Kernel estimates of nonparametric functional autoregression models and their bootstrap approximation

Tingyi Zhu and Dimitris N. Politis

Department of Mathematics
University of California, San Diego
La Jolla, CA 92093
e-mail: t8zhu@ucsd.edu*;* dpolitis@ucsd.edu

Abstract: This paper considers a nonparametric functional autoregression model of order one. Existing contributions addressing the problem of functional time series prediction have focused on the linear model and literatures are rather lacking in the context of nonlinear functional time series. In our nonparametric setting, we define the functional version of kernel estimator for the autoregressive operator and develop its asymptotic theory under the assumption of a strong mixing condition on the sample. The results are general in the sense that high-order autoregression can be naturally written as a first-order AR model. In addition, a component-wise bootstrap procedure is proposed that can be used for estimating the distribution of the kernel estimation and its asymptotic validity is theoretically justified. The bootstrap procedure is implemented to construct prediction regions that achieve good coverage rate. A supporting simulation study is presented in the end to illustrate the theoretical advances in the paper.

MSC 2010 subject classifications: 62G08, 62G09, 37M10, 60G25.
Keywords and phrases: Functional time series, nonparametric autoregression, α-mixing, regression-type bootstrap, prediction region.

Contents

1. Introduction

Popularized by the pioneering works of Ramsay and Silverman (1997) [29], (2002) [30], Functional Data Analysis (FDA) has emerged as a promising field of statistical research in the past decade. When functional data objects being collected sequentially over time that exhibit forms of dependence, such data are known as functional time series. The typical situation in which functional time series arise is when long continuous records of temporal sequence are segmented into curves over natural consecutive time intervals. Examples include daily price curves of financial transactions, daily electricity consumptions and daily patterns of environmental data. The primary goal of functional time series analysis is to provide reliable guesses for the future realizations. In this paper, we focus our attention on a first-order nonparametric functional autoregression–FAR(1) model which is defined by the recursion:

$$\mathcal{X}_{n+1} = \Psi(\mathcal{X}_n) + \mathcal{E}_{n+1}, \tag{1.1}$$

where the observations $\mathcal{X}_n$ and the error terms $\mathcal{E}_n$ are functions, and no linearity restrictions are imposed on the functional operator Ψ. Precise definitions and details of the model are stated in Section 2. Existing contributions have largely focused on the functional linear autoregression while the research addressing the nonlinear model is scarce. We approach this problem by merging the ideas in nonparametric time series and functional regression analysis, extending the theoretical study to the nonparametric model of functional autoregression.

    The research pertains to the FAR model can trace back to Bosq (2000) [4], in which the theory of linear processes in functional space was first developed. One of the major contribution of that book was the study of linear autoregressive processes in the Hilbert space. Under the assumption that the functional operator Ψ in (1.1) is linear, Bosq has derived a one-step ahead predictor $\hat{\Psi}$ based on a functional form of the Yule-Walker equation, which has been regarded as the classical benchmark in FAR(1) prediction. Since then, there has been abundant literatures on the study of the linear functional processes. We refer the readers Antoniadis and Sapatinas (2003) [2], Antoniadis et al. (2006) [1], Bosq (2007) [5], Kargin and Onatski (2008) [18], Gabrys et al. (2010) [14] and Horváth and Kokoszka (2011) [17], among other contributions. Bosq's predictor in [4] has a rather complicated form which makes it unrealistic to implement in practice. Aue et al. (2015) [3] proposed an alternative method of predicting linear FAR(1) process utilizing functional principal component analysis (FPCA). The method appears to be much more widely applicable under the idea that the dimension reduction with FPCA should lead to a vector-valued time series of FPC scores that can be predicted by existing multivariate methodologies. Hörmann et al.

(2015) [15] proposed a dynamic version of functional principal component analysis to address the problem of dimension reduction for functional time series. More recently, Klepsch et al. (2017) [20] used a similar dimension reduction technique to model the FARMA process with an application to traffic data. See also Klepsch and Klüppelberg (2017) [19] in which an innovation algorithm was proposed to obtain the best linear predictor of a functional moving average (FMA) process.

On the other hand, kernel methods have been a powerful tool when dealing with nonparametric models. Numerous early references have investigated its implementations in nonparametric univariate autoregression. To mention some, asymptotic study of the kernel smoother was presented in Robinson (1983) [32], and Masry (1996) [22]. The bootstrap procedures for this model and its validity were provided in Franke et al. (2002) [12]. Pan and Politis (2016) [25] developed a coherent methodology for the construction of bootstrap prediction intervals, which can be successfully applied to the nonlinear univariate autoregression models. Those results can be naturally extended to multivariate time series, but that is not the case for functional time series due to the infinite dimensional nature of functional data.

Nonparametric statistical methods for functional data analysis were established in Ferraty and Vieu (2006) [11]. The nonparametric functional regression model, i.e. $\mathcal{Y} = r(\mathcal{X}) + \varepsilon$, has been extensively studied since then. Ferraty et al. (2007) [10] concentrated on the situation where the response variable $\mathcal{Y}$ is scalar and $\mathcal{X}$ takes values in some functional space. Asymptotic properties concerning the kernel estimator $\hat{r}(\cdot)$ of the regressor $r(\cdot)$ have been investigated and the validity of its bootstrap approximation was proved in Ferraty et al. (2010) [8]. The results have been extended to the model with double functional setting (i.e. both $\mathcal{Y}$ and $\mathcal{X}$ are functionals); see Ferraty et al. (2012) [9]. Masry (2005) [23] and Delsol (2009) [6] investigated the same model taking into account dependent functional data.

