_{n+1}) \end{pmatrix}, \end{aligned} \tag{17}$$

where  $\hat{\Sigma}_0$  is the estimated correlation matrix of  $\varepsilon_t(s_n)$  and  $\varepsilon_t(s_{n+1})$ , with diagonal entry 1 and off-diagonal entry  $e^{-\hat{\rho}/N}$ . When  $s = s_n$ ,  $\tilde{X}_t^*(s) = X_t(s)$ . Here  $\tilde{f}_t(\cdot)$  is the estimated skeleton of  $X_t(\cdot)$  process, and  $X_t(s_n) - \tilde{f}_t(s_n)$  is the approximated residual process.

**Remark 5.** Plugging (17) into (16), we can fit  $\tilde{f}_{t+1}(\cdot)$ , and  $\tilde{X}_{t+1}^*(s)$  iteratively. However, empirical results show that its impact is minor with large  $N$ .

### 3.3. Prediction

Given  $X_1, \dots, X_t$ , the least squares prediction of  $X_{t+1}(s)$  is

$$\widehat{X}_{t+1}(s) = \sum_{i=1}^p \int_0^1 \widehat{\phi}_{k,i}(s-u) \widetilde{X}_{t+1-i}^*(u) du, \quad (18)$$

where  $\widetilde{X}_{t+1-i}^*(\cdot)$  is the fitted processes of  $X_{t+1-i}(\cdot)$  in (17). The residual is defined as

$$\widehat{\varepsilon}_{t+1}(s) = X_{t+1}(s) - \widehat{X}_{t+1}(s). \quad (19)$$

In addition, if  $X_{t+1}(\cdot)$  is partially observed at  $s = 0, \frac{1}{N}, \dots, s^*$ , the prediction of  $X_{t+1}(s)$ , for  $s \in (s^*, 1]$  can be obtained by

$$\widehat{X}_{t+1}^*(s) = \widehat{X}_{t+1}(s) + e^{-\widehat{\rho}(s-s^*)} (X_{t+1}(s^*) - \widehat{X}_{t+1}(s^*)), \quad (20)$$

where  $\widehat{X}_{t+1}(s)$  is from (18).

### 3.4. Determination of the B-spline approximation order

As in all nonparametric estimation, smoothness choice is crucial in balancing the estimation bias and variance (Ruppert et al., 1995; Fan and Gijbels, 1996). For spline methods, the tuning parameter is the number of degrees of freedom  $k$  (Zhou et al., 1998; Huang, 2003). We propose to choose  $k$  to minimize the out-sample rolling forecasting error instead of the typically used cross validation criterion due to the time series nature of our problem. Specifically, for a given  $k$ , and each  $t = T_0, \dots, T$ , we use data  $\{X_h(s_n), h = 0, \dots, t-1, n = 0, \dots, N\}$ , observed before time  $t$  to estimate the convolution functions using B-spline with  $k$  degrees of freedom, and obtain the prediction of  $X_t(s_n)$ ,  $\widehat{X}_{k,t}(s_n)$  as in (18). Define an overall squared rolling forecasting error

$$S(k) = \sum_{t=T_0}^T \sum_{n=0}^N (X_t(s_n) - \widehat{X}_{k,t}(s_n))^2. \quad (21)$$

The optimal  $k$  is chosen to be the one that minimizes  $S(k)$ .

### 3.5. CFAR order determination

F-test can be constructed for hypothesis testing of the significance of the convolution functions as well as for CFAR order determination. For testing  $H_0 : \phi_{r+1}(\cdot) = \dots = \phi_p(\cdot) = 0$  vs  $H_1 : \text{not } H_0$ , we reject  $H_0$  if

$$F = \frac{(SSE^{(r)} - SSE^{(f)})/[k(p-r)]}{SSE^{(f)}/[(N+1)(T-p)-pk]} > F_{\alpha, k(p-r), (N+1)(T-p)-pk}, \quad (22)$$

where  $SSE^{(f)}$  and  $SSE^{(r)}$  are sum of squared estimated residuals of the approximated model in (10) for full CFAR( $p$ ) and reduced CFAR( $r$ ) model respectively, for an optimally chosen  $k$ . Specifically,

$$SSE^{(f)} = \sum_{t=p+1}^T \widehat{\varepsilon}_{t,1}' \widehat{\Sigma}_0^{-1} \widehat{\varepsilon}_{t,1}, \quad SSE^{(r)} = \sum_{t=p+1}^T \widehat{\varepsilon}_{t,2}' \widehat{\Sigma}_0^{-1} \widehat{\varepsilon}_{t,2},$$

where  $\widehat{\Sigma}_0$  is the estimate of  $\Sigma_0$ , the correlation matrix of  $\varepsilon_t(s)$ , whose  $(i, j)$ th entry is  $e^{-\rho|i-j|/N}$ , from full and reduced models, respectively;  $\{\widehat{\varepsilon}_{t,i}(s_n), t = p+1, \dots, T, n = 0, \dots, N\}$ ,  $i = 1, 2$  are the residuals of the full and reduced models, respectively. This test can be used for model specification. We begin with the CFAR(1) model, and sequentially add more lags of  $X_t$ , until the newly introduced lag is not significant.

In addition, define the cross-sectional residuals of  $\varepsilon_t$ ,

$$e_t(s_n) = \varepsilon_t(s_n) - e^{-\widehat{\rho}/N} \varepsilon_t(s_{n-1}). \quad (23)$$

Under our model,  $\{e_t(s_n), n = 1, \dots, N\}$  is a white noise process, for  $t = p+1, \dots, T$ . Let  $\widehat{e}_t(s_n) = \widehat{\varepsilon}_t(s_n) - e^{-\widehat{\rho}/N} \widehat{\varepsilon}_t(s_{n-1})$ . Hence, the features of  $\{\widehat{e}_t(s_n), n = 1, \dots, N\}$ , can be studied with standard residual analysis for time series model validation.

## 4. Theoretical properties

We study the asymptotic properties of our estimator as both  $N$  and  $T$  go to infinity. Then we discuss how the bias and asymptotic variance of the point estimate of the convolution function  $\phi_t(s)$  would decrease and increase respectively, as the number of spline basis functions  $k$  approaches infinity.

### 4.1. Sieve framework

In this section we formulate our method under the sieve estimation framework.

The sieve method is designed for estimation of a parameter in an infinite-dimensional space  $\Theta$ . Often optimizing the objective function over the parameter space cannot be directly solved. Instead, we optimize the function over a sequence of subspaces  $\{\Theta_k, k \in \mathbb{Z}^+\}$ , which are called sieve spaces. If the sieve spaces satisfy certain conditions, the sequence of estimators, called sieve estimators, is expected to be consistent, as the complexity of sieve spaces goes to infinity with the sample size, see Chen (2008) and Halberstam and Richert (2013).

The parameter space  $\Theta$  for CFAR( $p$ ) models contains all the functions in  $L_2[-1, 1]$  that satisfy the stationary condition specified in Theorem 2. The sieve space,  $\Theta_k$ , used to approximate the parameter space  $\Theta$ , is defined as  $\Theta_k = \mathbf{S}_k^p \cap \Theta$ , where  $\mathbf{S}_k^p$  is the product of  $p$  copies of  $\mathbf{S}_k$ ,

$$\mathbf{S}_k = \left\{ \sum_{j=1}^k \beta_j B_{k,j}(\cdot), \beta = (\beta_1, \dots, \beta_k)' \in \mathbb{R}^k \right\},$$

and  $\{B_{k,j}(\cdot), j = 1, \dots, k\}$  are the uniform cubic B-spline functions defined on  $[-1, 1]$  with  $k$  degrees of freedom. In other words,  $\mathbf{S}_k$  is the cubic spline space with  $k-4$  uniform interior knots at  $\{-1, -1+1/m, \dots, 1-1/m, 1\}$ , where  $m = (k-3)/2$ .

The population objective function to maximize is,

$$Q(\phi, \sigma^2, \rho) = E \left[ -\frac{1}{2\sigma^2} \left( \rho^2 \int_0^1 \varepsilon_t^2(s) ds + \rho \varepsilon_t^2(0) + \rho \varepsilon_t^2(1) - \rho \right) \right], \quad (24)$$

where  $\varepsilon_t(s) = X_t(s) - \sum_{i=1}^p \int_0^1 \phi_i(s-u) X_{t-i}(u) du$ , and  $Q(\phi, \sigma^2, \rho)$  is the expectation of log-likelihood of the O-U process  $\varepsilon_t(\cdot)$ ; see Rao (1999). Define the population objective function  $Q_k(\beta, \sigma^2, \rho)$  in sieve space  $\Theta_k$ ,

$$Q_k(\beta, \sigma^2, \rho) = E \left[ -\frac{1}{2\sigma^2} \left( \rho^2 \int_0^1 \varepsilon_t^2(s) ds + \rho \varepsilon_t^2(0) + \rho \varepsilon_t^2(1) - \rho \right) \right],$$

where  $\varepsilon_t(s) = X_t(s) - \sum_{i=1}^p \sum_{j=1}^k \beta_{i,j} \int_0^1 B_{k,j}(s-u) X_{t-i}(u) du$ , and  $\phi_{i,k}(\cdot) = \sum_{j=1}^k \beta_{i,j} B_{k,j}(\cdot)$ ,  $i = 1, \dots, p$ .

When  $X_t(\cdot)$  is only observed at discrete points, with (8) we have the sample objective function  $Q_{k,T,N}(\beta, \sigma^2, \rho)$  as in (12), then the sieve estimator in space  $\Theta_k$  is

$$(\widehat{\beta}_k, \widehat{\sigma}_k^2, \widehat{\rho}_k) = \arg \max Q_{k,T,N}(\beta, \sigma^2, \rho).$$

### 4.2. Asymptotic properties for CFAR(1) models

Let  $\mathbf{Q}_k$  be the linear operator defined in Section 6.4 of Schumaker (1981), which maps  $C[-1, 1]$  onto the space  $\mathbf{S}_k$ . Let  $\widehat{\phi}_k = \mathbf{Q}_k \phi = \beta_{k,1} B_{k,1} + \dots + \beta_{k,k} B_{k,k}$ , and  $r_k(s) = \phi(s) - \widehat{\phi}_k(s)$ . Again,  $\widehat{\phi}_k = \sum_{j=1}^k \beta_{k,j} B_{k,j}(\cdot)$  is the estimated convolution function in  $\mathbf{S}_k$ .

**Theorem 3.**  $X_t$  is a stationary CFAR(1) process defined in (1).  $\widehat{\beta}_k$  and  $\widetilde{\beta}_k$  are the B-spline coefficients of  $\widehat{\phi}_k$  and  $\widetilde{\phi}_k$ . Assume  $\phi \in \text{Lip}_2^\zeta[-1, 1]$  with  $\zeta > 1$ . Then with given  $\sigma^2$  and  $\rho$ , as  $T \rightarrow \infty$  and  $N^2/T \rightarrow \infty$ ,

$$\sqrt{T}(\widehat{\beta}_k - \widetilde{\beta}_k - \mathbf{b}_k) \xrightarrow{d} N(\mathbf{0}, \Sigma_k),$$

where  $\Sigma_k = \Gamma_k^{-1} \Upsilon_k \Gamma_k^{-1}$ ,  $\Upsilon_k$  is the long-run variance of the process  $\mathbf{u}_t + \mathbf{z}_t + \mathbf{A}_t \mathbf{b}_k$ , where  $\mathbf{b}_k = \Gamma_k^{-1} \boldsymbol{\mu}_k$ , and the entries in  $\mathbf{u}_t$ ,  $\mathbf{z}_t$ ,  $\mathbf{A}_t$ ,  $\boldsymbol{\mu}_k$  and  $\Gamma_k$  are listed as follows

$$(\mathbf{u}_t)_i = \frac{1}{\sigma^2} \int_0^1 \int_0^1 \mathcal{A}(B_{k,i}, r_k)(u, v) X_{t-1}(u) X_{t-1}(v)(u, v) dudv,$$

$$(\mathbf{z}_t)_i = \int_0^1 \left[ \frac{2\rho}{\sigma^2} B_{k,i}(-u) \varepsilon_t(0) + \frac{1}{\sigma} \int_0^1 (\rho B_{k,i}(v - u) + B'_{k,i}(v - u)) dW_t(v) \right] X_{t-1}(u) du,$$

$$(\mathbf{A}_t)_{ij} = \frac{1}{\sigma^2} \int_0^1 \int_0^1 \mathcal{A}(B_{k,i}, B_{k,j})(u, v) X_{t-1}(u) X_{t-1}(v) dudv,$$

$$(\boldsymbol{\mu}_k)_i = \frac{1}{\sigma^2} \int_0^1 \int_0^1 \mathcal{A}(B_{k,i}, r_k)(u, v) \gamma(u, v) dudv,$$

$$(\Gamma_k)_{ij} = \frac{1}{\sigma^2} \int_0^1 \int_0^1 \mathcal{A}(B_{k,i}, B_{k,j})(u, v) \gamma(u, v) dudv,$$

where  $\gamma(u, v) = \text{Cov}(X_t(u), X_t(v))$ , and  $\mathcal{A}$  is a bivariate functional operator,  $\mathcal{A} : C^1[-1, 1] \otimes C^1[-1, 1] \rightarrow C[-1, 1]^2$ ,

$$\begin{aligned} \mathcal{A}(f, g)(u, v) &= \rho f(-u)g(-v) + \rho f(1-u)g(1-v) \\ &+ \int_0^1 f'(s-u)g'(s-v) ds + \rho^2 \int_0^1 f(s-u)g(s-v) ds. \end{aligned} \quad (25)$$

Theorem 3 provides the asymptotic distribution of the estimated B-spline coefficients when the dimension of the sieve space is fixed. We note that the asymptotic result does not involve  $N$  directly. As long as  $N$  goes to infinity fast enough with  $T$  such that  $N^2/T \rightarrow \infty$ , the error introduced by the linear interpolation is of a smaller order.

As a consequence of Theorem 3, we have the following central limit theorem for the pointwise estimate of the convolution function when  $k$  is fixed.

**Corollary 3.** Assume  $\phi \in \text{Lip}_2^\zeta[-1, 1]$  with  $\zeta > 1$ . For each fixed  $k$ , define  $\mathbf{B}_k(s) = (B_{k,1}(s), \dots, B_{k,k}(s))'$ , and

$$b_k(s) = \mathbf{B}_k(s)' \mathbf{b}_k - r_k(s), \quad \sigma_k^2(s) = \mathbf{B}_k(s)' \Sigma_k \mathbf{B}_k(s).$$

Then as  $T \rightarrow \infty$  and  $N^2/T \rightarrow \infty$ ,

$$\sqrt{T}(\widehat{\phi}_k(s) - \phi(s) - b_k(s)) \xrightarrow{d} N(0, \sigma_k^2(s)).$$

**Theorem 4.** Assume  $\phi \in C^\zeta[-1, 1]$ , where  $\zeta \geq 2$  is an integer. Then as  $k \rightarrow \infty$ , and

$$\|b_k(\cdot)\|_\infty = O(k^{-\zeta_0+3/2}), \quad \|\sigma_k^2(\cdot)\|_\infty = O(k),$$

where  $\zeta_0 = \min\{\zeta, 4\}$ .

Theorem 4 shows that the asymptotic bias converges with rate  $O(k^{-\zeta_0+3/2})$ , where  $k$  is the number of basis functions, which reflects the complexity of the sieve space; and  $\zeta$  indicates the smoothness of the convolution function. We note that even if  $\zeta > 4$ , the use of cubic splines is not able to take advantage of the additional smoothness. The asymptotic variance grows at the rate  $O(k)$  as  $k$  increases. This is a common phenomenon that the model complexity controls the trade-off between bias and variance. When  $k$  is small, the bias dominates, since there are not

enough knots to approximate the convolution function. When  $k$  is large, the variance dominates, and the estimation of extra B-spline coefficients introduces too much variability. Theorem 4 also hints on the selection principle of the number of knots. When the convolution function is smooth, a small  $k$  is favorable; when the sample size is large, we can afford to have more knots.

**Remark 6.** By balancing the bias and variance, the optimal convergence rate of the estimation error is attained at  $k \asymp O(T^{1/(2\zeta_0-2)})$ , then both the squared bias and variance will be  $O(T^{-\frac{2\zeta_0-3}{4\zeta_0-4}})$ . If the moduli of smoothness of  $\phi$  is greater than 4 and  $k = O(T^{-1/6})$ , the optimal convergence rate is  $O(T^{-5/12})$ . If B-splines with higher order are adopted, the estimator is expected to converge faster.

### 4.3. Asymptotic properties for CFAR(p)

The asymptotic properties of the estimators of CFAR(p) models are very similar to those of CFAR(1) models. Let  $\phi_{k,i} = \mathbf{Q}_k \phi_i = \widetilde{\beta}_{k,i,1} B_{k,1} + \dots + \widetilde{\beta}_{k,i,k} B_{k,k}$ , and  $r_{k,i}(s) = \phi(s) - \widetilde{\phi}_{k,i}(s)$ .

Theorem 5 provides the bound of the B-spline coefficients estimation error, and is a higher dimension version of Theorem 3.

