Journal of Econometrics 194 (2016) 263–282

Contents lists available at ScienceDirect

Journal of Econometrics

journal homepage: www.elsevier.com/locate/jeconom

Convolutional autoregressive models for functional time series

Xialu Liu [a], Han Xiao [b], Rong Chen [b,*]

[a] Management Information Systems Department, San Diego State University, San Diego, CA 92182, United States
[b] 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 became 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.

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 (Keselman and Keselman, 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; Lüetkepohl, 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 Δ_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 (X. Liu), hxiao@stat.rutgers.edu (H. Xiao), rongchen@stat.rutgers.edu (R. Chen).

http://dx.doi.org/10.1016/j.jeconom.2016.05.006

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 $\boldsymbol{\mu}$, $(\boldsymbol{\mu})_i$ denotes its ith 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 $L_2([0,1], \mathscr{B}_{[0,1]}, \lambda)$, abbreviated as $L_2[0,1]$, where $\mathscr{B}_{[0,1]}$ is the Borel σ-field, and λ 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:

$$\begin{aligned}
&\mathrm{Lip}^h[-1,1] = \{f \in [-1,1] : |f(x+\delta) - f(x)| \leqslant M\delta^h,\\
&\quad M < \infty,\ 0 < h \leqslant 1\},\\
&\mathrm{Lip}_2^\zeta[-1,1] = \{f \in C^r[-1,1] : f^{(r)} \in \mathrm{Lip}^h[-1,1],\\
&\quad \zeta > 1, r = \lfloor \zeta \rfloor, h = \zeta - r\},
\end{aligned}$$

If $f \in \mathrm{Lip}_2^\zeta[-1,1]$, then ζ 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

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

where $\phi_i \in L_2[-1,1]$ for $i = 1, \ldots, p$, are called convolution functions, and ε_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 ε_t follow the O–U process, which is Gaussian with the following covariance structure:

$$\varepsilon_t(s_1) \sim N\left(0, \frac{\sigma^2}{2\rho}\right), \qquad \mathrm{Corr}(\varepsilon_t(s_1), \varepsilon_t(s_2)) = e^{-\rho|s_1-s_2|},$$
$$\forall s_1, s_2 \in [0,1].$$

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

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

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 $\{\phi_i(\cdot), i = 1, \ldots, p\}$ 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 $\{\phi_i(\cdot), i = 1, \ldots, p\}$ 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