Motivated by the prior works aforementioned, this paper investigates the kernel estimator for nonparametric functional autoregression. We show the consistency of the estimator under the assumption of a strong mixing condition on the sample. The proof of its consistency involves a functional central limit theorem for dependent sequence in a triangular array setting. In addition, we propose a bootstrapping procedure in this functional dependent framework that can be used for estimating the distribution of the projection of the kernel estimation. The bootstrap prediction regions are constructed as a measurement of accuracy for the functional prediction. A regression bootstrap scheme is implemented in the procedure which provides a simplification for the bootstrap method in the autoregression case. Franke et al. (2002) [12] first applied the regression-type bootstrap in univariate nonlinear autoregression for inference of the kernel estimator, and Neumann and Kreiss (1998) [24] considered to what extent regression-type bootstrap procedures can be successfully applied as long as nonparametric estimators and tests for conditional mean in nonparametric autoregressions are considered. It was mentioned in Kreiss and Lahiri (2012) [21] that the regression-type bootstrap is also valid for the Yule-Walker esti-

mates for coefficients in a parametric AR(p) model, but it might not lead to asymptotically valid results for more general statistics.

The rest of the paper is organized as follows. In Section 2, the detailed mathematical background of the model is provided and the functional version kernel estimator of the autoregressive operator is defined. Some notations and necessary assumptions are stated in Section 3. Section 4 provides the asymptotic results of the proposed estimator. A componentwise bootstrap scheme is introduced and its validity is shown in Section 5. A simulated study is given in Section 6 while Section 7 presents the approach to construct the bootstrap prediction region. All proofs are gathered in Appendix (Section 8).

2. The FAR(1) model

Let $\{\mathcal{X}_n\}$ be a stationary and α-mixing functional sequence in some separable Hilbert space $\mathbb{H}$ with the usual definition of α-mixing coefficients introduced by Rosenblatt (1956) [33]. $\mathbb{H}$ is endowed with inner product $\langle\cdot,\cdot\rangle$ and corresponding norm $||\cdot||$ (i.e. $||g||^2 = \langle g, g\rangle$), and with orthonormal basis $\{e_k : k = 1, \cdots, \infty\}$. A semi-metric $d(\cdot,\cdot)$ is also defined on $\mathbb{H}$ to measure the proximity between two elements in $\mathbb{H}$. The semi-metric structure d will be the key tool for controlling the good behavior of the estimators whereas some separable Hilbert structure is necessary for studying the operator Ψ component by component. See more details on the two-topology framework in Ferraty et al. (2012) [9].

We consider the following FAR(1) model

$$\mathcal{X}_{i+1} = \Psi(\mathcal{X}_i) + \mathcal{E}_{i+1}, \quad i = 1, 2, \ldots, \tag{2.1}$$

where Ψ is the autoregressive operator mapping functions from $\mathbb{H}$ to $\mathbb{H}$, and the innovations $\mathcal{E}_i$'s are independent and identically distributed (i.i.d.) $\mathbb{H}$-valued random variables satisfy $E(\mathcal{E}_{i+1}|\mathcal{X}_i) = 0$ and $E(||\mathcal{E}_{i+1}||^2|\mathcal{X}_i) = \sigma_{\mathcal{E}}^2(\mathcal{X}_i) < \infty$. Assume here that the model is homoscedastic, that is, $\sigma_{\mathcal{E}}(\mathcal{X}_i) \equiv \sigma_{\mathcal{E}}$. The operator Ψ is not constrained to be linear; this is a Nonparametric Functional Autoregression model.

Remark 2.1. Because of the generality of the notation, a higher-order autoregression, say FAR(2), in which $\mathcal{X}_{i+1}$ depends on $\mathcal{X}_i$ and $\mathcal{X}_{i-1}$, can still be written as FAR(1) by redefining the $\mathcal{X}$ and Ψ; e.g. in the FAR(2) case, one may let $\mathcal{Y}_i = (\mathcal{X}_i, \mathcal{X}_{i-1})$ with an obvious choice for the FAR(1) operator relating $\mathcal{Y}_{i+1}$ to $\mathcal{Y}_i$ only.

Estimation of Ψ is given by the functional version of Nadaraya-Watson estimator of time series

$$\hat{\Psi}_h(\chi) = \frac{\sum\limits_{i=1}^{n-1} \mathcal{X}_{i+1} K(h^{-1}d(\mathcal{X}_i, \chi))}{\sum\limits_{i=1}^{n-1} K(h^{-1}d(\mathcal{X}_i, \chi))}, \tag{2.2}$$

where χ is a fixed element in $\mathbb{H}$, K is a kernel function and h is a bandwidth sequence tending to zero as n tends to infinity. The way of choosing a semi-metric $d(\cdot, \cdot)$ was discussed in Ferraty and Vieu (2006) [11]. For a fixed $k \in \mathbb{Z}^+$, applying $\langle \cdot, e_k \rangle$ on both sides of the Eq. (2.1) yields

$$\begin{aligned} \langle \mathcal{X}_{i+1}, e_k \rangle &= \langle \Psi(\mathcal{X}_i) + \mathcal{E}_{i+1}, e_k \rangle \\ &= \langle \Psi(\mathcal{X}_i), e_k \rangle + \langle \mathcal{E}_{i+1}, e_k \rangle \quad i = 1, 2, \ldots \end{aligned}$$

Let $X_{n,k}, \varepsilon_{n,k}$ be the jth component of the functional $\mathcal{X}_n$ and $\mathcal{E}_n$ respectively, i.e. $X_{n,k} = \langle \mathcal{X}_n, e_k \rangle$, $\varepsilon_{n,k} = \langle \mathcal{E}_n, e_k \rangle$. Also, define the functional ψ_k from $\mathbb{H}$ to $\mathbb{R}$ such that

$$\psi_k(\cdot) = \langle \Psi(\cdot), e_k \rangle. \tag{2.3}$$

When k is fixed, we will drop the index k for the simplicity of notations, using $\{X_n\}$ and $\{\varepsilon_n\}$ to denote the sequences $\{X_{n,k}\}$ and $\{\varepsilon_{n,k}\}$, respectively. Similarly, ψ can be used in place of ψ_k. For a fixed k, we obtain

$$X_{i+1} = \psi(\mathcal{X}_i) + \varepsilon_{i+1} \quad i = 1, 2, \ldots \tag{2.4}$$