**Theorem 5.** Assume  $X_t$  is a stationary CFAR(p) process defined in (1). Let  $\widehat{\beta}_k$  and  $\widetilde{\beta}_k$  be the B-spline coefficients of  $\{\widehat{\phi}_{k,1}, \dots, \widehat{\phi}_{k,p}\}$  and  $\{\widetilde{\phi}_{k,1}, \dots, \widetilde{\phi}_{k,p}\}$  respectively. Assume  $\phi_i \in \text{Lip}_2^{\zeta_i}[-1, 1]$  with  $\zeta_i > 1$ , for  $i = 1, \dots, p$ . Then with given  $\sigma^2$  and  $\rho$ , as  $T \rightarrow \infty$  and  $N^2/T \rightarrow \infty$ ,

$$\sqrt{T}(\widehat{\beta}_k - \widetilde{\beta}_k - \mathbf{b}_k) \xrightarrow{d} N(\mathbf{0}, \Sigma_k)$$

where  $\Sigma_k = \Gamma_k^{-1} \Upsilon_k \Gamma_k^{-1}$ ,  $\Upsilon_k$  is the long-run variance of the process  $\mathbf{u}_t + \mathbf{z}_t + \mathbf{A}_t \mathbf{b}_k$ , where  $\mathbf{b}_k = \Gamma_k^{-1} \boldsymbol{\mu}_k$ . Here  $\mathbf{u}_t = (\mathbf{u}'_{t,1}, \dots, \mathbf{u}'_{t,p})'$ ,  $\mathbf{z}_t = (\mathbf{z}'_{t,1}, \dots, \mathbf{z}'_{t,p})'$ ,  $\boldsymbol{\mu}_k = (\boldsymbol{\mu}'_{k,1}, \dots, \boldsymbol{\mu}'_{k,p})'$ ,  $\Gamma_k$  and  $\mathbf{A}_t$  can be partitioned into  $p \times p$  blocks.

$$\begin{aligned} \Gamma_k &= \begin{pmatrix} \Gamma_{k,0} & \Gamma_{k,-1} & \dots & \Gamma_{k,-p+1} \\ \Gamma_{k,1} & \Gamma_{k,0} & \dots & \Gamma_{k,-p+2} \\ \vdots & \vdots & \ddots & \vdots \\ \Gamma_{k,p-1} & \Gamma_{k,p-2} & \dots & \Gamma_{k,0} \end{pmatrix}, \\ \mathbf{A}_t &= \begin{pmatrix} \mathbf{A}_{t,1,1} & \mathbf{A}_{t,1,2} & \dots & \mathbf{A}_{t,1,p} \\ \mathbf{A}_{t,2,1} & \mathbf{A}_{t,2,2} & \dots & \mathbf{A}_{t,2,p} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{A}_{t,p,1} & \mathbf{A}_{t,p,2} & \dots & \mathbf{A}_{t,p,p} \end{pmatrix}. \end{aligned}$$

The entries in  $\mathbf{u}_t$ ,  $\mathbf{z}_t$ ,  $\mathbf{A}_t$ ,  $\mathbf{b}_k$  and  $\Gamma_k$  are listed as follows

$$(\mathbf{u}_{t,h})_i = \frac{1}{\sigma^2} \sum_{q=1}^p \int_0^1 \int_0^1 \mathcal{A}(B_{k,i}, r_{k,q})(u, v) X_{t-h}(u) \times X_{t-q}(v)(u, v) dudv,$$

$$(\mathbf{z}_{t,h})_i = \int_0^1 \left[ \frac{2\rho}{\sigma^2} B_{k,i}(-u) \varepsilon_t(0) + \frac{1}{\sigma} \int_0^1 (\rho B_{k,i}(v - u) + B'_{k,i}(v - u)) dW_t(v) \right] X_{t-h}(u) du,$$

$$(\mathbf{A}_{t,l,h})_{ij} = \frac{1}{\sigma^2} \int_0^1 \int_0^1 \mathcal{A}(B_{k,i}, B_{k,j})(u, v) X_{t-l}(u) X_{t-h}(v) dudv,$$

$$(\boldsymbol{\mu}_{k,h})_i = \frac{1}{\sigma^2} \sum_{q=1}^p \int_0^1 \int_0^1 \mathcal{A}(B_{k,i}, r_{k,q})(u, v) \gamma_{q-h}(u, v) dudv,$$

$$(\Gamma_{k,h})_{ij} = \frac{1}{\sigma^2} \int_0^1 \int_0^1 \mathcal{A}(B_{k,i}, B_{k,j})(u, v) \gamma_h(u, v) dudv,$$

for  $i, j = 1, \dots, k$  and  $l, h = 1, \dots, p$ , where  $\gamma_q(u, v) = \text{Cov}(X_t(u), X_{t+q}(v))$ ,  $q \in \mathbb{Z}$ .

**Table 1**

Simulation results with different combinations of  $(k, N, T)$  for [Example 1](#). Means and standard deviations (in brackets) of the estimation errors  $E$  and the sums of squared one-step-ahead out-of-sample forecasting errors, and the frequency of  $\hat{p} = p$ .

Model			Estimation error $E$	Forecasting MSE ( $\frac{\sigma^2}{2\rho} = 1$ )		Frequency of $\hat{p} = p$
CFAR(1)			CFAR	CFAR	FAR	CFAR
$N = 100$	$T = 100$	$k = 7$	0.43 (0.10)	1.06(0.12)	1.41 (0.38)	96%
		$k = 11$	0.17 (0.10)	1.05(0.12)	1.41 (0.38)	97%
		$k = 19$	0.33 (0.14)	1.05(0.12)	1.41 (0.38)	98%
$N = 100$	$T = 10$	$k = 11$	4.37 (3.35)	1.76 (0.89)	6.66 (7.28)	91%
	$T = 100$		0.17 (0.10)	1.05 (0.12)	1.41 (0.38)	97%
	$T = 1000$		0.03 (0.02)	1.00 (0.03)	1.02 (0.03)	97%
$N = 10$	$T = 100$	$k = 11$	0.20 (0.10)	1.08 (0.13)	1.45 (0.40)	96%
$N = 100$			0.17 (0.10)	1.05 (0.12)	1.41 (0.38)	97%
$N = 1000$			0.17 (0.10)	1.05 (0.12)	1.42 (0.40)	97%

The central limit theorem for the point estimate of the convolutions functions is obtained as a consequence of [Theorem 5](#).

**Corollary 4.** Assume  $\phi_i \in \text{Lip}^\zeta[-1, 1]$ , with  $\zeta_i > 1, i = 1, \dots, p$ . For each fixed  $k$ , define  $\mathbf{B}_{k,i}(s) = (\mathbf{B}'_{k,i,1}, \dots, \mathbf{B}'_{k,i,k})'$ ,  $\mathbf{B}_{k,i,i} = (B_{k,1}(s), \dots, B_{k,k}(s))'$ , and  $\mathbf{B}_{k,i,j}$  is a  $p$ -dimensional vector whose entries are all 0, for  $j \neq i$ . Then

$$b_{k,i}(s) = \mathbf{B}_{k,i}(s)' \mathbf{b}_k - r_{k,i}(s), \quad \sigma_{k,i}^2(s) = \mathbf{B}_{k,i}(s)' \boldsymbol{\Sigma}_k \mathbf{B}_{k,i}(s),$$

for  $i = 1, \dots, p$ ,

and as  $T \rightarrow \infty$  and  $N^2/T \rightarrow \infty$ .

$$\sqrt{T}(\hat{\phi}_{k,i}(s) - \phi(s) - b_{k,i}(s)) \xrightarrow{d} N(0, \sigma_{k,i}^2(s)), \quad \text{for } i = 1, \dots, p.$$

The behavior of the asymptotic bias and variance as  $k$  increases is similar with the CFAR(1) case, as shown in the next theorem.

**Theorem 6.** Assume  $\phi_i \in C^{\zeta_i}[-1, 1]$  with  $\zeta_i \geq 2$ . It holds that as  $k \rightarrow \infty$ ,

$$\|b_{k,i}(\cdot)\|_\infty = O(k^{-\zeta_0+3/2}), \quad \|\sigma_{k,i}^2(\cdot)\|_\infty = O(k),$$

for  $i = 1, \dots, p$ ,

where  $\zeta_0 = \min\{\zeta_1, \dots, \zeta_p, 4\}$ .

### 5. Simulations

In this section, we illustrate the performance of the proposed estimators, demonstrate the impacts of  $k, T$  and  $N$ , and compare our models with FAR models through two simulated examples.

For each model, we generate the functional time series and estimate the convolution functions, with 100 repetitions. The performance of estimation is evaluated by the integrated squared error,  $E_i = \|\hat{\phi}_{k,i} - \phi_i\|_2^2$ , for  $i = 1, \dots, p$ .

**Example 1.** Consider a simple CFAR(1) model with the convolution function  $\phi(\cdot)$  being the normal density function with mean 0 and standard deviation 0.1, truncated at  $[-1, 1]$ . For the noise process, we use  $\rho = 5$  and  $\sigma^2 = 10$ , and assume that they are known. We summarize the results for different combination of  $(k, N, T)$  in [Table 1](#). We also use figures to demonstrate various features of the model and the method. Most of the figures are based on a typical data set, whose estimation error is the median value among the 100 simulated sets. For the plots we set  $N = 100, T = 100$ , and consider various  $k$ .

The second column of [Table 1](#) shows the estimation performance of  $\phi(\cdot)$  with different combination of  $(k, N, T)$ . If  $T$  and  $N$  are fixed, as  $k$  increases, the average of  $E$  decreases first, due to the improvement of sieve approximation, and then increases because of the introduction of extra B-spline coefficients. If  $k$  and  $N$  are fixed, as  $T$  increases, the estimates become more accurate and

stable since the average and standard error of  $E$  decrease. The error  $E$  decreases approximately at the rate of  $1/T$  which confirms the theoretical results. If  $k$  and  $T$  are fixed and  $N$  is small, increasing  $N$  improves the performance of the estimators. However, when  $N$  is large enough,  $E$  stays at the same level. Because of the strong spatial correlation in the error process, denser observations do not provide much extra information.

The third and fourth columns of [Table 1](#) display the sums of squared out-of-sample forecasting errors for CFAR and FAR models, respectively. It is seen that our method has a better prediction accuracy when model is correctly specified. The last column reports the frequency of the determined order  $\hat{p}$  using the  $F$  test in [Section 3.5](#) is the correct order  $\hat{p} = p$  for 100 data sets. It is seen that the sequential  $F$ -tests perform reasonably well.

[Figs. 2–4](#) demonstrate the results for the chosen typical data set. [Fig. 2](#) displays predicted  $X_t(\cdot)$  using (18) from time  $t = 2$  to 7, using  $k = 15, T = 100$  and  $N = 100$ . The predictions (dotted lines) are close to the skeletons (dashed lines). Also note that the noise process is large in magnitude with strong spatial correlation, hence a direct smoothing of  $X_t(\cdot)$  will be far away from the true skeleton  $f_t(\cdot)$ .

[Fig. 3](#) demonstrates the use of the out-of-sample rolling forecasting in choosing  $k$ , for  $T = 100, T_0 = T/2$  and  $N = 100$ . The left panel of [Fig. 3](#) displays the sum of squared forecasting error across  $k$  for the typical data set. For small  $k$ , the forecasting error is very large, due to the lack of knots to approximate  $\phi(\cdot)$  well. As the complexity of the sieve space  $k$  increases, forecasting error reduces very quickly, until the minimum prediction error is reached. After that, the error increases gradually with  $k$  because of the variability introduced by the extra splines. It is interesting to note that when  $k$  is odd, the prediction error is smaller than that when  $k$  is even. This is due to the fact that  $\phi(\cdot)$  and its second derivative reach their maximum value at 0. The estimate benefits from a knot at 0 when  $k$  is odd. The right panel of [Fig. 3](#) shows a histogram of the 100 optimal  $k$  of the 100 data sets.

[Fig. 4](#) displays the estimated  $\phi(\cdot)$  for the typical data set when  $k$  takes different values. As expected, the accuracy of the estimate improves first and then becomes worse as  $k$  increases. The top panels show the estimates when the bias dominates, and the bottom panels show the estimates when the variance dominates.

**Example 2.** This simulated example is based on the estimated model obtained for the real data example presented in [Section 6](#). Here we consider a CFAR(2) model, and an irregular setting with different observation locations at different times and a spatially nonstationary noise process. We set  $\phi_1(s) = 15(s + 0.7)(s + 0.3)(s - 0.3)I_{[s \leq 0.3]} + 3(s - 0.3)(s - 1.3)I_{[s > 0.3]}$  and  $\phi_2(s) = 15(s + 0.6)(s + 0.4)(s - 0.2)I_{[s \leq 0.2]} + 7(s - 0.2)(s - 0.5)(s - 0.8)I_{[s > 0.2]}$ . By examining the variance of the real data in [Section 6](#),

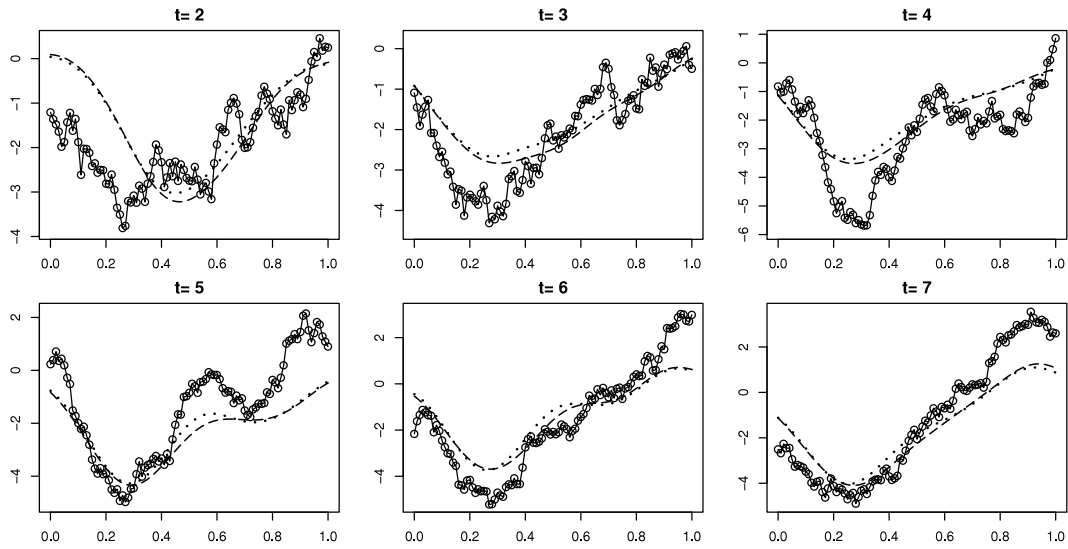


Fig. 2. Plots of predicted  $X_t(\cdot)$  when  $k = 15$ ,  $N = 100$ , and  $T = 100$ ,  $t = 2-7$ , for the typical data set for Example 1. The solid lines and circles are  $X_t(\cdot)$ ; the dashed lines are  $f_t(\cdot)$ ; the dotted lines are the predictions.

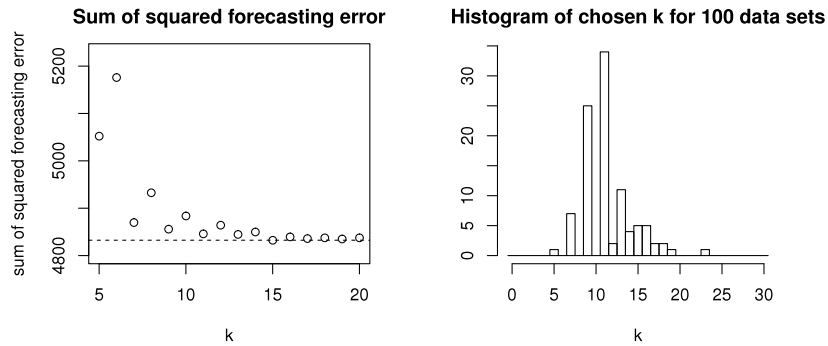


Fig. 3. Left panel: plot of sum of squared out-of-sample forecasting error against  $k$  for the typical data set when  $T = 100$  and  $N = 100$  for Example 2; right panel: histogram of the optimal  $k$  of the 100 data sets when  $T = 100$  and  $N = 100$  for Example 2.

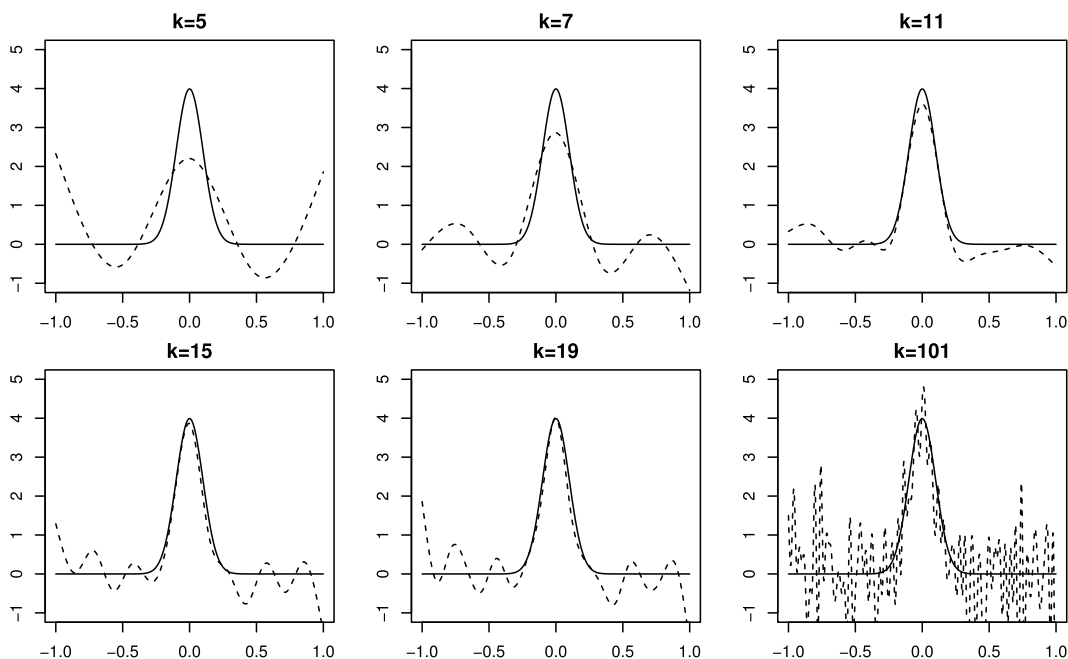


Fig. 4. Plots of  $\phi(\cdot)$  (solid line) and  $\hat{\phi}(\cdot)$  (dashed line) with  $T = 100$ ,  $N = 100$ , and  $k = 5, 7, 11, 15, 19, 101$  for the typical data set for Example 2.



**Table 2**

Simulation results with different  $T$  for Example 2. Means and standard deviations (in brackets) of estimation errors  $E$  and the sums of squared one-step-ahead out-of-sample forecasting errors, and the frequency of  $\hat{p} = p$ .

Model	Estimation error $E$		Estimation MSE		Forecasting MSE ( $\frac{\sigma^2}{2p} = 2$ )		Frequency of $\hat{p} = p$
	CFAR $\phi_1$	CFAR $\phi_2$	$\rho$	$\sigma^2$	CFAR	FAR	CFAR
$T = 50$	0.91 (1.41)	0.89 (1.15)	0.0213(0.0347)	0.0057(0.0092)	2.19(0.59)	4.70 (3.90)	90%
$T = 100$	0.97 (0.86)	0.42 (0.45)	0.0113(0.0145)	0.0022(0.0032)	2.13(0.44)	4.08 (2.05)	92%
$T = 200$	0.82 (0.24)	0.21 (0.24)	0.0046(0.0065)	0.0013(0.0019)	2.16(0.32)	3.91 (1.47)	85%

**Table 3**

The  $p$ -values for testing  $H_0 : \phi_i = 0$ , when  $k = 11$  for real data analysis.

$H_0$	$\phi_1 = 0$	$\phi_2 = 0$	$\phi_3 = 0$
$p$ -value	0.0061	0.0076	0.8651

the heteroscedasticity weight function of the noise process as defined in (2) is chosen as

$$w(s) \propto \begin{cases} e^{-10w} + 0.1, & 0 \leq s \leq 0.6, \\ 0.2(w - 0.6) + e^{-6} + 0.1, & 0.6 < s \leq 1. \end{cases} \quad (26)$$

Let  $N_t = 40 + m_t$ , where  $N_t$  is the number of observations at time  $t$ , and  $\{m_t\}$  are generated independently from the Poisson distribution with mean 5. The observation location  $\{s_0, \dots, s_{N_t}\}$  at time  $t$  is generated independently from the uniform distribution on  $[0, 1]$ . We set  $\rho = 1$ , and  $\sigma^2 = 2$ , but assume they are unknown. We also choose  $k = 6$  for the estimation.

The simulation results are summarized in Table 2. The estimates of convolution functions,  $\rho$  and  $\sigma^2$  improve as  $T$  increases, as shown in the columns 2–5 of Table 2, respectively. Comparing the out-of-sample forecasting errors of CFAR and FAR models, we see that CFAR model provides more accurate predictions even when sample size is small. From the last column of Table 2, it is seen that the sequential  $F$ -tests perform well in determining the CFAR order.

**6. Data analysis**

A European call option is a contract which gives the holder the right to buy an underlying asset at a specified strike price on a specified expiration date; see Hull (2006). Its value is determined by the volatility of the underlying asset and other directly observed parameters including risk-free rate and maturity, in the most widely used Black–Scholes pricing formula; see Black and Scholes (1973) and Merton (1973). The volatility derived from option pricing methods with the observed option price is called the implied volatility. A plot of the implied volatility of an option as a function of its moneyness is known as a volatility curve, where moneyness is the relative strike price with respect to the price of the underlying asset; see Hull (2006). In this section, we apply our method to the S&P 500 index European call option data and study the dynamics of volatility curves. The option data is collected from July 9, 2004 to September 20, 2004, and the expiration date is December 18, 2004. Hence, the time series is  $T = 51$  long. We select the options with strikes between 900 and 1550, which were actively traded within this period and have solutions for implied volatilities derived from the Black–Scholes model most of the time. The number of observations at each time varies from 43 to 48, and the total number of observations is 2241. Our aim is to construct the volatility curve against moneyness, study its evolution dynamics, and make forecasts.

The left panel of Fig. 5 displays the implied volatility curves as functions of moneyness over time, where the volatility smiles are clearly demonstrated, since deep in-the-money or out-the-money options have larger implied volatility than others. As the expiration date approaches, the implied volatilities of at-the-money options decrease, those of out-the-money options increase,

and the location where the volatility curve reaches a minimum value approaches to 1 from the right side, as shown in the top left panel of Fig. 5. It indicates that the process may not be stationary, so we eliminate this trend before fitting the CFAR model by taking  $Z_t(s) = X_t(s + \tau_t/500)$ , and  $Y_t(s) = Z_t(s) - Z_{t-1}(s)$ ,  $t = 2, \dots, T$ , where  $\tau_t$  is the time to maturity,  $X_t(s)$  is the implied volatility with moneyness  $s$ , and  $\tilde{Z}_t(\cdot)$  is the linear interpolation of  $Z_t(\cdot)$  as defined in (8). Many empirical results show that the implied volatility curves are unit-root-nonstationary (Fengler et al., 2005; Gonçalves and Guidolin, 2006; Lam et al., 2011). All the curves  $Y_t(\cdot)$  are transformed to have the common support  $[0, 1]$ , and they are plotted in the right panels of Fig. 5. The data for deep in-the-money and out-the-money options are much more volatile than those of at-the-money options, as shown in Fig. 6 (top left panel). Therefore, we set a heteroscedasticity function (26) for the noise process when modeling  $\{Y_t\}$ .

The  $p$ -values of the sequential  $F$ -tests for testing  $H_0 : \phi_i = 0$  vs  $H_1 : \phi_i \neq 0$  for  $i = 1, 2, 3$  in Table 3 indicate that CFAR(2) model should be used to fit the data. The left panel of Fig. 6 displays the sums of squared out-of-sample forecasting errors against  $k$  where outsample forecasting period is from  $T_0 = (4/5)T$  to  $T$ . The number of basis functions  $k$  is chosen to be 6, where the error reaches its minimum value. The middle and right panels of Fig. 6 display the estimated  $\phi_1$  and  $\phi_2$ , respectively.

Set the residuals as

$$\hat{\varepsilon}_t(s_i) = \hat{\varepsilon}_t(s_i) - e^{-\hat{\rho}|s_i - s_{i-1}|} \hat{\varepsilon}_t(s_{i-1}), \quad i = 2, \dots, N_t.$$

The sample spatial autocorrelation functions of  $\{\hat{\varepsilon}_t(s_i), i = 2, \dots, N_t\}$  for  $t = 3, \dots, 38$  are shown in Fig. 7. Overall, most of the sample autocorrelations of  $\hat{\varepsilon}_t(\cdot)$  are not significant, which implies that the CFAR(2) model fits data well.

One-step-ahead out-of-sample forecasting error of  $X_t(\cdot)$  is used to compare the prediction performance of the estimated CFAR(2) model with the FAR model by Bosq (2000) and the functional random walk model, where  $\hat{Y}_{t+1}(s_i) = 0$  (i.e.  $\hat{X}_{t+1}(s_i) = X_t(s_i)$ ),  $\forall s_i \in [0, 1]$ . Table 4 shows the estimated  $\rho$  and  $\sigma^2$  using the CFAR(2) model, and the mean squared out-sample forecasting errors of different models. CFAR(2) outperforms the other two models, reducing the forecasting MSE by more than 20%.

**Acknowledgments**

Xiao’s research is partially supported by NSF grant DMS 1209091. Chen’s research is partially supported by NSF grants DMS 1209085 and DMS 1513409. We would like to thank the editor and two anonymous referees for their valuable comments and suggestions which led to a substantial improvement of the paper.

**Appendix**

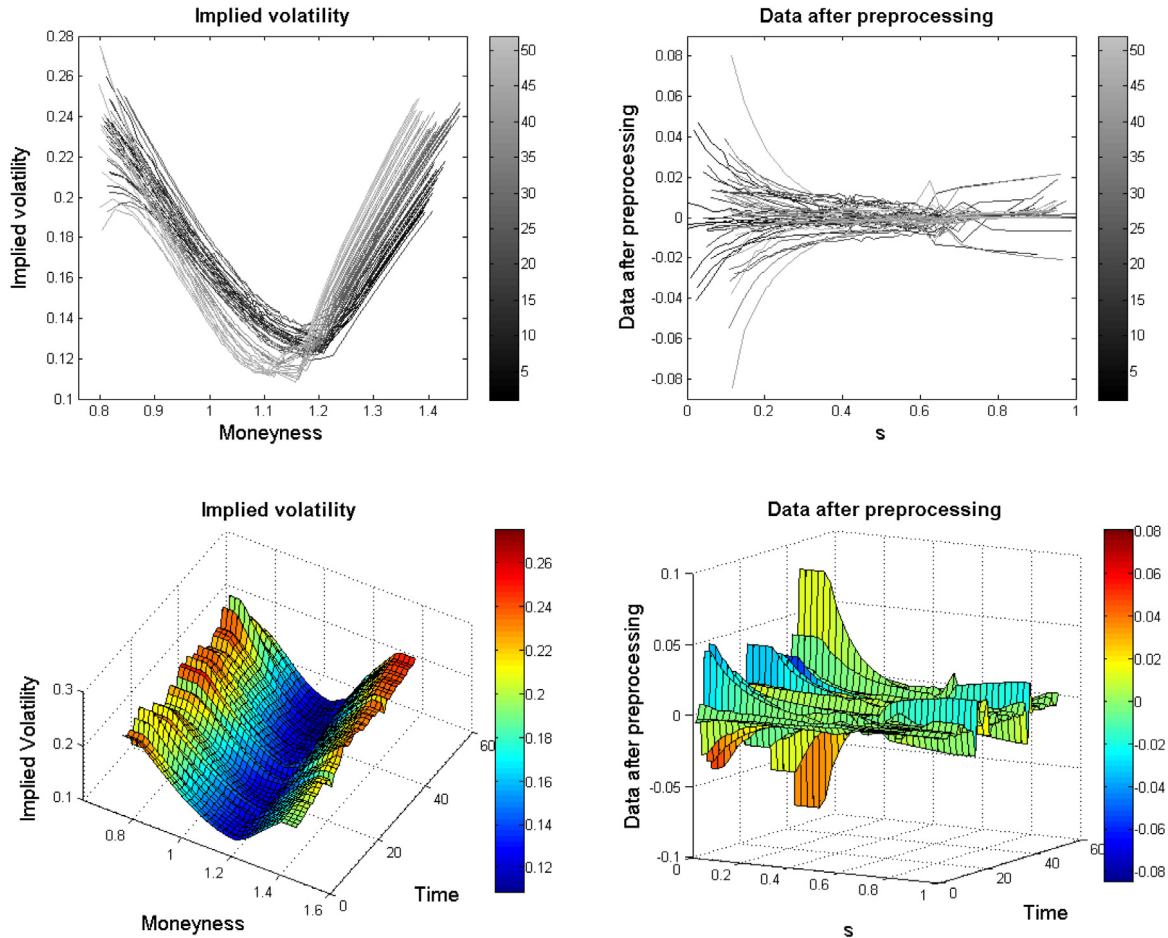
The detailed proofs are provided in the Appendix.

**Proof of Theorem 1.** By Cauchy–Schwarz inequality, for any function  $x \in L_2[0, 1]$ ,

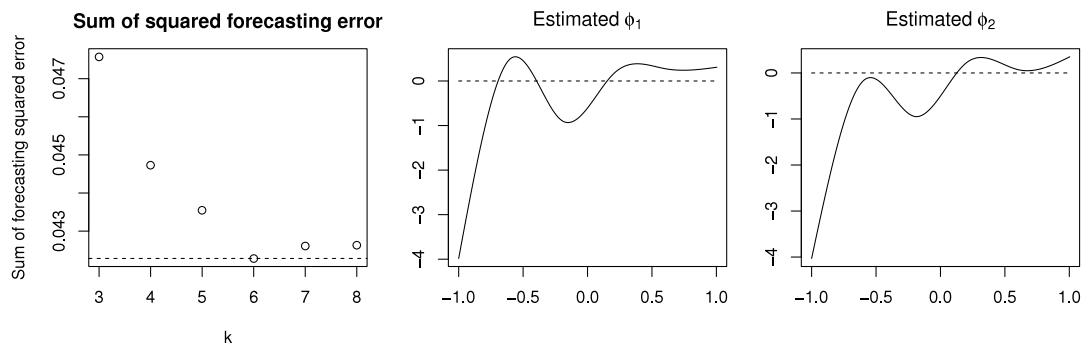
$$\begin{aligned} \left\| \int_0^1 \phi(s - u)x(u)du \right\|_2^2 &= \int_0^1 \left( \int_0^1 \phi(s - u)x(u)du \right)^2 ds \\ &\leq \int_0^1 \left( \int_0^1 \phi^2(s - u)du \right) \left( \int_0^1 x^2(u)du \right) ds \leq \kappa^2 \|x\|. \end{aligned}$$

**Table 4**  
Parameter estimates of the CFAR model and forecasting MSEs of different models for real data analysis.

$\hat{\rho}$	$\hat{\sigma}^2$	MSE(CFAR)	MSE(FAR)	MSE (random walk)
1.0597	2.2279e-04	8.7134e-05	1.1010e-04	1.1938e-04



**Fig. 5.** Plots of implied volatilities against moneyiness (top left panel) for different time (bottom left panel) and plots of data after preprocessing (top right panel) across time (bottom right panel). The colorbars on the right show the time (top panels) and the implied volatility values (bottom panels) corresponding to different color scales. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)



**Fig. 6.** Plot of the sum of squared out-of-sample forecasting errors against  $k$  (left panel) and plots of  $\hat{\phi}_1(\cdot)$  (middle panel) and  $\hat{\phi}_2(\cdot)$  (right panel) for real data analysis.

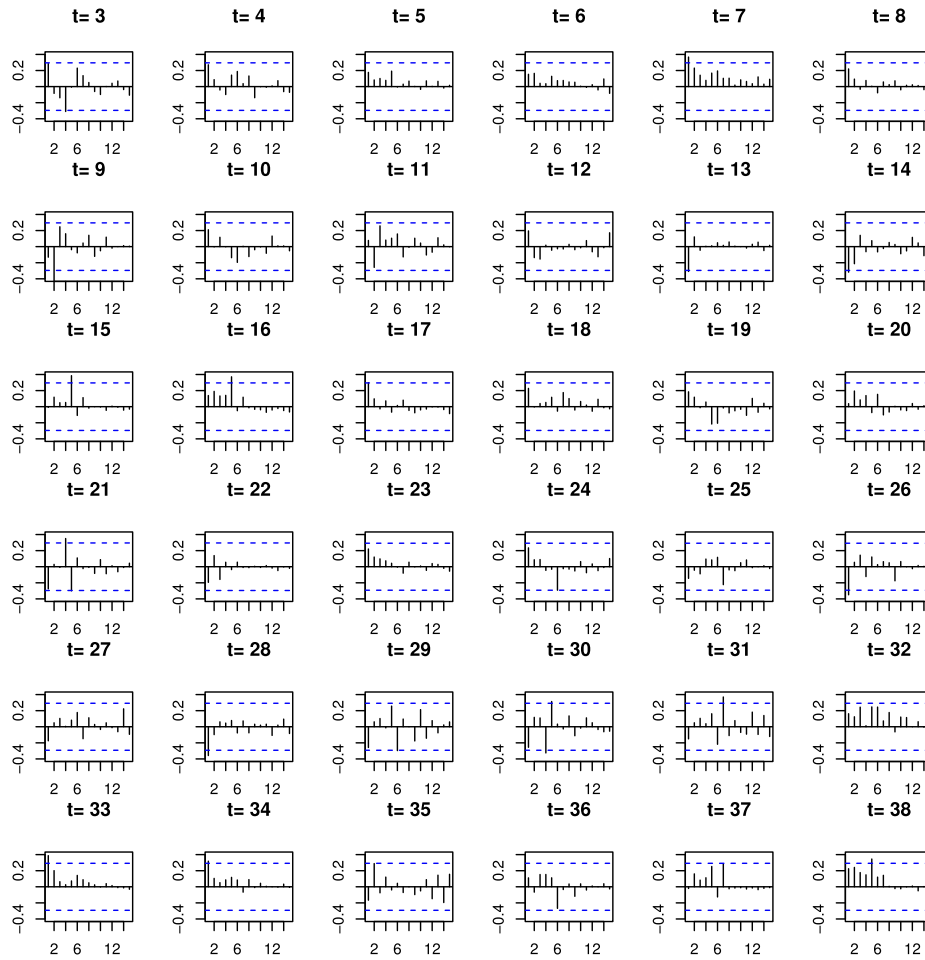


Fig. 7. Sample autocorrelations of  $\hat{e}_t(\cdot)$  with lag 0 autocorrelation removed when  $t = 3, \dots, 38$  for real data analysis.

It follows that

$$\sup_{\|x\| \leq 1} \left\| \int_0^1 \phi(s-u)x(u)du \right\|_2^2 \leq \kappa^2 < 1.$$

By Lemma 3.1 in Bosq (2000),  $X_t$  is stationary. ■

**Proof of Theorem 2.** Let  $\mathbf{H}^p[0, 1]$  be the Cartesian product of  $p$  copies of the random function space  $\mathbf{H}[0, 1]$ . The norm in  $\mathbf{H}^p[0, 1]$  is defined as

$$\|(X_1, \dots, X_p)\|_p = \sqrt{\sum_{i=1}^p \|X_i\|_2^2}, \quad \text{where } X_1, \dots, X_p \in \mathbf{H}[0, 1].$$

Consider

$$\mathbf{A} = \begin{bmatrix} \Delta_1 & \Delta_2 & \dots & \Delta_p \\ I & 0 & \dots & 0 \\ 0 & I & \dots & 0 \\ 0 & \dots & I & 0 \end{bmatrix}, \quad \mathbf{K} = \begin{bmatrix} \kappa_1 & \kappa_2 & \dots & \kappa_p \\ 1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ 0 & \dots & 1 & 0 \end{bmatrix}, \quad (\text{A.1})$$

where  $I$  denotes the identity operator,  $\Delta_i$  is a convolution operator associating with  $\phi_i$ , i.e.  $\Delta_i X = \int_0^1 \phi_i(\cdot - u)X(u)du$ . Note that  $\|\mathbf{A}X\|_p \leq \|\mathbf{K}X\|_p$ . And all the roots of the characteristic function are eigenvalues of the matrix  $\mathbf{K}$ . Let  $\lambda_{\max}$  be the maximum eigenvalue in modulus of  $\mathbf{K}$ , and  $|\lambda_{\max}| < 1$ . Hence, there exists an integer  $j$ , such that  $\|\mathbf{A}^j\| \leq \|\mathbf{K}^j\| < 1$ . By Theorem 5.1 in Bosq (2000), CFAR( $p$ ) model is stationary if (4) is satisfied. ■

**Lemma 1.** Let  $f, g \in \text{Lip}_2^\zeta[-1, 1]$ , and  $\zeta > 1$ ,  $\{s_0 = 0, s_1, \dots, s_N = 1\}$  is an equally spaced partition of  $[0, 1]$ . For any  $u, v \in [0, 1]$ ,

let  $\mathbf{f}_u$  and  $\mathbf{g}_v$  be  $(N + 1) \times 1$  vectors,  $\mathbf{f}_u = (f(s_0 - u), f(s_1 - u), \dots, f(s_N - u))'$ ,  $\mathbf{g}_v = (g(s_0 - v), g(s_1 - v), \dots, g(s_N - v))'$ .  $\Sigma$  is defined in Section 3.1, where  $(\Sigma)_{ij} = e^{-\rho|i-j|/N} \sigma^2 / 2\rho$ . Define  $A_N(f, g)(u, v) = \mathbf{f}_u \Sigma^{-1} \mathbf{g}_v$ , then

$$A_N(f, g)(u, v) = \frac{1}{\sigma^2} \mathcal{A}(f, g)(u, v) + O(N^{-\zeta_0+1}), \quad (\text{A.2})$$

where  $\zeta_0 = \min\{\zeta, 2\}$ , and  $\mathcal{A}$  is the binary functional operator defined in (25).