Eq. (2.4) can be treated as an auxiliary functional autoregressive model with scalar response. The scalar innovations ε_i's are i.i.d. random variables satisfy $E(\varepsilon_{i+1} | \mathcal{X}_i) = 0$ and $E(\varepsilon_{i+1}^2 | \mathcal{X}_i) = \sigma_\varepsilon^2 < \infty$. Again, the operator ψ in (2.4) is not constrained to be linear. Accordingly, its kernel estimation is given by

$$\hat{\psi}_h(\chi) = \frac{\sum_{i=1}^{n-1} X_{i+1} K(h^{-1} d(\mathcal{X}_i, \chi))}{\sum_{i=1}^{n-1} K(h^{-1} d(\mathcal{X}_i, \chi))}, \tag{2.5}$$

and the two estimators, $\hat{\Psi}_h$ and $\hat{\psi}_h$, are connected in the following way

$$\hat{\psi}_h(\chi) = \langle \hat{\Psi}_h(\chi), e_k \rangle. \tag{2.6}$$

Consistency of both $\hat{\psi}_h(\chi)$ and $\hat{\Psi}_h(\chi)$ will be addressed in Section 4. While it is more of the interest to study the estimator $\hat{\Psi}_h(\chi)$, the need for model (2.4) and the estimator $\hat{\psi}_h(\chi)$ will be seen in Section 5 where a componentwise bootstrap approximation is proposed.

3. Assumptions and notations

In the sequel, χ is a fixed element and $\mathcal{X}$ is a random element of the functional space $\mathbb{H}$. For $k = 1, 2, \ldots$, let $\varphi_{\chi,k}$ be a real-valued function defined as

$$\begin{aligned} \varphi_{\chi,k}(s) &= E[\psi(\mathcal{X}) - \psi(\chi) | d(\mathcal{X}, \chi) = s] \\ &= E[\langle \Psi(\mathcal{X}) - \Psi(\chi), e_k \rangle | d(\mathcal{X}, \chi) = s]. \end{aligned}$$

Again, for the simplicity of notations, we drop the index k to use φ_χ in place of $\varphi_{\chi,k}$ when k is fixed. Let F be the distribution of the random variable $d(\mathcal{X}, \chi)$,

$$F_\chi(t) = P(d(\mathcal{X}, \chi) \le t),$$

which is usually called the small ball probability function in functional data analysis. Also define for $s \in [0, 1]$,

$$\tau_{h\chi}(s) = \frac{F_\chi(hs)}{F_\chi(h)} = P(d(\mathcal{X}, \chi) \le hs | d(\mathcal{X}, \chi) \le h).$$

Technical aspects of the functions φ_χ, F_χ and $\tau_{h\chi}$ have been discussed in Ferraty et al.(2007) [10]. The following assumptions are needed:

(A1) ψ is continuous in a neighborhood of χ with respect to the semi-metric d, and $F_\chi(0) = 0$.

(A2) $\varphi_\chi(0) = 0$ and $\varphi_\chi'(0)$ exists.

(A3) $h \to 0$ and $nF_\chi(h) \to \infty$, as $n \to \infty$.

(A4) The kernel function K is supported on $[0, 1]$ and has a continuous derivative with $K'(s) \le 0$, and $K(1) > 0$.

(A5) For $s \in [0, 1]$, $\lim_{h \downarrow 0} \tau_{h\chi}(s) = \tau_{0\chi}(s)$.

(A1)-(A5) are the standard assumptions inherited from those introduced in the independent case in the setting of nonparametric functional regression. For our autoregressive model, additional assumptions (A6)-(A10) below are necessary:

(A6) $\exists\, p > 2, E(|\varepsilon_i|^p | \mathcal{X}) < \infty$.

(A7) $\max(E(|X_{i+1} X_{j+1}| | \mathcal{X}_i, \mathcal{X}_j), E(|X_{i+1}| | \mathcal{X}_i, \mathcal{X}_j)) < \infty \quad \forall i, j \in \mathbb{Z}$.

(A8) $\{\mathcal{X}_n\}$ is α-mixing process with mixing coefficients $\alpha(n) \le Cn^{-a}$.

$(A9) \quad \exists\, \nu > 0$, such that $\Theta(h) = O(F_\chi(h)^{1+\nu})$, with $a > \dfrac{(1+\nu)p - 2}{\nu(p-2)}$,

where p and a are defined in (A6) and (A8) respectively, and

$$\Theta(s) := \max\{\max_{i \ne j} P(d(\mathcal{X}_i, \chi) \le s, d(\mathcal{X}_j, \chi) \le s), F_\chi^2(s)\}.$$

(A10) $\exists\, \gamma > 0$ such that $nF_\chi(h)^{1+\gamma} \to \infty$ and $a > \max\left\{\dfrac{4}{\gamma}, \dfrac{p}{p-2} + \dfrac{2(p-1)}{\gamma(p-2)}\right\}$

where p and a are defined in (A6) and (A8), respectively.

Remark 3.1. Delsol (2009) [6] considered a functional regression model with functional variables under dependence. (A6)-(A10) are inherited from the additional assumptions Delsol made to control the dependence between variables. These assumptions enable us to obtain asymptotic results of the estimator in

our autoregressive model, which has a formal resemblance to the analogous regression model. In particular, (A8)-(A10) is a set of conditions on the mixing coefficients, and (A8) is the so-called arithmetically α-mixing condition which is typically satisfied in the case of an autoregressive model. From a theoretical point of view, a more general set of conditions on the mixing coefficients ((H1)-(H2) in Delsol (2009) [6]) is available, see the details therein.