**Proof.** Define  $\varphi = e^{-\rho/N}$ . From Sutradhar and Kumar (2003), it is known that

$$\Sigma_0^{-1} = \frac{1}{1-\varphi^2} \begin{pmatrix} 1 & -\varphi & 0 & \dots & 0 \\ -\varphi & 1+\varphi^2 & -\varphi & \dots & 0 \\ 0 & \dots & -\varphi & 1+\varphi^2 & -\varphi \\ 0 & \dots & \dots & -\varphi & 1 \end{pmatrix}.$$

Using Cholesky decomposition and rearranging the rows and columns, we are able to find  $\mathbf{U}$  and prove that  $\Sigma^{-1} = \frac{\sigma^2}{2\rho} \mathbf{U}'\mathbf{U}$ , where

$$\mathbf{U} = \frac{1}{\sqrt{1-\varphi^2}} \begin{pmatrix} \sqrt{1-\varphi^2} & 0 & \dots & 0 \\ -\varphi & 1 & \dots & 0 \\ 0 & \dots & -\varphi & 1 \end{pmatrix}. \quad (\text{A.3})$$

And we have the equation given in Box 1.

$$\mathbf{Uf}_u = \frac{1}{\sqrt{1-\varphi^2}} \left( \sqrt{1-\varphi^2}f(s_0-u) \quad f(s_1-u) - \varphi f(s_0-u) \quad \dots \quad f(s_N-u) - \varphi f(s_{N-1}-u) \right)'$$

**Box I.**

$\mathbf{Ug}_v$  has a similar expression, hence

$$\begin{aligned} \mathbf{f}'_u \Sigma^{-1} \mathbf{g}_v &= \frac{2\rho}{\sigma^2} f(-u)g(-v) + \frac{2\rho}{(1-\varphi^2)\sigma^2} \sum_{n=1}^N (f(s_n-u) \\ &\quad - \varphi f(s_{n-1}-u)) (g(s_n-v) - \varphi g(s_{n-1}-v)) \\ &= \frac{2\rho}{\sigma^2} f(-u)g(-v) + \frac{2\rho}{(1-\varphi^2)\sigma^2} \sum_{n=1}^N [f(s_n-u) \\ &\quad - f(s_{n-1}-u)] [g(s_n-v) - g(s_{n-1}-v)] \\ &\quad + \frac{2\rho}{(1-\varphi^2)\sigma^2} \sum_{n=1}^N (1-\varphi) [f(s_n-u) \\ &\quad - f(s_{n-1}-u)] g(s_{n-1}-v) \\ &\quad + \frac{2\rho}{(1-\varphi^2)\sigma^2} \sum_{n=1}^N (1-\varphi) f(s_{n-1}-v) \\ &\quad \times [g(s_n-v) - g(s_{n-1}-v)] \\ &\quad + \frac{2\rho}{(1-\varphi^2)\sigma^2} \sum_{n=1}^N (1-\varphi)^2 f(s_{n-1}-v) g(s_{n-1}-v) \\ &= L_1 + L_2 + L_3 + L_4 + L_5. \end{aligned}$$

Since  $f, g \in \text{Lip}_2^\zeta[-1, 1]$ ,  $|f'(u) - f'(v)| \leq M|u - v|$ , if  $\zeta \geq 2$ ;  $|f'(u) - f'(v)| \leq M|u - v|^{-\zeta+1}$ , if  $\zeta < 2$ , for some positive constant  $M$ . It follows that

$$\begin{aligned} L_2 &= \frac{2\rho}{N(1-\varphi^2)\sigma^2} \sum_{n=1}^N \frac{1}{N} \left( \frac{f(s_n-u) - f(s_{n-1}-u)}{1/N} \right) \\ &\quad \times \left( \frac{g(s_n-v) - g(s_{n-1}-v)}{1/N} \right) \\ &= \frac{2\rho}{N(1-\varphi^2)\sigma^2} \sum_{n=1}^N \frac{1}{N} (f'(s_{n-1}-u) + O(N^{-\zeta_0+1})) \\ &\quad \times (g'(s_{n-1}-v) + O(N^{-\zeta_0+1})) \\ &= \frac{1}{\sigma^2} \int_0^1 f'(s-u)g'(s-v) ds + O(N^{-\zeta_0+1}). \end{aligned}$$

Similarly, we obtain  $L_3 = \frac{\rho}{\sigma^2} \int_0^1 f'(s-u)g(s-v) ds + O(N^{-\zeta_0+1})$ ,  $L_4 = \frac{\rho}{\sigma^2} \int_0^1 f(s-u)g'(s-v) ds + O(N^{-\zeta_0+1})$ , and  $L_5 = \frac{\rho^2}{\sigma^2} \int_0^1 f(s-u)g(s-v) ds + O(N^{-\zeta_0+1})$ .

Using integration by parts, we have

$$\begin{aligned} \mathbf{f}'_u \Sigma^{-1} \mathbf{g}_v &= \frac{1}{\sigma^2} \left( \rho f(-u)g(-v) + \rho f(1-u)g(1-v) \right. \\ &\quad \left. + \int_0^1 f'(s-u)g'(s-v) ds + \rho^2 \int_0^1 f(s-u)g(s-v) ds \right) \\ &\quad + O(N^{-\zeta_0+1}). \quad \blacksquare \end{aligned}$$

**Lemma 2.** If the CFAR(1) process  $X = (X_t, t \in \mathbb{Z})$  satisfies (3) and  $\phi \in C^2[-1, 1]$ , we define  $\gamma_h(u, v) = \text{Cov}(X_t(u), X_{t+h}(v))$ , then

there exists a positive constant  $\iota$  which only depends on  $\phi$  and  $\rho$  such that:

- (i)  $\|\Psi_h\|_\infty \leq \kappa^h$  for  $h \geq 2$ .
- (ii)  $\|\gamma_h\|_\infty \leq \sigma^2 \iota \kappa^{|h|}$  for  $h \in \mathbb{Z}$ .
- (iii) For  $h \geq 1$ ,

$$\max_{u,v} \left\{ \left| \frac{\partial \gamma_h(u, v)}{\partial u} \right|, \left| \frac{\partial \gamma_h(u, v)}{\partial v} \right| \right\} \leq \sigma^2 \iota \kappa^h,$$

and when  $h = 0$ , for any  $u_1, u_2, v \in [0, 1]$ ,

$$|\gamma(u_1, v) - \gamma(u_2, v)| \leq \sigma^2 \iota |u_1 - u_2|.$$

- (iv) For  $h \geq 1$ ,

$$\max_{u,v} \left\{ \left| \frac{\partial^2 \gamma_h(u, v)}{\partial u^2} \right|, \left| \frac{\partial^2 \gamma_h(u, v)}{\partial u \partial v} \right|, \left| \frac{\partial^2 \gamma_h(u, v)}{\partial v^2} \right| \right\} \leq \sigma^2 \iota \kappa^h,$$

and when  $h = 0$ ,

$$\max_{u \neq v} \left\{ \left| \frac{\partial^2 \gamma(u, v)}{\partial u^2} \right|, \left| \frac{\partial^2 \gamma(u, v)}{\partial u \partial v} \right|, \left| \frac{\partial^2 \gamma(u, v)}{\partial v^2} \right| \right\} \leq \sigma^2 \iota.$$

**Proof.** By Cauchy–Schwarz inequality, for any  $u, v \in [0, 1]$ ,

$$\begin{aligned} |\Psi_2(u, v)| &= \left| \int_0^1 \phi(u-s)\phi(s-v) ds \right| \leq \left( \int_0^1 \phi^2(u-s) ds \right)^{1/2} \\ &\quad \cdot \left( \int_0^1 \phi^2(s-v) ds \right)^{1/2} \leq \kappa^2, \\ |\Psi_h(u, v)| &\leq \left( \int_0^1 \Psi_{h-1}^2(u, s) ds \right)^{1/2} \left( \int_0^1 \phi^2(s-v) ds \right)^{1/2} \\ &\leq \kappa \|\Psi_{h-1}\|_\infty, \quad \text{for } h \geq 3. \end{aligned}$$

Hence, Lemma 2(i) can be proved inductively.

By the definition of  $\Psi$ , for  $h \geq 2$ ,

$$\Psi_h(u, v) = \int_{[0,1]^h} \phi(u-s_1)\phi(s_1-s_2) \cdots \phi(s_{h-1}-v) ds_1 \cdots ds_{h-1},$$

then we have

$$\frac{\partial \Psi_h(u, v)}{\partial u} = \int_{[0,1]^h} \phi'(u-s)\Psi_{h-1}(s, v) ds,$$

$$\frac{\partial \Psi_h(u, v)}{\partial v} = \int_{[0,1]^h} \Psi_{h-1}(u, s)\phi'(s-v) ds,$$

$$\frac{\partial^2 \Psi_h(u, v)}{\partial u^2} = \int_{[0,1]^h} \phi''(u-s)\Psi_{h-1}(s, v) ds,$$

$$\frac{\partial^2 \Psi_h(u, v)}{\partial v^2} = \int_{[0,1]^h} \Psi_{h-1}(u, s)\phi''(s-v) ds.$$

Let  $d_1 = \max_{-1 \leq s \leq 1} |\phi'(s)|$  and  $d_2 = \max_{-1 \leq s \leq 1} |\phi''(s)|$ , and it holds that

$$\max_{u,v} \left\{ \left| \frac{\partial \Psi_h(u, v)}{\partial u} \right|, \left| \frac{\partial \Psi_h(u, v)}{\partial v} \right| \right\} \leq d_1 \kappa^{h-1}, \quad \text{for } h \geq 1,$$

$$\max_{u \neq v} \left\{ \left| \frac{\partial^2 \Psi_h(u, v)}{\partial u^2} \right|, \left| \frac{\partial^2 \Psi_h(u, v)}{\partial v^2} \right| \right\} \leq d_2 \kappa^{h-1}, \quad \text{for } h \geq 1.$$

By Corollary 1, the covariances  $\gamma(u, v)$  have the expression:

$$\begin{aligned} \gamma(u, v) &= \frac{\sigma^2}{2\rho} e^{-\rho|u-v|} \\ &+ \frac{\sigma^2}{2\rho} \sum_{\ell=1}^{\infty} \int_0^1 \int_0^1 \Psi_{\ell}(u, w) \Psi_{\ell}(v, z) e^{-\rho|w-z|} dw dz. \end{aligned}$$

It follows that, for any  $u, v \in [0, 1]$ ,

$$\begin{aligned} \|\gamma\|_{\infty} &\leq \sigma^2/[2\rho(1-\kappa^2)], \\ |\gamma(u_1, v) - \gamma(u_2, v)| &\leq \frac{\sigma^2}{2\rho} \left( \rho + \frac{d_1\kappa}{1-\kappa^2} \right) |u_1 - u_2|, \\ \max_{u \neq v} \left\{ \left| \frac{\partial^2 \gamma(u, v)}{\partial u^2} \right|, \left| \frac{\partial^2 \gamma(u, v)}{\partial u \partial v} \right|, \left| \frac{\partial^2 \gamma(u, v)}{\partial v^2} \right| \right\} \\ &\leq \frac{\sigma^2}{2\rho} \left( \rho^2 + \frac{d}{1-\kappa^2} \right), \end{aligned}$$

where  $d = \max\{d_1^2, d_2\kappa\}$ .

For  $h \geq 1$ , the autocovariances  $\gamma_h(u, v) = \text{Cov}(X_t(u), X_{t+h}(v))$  are given by

$$\begin{aligned} \gamma_h(u, v) &= \frac{\sigma^2}{2\rho} \int_0^1 e^{-\rho|u-s|} \Psi_h(v, s) ds \\ &+ \frac{\sigma^2}{2\rho} \sum_{\ell=1}^{\infty} \int_0^1 \int_0^1 \Psi_{\ell}(u, w) \Psi_{\ell+h}(v, z) e^{-\rho|w-z|} dw dz. \end{aligned}$$

Hence, for  $h \geq 1$ , we have

$$\begin{aligned} \|\gamma_h\|_{\infty} &\leq \sigma^2 \kappa^{|h|} / [2\rho(1-\kappa^2)], \\ \left\| \frac{\partial \gamma_h(u, v)}{\partial u} \right\|_{\infty} &\leq \frac{\sigma^2}{2\rho} \left( \rho + \frac{d_1\kappa}{1-\kappa^2} \right) \kappa^h, \\ \left\| \frac{\partial \gamma_h(u, v)}{\partial v} \right\|_{\infty} &\leq \frac{\sigma^2}{2\rho} \cdot \frac{d_1\kappa^{h-1}}{1-\kappa^2}, \\ \max_{u \neq v} \left\{ \left| \frac{\partial^2 \gamma_h(u, v)}{\partial u^2} \right|, \left| \frac{\partial^2 \gamma_h(u, v)}{\partial u \partial v} \right|, \left| \frac{\partial^2 \gamma_h(u, v)}{\partial v^2} \right| \right\} \\ &\leq \frac{\sigma^2}{2\rho} \left( \rho^2 + \frac{d}{1-\kappa^2} \right) \kappa^h. \end{aligned}$$

Since  $\gamma_h(u, v) = \gamma_{-h}(v, u)$ , the proof of Lemma 2(ii), 2(iii) and 2(iv) is complete. ■

**Lemma 3.** The CFAR(1) process  $X = (X_t, t \in \mathbb{Z})$  satisfies (3). Let  $\varepsilon_0^*$  be an i.i.d. copy of  $\varepsilon_0$ , and  $X_t^*$  be obtained by replacing  $\varepsilon_0$  with  $\varepsilon_0^*$  in the definition of  $X_t$ . If  $f : [0, 1]^2 \rightarrow \mathbb{R}$  be a continuous function, and set  $\|f\|_{\infty} := \max_{u,v} |f(u, v)|$ . Define

$$\begin{aligned} Y_t &= \int_0^1 \int_0^1 f(u, v) X_t(u) X_t(v) dudv, \quad \text{and} \\ Y_t^* &= \int_0^1 \int_0^1 f(u, v) X_t^*(u) X_t^*(v) dudv. \end{aligned}$$

Then, for each  $q \in \mathbb{N}$  and  $q \geq 1$ , and  $t \geq 0$ ,

$$\|Y_t - Y_t^*\|_q \leq \frac{\sqrt{2}\sigma^2}{\rho\sqrt{1-\kappa^2}} [(2q-1)!!]^{1/q} \|f\|_{\infty} \cdot \kappa^t.$$

**Proof.** By Cauchy–Schwarz inequality,

$$\begin{aligned} \|Y_t - Y_t^*\|_q &\leq 2 \left\| \int_0^1 \int_0^1 f(u, v) [X_t(u) - X_t^*(u)] X_t(v) dudv \right\|_q \\ &\leq 2 \|f\|_{\infty} \int_0^1 \int_0^1 \| [X_t(u) - X_t^*(u)] X_t(v) \|_q dudv \end{aligned}$$

$$\begin{aligned} &\leq 2 \|f\|_{\infty} \int_0^1 \int_0^1 \left\| \left[ \int_0^1 \Psi_t(u, w) (\varepsilon_t(w) - \varepsilon_t^*(w)) \right] dw \right\|_{2q} \\ &\quad \cdot \|X_t(v)\|_{2q} dudv \\ &\leq \frac{\sqrt{2}\sigma^2}{\rho\sqrt{1-\kappa^2}} [(2q-1)!!]^{1/q} \|f\|_{\infty} \cdot \kappa^t. \quad \blacksquare \end{aligned}$$

**Proof of Theorem 3.** Define  $\delta_t(\cdot) = X_t(\cdot) - \tilde{X}_t(\cdot)$ . For any function  $f \in C[-1, 1]$ , denote  $(f * g)(u) = \int_0^1 f(u-v)g(v)dv$ . We first observe that  $X_t$  can be decomposed as

$$\begin{aligned} X_t(s) &= \sum_{j=1}^k \beta_{k,j} (B_{k,j} * \tilde{X}_{t-1})(s) + (r_k * \tilde{X}_{t-1})(s) \\ &\quad + \varepsilon_t(s) + (\phi * \delta_t)(s). \end{aligned}$$

Let  $\tilde{\mathbf{v}}_t$  and  $\tilde{\mathbf{w}}_t$  be two  $(N+1)$ -dimensional vectors whose  $i$ th entries are  $(r_k * \tilde{X}_{t-1})((i-1)/N)$  and  $(\phi * \delta_t)((i-1)/N)$  respectively. Let  $\boldsymbol{\varepsilon}_t$  be the  $(N+1)$ -dimensional vector whose  $i$ th entry is  $\varepsilon_t((i-1)/N)$ . The estimator  $\hat{\boldsymbol{\beta}}_k$  can be decomposed as

$$\hat{\boldsymbol{\beta}}_k = \tilde{\boldsymbol{\beta}}_k + \left( \sum_{t=2}^T \mathbf{M}'_t \boldsymbol{\Sigma}^{-1} \mathbf{M}_t \right)^{-1} \left( \sum_{t=2}^T \mathbf{M}'_t \boldsymbol{\Sigma}^{-1} (\tilde{\mathbf{v}}_t + \boldsymbol{\varepsilon}_t + \tilde{\mathbf{w}}_t) \right).$$

We claim that under the condition  $T = o(N^2)$

$$\frac{1}{\sqrt{T}} \sum_{t=2}^T \mathbf{M}'_t \boldsymbol{\Sigma}^{-1} \tilde{\mathbf{w}}_t = o_p(1). \tag{A.4}$$

Let  $\boldsymbol{\mu}_{k,N} = E(\mathbf{M}'_t \boldsymbol{\Sigma}^{-1} \tilde{\mathbf{v}}_t)$ ,  $\boldsymbol{\Gamma}_{k,N} = E(\mathbf{M}'_t \boldsymbol{\Sigma}^{-1} \mathbf{M}_t)$ , and  $\mathbf{b}_{k,N} = \boldsymbol{\Gamma}_{k,N}^{-1} \boldsymbol{\mu}_{k,N}$ , then

$$\begin{aligned} \hat{\boldsymbol{\beta}}_k &= \tilde{\boldsymbol{\beta}}_k + \mathbf{b}_{k,N} + \boldsymbol{\Gamma}_{k,N}^{-1} \frac{1}{T} \sum_{t=2}^T (\mathbf{M}'_t \boldsymbol{\Sigma}^{-1} \tilde{\mathbf{v}}_t - \boldsymbol{\mu}_{k,N}) \\ &\quad + \boldsymbol{\Gamma}_{k,N}^{-1} \frac{1}{T} \sum_{t=2}^T \mathbf{M}'_t \boldsymbol{\Sigma}^{-1} \boldsymbol{\varepsilon}_t \\ &\quad - \boldsymbol{\Gamma}_{k,N}^{-1} \frac{1}{T} \sum_{t=2}^T (\mathbf{M}'_t \boldsymbol{\Sigma}^{-1} \mathbf{M}_t - \boldsymbol{\Gamma}_{k,N}) \mathbf{b}_{k,N} + o_p(T^{-1/2}). \end{aligned} \tag{A.5}$$

We claim that

$$\begin{aligned} \hat{\boldsymbol{\beta}}_k &= \tilde{\boldsymbol{\beta}}_k + \mathbf{b}_k + \boldsymbol{\Gamma}_k^{-1} \frac{1}{T} \sum_{t=2}^T (\mathbf{u}_t - \boldsymbol{\mu}_k) + \boldsymbol{\Gamma}_k^{-1} \frac{1}{T} \sum_{t=2}^T \mathbf{z}_t \\ &\quad - \boldsymbol{\Gamma}_k^{-1} \frac{1}{T} \sum_{t=2}^T (\mathbf{A}_t - \boldsymbol{\Gamma}_k) \mathbf{b}_k + o_p(T^{-1/2}). \end{aligned} \tag{A.6}$$

If we represent  $\varepsilon_t(u)$  as

$$\varepsilon_t(u) = \varepsilon_t(0) e^{-\rho u} + \sigma \int_0^u e^{-\rho(u-v)} dW_t(v),$$

then similar to the proof of Lemma 1, we get

$$\begin{aligned} (\mathbf{z}_t)_j &= \int_0^1 \left[ \frac{2\rho}{\sigma^2} B_{k,j}(-u) \varepsilon_t(0) + \frac{1}{\sigma} \int_0^1 (\rho B_{k,j}(v-u) \right. \\ &\quad \left. + B'_{k,j}(v-u)) dW_t(v) \right] X_{t-1}(u) du. \end{aligned} \tag{A.7}$$

The vector-valued process  $\mathbf{u}_t + \mathbf{z}_t + \mathbf{A}_t \mathbf{b}_k$  is a stationary process. For any fixed unit vector  $\boldsymbol{\theta} \in \mathbb{R}^k$ , by Lemma 3, the physical dependence measures defined in Wu (2005) of the process  $\boldsymbol{\theta}'(\mathbf{u}_t + \mathbf{z}_t + \mathbf{A}_t \mathbf{b}_k)$  decay geometrically fast, and therefore the martingale approximation used in Wu (2005) can be applied to

obtain the central limit theorem for  $\sum_{t=2}^T \theta'(\mathbf{u}_t + \mathbf{z}_t + \mathbf{A}_t \mathbf{b}_k)$ . By the Cramer–Wold device, we have

$$\sqrt{T}(\widehat{\boldsymbol{\beta}}_k - \widetilde{\boldsymbol{\beta}}_k - \mathbf{b}_k) \xrightarrow{d} N(\mathbf{0}, \boldsymbol{\Gamma}_k^{-1} \boldsymbol{\Upsilon}_k \boldsymbol{\Gamma}_k^{-1}),$$

where  $\boldsymbol{\Upsilon}_k$  is the long-run variance of the process  $\mathbf{u}_t + \mathbf{z}_t + \mathbf{A}_t \mathbf{b}_k$ . Now we prove the claim (A.4). We proceed by showing that each entry of the vector converges to zero in probability. The  $j$ th entry of  $\mathbf{M}_t \boldsymbol{\Sigma}^{-1} \widetilde{\mathbf{w}}_t$  can be written as

$$\int_0^1 \int_0^1 A_N(B_{k,j}, \phi)(u, v) \widetilde{X}_{t-1}(u) \delta_t(v) dudv.$$

By Lemma 1, there exists a constant  $c_1 > 0$  such that

$$|A_N(B_j, \phi)(u, v)| \leq c_1, \quad \text{for all } u, v, N. \tag{A.8}$$

Let

$$\widetilde{\gamma}_h(u, v) = \text{Cov}(X_t(u), \delta_{t+h}(v)),$$

$$\widetilde{\gamma}_h(u, v) = \text{Cov}(\delta_t(u), \delta_{t+h}(v)).$$

By Lemma 2, there exists a constant  $c_2 > 0$  such that

$$\|\widetilde{\gamma}_h\|_\infty \leq c_2 \kappa^h / N, \quad \|\widetilde{\gamma}_h\|_\infty \leq c_2 \kappa^h / N.$$

It follows that

$$E \int_0^1 \int_0^1 |A_N(B_{k,j}, \phi)(u, v) \delta_t(u) \delta_t(v)| dudv = O(N^{-1}),$$

so it suffices to prove  $\sum_{t=2}^T y_{jt} = o_p(T^{1/2})$ , where

$$y_{jt} = \int_0^1 \int_0^1 A_N(B_{k,j}, \phi)(u, v) X_{t-1}(u) \delta_t(v) dudv.$$

By (A.8),

$$E y_{jt} = \int_0^1 \int_0^1 A_N(B_{k,j}, \phi)(u, v) [\gamma_1(u, v) - \widetilde{\gamma}_1(u, v)] dudv = O(1/N). \tag{A.9}$$

The autocovariance is

$$\begin{aligned} \text{Cov}(y_{jt}, y_{j,t+h}) &= \int_{[0,1]^4} A_N(B_{k,j}, \phi)(u_1, v_1) A_N(B_{k,j}, \phi)(u_2, v_2) \\ &\quad \times [\gamma_h(u_1, u_2) \widetilde{\gamma}_h(v_1, v_2) \\ &\quad + \widetilde{\gamma}_{h+1}(u_1, v_2) \widetilde{\gamma}_{-h+1}(u_2, v_1)] du_1 du_2 du_3 du_4. \end{aligned}$$

By (A.8) and Lemma 2, we know there exists a constant  $c_3 > 0$  such that

$$|\text{Cov}(w_{jt}, w_{j,t+h})| \leq c_3 \kappa^{2h} / N.$$

It follows that

$$\text{Var} \left( \sum_{t=2}^T y_{jt} \right) = O(T/N). \tag{A.10}$$

Combining (A.9) and (A.10), and using the condition  $T = o(N^2)$ , we have  $\sum_{t=2}^T y_{jt} = o_p(T^{1/2})$ , and the proof of claim (A.4) is complete.

The difference between the two expressions in (A.5) and (A.6) is due to the approximation of  $X_t$  by  $\widetilde{X}_t$ , so the proof of claim (A.6) is in the same fashion of that of claim (A.4). We omit the details. ■

**Lemma 4.** *The distance of two adjacent knots for uniform cubic B-spline functions defined on  $[-1, 1]$  with  $k$  degrees of freedom is  $1/m$ , and  $m = (k - 3)/2$ . Define two  $(m + 4) \times (m + 4)$  matrices  $\mathbf{P}_s$  and  $\mathbf{V}_s$ :*

$$(\mathbf{P}_s)_{qj} = \int_{s/m}^{1+s/m} \int_{s/m}^{1+s/m} e^{-\rho|u-v|} B'_{k,q+m-1}(u) B'_{k,j+m-1}(v) dudv,$$

$$(\mathbf{V}_s)_{qj} = \rho^2 \int_{s/m}^{1+s/m} \int_{s/m}^{1+s/m} e^{-\rho|u-v|} B_{k,q+m-1}(u) B_{k,j+m-1}(v) dudv,$$

for  $q, j = 1, \dots, m + 4$ . Let  $\mathbf{b} = (b_0, \dots, b_{m+3})'$  be a unit vector. There exist constants  $c_1$  and  $c_2$  such that if  $s \in I_\alpha$ ,  $\sum_{j=0}^3 b_j^2 \geq \frac{1}{3m}$ , and  $m \geq c_1$ , then

$$\mathbf{b}'(\mathbf{P}_s + \mathbf{V}_s)\mathbf{b} \geq c_2 \sum_{j=0}^3 b_j^2.$$

**Proof.** Rescale the uniform B-spline functions,

$$B_q(s) = B_{k,m+q}(s/m) = [q, q + 1, \dots, q + 4](\cdot - s)_+^3,$$

for  $q = 0, \dots, m + 3$ .

The support of  $B_q(\cdot)$  is  $[q, q + 4]$  and we denote  $B(u) \equiv B_{-2}(u)$ .  $\mathbf{P}_s$  and  $\mathbf{V}_s$  can be written as

$$(\mathbf{P}_s)_{qj} = \int_s^{m+s} \int_s^{m+s} e^{-\rho|u-v|/m} B'_{q-1}(u) B'_{j-1}(v) dudv,$$

$$(\mathbf{V}_s)_{qj} = \frac{\rho^2}{m^2} \int_s^{m+s} \int_s^{m+s} e^{-\rho|u-v|/m} B_{q-1}(u) B_{j-1}(v) dudv,$$

for  $q, j = 1, \dots, m + 4$ . For a fixed  $3 < s < 4$ , there are  $m + 4$  spline functions which are not identically zero on the interval  $[s, m + s]$ :  $B_0(u), B_1(u), \dots, B_{m+3}(u)$ . For any numbers  $b_0, b_1, b_2, b_3$ , there exists a constant  $c_3 > 0$  such that  $\int_s^4 \left[ \sum_{j=0}^3 b_j B_j(s) \right]^2 ds \geq c_3 \sum_{j=0}^3 b_j^2$ . Thus there exists an  $s \in [3, 4]$  such that

$$\left| \sum_{j=0}^3 b_j B_j(s) \right| \geq \sqrt{c_3} \cdot \left( \sum_{j=0}^3 b_j^2 \right)^{1/2}.$$

On the other hand, there exists a constant  $c_4$  such that for all  $s \in [3, 4]$ ,  $\left| \sum_{j=0}^3 b_j B'_j(s) \right| \leq c_4 \cdot \left( \sum_{j=0}^3 b_j^2 \right)^{1/2}$ . As a result, there exists an interval  $I_\alpha$  of length  $c_5 > 0$ , which is contained in  $[3, 4]$ , such that for each  $s$  in this interval

$$\left| \sum_{j=0}^3 b_j B_j(s) \right| \geq c_6 \cdot \left( \sum_{j=0}^3 b_j^2 \right)^{1/2},$$

where  $0 < c_6 < 1$  is an absolute constant.

Define the functions:

$$f_1(u) = \sum_{j=0}^3 b_j B_j(u), \quad f_2(u) = \sum_{j=4}^{m-1} b_j B_j(u),$$

$$f_3(u) = \sum_{j=m}^{m+3} b_j B_j(u),$$

and let  $f(u) = f_1(u) + f_2(u) + f_3(u)$ . Using the identity

$$e^{-\rho|u|/m} = \frac{1}{\pi} \int_{-\infty}^{\infty} e^{iu\lambda} \cdot \frac{m/\rho}{1 + (m\lambda/\rho)^2} d\lambda,$$

we have

$$\begin{aligned} \mathbf{b}'\mathbf{P}_s\mathbf{b} &= \frac{1}{\pi} \int_s^{m+s} \int_s^{m+s} f'(u) f'(v) \int_{-\infty}^{\infty} e^{i(u-v)\lambda} \\ &\quad \cdot \frac{m/\rho}{1 + (m\lambda/\rho)^2} d\lambda dudv \\ &= \frac{1}{\pi} \int_{-\infty}^{\infty} \left| \int_s^{m+s} f'(u) e^{iu\lambda} du \right|^2 \frac{m/\rho}{1 + (m\lambda/\rho)^2} d\lambda \\ &= \frac{1}{\pi} \int_{-\infty}^{\infty} \left| -f_1(s) e^{is\lambda} + f_3(m+s) e^{i(m+s)\lambda} \right. \\ &\quad \left. - (i\lambda) \int_s^{m+s} f(u) e^{iu\lambda} du \right|^2 \frac{m/\rho}{1 + (m\lambda/\rho)^2} d\lambda. \end{aligned}$$

Similarly,

$$\mathbf{b}'\mathbf{V}_s\mathbf{b} = \frac{\rho^2}{\pi m^2} \int_{-\infty}^{\infty} \left| \int_s^{m+s} f(u)e^{iu\lambda} du \right|^2 \frac{m/\rho}{1 + (m\lambda/\rho)^2} d\lambda.$$

Neuman (1981) showed the Fourier transform of the central spline  $B_{-2}(u)$  is

$$\int_{-2}^2 B_{-2}(u)e^{iu\lambda} du = \left[ \frac{2 \sin(\lambda/2)}{\lambda} \right]^4.$$

Let  $g(\lambda) = \sum_{j=4}^{m-1} b_j e^{i(j+2)\lambda}$ , we have

$$\begin{aligned} \int_s^{m+s} f(u)e^{iu\lambda} du &= \int_s^{m+s} f_1(u)e^{iu\lambda} du + \int_s^{m+s} f_3(u)e^{iu\lambda} du \\ &\quad + g(\lambda) \int_0^4 B_{-2}(u)e^{iu\lambda} du \\ &= \int_s^{m+s} f_1(u)e^{iu\lambda} du + \int_s^{m+s} f_3(u)e^{iu\lambda} du \\ &\quad + g(\lambda) \left[ \frac{2 \sin(\lambda/2)}{\lambda} \right]^4. \end{aligned} \tag{A.11}$$

Note that

$$\left| \int_s^{m+s} f_1(u)e^{iu\lambda} du \right| \leq \int_3^7 \sum_{j=0}^3 |b_j| B_j(u) du \leq 2, \tag{A.12}$$

and a similar inequality holds for the Fourier transform of  $f_3(u)$ .

We will consider two cases.

Case 1:  $\max_{\lambda \in [0, 2/m]} |g(\lambda)| \leq m |f_1(s)|/3$ . On the interval  $\lambda \in [\pi/(6m), 11\pi/(6m)]$ , by law of cosines, it holds that

$$|-f_1(s)e^{is\lambda} + f_3(m+s)e^{i(m+s)\lambda}| \geq |f_1(s) \sin(m\lambda)| \geq |f_1(s)|/2,$$

and with (A.11) and (A.12), we have

$$\begin{aligned} &\left| -f_1(s)e^{is\lambda} + f_3(m+s)e^{i(m+s)\lambda} - (i\lambda) \int_s^{m+s} f(u)e^{iu\lambda} du \right| \\ &\geq \frac{|f_1(s)|}{2} - \frac{22\pi}{3m} - \frac{|f_1(s)|}{3} \\ &= \frac{|f_1(s)|}{6} - \frac{22\pi}{3m}. \end{aligned}$$

It follows that

$$\begin{aligned} \mathbf{b}'\mathbf{P}_s\mathbf{b} &\geq \frac{1}{\pi} \int_{\pi/(6m)}^{11\pi/(6m)} \left( \frac{|f_1(s)|}{6} - \frac{22\pi}{3m} \right)^2 \frac{m/\rho}{1 + (m\lambda/\rho)^2} d\lambda \\ &= \frac{1}{\pi} \int_{\pi/6}^{11\pi/6} \left( \frac{|f_1(s)|}{6} - \frac{22\pi}{3m} \right)^2 \frac{\rho}{\rho^2 + \lambda^2} d\lambda \\ &\geq \frac{60\rho}{36\rho^2 + 121\pi^2} \left( \frac{|f_1(s)|}{6} - \frac{22\pi}{3m} \right)^2. \end{aligned}$$

Recall that  $\sum_{j=0}^3 b_j^2 \geq \frac{1}{3m}$ , so

$$|f_1(s)| \geq c_4 \left( \sum_{j=0}^3 b_j^2 \right)^{1/2} \geq c_6/\sqrt{3m}.$$

Therefore, for Case 1, there exists a constant  $c_7 > 0$ , such that when  $m \geq c_7$ , it holds that

$$\mathbf{b}'\mathbf{P}_s\mathbf{b} \geq \frac{5\rho}{109\rho^2 + 364\pi^2} |f_1(s)|^2 \geq \frac{5\rho c_6^2}{109\rho^2 + 364\pi^2} \sum_{j=0}^3 b_j^2.$$

Case 2:  $\max_{\lambda \in [0, 2/m]} |g(\lambda)| > m |f_1(s)|/3$ . This implies that there exists a  $\lambda_0 \in [0, 2/m]$  such that  $|g(\lambda_0)| \geq m |f_1(s)|/3 \geq$

$c_6\sqrt{m}/6$ . Since  $\mathbf{b}$  is a unit vector,  $|g(\lambda)| \leq \sqrt{m}$ , and it follows that  $f_1(s) \leq 3/\sqrt{m}$ . By Zygmund (2002), the derivative of  $g(\lambda)$  satisfies  $|g'(\lambda)| \leq m^{3/2}$ . Therefore, on an interval  $I_1 \subset [0, 3/m]$  of length  $c_6/(12m)$ ,  $|g(\lambda)| \geq c_6\sqrt{m}/12$ . On this interval, it holds that

$$\left| \int_s^{m+s} f(u)e^{iu\lambda} du \right| \geq \frac{c_6\sqrt{m}}{12} \left[ \frac{2 \sin(\lambda/2)}{\lambda} \right]^4 - 4.$$

There exists a  $c_8 > 0$  such that when  $m \geq c_8$ ,

$$\left| \int_s^{m+s} f(u)e^{iu\lambda} du \right| \geq \frac{c_6\sqrt{m}}{13}.$$