The semi-metric d will act on the asymptotic behavior of the estimator through φ_χ, F_χ and $\tau_{h\chi}$, and the following quantities:

$$M_0 = K(1) - \int_0^1 (sK(s))' \tau_{0\chi}(s) ds,$$

$$M_1 = K(1) - \int_0^1 K'(s) \tau_{0\chi}(s) ds,$$

$$M_2 = K^2(1) - \int_0^1 (K^2)'(s) \tau_{0\chi}(s) ds.$$

4. Asymptotic study

4.1. Consistency of estimator $\hat{\psi}_h(\chi)$

First, we have the following point-wise asymptotic results for the estimator $\hat{\psi}_h(\chi)$:

Theorem 4.1. *Assume (A1)-(A6), then*

$$E[\hat{\psi}_h(\chi)] - \psi(\chi) = \varphi_\chi'(0) \frac{M_0}{M_1} h + O(\frac{1}{nF_\chi(h)}) + o(h), \qquad (4.1)$$

Theorem 4.2. *Assume (A1)-(A10), then*

$$Var(\hat{\psi}_h(\chi)) = \frac{\sigma_\varepsilon^2}{M_1^2} \frac{M_2}{nF_\chi(h)} + o(\frac{1}{nF_\chi(h)}). \qquad (4.2)$$

Theorem 4.1 and 4.2 can be obtained following Lemma 8.1 and 8.2, and along the similar lines of the proof of Theorem 1 in Ferraty et al. (2007) [10]. See details of the proof in Appendix. Combining 4.1 and 4.2, we immediately obtain the following corollary:

Corollary 4.1. *Assume (A1)-(A10), then*

$$\hat{\psi}_h(\chi) \xrightarrow{p} \psi(\chi). \qquad (4.3)$$

The pointwise asymptotic normality is given in Theorem 4.3 below:

Theorem 4.3. *Assume (A1)-(A10), then*

$$\sqrt{n\hat{F}_\chi(h)} \left(\hat{\psi}_h(\chi) - \psi(\chi) - B_n \right) \frac{M_1}{\sqrt{\sigma_\varepsilon^2 M_2}} \xrightarrow{d} N(0,1), \qquad (4.4)$$

where $B_n = h\varphi_\chi{}'(0)M_0/M_1$, and $\hat{F}_\chi(h)$ is the empirical estimation of $F_\chi(h)$:

$$\hat{F}_\chi(h) = \frac{\#(i : d(\mathcal{X}_i, \chi) \le h)}{n}.$$

Proof. See Appendix.   □

The bias term can be ignored under the following additional assumption:

(A11) $lim_{n\to\infty} h\sqrt{nF_\chi(h)} = 0$.

Corollary 4.2. *Assume (A1)-(A11), then*

$$\sqrt{nF_\chi(h)} \left(\hat{\psi}_h(\chi) - \psi(\chi) \right) \frac{M_1}{\sqrt{\sigma_\varepsilon^2 M_2}} \xrightarrow{d} N(0,1). \tag{4.5}$$

4.2. Consistency of estimator $\hat{\Psi}_h(\chi)$

Corollary 4.1, together with Eq. (2.6), implies the following componentwise consistency of $\hat{\Psi}_h$:

$$\langle \hat{\Psi}_h(\chi) - \Psi(\chi), e_k \rangle \xrightarrow{p} 0, \quad k = 1, \dots, \infty \tag{4.6}$$

However, (4.6) does not guarantee the consistency of the estimator $\hat{\Psi}_h$ in an infinite-dimensional space. A more general consistency result is desired. We consider the following regularity conditions:

(C1) For each $k \ge 1$, ψ_k is continuous in a neighborhood of χ with respect to semi-metric d, and $F_\chi(0) = 0$.

(C2) For some $\beta > 0$, all $0 \le s \le \beta$ and all $k \ge 1$, $\varphi_{\chi,k}(0) = 0$, $\varphi'_{\chi,k}(s)$ exists, and $\varphi'_{\chi,k}(s)$ is Hölder continuous of order $0 < \alpha \le 1$ at 0, i.e. there exists a $0 < L_k < \infty$ such that $|\varphi'_{\chi,k}(s) - \varphi'_{\chi,k}(0)| \le L_k s^\alpha$ for all $0 \le s \le \beta$. Moreover, $\sum_{k=1}^\infty L_k^2 < \infty$ and $\sum_{k=0}^\infty \varphi'_{\chi,k}(0) < \infty$.

(C3) The bandwidth h satisfies $h \to 0$, $nF_\chi(h) \to \infty$, and $(nF_\chi(h))^{1/2}h^{1+\alpha} = o(1)$.

With additional assumptions on the mixing coefficients and moments, we are able to prove the following limit result for $\hat{\Psi}_h(\chi)$.

Theorem 4.4. *For some fixed $\chi \in \mathbb{H}$, assume $\exists\, \delta' > \delta > 0$ such that*

(i) $\frac{2+\delta}{2+\delta'} + \frac{(1-\delta)(2+\delta)}{2} \le 1$,

(ii) $E\|\mathcal{X}_i - \Psi(\chi)\|^{2+\delta'} < \infty$,

(iii) $\sum_n \alpha(n)^{\frac{\delta}{2+\delta}} < \infty$,

where $\alpha(\cdot)$ is the mixing coefficient of the functional sequence $\{\mathcal{X}_t, t \in \mathbb{N}\}$. Also assume (A4), (C1)-(C3). Then

$$\hat{\Psi}_h(\chi) = \Psi(\chi) - \mathcal{B}_n + O_p(\frac{1}{\sqrt{nF_\chi(h)^{1+\delta}}}) \tag{4.7}$$

where

$$\mathcal{B}_n = h \frac{M_0}{M_1} \sum_{k=1}^{\infty} \varphi'_{\chi,k}(0) e_k.$$

Remark 4.1. To prove Theorem 4.4, we need a functional central limit theorem for dependent sequence in a triangular array setting such as Theorem 2.3 in Politis and Romano (1994) [28]. See Appendix for details of the proof. The assumptions (i)-(iii) show a tradeoff between the moment assumptions and the mixing conditions. The conditions on mixing coefficients can be less stringent if higher moments are assumed. The parameter δ' controls the moment while δ controls the mixing condition and they can be chosen under the constraint (i), for example, $\delta = 0.5$ and $\delta' = 5$.