It follows that

$$\begin{aligned} \mathbf{b}'\mathbf{V}_s\mathbf{b} &\geq \frac{\rho^2}{\pi m^2} \int_{I_1} \left| \int_s^{m+s} f(u)e^{iu\lambda} du \right|^2 \frac{m/\rho}{1 + (m\lambda/\rho)^2} d\lambda \\ &\geq \frac{\rho^2}{\pi m^2} \int_{m/1}^{c_4^2 m} \frac{\rho}{\rho^2 + \lambda^2} d\lambda \\ &\geq \frac{c_6^3 \rho^3}{2028\pi(\rho^2 + 9)} \cdot \frac{1}{m} \geq \frac{c_6^3 \rho^3}{18252\pi(\rho^2 + 9)} \cdot |f_1(s)|^2 \\ &\geq \frac{c_6^5 \rho^3}{18252\pi(\rho^2 + 9)} \cdot \sum_{j=0}^3 b_j^2. \end{aligned}$$

Note that

$$\begin{aligned} \mathbf{b}'\mathbf{P}_s\mathbf{b} &= \sum_{q=0}^{m+3} \sum_{j=0}^{m+3} \int_{s/m}^{1+s/m} \int_{s/m}^{1+s/m} b_i b_j e^{-\rho|u-v|/m} B'_{q+m}(u) B'_{j+m}(v) dudv \\ &= E \left( \sum_{q=0}^{m+3} \int_s^{1+s/m} b_q e_t(u) B'_{q+m}(u) du \right)^2 \geq 0. \end{aligned}$$

Similarly, we have  $\mathbf{b}'\mathbf{V}_s\mathbf{b} \geq 0$ , so both  $\mathbf{P}_s$  and  $\mathbf{V}_s$  are non-negative definite. Hence, the proof is completed by setting  $c_1 = \max\{c_7, c_8\}$ , and

$$c_2 = \frac{c_6^5 \rho \min\{\rho^2, 1\}}{18252\pi(\rho^2 + 9)}. \blacksquare$$

**Lemma 5.** The minimum eigenvalue of  $\mathbf{\Gamma}_k$ , defined in Theorem 3 is  $O(k^{-1})$ .

**Proof.**  $\mathbf{\Gamma}_k$  could be written as sum of the following four matrices.

$$\mathbf{\Gamma}_k = \sum_{i=1}^4 \mathbf{G}_{k,i}.$$

$$(\mathbf{G}_{k,1})_{ij} = \int_0^1 \int_0^1 \int_0^1 \gamma_0(u, v) B'_{k,i}(s-u) B'_{k,j}(s-v) dudvds,$$

$$(\mathbf{G}_{k,2})_{ij} = \rho^2 \int_0^1 \int_0^1 \int_0^1 \gamma_0(u, v) B_{k,i}(s-u) B_{k,j}(s-v) dudvds,$$

$$(\mathbf{G}_{k,3})_{ij} = \rho \int_0^1 \int_0^1 \gamma_0(u, v) B_{k,i}(-u) B_{k,j}(-v) dudv,$$

$$(\mathbf{G}_{k,4})_{ij} = \rho \int_0^1 \int_0^1 \gamma_0(u, v) B_{k,i}(1-u) B_{k,j}(1-v) dudv,$$

for  $i, j = 1, \dots, k$ . Define  $\mathbf{W}_{k,1}$  and  $\mathbf{W}_{k,2}$ .

$$(\mathbf{W}_{k,1})_{ij} = \frac{\sigma^2}{2\rho} \int_0^1 \int_0^1 \int_0^1 e^{-\rho|u-v|} B'_{k,i}(s-u) B'_{k,j}(s-v) dudvds,$$

$$(\mathbf{W}_{k,2})_{ij} = \frac{\rho\sigma^2}{2} \int_0^1 \int_0^1 \int_0^1 e^{-\rho|u-v|} B_{k,i}(s-u) B_{k,j}(s-v) dudvds,$$

for  $i, j = 1, \dots, k$ . For any  $\mathbf{b} = \{b_1, \dots, b_k\} \in \mathbb{R}^k$ ,

$$\begin{aligned} \mathbf{b}'\mathbf{G}_{k,1}\mathbf{b} &= \int_0^1 \mathbb{E} \left( \sum_{i=1}^k \int_0^1 b_i X_t(u) B'_{k,i}(s-u) du \right)^2 ds \\ &= \int_0^1 \mathbb{E} \left[ \sum_{i=1}^k \int_0^1 b_i \left( \int_0^1 \phi(u-v) X_{t-1}(v) dv \right. \right. \\ &\quad \left. \left. + \varepsilon_{t-1}(u) B'_{k,i}(-u) \right) du \right]^2 ds \\ &\geq \int_0^1 \mathbb{E} \left[ \sum_{i=1}^k \int_0^1 b_i \varepsilon_{t-1}(u) B'_{k,i}(-u) du \right]^2 ds \\ &\geq \mathbf{b}'\mathbf{W}_{k,1}\mathbf{b} \geq 0. \end{aligned}$$

So  $\mathbf{G}_{k,1} \geq \mathbf{W}_{k,1} \geq \mathbf{0}$ . In a similar way, we can show that  $\mathbf{G}_{k,2} \geq \mathbf{W}_{k,2} \geq \mathbf{0}$ . Thus  $\mathbf{\Gamma}_k \geq \mathbf{W}_{k,1} + \mathbf{W}_{k,2}$ . Let  $\mathbf{b} = (b_0, \dots, b_{2m+2})$  be a unit vector, and

$$\mathcal{D}_1 = \left\{ 0 \leq j \leq 2m-1 : \sum_{l=j}^{j+3} b_l^2 \geq 1/(3m) \right\},$$

and  $\mathcal{D}_2 = \{0, 1, \dots, 2m-1\} \setminus \mathcal{D}_1$ . Since  $\sum_{j \in \mathcal{D}_2} \sum_{l=j}^{j+3} b_l^2 \leq 2m \cdot 1/(3m) = 2/3$ , it holds that  $\sum_{j \in \mathcal{D}_1} \sum_{l=j}^{j+3} b_l^2 \geq 1/3$ . For each  $j \in \mathcal{D}_1$  and  $j \leq m-1$ , by Lemma 4, There exists an interval  $I_j \subset [-1 + j/m, -1 + (j+1)/m]$  of length  $c_5/m$ , such that for each  $s \in I_j$

$$\begin{aligned} &\int_0^1 \int_0^1 \sum_{h,l=0}^{2m+2} b_h b_l [B'_{k,h}(s-u) B'_{k,l}(s-v) \\ &\quad + \rho^2 B_{k,h}(s-u) B_{k,l}(s-v)] e^{-\rho|u-v|} dudv \geq c_6 \sum_{l=j}^{j+3} b_l^2. \end{aligned}$$

It follows that

$$\begin{aligned} &\int_0^1 \int_0^1 \int_0^1 \sum_{h,l=0}^{2m+2} b_h b_l [B'_{k,h}(s-u) B'_{k,l}(s-v) \\ &\quad + \rho^2 B_{k,h}(s-u) B_{k,l}(s-v)] e^{-\rho|u-v|} dudvds \\ &\geq \frac{c_3 c_6}{m} \sum_{j \in \mathcal{D}_1, j \leq m-1} \sum_{l=j}^{j+3} b_l^2. \end{aligned}$$

By applying reverting the order of the rows and the columns, we can show that for each  $j \in \mathcal{D}_1$  and  $j \geq m$ , there exists an interval  $I_j \subset [-1 + j/m, -1 + (j+1)/m]$  of length  $c_5/m$ , such that for each  $s \in I_j$

$$\begin{aligned} &\int_0^1 \int_0^1 \sum_{h,l=0}^{2m+2} b_h b_l [B'_{k,h}(s-u) B'_{k,l}(s-v) \\ &\quad + \rho^2 B_{k,h}(s-u) B_{k,l}(s-v)] e^{-\rho|u-v|} dudv \geq c_6 \sum_{l=j}^{j+3} b_l^2. \end{aligned}$$

So similarly,

$$\begin{aligned} &\int_0^1 \int_0^1 \int_0^1 \sum_{h,l=0}^{2m+2} b_h b_l [B'_{k,h}(s-u) B'_{k,l}(s-v) \\ &\quad + \rho^2 B_{k,h}(s-u) B_{k,l}(s-v)] e^{-\rho|u-v|} dudvds \\ &\geq \frac{c_3 c_6}{m} \sum_{j \in \mathcal{D}_1, j \geq m} \sum_{l=j}^{j+3} b_l^2. \end{aligned}$$

Therefore,

$$\mathbf{b}'(\mathbf{W}_{k,1} + \mathbf{W}_{k,2})\mathbf{b} \geq \frac{c_3 c_6}{6m},$$

and the proof is complete. ■

**Lemma 6.** Assume  $\mathcal{E} : [0, 1]^2 \rightarrow [0, 1]$  is a function, and satisfies

$$\begin{aligned} \|\mathcal{E}^{(1)}\|_\infty &= \sup_{0 \leq u_1, u_2, v \leq 1, u_1 \neq u_2} \left\{ \left| \frac{\mathcal{E}(u_1, v) - \mathcal{E}(u_2, v)}{u_1 - u_2} \right|, \right. \\ &\quad \left. \left| \frac{\mathcal{E}(v, u_1) - \mathcal{E}(v, u_2)}{u_1 - u_2} \right| \right\} < \infty; \\ \|\mathcal{E}^{(2)}\|_\infty &= \sup_{0 \leq u, v \leq 1, u \neq v} \left\{ \left| \frac{\partial^2 \mathcal{E}(u, v)}{\partial u^2} \right|, \left| \frac{\partial^2 \mathcal{E}(u, v)}{\partial u \partial v} \right|, \right. \\ &\quad \left. \left| \frac{\partial^2 \mathcal{E}(u, v)}{\partial v^2} \right| \right\} < \infty. \end{aligned} \tag{A.13}$$

Define the  $k \times k$  matrix  $\mathbf{A}$  with the  $(j, l)$ th entry

$$\int_0^1 \int_0^1 \int_0^1 B'_{k,j}(s-u) B'_{k,l}(s-v) \mathcal{E}(u, v) dudvds. \tag{A.14}$$

Then  $\|\mathbf{A}\| = O(k^{-1})$ .

**Proof.** Let  $\mathcal{J}_j$  be the support of  $B_{k,j}$ , and let  $\mathcal{J}_l$  be the set of  $s$  which makes the integral

$$\int_0^1 \int_0^1 B'_{k,j}(s-u) B'_{k,l}(s-v) \mathcal{E}(u, v) dudv \tag{A.15}$$

nonzero. Define two subsets of  $\mathcal{J}_j$ ,

$$\mathcal{J}_{j1} := \{s : s \in \mathcal{J}_j, (\mathcal{J}_j \cup \mathcal{J}_l) \not\subset [s-1, s]\},$$

and  $\mathcal{J}_{j2} := \mathcal{J}_j \setminus \mathcal{J}_{j1}$ . Use  $\lambda$  to denote the Lebesgue measure. Let  $k = 2m + 3$ , and note that  $1/m$  is the mesh size. For each  $1 \leq j \leq k$ , we consider the entries in the  $j$ th row of the matrix  $\mathbf{A}$ . There are three cases.

Case 1. If  $||l - j| - m| \leq 3$ , then  $\lambda(\mathcal{J}_{j1}) \leq 7/m$ , and it follows that  $|(\mathbf{A})_{jl}| \leq C_1/k$ , where  $C_1$  is a constant depending only on  $\mathcal{E}$ . Observe that for each  $j$ , the inequality  $||l - j| - m| \leq 3$  only holds for at most 14 different values of  $l$ .

Case 2. If  $|j - l| \leq 3$ , then  $\lambda(\mathcal{J}_{j1}) \leq 7/m$ , and integrating (A.15) over  $s \in \mathcal{J}_{j1}$  leads to a value bounded by  $C_1/(2k)$ . For  $s \in \mathcal{J}_{j2}$ , the integral (A.15) can be written as

$$\begin{aligned} &\int_0^1 \int_0^1 B'_{k,j}(s-u) B'_{k,l}(s-v) \mathcal{E}(u, v) dudv \\ &= \int_{\mathcal{J}_j} \int_{\mathcal{J}_l} B'_{k,j}(u) B'_{k,l}(v) \mathcal{E}(s-u, s-v) dudv. \end{aligned}$$

Let  $u_j^*$  be the left boundary of  $\mathcal{J}_j$ , then

$$\begin{aligned} &\left| \int_{\mathcal{J}_j} \int_{\mathcal{J}_l} B'_{k,j}(u) B'_{k,l}(v) \mathcal{E}(s-u, s-v) dudv \right| \\ &= \left| \int_{\mathcal{J}_j} \int_{\mathcal{J}_l} B'_{k,j}(u) B'_{k,l}(v) [\mathcal{E}(s-u, s-v) \right. \\ &\quad \left. - \mathcal{E}(s-u_j^*, s-v)] dudv \right| \\ &\leq \int_{\mathcal{J}_j} \int_{\mathcal{J}_l} |B'_{k,j}(u) B'_{k,l}(v)| \cdot \|\mathcal{E}^{(1)}\|_\infty (u - u_j^*) dudv \\ &\leq C_1/(2k). \end{aligned} \tag{A.16}$$

Since  $\lambda(\mathcal{J}_{j2}) \leq 1$ , we have  $|(\mathbf{A})_{jl}| \leq C_1/k$ .



Case 3. If  $3 < |j - l| < m - 3$ , then for each  $s \in \mathcal{S}_{j1}$ , it must hold that either  $\mathcal{S}_j \subset [s - 1, s]$  or  $\mathcal{S}_l \subset [s - 1, s]$ . Similar as (A.16), it holds that for each  $s \in \mathcal{S}_{j1}$ , the integral (A.15) has an absolute value less than  $C_1/k$ . For  $s \in \mathcal{S}_{j2}$ , since  $\mathcal{S}_j$  and  $\mathcal{S}_l$  have no intersection, we have

$$\begin{aligned} \mathcal{E}(s - u, s - v) &= \mathcal{E}(s - u_j^*, s - v) \\ &\quad - \frac{\partial \mathcal{E}(s - u_j^*, s - v_l^*)}{\partial u} (u - u_j^*) \\ &\quad - \frac{\partial \mathcal{E}(s - u_j^*, s - v_l^*)}{\partial v} (v - v_l^*) + R(u, v) \end{aligned}$$

where

$$|R(u, v)| \leq \|\mathcal{E}^{(2)}\|_\infty (u - u_j^* + v - v_l^*)^2, \quad u \in \mathcal{S}_j, \quad v \in \mathcal{S}_l.$$

Therefore

$$\begin{aligned} &\left| \int_{\mathcal{S}_j} \int_{\mathcal{S}_l} B'_{k,j}(u) B'_{k,l}(v) \mathcal{E}(s - u, s - v) \, dudv \right| \\ &= \left| \int_{\mathcal{S}_j} \int_{\mathcal{S}_l} B'_{k,j}(u) B'_{k,l}(v) R(u, v) \, dudv \right| \\ &\leq \int_{\mathcal{S}_j} \int_{\mathcal{S}_l} |B'_{k,j}(u) B'_{k,l}(v)| \cdot \|\mathcal{E}^{(2)}\|_\infty (u - u_j^* + v - v_l^*)^2 \, dudv \\ &\leq C_1/k^2. \end{aligned} \tag{A.17}$$

Since  $\lambda(\mathcal{S}_{j1}) \leq 8/m$  and  $\lambda(\mathcal{S}_{j2}) \leq 1$ , it holds that

$$|(\mathbf{A})_{jl}| \leq \frac{C_1}{k} \cdot \frac{8}{m} + \frac{C_1}{k^2} \leq \frac{C_2}{k^2},$$

where  $C_2$  is a constant which only depends on  $\mathcal{E}$ .

Combining these three cases, we see that there exists a constant  $C_3$  which only depends on  $\mathcal{E}$ , such that  $\|\mathbf{A}\|_\infty \leq C_3/k$ , and  $\|\mathbf{A}\|_1 \leq C_3/k$ . Therefore,

$$\|\mathbf{A}\| \leq \|\mathbf{A}\|_\infty^{1/2} \|\mathbf{A}\|_1^{1/2} \leq C_3/k. \quad \blacksquare$$

**Proof of Theorem 4.** By Corollary 6.26 of Schumaker (1981),

$$\|\phi - \tilde{\phi}_k\|_\infty \leq c_1 k^{-\zeta}, \quad \|\phi'(\cdot) - \tilde{\phi}'_k(\cdot)\|_\infty \leq c_1 k^{-\zeta+1}, \tag{A.18}$$

where  $c_1$  is an absolute constant. There are four terms in the definition of  $\mathcal{A}(B_{k,j}, r_k)$ . We first consider

$$\int_0^1 \int_0^1 \int_0^1 B'_{k,j}(s - u) r'_k(s - v) \gamma(u, v) \, dudvds.$$

Let

$$\mathbb{S}_j = \{s : \lambda(\mathcal{S}_j \cap [s - 1, s]) > 0\},$$

and  $\mathbb{S}_{j1} = \{s \in \mathbb{S}_j : \mathcal{S}_j \not\subset [s - 1, s]\}$ ,  $\mathbb{S}_{j2} = \mathbb{S}_j \setminus \mathbb{S}_{j1}$ . When  $s \in \mathbb{S}_{j1}$ , it holds that

$$\left| \int_0^1 \int_0^1 B'_{k,j}(s - u) r'_k(s - v) \gamma(u, v) \, dudv \right| \leq C_\rho k^{-\zeta+1},$$

where  $C$  is constant which only depends on  $\rho, \sigma^2$  and  $\phi$ . When  $s \in \mathbb{S}_{j2}$  and  $3 < j < k - 2$ , using the fact that  $\int_{\mathcal{S}_j} B'_{k,j}(u) \, du = 0$  and Lemma 2, we have

$$\begin{aligned} &\left| \int_0^1 \int_0^1 B'_{k,j}(s - u) r'_k(s - v) \gamma(u, v) \, dudv \right| \\ &= \left| \int_0^1 \int_{\mathcal{S}_j} B'_{k,j}(u) r'_k(s - v) [\gamma(s - u, v) - \gamma(s - u_j^*, v)] \, dudv \right| \\ &\leq Ck^{-\zeta}. \end{aligned}$$

Noticing that  $\lambda(\mathbb{S}_{j1}) \leq 4/m$ , and combining the previous two cases, we see that

$$\left| \int_0^1 \int_0^1 \int_0^1 B'_{k,j}(s - u) r'_k(s - v) \gamma(u, v) \, dudvds \right| \leq Ck^{-\zeta}.$$

It can be shown that the other three terms in the definition of  $\mathcal{A}(B_{k,j}, r_k)$  have the same order  $O(k^{-\zeta})$ , and hence

$$|(\boldsymbol{\mu}_k)_j| \leq Ck^{-\zeta}.$$

By Lemma 5,

$$\|\mathbf{b}_k\| = \|\boldsymbol{\Gamma}_k^{-1} \boldsymbol{\mu}_k\| = O(k^{-\zeta+3/2}).$$

It follows that

$$\|\mathbf{b}_k(\cdot)\|_\infty \leq \|\mathbf{B}_k(\cdot)\| \|\mathbf{b}_k\|_\infty + \|r_k(\cdot)\|_\infty = O(k^{-\zeta+3/2}).$$

For the statement regarding  $\|\sigma_k^2(\cdot)\|_\infty$ , it suffices to show that

$$\|\boldsymbol{\Sigma}_k\| = O(k).$$

We proceed by calculating the long-run variances of the three processes  $\mathbf{u}_t, \mathbf{z}_t$  and  $\mathbf{A}_t \mathbf{b}_k$ . First we observe that  $\mathbf{z}_t$  is a martingale, and

$$\text{Var}(\mathbf{z}_t) = \boldsymbol{\Gamma}_k.$$

The covariance between  $u_{tj}$  and  $u_{t+h,l}$  is

$$\begin{aligned} \text{Cov}(u_{tj}, u_{t+h,l}) &= \int_{[0,1]^4} \mathcal{A}(B_{k,j}, r_k)(u_1, v_1) \mathcal{A}(B_{k,l}, r_k)(u_2, v_2) \\ &\quad \times [\gamma_h(u_1, u_2) \gamma_h(v_1, v_2) \\ &\quad + \gamma_h(u_1, v_2) \gamma_h(v_1, u_2)] \, du_1 du_2 du_3 du_4. \end{aligned}$$

By (A.18), we have

$$\mathcal{A}(B_{k,j}, r_k)(u, v) \leq C_\rho k^{-\zeta+1}$$

where  $C_\rho$  is constant which only depends on  $\rho$ . The preceding bound, together with Lemma 2 implies that,

$$|\text{Cov}(u_{tj}, u_{t+h,l})| \leq 2C_\rho^2 / (1 - \kappa_2^2) \sigma^4 \cdot k^{-2\zeta+2} \kappa^{2|h|}.$$

Therefore, the operator norm of the long-run variance of  $\mathbf{u}_t$  is of the order  $O(k^{-2\zeta+3})$ . The  $(j_1, j_2)$ th entry of the matrix  $\mathbf{E}(\mathbf{A}_t \mathbf{b}_k \mathbf{b}'_k \mathbf{A}'_{t+h})$  is

$$\begin{aligned} &\sum_{l_1, l_2=1}^k \int_{[0,1]^4} \mathcal{A}(B_{k,l_1}, B_{k,j_1})(u_1, v_1) (\mathbf{b}_k)_{l_1} \\ &\quad \times (\mathbf{b}_k)_{l_2} \mathcal{A}(B_{k,l_2}, B_{k,j_2})(u_2, v_2) \\ &\quad \times [\gamma_h(u_1, u_2) \gamma_h(v_1, v_2) \\ &\quad + \gamma_h(u_1, v_2) \gamma_h(v_1, u_2)] \, du_1 du_2 du_3 du_4. \end{aligned} \tag{A.19}$$

By the definition of  $\mathcal{A}(\cdot, \cdot)$ , the product  $\mathcal{A}(B_{k,j_1}, B_{k,h_1}) \mathcal{A}(B_{k,j_2}, B_{k,h_2})$  can be expanded to 16 terms. We first consider the term

$$\begin{aligned} &\int_{[0,1]^6} B'_{k,l_1}(s_1 - u_1) B'_{k,j_1}(s_1 - v_1) B'_{k,l_2}(s_2 - u_2) B'_{k,j_2}(s_2 - v_2) \\ &\quad \times [\gamma_h(u_1, u_2) \gamma_h(v_1, v_2) \\ &\quad + \gamma_h(u_1, v_2) \gamma_h(v_1, u_2)] \, ds_1 ds_2 du_1 dv_1 du_2 dv_2. \end{aligned} \tag{A.20}$$

Let  $\mathcal{B}_k$  be the set of the pairs  $(j, l)$  such that either  $|j - l| \leq 3$ , or  $|j - l - m| \leq 3$ . Lemma 2 gives bounds on derivatives of  $\gamma_h(u, v)$ . Using these bounds, and a similar argument as the proof of Lemma 6, we can show that there exists a constant  $c_2 > 0$  such that the absolute value of (A.20) can be controlled as

$$\begin{cases} c_2 \kappa^{2h} / k^2 & \text{if } (j_1, l_1) \in \mathcal{B}_k \text{ and } (j_2, l_2) \in \mathcal{B}_k; \\ c_2 \kappa^{2h} / k^3 & \text{if one and only one of } (j_1, l_1) \text{ and } \\ & (j_2, l_2) \text{ belongs to } \mathcal{B}_k; \\ c_2 \kappa^{2h} / k^4 & \text{if } (j_1, l_1) \notin \mathcal{B}_k \text{ and } (j_2, l_2) \notin \mathcal{B}_k. \end{cases}$$

It can be shown that the other 15 terms have the same bounds, and we omit the details. Let  $\mathbf{C}$  be a  $k \times k$  matrix whose  $(j, l)$ th entry is  $1/k$  when  $(j, l) \in \mathcal{B}_k$ , and  $1/k^2$  when  $(j, l) \notin \mathcal{B}_k$ , then

$$\|E(\mathbf{A}_t \mathbf{b}_k \mathbf{b}'_k \mathbf{A}'_{t+h})\| \leq c_3 \kappa_2^{2h} \cdot \|\mathbf{C}\| \cdot \|\mathbf{b}_k\| \cdot \|\mathbf{b}'_k\| \cdot \|\mathbf{C}\| \leq c_4 \kappa^{2h} k^{-2\xi+3},$$

where  $c_3$  and  $c_4$  are absolute constants, and  $\|\mathbf{b}_k\|$  consists of entry-wise absolute values of  $\mathbf{b}_k$ . Therefore, the long-run variance of  $\mathbf{A}_t \mathbf{b}_k$  is of the order  $O(k^{-2\xi+3})$ , and the proof is complete. Since the population objective function is the expectation of log-likelihood function and  $\hat{\phi}_k(\cdot)$  is consistent, then  $\hat{\sigma}_k^2$  and  $\hat{\rho}_k$  are also consistent. ■

**Lemma 7.** *If the CFAR(p) process  $X = (X_t, t \in \mathbb{Z})$  satisfies (4) and  $\phi_i \in C^2[-1, 1]$  for  $i = 1, \dots, p$ , then  $\mathbf{X}_t = \sum_{j=0}^{\infty} \Delta^j \boldsymbol{\varepsilon}_{t-j}$  is a stationary process, where  $\mathbf{X}_t = (X_t, X_{t-1}, \dots, X_{t-p+1})'$ ,  $\boldsymbol{\varepsilon}_t = (\varepsilon_t, 0, \dots, 0)'$ , and  $\Delta$  is defined in (A.1). There exist three constants  $\iota, t_0$  and  $\lambda$  which only depend on  $\{\phi_i, i = 1, \dots, p\}$  and  $\rho$  such that:*

- (i)  $\|\Delta^h\| \leq \iota \lambda^h$  for  $h \geq 2$ .
- (ii)  $\|\gamma_h\|_{\infty} \leq \sigma^2 \iota \lambda^{|h|}$  for  $h \in \mathbb{Z}$ .
- (iii) For  $h \geq 1$ ,

$$\max_{u,v} \left\{ \left| \frac{\partial \gamma_h(u, v)}{\partial u} \right|, \left| \frac{\partial \gamma_h(u, v)}{\partial v} \right| \right\} \leq \sigma^2 \iota \kappa^h.$$

- (iv) For  $h \geq 1$ ,

$$\max_{u,v} \left\{ \left| \frac{\partial^2 \gamma_h(u, v)}{\partial u^2} \right|, \left| \frac{\partial^2 \gamma_h(u, v)}{\partial u \partial v} \right|, \left| \frac{\partial^2 \gamma_h(u, v)}{\partial v^2} \right| \right\} \leq \sigma^2 \iota \kappa^h.$$

**Proof.** By Lemma 5.1 and Theorem 5.1 in Bosq (2000), it is straightforward to see that  $\mathbf{X}_t$  has the following expression:  $\mathbf{X}_t = \sum_{j=0}^{\infty} \Delta^j \boldsymbol{\varepsilon}_{t-j}$ . Since we show that  $\|\Delta^j\|^{1/j} \rightarrow \lambda_{\max} < 1$  in the proof of Theorem 2, there exists an integer  $t_0$  such that  $\|\Delta^h\| \leq \lambda^h$ , for  $h \geq t_0$ , where  $\lambda = (1 + \lambda_{\max})/2$ .

For  $h \geq t_0$ , any  $u, v \in [0, 1]$ ,

$$\begin{aligned} \gamma_h(u, v) &= \frac{1}{p} E(\mathbf{X}_t(u)' \mathbf{X}_{t+h}(v)) \\ &= \frac{1}{p} \sum_{j=0}^{\infty} E[(\Delta^j \boldsymbol{\varepsilon}_{t-j})(u)' (\Delta^{j+h} \boldsymbol{\varepsilon}_{t-j})(v)] \\ &\leq \frac{1}{p} \sum_{j=0}^{\infty} \|(\Delta^j \boldsymbol{\varepsilon}_{t-j})(u)\|_2 \cdot \|(\Delta^{j+h} \boldsymbol{\varepsilon}_{t-j})(v)\|_2 \\ &\leq \frac{\delta \sigma^2 \lambda^h}{2\rho p(1-\lambda)}, \end{aligned}$$

where  $\delta = \max_{j \geq 0} \|\Delta^j\|$ . By the definition of  $\Delta_i$ , we have

$$\begin{aligned} \Delta_{q_1} \Delta_{q_2} \dots \Delta_{q_n} f(s) &= \int_{[0,1]^n} \phi_{q_1}(s-u_1) \phi_{q_2}(u_1-u_2) \dots \phi_{q_n} \\ &\quad \times (u_{q_{n-1}} - u_{q_n}) f(u_n) du_1 \dots du_{q_n}. \end{aligned}$$

The entries in  $i$ th row of  $\Delta^p$  can be regarded a polynomial in  $\{\Delta_1, \dots, \Delta_p\}$  of degree  $p-i+1$ , without intercept (the identity operator). Hence, we can define two linear operators  $\Delta^{(1)}$  and  $\Delta^{(2)}$  in  $\mathbf{H}^p[0, 1]$  such that  $\Delta^{(1)} f(s) = [\Delta^p f(s)]'$ , and  $\Delta^{(2)} f(s) = [\Delta^p f(s)]''$ . Specifically, assume that the  $(i, j)$ th entry in  $\Delta$  is  $\Delta_{q_1} \Delta_{q_2} \dots \Delta_{q_n}$ , and we define  $(i, j)$ th entries in  $\Delta^{(1)}$  and  $\Delta^{(2)}$  as follows

$$\begin{aligned} (\Delta^{(1)} f(s))_{ij} &= \int_{[0,1]^n} \phi'_{q_1}(s-u_1) \phi_{q_2}(u_1-u_2) \dots \phi_{q_n} \\ &\quad \times (u_{q_{n-1}} - u_{q_n}) f(u_n) du_1 \dots du_{q_n}, \\ (\Delta^{(2)} f(s))_{ij} &= \int_{[0,1]^n} \phi''_{q_1}(s-u_1) \phi_{q_2}(u_1-u_2) \dots \phi_{q_n} \\ &\quad \times (u_{q_{n-1}} - u_{q_n}) f(u_n) du_1 \dots du_{q_n}. \end{aligned}$$

Then, the operator norms of  $\Delta^{(1)}$  and  $\Delta^{(2)}$  are bounded by  $pd_{p,1} \kappa_{\max}^{p-1}$  and  $pd_{p,2} \kappa_{\max}^{p-1}$  respectively, where  $\kappa_{\max} = \max\{\kappa_1, \dots, \kappa_p, 1\}$ ,  $d_{p,1} = \max_{1 \leq j \leq p, 0 \leq s \leq 1} |\phi'_j(s)|$ , and  $d_{p,2} = \max_{1 \leq j \leq p, 0 \leq s \leq 1} |\phi''_j(s)|$ .

On the other hand, by the definition of  $\Delta$ , it suffices to show that the operator norms of  $\partial(\Delta^j f(s))/\partial s$  and  $\partial(\Delta^j f(s))/\partial s^2$  are bounded by  $pd_{p,1} \kappa_{\max}^{p-1} \|f\|_{\infty}$  and  $pd_{p,2} \kappa_{\max}^{p-1} \|f\|_{\infty}$  respectively, for  $1 \leq j < p$ .

It follows that,

$$\begin{aligned} \frac{\partial \gamma_h(u, v)}{\partial u} &\leq \frac{\sigma^2 \delta_1 \lambda^h}{2\rho p(1-\lambda)}, & \frac{\partial \gamma_h(u, v)}{\partial v} &\leq \frac{\sigma^2 \delta d_{p,1} \kappa_{\max}^{p-1} \lambda^{h-p}}{2\rho(1-\lambda)}, \\ \frac{\partial^2 \gamma_h(u, v)}{\partial u^2} &\leq \frac{\sigma^2 d \delta_2 \lambda^h}{2\rho p(1-\lambda)}, & \frac{\partial^2 \gamma_h(u, v)}{\partial v^2} &\leq \frac{\sigma^2 \delta d_{p,2} \kappa_{\max}^{p-1} \lambda^{h-p}}{2\rho(1-\lambda)}, \\ \frac{\partial^2 \gamma_h(u, v)}{\partial u \partial v} &\leq \frac{\sigma^2 \delta_2 \delta d_{p,2} \kappa_{\max}^{p-1} \lambda^{h-p}}{2\rho(1-\lambda)}, \end{aligned}$$

for  $h \geq t_0 + p$ , where  $\delta_1 = \max\{\rho, pd_{p,1} \kappa_{\max}^{p-1} \delta\}$ , and  $\delta_2 = \max\{\rho^2, pd_{p,2} \kappa_{\max}^{p-1} \delta\}$ .

The proof is complete. ■

**Lemma 8.** *The CFAR(p) process  $X = (X_t, t \in \mathbb{Z})$  satisfies (4). Let  $\varepsilon_0^*$  be an i.i.d. copy of  $\varepsilon_0$ , and  $X_t^*$  be obtained by replacing  $\varepsilon_0$  with  $\varepsilon_0^*$  in the definition of  $X_t$ . If  $f : [0, 1]^2 \rightarrow \mathbb{R}$  be a continuous function, and set  $\|f\|_{\infty} := \max_{u,v} |f(u, v)|$ , then for  $-p \leq h \leq p$ ,*

$$Y_{t,h} = \int_0^1 \int_0^1 f(u, v) X_t(u) X_{t-h}(v) dudv, \quad \text{and}$$

$$Y_{t,h}^* = \int_0^1 \int_0^1 f(u, v) X_t^*(u) X_{t-h}^*(v) dudv.$$

Then, for each  $q \in \mathbb{N}$ , there exists a constant  $\iota$  which depends on  $\{\phi_i, i = 1, \dots, p\}$ ,  $\rho, \sigma^2, p$  and  $f(u, v)$ , such that

$$\|Y_{t,h} - Y_{t,h}^*\|_q \leq \iota \lambda^t.$$

**Proof.** By Cauchy-Schwarz inequality and Lemma 7, for  $t \geq t_0 + p$ ,

$$\begin{aligned} \|Y_{t,h} - Y_{t,h}^*\|_q &\leq \left\| \int_0^1 \int_0^1 f(u, v) [X_t(u) - X_t^*(u)] X_{t-h}^*(v) dudv \right\|_q \\ &\quad + \left\| \int_0^1 \int_0^1 f(u, v) [X_{t-h}(v) - X_{t-h}^*(v)] X_t(u) dudv \right\|_q \\ &\leq \|f\|_{\infty} \int_0^1 \int_0^1 \|\Delta^t \boldsymbol{\varepsilon}_t(u) - \Delta^t \boldsymbol{\varepsilon}_t^*(u)\|_{2q} \cdot \|X_{t-h}^*(v)\|_{2q} dudv \\ &\quad + \|f\|_{\infty} \int_0^1 \int_0^1 \|\Delta^{t-h} \boldsymbol{\varepsilon}_{t-h}(u) - \Delta^{t-h} \boldsymbol{\varepsilon}_{t-h}^*(u)\|_{2q} \\ &\quad \cdot \|X_t(v)\|_{2q} dudv \\ &\leq 2\|f\|_{\infty} \cdot \lambda^{t-p} \left( \left( \frac{\sigma^2}{\rho} \right)^q [(2q-1)!!] \right)^{1/(2q)} \\ &\quad \times (\gamma_{\max}^q [(2q-1)!!])^{1/(2q)} \\ &\leq \frac{2\sigma \gamma_{\max}^{1/2}}{\rho^{1/2}} [(2q-1)!!]^{1/q} \|f\|_{\infty} \cdot \lambda^{t-p} \end{aligned}$$

where  $\gamma_{\max} = \max_{s \in [0,1]} \text{Var}(X_t(s))$ . ■

**Proof of Theorem 5.** The proof is similar to that of Theorem 3.  $X_t$  can be decomposed as

$$\begin{aligned} X_t(s) &= \sum_{i=1}^p \sum_{j=1}^k \tilde{\beta}_{k,i,j} (B_{k,j} * \tilde{X}_{t-i})(s) + \sum_{i=1}^p (r_{k,i} * \tilde{X}_{t-i})(s) \\ &\quad + \varepsilon_t(s) + \sum_{i=1}^p (\phi_i * \delta_t)(s). \end{aligned}$$