5. Bootstrap approximation

Ferraty et al. (2010, 2012) [8, 9] have employed both naive and wild bootstrap to approximate the asymptotic distribution of the kernel estimators for nonparametric functional regressions. Their first result showed the validity of bootstrap when the explanatory variable is functional and the response is real. To extend the bootstrap approach to the double functional setting, i.e. when both variables are functional, they introduced the notion of "componentwise bootstrap", in which the idea is to show that the bootstrap approximation has good theoretical behaviors when functionals are projected to a fixed basis element e_k. Here we take advantage of the auxiliary univariate model in (2.4), extending this componentwise bootstrap idea to the functional autoregression.

First, we propose a bootstrap procedure to approximate the distribution of $\hat{\psi}_h(\chi) - \psi(\chi)$ under the AR model (2.4), which consists of the following steps:

Algorithm 5.1.

1. *For $i = 1, \ldots, n$, define $\hat{\varepsilon}_{i,b} = X_{i+1} - \hat{\psi}_b(\mathcal{X}_i)$, where b is a second smoothing parameter.*
2. *Draw n i.i.d. random variables $\varepsilon_1^*, \ldots, \varepsilon_n^*$ from the empirical distribution of $(\hat{\varepsilon}_{1,b} - \bar{\hat{\varepsilon}}_b, \ldots, \hat{\varepsilon}_{n,b} - \bar{\hat{\varepsilon}}_b)$ where $\bar{\hat{\varepsilon}}_b = n^{-1} \sum_{i=1}^{n} \hat{\varepsilon}_{i,b}$.*
3. *For $i = 1, \ldots, n-1$, let $X_{i+1}^* = \hat{\psi}_b(\mathcal{X}_i) + \varepsilon_{i+1}^*$.*
4. *Define*

$$\hat{\psi}_{hb}^*(\chi) = \frac{\sum_{i=1}^{n-1} X_{i+1}^* K(h^{-1} d(\mathcal{X}_i, \chi))}{\sum_{i=1}^{n-1} K(h^{-1} d(\mathcal{X}_i, \chi))}. \tag{5.1}$$

Remark 5.1. From a theoretical point of view, the second smoothing parameter b has to be asymptotically larger than h (see condition (D5)), so over-smoothing is needed to make the bootstrap procedure work, as is the case in the functional regression. However, the two bandwidths have to be fixed in practice and a cross-validation procedure is used to determine the bandwidths in the simulation study in Section 6.

Remark 5.2. It is also worth being noted that instead of generating a real bootstrap sequence, here we are actually generating a scatter plot. Every bootstrap point X_{i+1}^* is generated by the prior true point $\mathcal{X}_i$—see step 3 above—and the desired estimation in the bootstrap world comes from fitting the pairs $(X_{i+1}^*, \mathcal{X}_i)$. This is so called regression bootstrap scheme, introduced in Franke et al. (2002) [12] as one of the three bootstrap schemes proposed for the scalar-valued nonlinear autoregression. The main reason we use the regression bootstrap scheme here is in the fact that when conditioning on the sample $\{\mathcal{X}_1, \ldots, \mathcal{X}_n\}$, it eliminates the random element in the denominator of $\hat{\psi}_{hb}^*(\chi)$—see Eq. (5.1)— which makes the proof of Theorem 5.1 proceed (see details in appendix). The regression-type bootstrap is considered as an important simplification for the bootstrap method in autoregression. We refer the readers Neumann and Kreiss (1998) [24] for its applications in nonparametric autoregressions, and Kreiss and Lahiri (2012) [21] for its extensions in parametric time series models.

Theorem 5.1. *If conditions of Theorem 4.3 hold, as well as assumptions (D1)-(D7) in Appendix, we have*

$$\sup_{y \in \mathbb{R}} \left| P^* \left(\sqrt{nF_\chi(h)} \{ \hat{\psi}_{hb}^*(\chi) - \hat{\psi}_b(\chi) \} \leq y \right) \right.$$
$$\left. - P \left(\sqrt{nF_\chi(h)} \{ \hat{\psi}_h(\chi) - \psi(\chi) \} \leq y \right) \right| \stackrel{a.s.}{\to} 0,$$

$$(5.2)$$

where P^* denotes probability conditioned on the sample $\{\mathcal{X}_1, \ldots, \mathcal{X}_n\}$.

Theorem 5.1 shows the validity of the bootstrap procedure for $\hat{\psi}_h$. Then, the bootstrap procedure for $\hat{\Psi}_h$ is proposed as follows:

Algorithm 5.2.

1. *For $i = 1, \ldots, n$, define $\hat{\mathcal{E}}_{i,b} = \mathcal{X}_{i+1} - \hat{\Psi}_b(\mathcal{X}_i)$, where b is a second smoothing parameter.*
2. *Draw n i.i.d. random variables $\mathcal{E}_1^*, \ldots, \mathcal{E}_n^*$ from the empirical distribution of $(\hat{\mathcal{E}}_{1,b} - \bar{\hat{\mathcal{E}}}_b, \ldots, \hat{\mathcal{E}}_{n,b} - \bar{\hat{\mathcal{E}}}_b)$ where $\bar{\hat{\mathcal{E}}}_b = n^{-1} \sum_{i=1}^n \hat{\mathcal{E}}_{i,b}$.*
3. *For $i = 1, \ldots, n-1$, let $\mathcal{X}_{i+1}^* = \hat{\Psi}_b(\mathcal{X}_i) + \mathcal{E}_{i+1}^*$.*
4. *Define*

$$\hat{\Psi}_{hb}^*(\chi) = \frac{\sum_{i=1}^{n-1} \mathcal{X}_{i+1}^* K(h^{-1}d(\mathcal{X}_i, \chi))}{\sum_{i=1}^{n-1} K(h^{-1}d(\mathcal{X}_i, \chi))}. \tag{5.3}$$

Theorem 5.2. *For any $\eta \in \mathbb{H}$ with $||\eta|| = 1$ and any bandwidth h and b, let $\hat{\Psi}_{\eta,h}(\chi) = \langle \hat{\Psi}_h(\chi), \eta \rangle$, $\hat{\Psi}_{\eta,hb}^*(\chi) = \langle \hat{\Psi}_{hb}^*(\chi), \eta \rangle$ and $\Psi_\eta(\chi) = \langle \Psi(\chi), \eta \rangle$. If, in addition to conditions of Theorem 4.3, (D1)-(D7) in Appendix hold, we have for every χ*

$$\sup_{y \in \mathbb{R}} \left| P^* \left(\sqrt{nF_\chi(h)} \{ \hat{\Psi}_{\eta,hb}^*(\chi) - \hat{\Psi}_{\eta,b}(\chi) \} \leq y \right) \right.$$

$$- P^* \left(\sqrt{nF_\chi(h)} \{ \hat{\Psi}_{\eta,h}(\chi) - \Psi_\eta(\chi) \} \leq y \right) \Bigg| \overset{a.s.}{\to} 0, \tag{5.4}$$

where P^* denotes probability conditioned on the sample $\{\mathcal{X}_1, \ldots, \mathcal{X}_n\}$.