Let  $\tilde{\mathbf{v}}_t$  and  $\tilde{\mathbf{w}}_t$  be two  $(N + 1)$ -dimensional vectors whose  $j$ th entries are  $\sum_{i=1}^p (r_{k,i} * \tilde{X}_{t-i})(j - 1)/N$  and  $\sum_{i=1}^p (\phi_i * \delta_t)(j - 1)/N$  respectively. Let  $\boldsymbol{\varepsilon}_t$  be the  $(N + 1)$ -dimensional vector whose  $j$ th entry is  $\varepsilon_t(j - 1)/N$ . The estimate  $\hat{\boldsymbol{\beta}}_k$  can be decomposed as

$$\hat{\boldsymbol{\beta}}_k = \tilde{\boldsymbol{\beta}}_k + \left( \sum_{t=p+1}^T \mathbf{M}'_t \boldsymbol{\Sigma}^{-1} \mathbf{M}_t \right)^{-1} \left( \sum_{t=p+1}^T \mathbf{M}'_t \boldsymbol{\Sigma}^{-1} (\tilde{\mathbf{v}}_t + \boldsymbol{\varepsilon}_t + \tilde{\mathbf{w}}_t) \right).$$

We claim that under the condition  $T = o(N^2)$

$$\frac{1}{\sqrt{T}} \sum_{t=p+1}^T \mathbf{M}'_t \boldsymbol{\Sigma}^{-1} \tilde{\mathbf{w}}_t = o_p(1). \tag{A.21}$$

Let  $\boldsymbol{\mu}_{k,N} = E(\mathbf{M}'_t \boldsymbol{\Sigma}^{-1} \tilde{\mathbf{v}}_t)$ ,  $\boldsymbol{\Gamma}_{k,N} = E(\mathbf{M}'_t \boldsymbol{\Sigma}^{-1} \mathbf{M}_t)$ , and  $\mathbf{b}_{k,N} = \boldsymbol{\Gamma}_{k,N}^{-1} \boldsymbol{\mu}_{k,N}$ , then

$$\begin{aligned} \hat{\boldsymbol{\beta}}_k &= \tilde{\boldsymbol{\beta}}_k + \mathbf{b}_{k,N} + \boldsymbol{\Gamma}_{k,N}^{-1} \frac{1}{T} \sum_{t=p+1}^T (\mathbf{M}'_t \boldsymbol{\Sigma}^{-1} \tilde{\mathbf{v}}_t - \boldsymbol{\mu}_{k,N}) \\ &\quad + \boldsymbol{\Gamma}_{k,N}^{-1} \frac{1}{T} \sum_{t=p+1}^T \mathbf{M}'_t \boldsymbol{\Sigma}^{-1} \boldsymbol{\varepsilon}_t \\ &\quad - \boldsymbol{\Gamma}_{k,N}^{-1} \frac{1}{T} \sum_{t=p+1}^T (\mathbf{M}'_t \boldsymbol{\Sigma}^{-1} \mathbf{M}_t - \boldsymbol{\Gamma}_{k,N}) \mathbf{b}_{k,N} + o_p(T^{-1/2}). \end{aligned} \tag{A.22}$$

We claim that

$$\begin{aligned} \hat{\boldsymbol{\beta}}_k &= \tilde{\boldsymbol{\beta}}_k + \mathbf{b}_k + \boldsymbol{\Gamma}_k^{-1} \frac{1}{T} \sum_{t=p+1}^T (\mathbf{u}_t - \boldsymbol{\mu}_k) + \boldsymbol{\Gamma}_k^{-1} \frac{1}{T} \sum_{t=p+1}^T \mathbf{z}_t \\ &\quad - \boldsymbol{\Gamma}_k^{-1} \frac{1}{T} \sum_{t=p+1}^T (\mathbf{A}_t - \boldsymbol{\Gamma}_k) \mathbf{b}_k + o_p(T^{-1/2}). \end{aligned} \tag{A.23}$$

By Lemma 8, we can finish this proof similar to that of Theorem 3. ■

**Lemma 9.** Let  $q \geq 2$  be an integer. Consider two  $(qk) \times (qk)$  symmetric positive semi-definite matrices:

$$\boldsymbol{\Omega}_1 = \begin{pmatrix} \boldsymbol{\Omega} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix}, \quad \boldsymbol{\Omega}_2 = \begin{pmatrix} \boldsymbol{\Omega}_{11} & \boldsymbol{\Omega}_{12} \\ \boldsymbol{\Omega}_{21} & \boldsymbol{\Omega}_{22} \end{pmatrix},$$

where  $\boldsymbol{\Omega}$  and  $\boldsymbol{\Omega}_{11}$  are  $(q - 1)k$ -dimensional square matrices. Assume there exist positive constants  $c_1 \leq c_2$  such that

$$\|\boldsymbol{\Omega}_{11}\| \leq c_2, \quad \boldsymbol{\Omega} \geq c_1 \mathbf{I}_{(q-1)k}, \quad \boldsymbol{\Omega}_{22} \geq c_1 \mathbf{I}_k.$$

Then

$$\boldsymbol{\Omega}_1 + \boldsymbol{\Omega}_2 \geq \frac{c_1^2}{c_1 + 2c_2} \mathbf{I}_{qk}.$$

**Proof.** Let  $0 < c < 1$ , and define

$$\begin{aligned} \boldsymbol{\Omega}_3 &= \begin{pmatrix} c^{-1/2} \mathbf{I}_{(q-1)k} & \mathbf{0} \\ \mathbf{0} & c^{1/2} \mathbf{I}_k \end{pmatrix} \begin{pmatrix} \boldsymbol{\Omega}_{11} & \boldsymbol{\Omega}_{12} \\ \boldsymbol{\Omega}_{21} & \boldsymbol{\Omega}_{22} \end{pmatrix} \\ &\quad \times \begin{pmatrix} c^{-1/2} \mathbf{I}_{(q-1)k} & \mathbf{0} \\ \mathbf{0} & c^{1/2} \mathbf{I}_k \end{pmatrix} = \begin{pmatrix} c^{-1} \boldsymbol{\Omega}_{11} & \boldsymbol{\Omega}_{12} \\ \boldsymbol{\Omega}_{21} & c \boldsymbol{\Omega}_{22} \end{pmatrix}. \end{aligned}$$

We see that  $\boldsymbol{\Omega}_3$  is positive semi-definite, and

$$\boldsymbol{\Omega}_1 + \boldsymbol{\Omega}_2 = \boldsymbol{\Omega}_3 + \begin{pmatrix} \boldsymbol{\Omega} - (c^{-1} - 1)\boldsymbol{\Omega}_{11} & \mathbf{0} \\ \mathbf{0} & (1 - c)\boldsymbol{\Omega}_{22} \end{pmatrix}.$$

Since  $\boldsymbol{\Omega} \geq c_1 \mathbf{I}_{(q-1)k}$  and  $\|\boldsymbol{\Omega}_{11}\| \leq c_2$ , it holds that

$$\boldsymbol{\Omega} - (c^{-1} - 1)\boldsymbol{\Omega}_{11} \geq (c_1 - (c^{-1} - 1)c_2)\mathbf{I}_{(q-1)k}.$$

By taking  $c = 2c_2/(c_1 + 2c_2)$ , we have

$$\boldsymbol{\Omega}_1 + \boldsymbol{\Omega}_2 \geq \boldsymbol{\Omega}_3 + \frac{c_1^2}{c_1 + 2c_2} \mathbf{I}_{qk},$$

and the proof is complete. ■

**Lemma 10.** Assume that  $\phi_1, \dots, \phi_p \in C^2[-1, 1]$ , the minimum eigenvalue of  $\boldsymbol{\Gamma}_k$ , defined in Theorem 5 is  $O(k^{-1})$ .

**Proof.** For any function  $\Psi_i$ , denote by  $(\Psi_i \varepsilon_t)(u)$  the random variable  $\int_0^1 \Psi_i(u, v) \varepsilon_t(v) dv$ . Define  $\gamma_{11}(u, v) := \sigma^2 e^{-\rho|u-v|}$ . For each  $2 \leq i \leq p$ , let  $\gamma_{1i}(u, v) := \text{Cov}[\varepsilon_t(u), (\Psi_{i-1} \varepsilon_t)(v)]$  and  $\gamma_{i1}(u, v) := \gamma_{1i}(v, u)$ . For every pair  $2 \leq i, j \leq p$ , define  $\gamma_{ij}(u, v) = \text{Cov}[(\Psi_{i-1} \varepsilon_t)(u), (\Psi_{j-1} \varepsilon_t)(v)]$ . Now for each pair  $1 \leq i, j \leq p$ , define a  $k \times k$  matrix  $\boldsymbol{\Xi}_{ij}$ , with  $(h, l)$ th entry

$$\begin{aligned} (\boldsymbol{\Xi}_{ij})_{hl} &= \int_0^1 \int_0^1 \int_0^1 B'_{k,h}(s-u) B'_{k,l}(s-v) \\ &\quad + \rho^2 B_{k,h}(s-u) B_{k,l}(s-v) ds \gamma_{ij}(u, v) dudv. \end{aligned}$$

For each  $1 \leq i \leq p$ , define the  $p \times p$  block matrix

$$\boldsymbol{\Gamma}_{p,i} = \begin{pmatrix} \boldsymbol{\Xi}_{ii} & \boldsymbol{\Xi}_{i,i-1} & \dots & \boldsymbol{\Xi}_{i,1} & \mathbf{0} \\ \boldsymbol{\Xi}_{i-1,i} & \boldsymbol{\Xi}_{i-1,i-1} & \dots & \boldsymbol{\Xi}_{i-1,1} & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \boldsymbol{\Xi}_{1,i} & \boldsymbol{\Xi}_{1,i-1} & \dots & \boldsymbol{\Xi}_{11} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \dots & \mathbf{0} & \mathbf{0} \end{pmatrix}.$$

Using the representation (7), we know

$$\boldsymbol{\Gamma}_k \geq \sum_{i=1}^p \boldsymbol{\Gamma}_{p,i}.$$

By Lemma 5, there exists a constant  $c_1 > 0$  such that

$$\boldsymbol{\Xi}_{11} \geq c_1/k \cdot \mathbf{I}_k.$$

The condition that  $\phi_1, \dots, \phi_p \in C^2[-1, 1]$  implies that all functions  $\gamma_{ij}(u, v)$  satisfy the assumption (A.13), and then by Lemma 6, the operator norms of all matrices  $\boldsymbol{\Xi}_{ij}$  have the order  $O(k^{-1})$  uniformly. Then we can apply Lemma 9 inductively to complete the proof. ■

**Proof of Theorem 6.** By Lemmas 9 and 10, we can complete the proof of Theorem 6, following the proof of Theorem 4. ■

### References

Aue, A., Noriho, D.D., Hörmann, S., 2012. On the Prediction of Functional Time Series. Technical Report.

Berkes, I., Hörvath, L., Rice, G., 2013. Weak invariance principles for sums of dependent random function. *Stochastic Process. Appl.* 123, 385–403.

Black, F., Scholes, M., 1973. The pricing of options and corporate liabilities. *J. Polit. Econ.* 81, 637–654.

Bosq, D., 2000. *Linear Processes in Function Spaces, Theory and Applications*. Springer-Verlag, New York.

Brumback, B., Rice, J.A., 1998. Smoothing spline models for the analysis of nested and crossed samples of curves. *J. Amer. Statist. Assoc.* 93, 961–994.

Cai, Z., Fan, J., Yao, Q., 2000. Functional-coefficient regression models for non-linear time series. *J. Amer. Statist. Assoc.* 95, 941–956.

Chan, K.S., 1993. Consistency and limiting distribution of the least squares estimator of a threshold autoregressive model. *Ann. Statist.* 21, 520–533.

Chen, X., 2008. Large sample sieve estimation of semi-nonparametric models. *Handb. Econom.* 6, 5549–5632.

Chen, R., Tsay, R.S., 1993a. Functional coefficient autoregressive models. *J. Amer. Statist. Assoc.* 88, 298–308.

Chen, R., Tsay, R.S., 1993b. Nonlinear additive ARX models. *J. Amer. Statist. Assoc.* 88, 955–967.

- Cressie, H., Huang, H., 1999. Classes of nonseparable, spatio-temporal stationary covariance function. *J. Amer. Statist. Assoc.* 94, 1330–1339.
- Diebold, F.X., Li, C., 2006. Forecasting the term structure of government bond yields. *J. Econometrics* 130, 337–364.
- Fan, J., Gijbels, I., 1996. *Local Polynomial Modelling and its Applications*. Chapman and Hall, London.
- Fan, J., Yao, Q., 2003. *Nonlinear Time Series: Nonparametric and Parametric Methods*. Springer-Verlag, New York.
- Fengler, M.R., Härdle, W., Mammen, E., 2005. A dynamic semiparametric factor model for implied volatility string dynamics. *J. Econometrics* 5, 189–218.
- Ferraty, F., Vieu, P., 2006. *Nonparametric Functional Data Analysis*. Springer, New York.
- Gasser, T., Müller, H.G., Köhler, W., Molinari, L., Prader, A., 1984. Nonparametric regression analysis of growth curves. *Ann. Statist.* 12, 210–229.
- Gneiting, T., 2002. Nonseparable, stationary covariance functions for space–time data. *J. Amer. Statist. Assoc.* 97, 590–600.
- Goncalves, S., Guidolin, M., 2006. Predictable dynamics in the S&P 500 index options implied volatility surface. *J. Bus.* 79, 1591–1635.
- Halberstam, H., Richert, H.E., 2013. *Sieve Methods*. Dover Publications, New York.
- Handcock, M.S., Wallis, J.R., 1994. An approach to statistical spatial–temporal modeling of meteorological fields. *J. Amer. Statist. Assoc.* 89, 368–378.
- Härdle, W., Chen, R., Lütkepohl, H., 1997. A review of nonparametric time series analysis. *Internat. Statist. Rev.* 65, 49–72.
- Hörmann, S., Horváth, L., Reeder, R., 2013. A functional version of ARCH model. *Econometric Theory* 29, 267–288.
- Hörmann, S., Kokoszka, P., 2010. Weakly dependent functional data. *Ann. Statist.* 38, 1845–1884.
- Horváth, L., Huskova, M., Kokoszka, P., 2010. Testing the stability of the functional autoregressive process. *J. Multivariate Anal.* 101, 352–367.
- Horváth, L., Kokoszka, P., 2012. *Inference for Functional Data with Applications*. Springer.
- Horváth, L., Kokoszka, P., Reeder, R., 2012. Estimation of the mean of functional time series and a two-sample problem. *Journal of Royal Statistical Society B* 75, 103–122.
- Houghton, A.N., Flannery, J., Viola, M.V., 1980. Malignant melanoma in CT and Denmark. *Int. J. Cancer* 25, 95–104.
- Huang, J.Z., 2003. Local asymptotics for polynomial spline regression. *Ann. Statist.* 31, 1600–1635.
- Hull, J.C., 2006. *Options, Futures, and Other Derivatives*. Pearson Education, India.
- James, G.M., Hastie, T.J., Sugar, C.A., 2000. Principal component models for sparse functional data. *Biometrika* 87, 587–602.
- James, G.M., Sugar, C.A., 2003. Clustering for sparsely sampled functional data. *J. Amer. Statist. Assoc.* 98 (462), 397–408.
- Keselman, H.J., Keselman, J.C., 1993. Analysis of repeated measurements. *Appl. Anal. Var. Behav. Sci.* 105–145.
- Lam, C., Yao, Q., Bathia, N., 2011. Estimation of latent factors for high-dimensional time series. *Biometrika* 98, 901–918.
- Lütkepohl, H., 2005. *New Introduction to Multiple Time Series Analysis*. Springer, Berlin.
- Mas, A., 2002. Weak convergence for the covariance operators of a Hilbertian linear process. *Stochastic Process. Appl.* 99 (1), 117–135.
- Merton, R., 1973. Theory of rational option pricing. *Bell J. Econ. Manage. Sci.* 4 (1), 141–183.
- Nadaraya, E.A., 1964. On estimating regression. *Theory Probab. Appl.* 9, 141–142.
- Neuman, E., 1981. Moments and Fourier transforms of B-splines. *J. Comput. Math.* 7, 51–62.
- Ramsay, J.O., Silverman, B.W., 2005. *Functional Data Analysis*. Springer, New York.
- Rao, B.P., 1999. *Semimartingales and Statistical Inference*. Chapman and Hall, London.
- Ratcliffe, S.J., Leader, L.R., Heller, G.Z., 2002a. Functional data analysis with application to periodically stimulated foetal heart rate data. I: Functional regression. *Stat. Med.* 21, 1103–1114.
- Ratcliffe, S.J., Leader, L.R., Heller, G.Z., 2002b. Functional data analysis with application to periodically stimulated fetal heart rate data. II: Functional logistic regression. *Stat. Med.* 21, 1115–1127.
- Roberts, J.M., 1995. New Keynesian economics and the Phillips curve. *J. Money, Credit Banking* 27, 975–984.
- Ruppert, D., Sheather, S., Wand, M.P., 1995. An effective bandwidth selector for local least squares regression. *J. Amer. Statist. Assoc.* 90, 1257–1270.
- Schumaker, L., 1981. *Spline Functions: Basic Theory*. John Wiley & Sons, New York.
- Silverman, B.W., 1984. Spline smoothing: the equivalent variable kernel method. *Ann. Statist.* 898–916.
- Sutradhar, B.C., Kumar, P., 2003. The inversion of correlation matrix for MA (1) process. *Appl. Math. Lett.* 163, 317–321.
- Tiao, G.C., Tsay, R.S., 1983. Multiple time series modeling and extended sample cross-correlations. *J. Bus. Econom. Statist.* 1, 43–56.
- Tiao, G.C., Tsay, R.S., 1989. Model specification in multivariate time series. *J. R. Stat. Soc. Ser. B Stat. Methodol.* 51, 157–213.
- Tong, H., Lim, K.S., 1980. Threshold autoregression, limit cycles and cyclical data. *J. Roy. Statist. Soc. Ser. B* 42 (3), 245–292.
- Watson, G.S., 1964. Smooth regression analysis. *Sankhyā* 26, 359–372.
- Wu, W.B., 2005. Nonlinear system theory: Another look at dependence. *Proceedings of the National Academy of Sciences of the United States of America* 102, 14150–14154.
- Xia, Y., Li, W.K., 1999. On single-index coefficient regression models. *J. Amer. Statist. Assoc.* 94, 1275–1285.
- Zhou, S., Shen, X., Wolfe, D.A., 1998. Local asymptotics for regression splines and confidence regions. *Ann. Statist.* 26, 1760–1782.
- Zygmund, A., 2002. *Trigonometric Series*. Cambridge University Press.