Proof. Choosing a basis with $e_1 = \eta$, this theorem is a direct consequence of Theorem 5.1. □

6. Simulations

In this section, the theoretical results of the previous sections are illustrated through a simulation study. First we provide the details of the process of simulating a FAR(1) series. We use a linear functional series here since its stationarity can be guaranteed by existing theories. The performances of the kernel estimation and the behaviors of the bootstrap procedure will be shown through the experiments performed on the simulated data.

Data Generating Process. The simulated realization of a linear FAR(1) series has been discussed in Didericksen (2012) [7]. Curves in the series are assumed to be elements of the Hilbert space $L^2[0,1]$. The linear operator $\Psi(\mathcal{X}) = \int_0^1 \psi(s,t)\mathcal{X}(s)ds$ is acted on the functions $\mathcal{X}_i$'s, thereby the series are generated according to the model

$$\mathcal{X}_{n+1}(t) = \int_0^1 \psi(t,s)\mathcal{X}_n(s)ds + \mathcal{E}_{n+1}(t). \tag{6.1}$$

We use the kernel

$$\psi(s,t) = C \cdot s\mathbb{1}\{s \leq t\},$$

such that (6.1) becomes

$$\mathcal{X}_{n+1}(t) = C \int_0^t s\mathcal{X}_n(s)ds + \mathcal{E}_{n+1}(t), \tag{6.2}$$

Here, C is a normalizing constant to be chosen such that $||\Psi|| < 1$, which ensures the existence of a stationary causal solution to FAR(1) model; see Bosq (2000) [4]. We pick $C = 3$, such that $||\Psi|| = 0.5$.

We use the Brownian bridge process as the error process $\mathcal{E}(t)$ (see Didericksen (2012) [7]), which is defined by

$$\mathcal{E}(t) = W(t) - tW(t), \tag{6.3}$$

where $W(\cdot)$ is the standard Wiener process

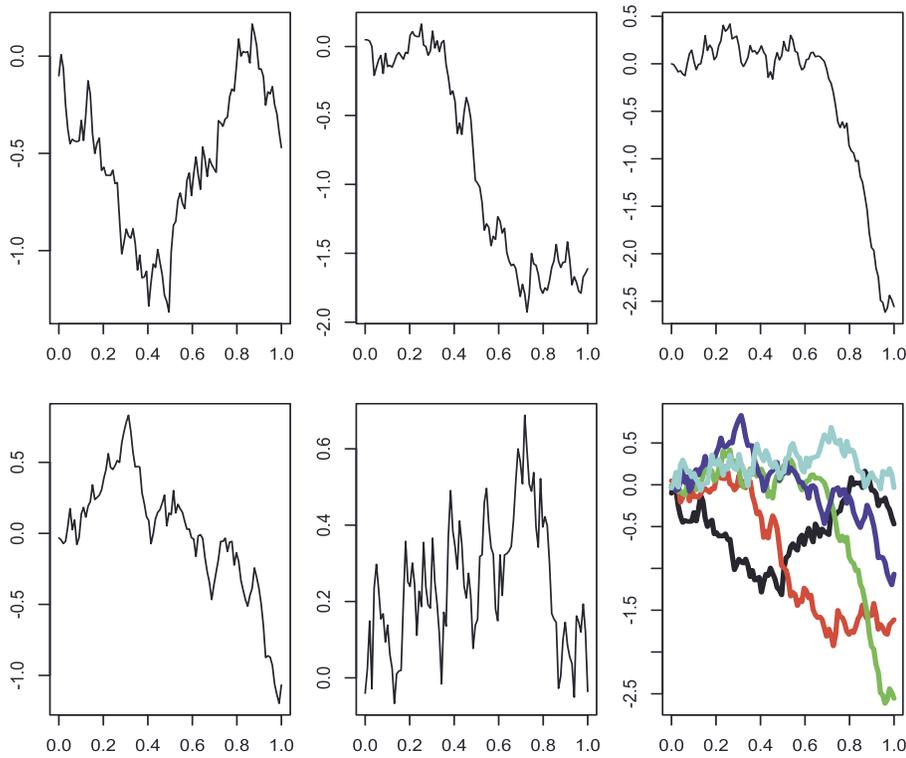$$W\left(\frac{k}{K}\right) = \frac{1}{\sqrt{K}} \sum_{j=1}^{k} Z_j, \quad k = 0, 1, \ldots, K,$$

FIG 1. *5 Curves $\mathcal{X}_{101}, \mathcal{X}_{102}, ..., \mathcal{X}_{105}$ from the sample.*

where Z_k's are standard independent normal and $Z_0 = 0$. The interval $[0, 1]$ is equally partitioned such that $0 = t_1 < t_2 < \cdots < t_{99} < t_p = 1$ with $p = 100$. We choose the initial curve $\mathcal{X}_1 = cos(t)$ for $t \in [0, 1]$, and build the series $\mathcal{X}_1, \ldots, \mathcal{X}_{250}$ according to the following scheme for $j = 1, \ldots, 100$:

$$\mathcal{X}_1(t_j) = cos(t_j),$$
$$\mathcal{X}_i(t_j) = 3 \int_0^{t_j} s\mathcal{X}_{i-1}(s)ds + \mathcal{E}_i(t_j), \quad i = 2, \ldots, 250.$$

Figure 1 displays the curves $\mathcal{X}_{101}, \mathcal{X}_{102}, ..., \mathcal{X}_{105}$. The last panel is the five curves combined. Out of the 250 curves we generate, the first 200 are used as the learning sample (i.e. $\{\mathcal{X}_i\}_{i=1,\ldots,200}$), and the last 50 make up the testing sample (i.e. $\{\mathcal{X}_i\}_{i=201,\ldots,250}$). The learning sample allows us to compute the kernel estimator while the testing sample will be accessed to assess the performance of the estimator and the behavior of bootstrap approximation.

Computing Kernel Estimator. With the simulated data, we use the learning sample to compute the kernel estimator by Eq. (2.2). The kernel function is chosen to be $K(u) = 1.5(1 - u^2)1_{[0,1]}(u)$, while the bandwidth h is selected through
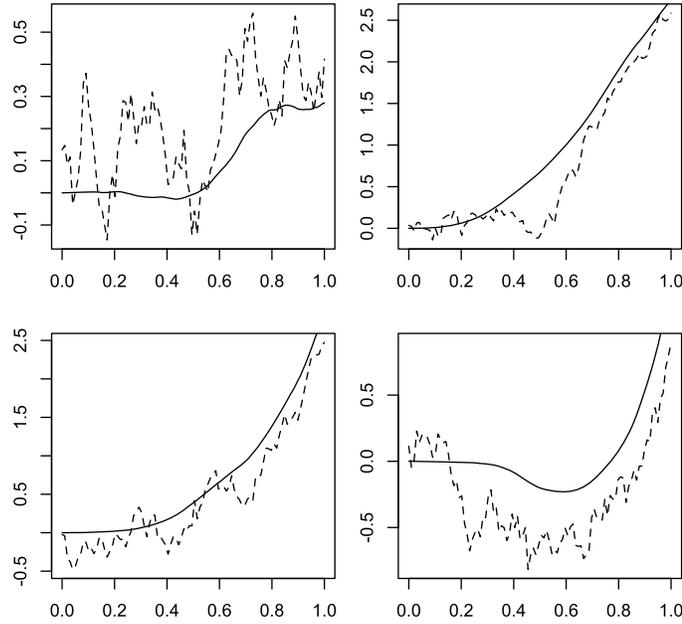
FIG 2. *Kernel estimations* $\hat{\Psi}_h(\chi)$ *(dashed lines); true operator* $\Psi(\chi)$ *(solid lines).*

a cross-validation procedure. A projection-based semi-metric is considered:

$$d(\chi_1, \chi_2) = \sqrt{\sum_{k=1}^{J} \langle \chi_1 - \chi_2, v_{k,n} \rangle^2},$$

where $v_{k,n}$, $k = 1, \ldots, J$ are orthonormal eigenfunctions associated with the largest J eigenvalues of the empirical covariance operator of the learning sample:

$$\mathcal{C}(\cdot) = \frac{1}{200} \sum_{i=1}^{200} \langle \mathcal{X}_i, \cdot \rangle \mathcal{X}_i.$$

Figure 2 compares the kernel estimation (i.e. $\hat{\Psi}_h(\chi)$) with the true operator (i.e. $\Psi(\chi)$) at $\chi = \mathcal{X}_{201}, \mathcal{X}_{202}, \mathcal{X}_{203}, \mathcal{X}_{204}$, which shows the quality of the kernel estimation (the relatively higher volatility of panel 1 and 4 is due to the scaling difference of y axis).

Bootstrap Approximation. To illustrate Theorem 5.2, we compare, for $k = 1, 2, 3, 4$, the density function of the componentwise bootstrapped error $f_{k,\chi}^*$

$$\langle \hat{\Psi}_{hb}^*(\chi) - \hat{\Psi}_b(\chi), v_{k,n} \rangle,$$

with the density function of the component wise true error $f_{k,\chi}^{true}$

$$\langle \hat{\Psi}_h(\chi) - \Psi(\chi), v_{k,n} \rangle.$$

To estimate $f_{k,\chi}^*$, we use the bootstrap procedure as described in Section 5:

1. compute $\hat{\Psi}_b(\chi)$ over the learning sample $\mathcal{X}_1, \ldots, \mathcal{X}_{200}$,
2. repeat 200 times the bootstrap algorithm introduced in the previous section to obtain

$$\hat{\Psi}_{hb}^{*1}(\chi), \ldots, \hat{\Psi}_{hb}^{*200}(\chi),$$

3. estimate the density $f_{k,\chi}^*$ over the 200 values

$$\langle \hat{\Psi}_{hb}^{*1}(\chi) - \hat{\Psi}_b(\chi), v_{k,n} \rangle, \ldots, \langle \hat{\Psi}_{hb}^{*200}(\chi) - \hat{\Psi}_b(\chi), v_{k,n} \rangle.$$

The Monte-Carlo scheme is used to estimate $f_{k,\chi}^{true}$:

1. build 200 samples $\{\mathcal{X}_1^s, \ldots, \mathcal{X}_{200}^s\}_{s=1,\ldots,200}$,
2. for the sth sample $\{\mathcal{X}_1^s, \ldots, \mathcal{X}_{200}^s\}$, compute $\hat{\Psi}_h^s$ to obtain

$$\hat{\Psi}_h^1(\chi), \ldots, \hat{\Psi}_h^{200}(\chi),$$

3. estimate the density $f_{k,\chi}^{true}$ over the 200 values

$$\langle \hat{\Psi}_h^1(\chi) - \Psi(\chi), v_{k,n} \rangle, \ldots, \langle \hat{\Psi}_h^{200}(\chi) - \Psi(\chi), v_{k,n} \rangle.$$

The bandwidth $h = h_{CV}$ is selected through a cross-validation procedure and we set $b = h$, which was the same setting used in the context of functional regression in Ferraty et al. (2010) [8]. They have studied the influence of both bandwidths on the behavior of the bootstrap procedure by varying b and h around h_{CV}. Their simulation showed the bootstrap works well for any combination of b and h, which leads to the conclusion that bootstrap results are not sensitive to the bandwidth choice.

Figure 3 presents the comparisons between the estimated $f_{k,\chi}^*$ and the estimated $f_{k,\chi}^{true}$ for the first four components, at the curves $\chi = \mathcal{X}_{201}, \ldots, \mathcal{X}_{205}$. The density estimation is performed with Gaussian kernel and the bandwidth being chosen three times the rule-of-thumb bandwidth estimator $(1.06\hat{\sigma}n^{-1/5})$ where $\hat{\sigma}$ is the sample standard deviation.

7. Bootstrap prediction regions

7.1. Construction of bootstrap prediction regions

Politis (2013) [26] constructed bootstrap prediction intervals in regression based on a bootstrap approximation to the distribution of the error in prediction—also called a 'predictive root'. This method was later extended to autoregression by Pan and Politis (2016) [25] who constructed bootstrap prediction intervals in the setting of a (univariate) nonparametric autoregression. Given a valid resampling procedure as developed in Section 5, we are able to construct bootstrap prediction regions in the functional autoregression model in the spirit of Pan and Politis (2016) [25]. Two bootstrap ideas, forward and backward, were considered in Pan and Politis (2016) [25] for the construction of prediction intervals with conditional validity. It is unrealistic, however, to generate the bootstrap pseudo-data backward in the functional setting. Therefore, we restrict to forward bootstrap scheme in our functional model.
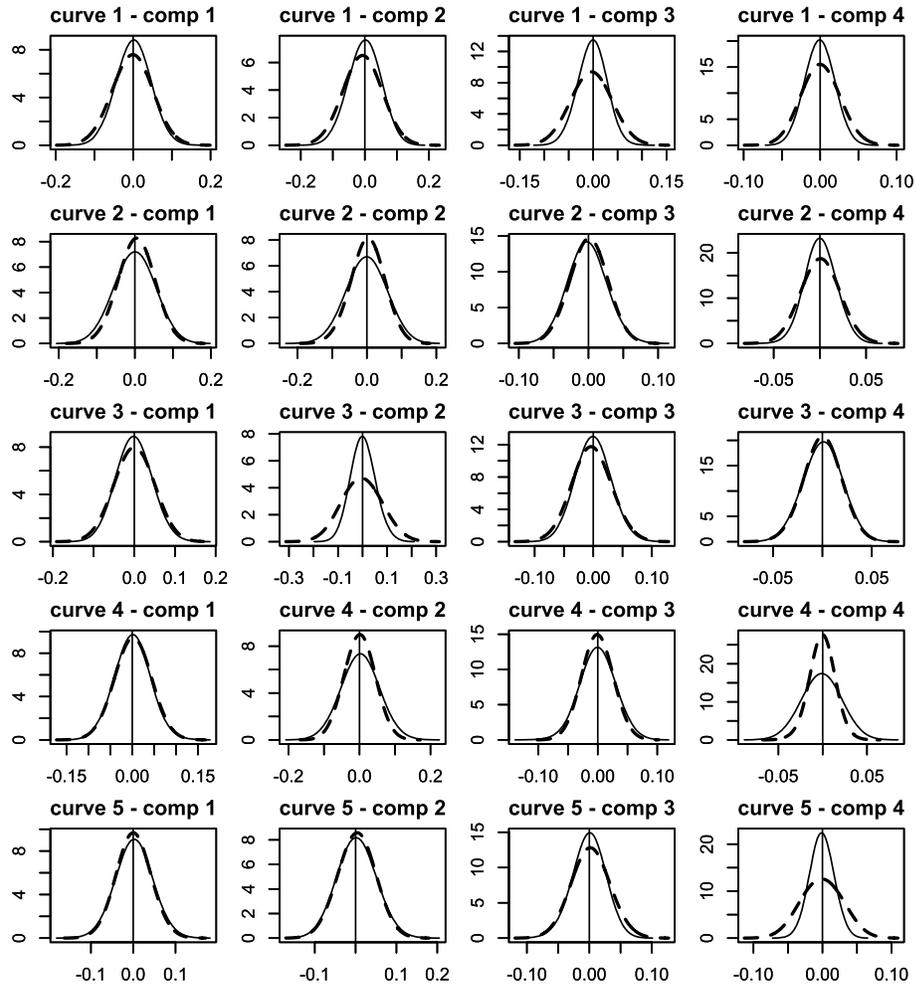
FIG 3. *Solid line: true error, dashed line: bootstrap error.*

Algorithm 7.1. *Bootstrap with fitted residual*

1. *Construct the Nadaraya-Watson kernel estimator $\hat{\Psi}_h(\cdot)$ by*

$$\hat{\Psi}_h(\chi) = \frac{\sum\limits_{i=1}^{n-1} \mathcal{X}_{i+1} K(h^{-1}d(\mathcal{X}_i, \chi))}{\sum\limits_{i=1}^{n-1} K(h^{-1}d(\mathcal{X}_i, \chi))}. \tag{7.1}$$

2. *Compute the fitted residuals: $\hat{\mathcal{E}}_{i,b} = \mathcal{X}_i - \hat{\Psi}_b(\mathcal{X}_{i-1})$ for $i = 1, \ldots, n-1$, where b is a second smoothing parameter.*
3. *Center the residuals: $\hat{r}_{i,b} = \hat{\mathcal{E}}_{i,b} - (n-1)^{-1} \sum_{i=1}^{n-1} \hat{\mathcal{E}}_{i,b}$*

(a) *Let the empirical distribution of $r_{i,b}$ be denoted by $\hat{F}_n$, and draw bootstrap pseudo-residuals $\mathcal{E}_1^*, \ldots, \mathcal{E}_n^*$ i.i.d. from $\hat{F}_n$.*

(b) *Use regression bootstrap scheme to generate the pseudo-data:*

$$\mathcal{X}_{i+1}^* = \hat{\Psi}_b(\mathcal{X}_i) + \mathcal{E}_{i+1}^* \quad for \ i = 1, \ldots, n-1$$

(c) *Re-estimate Ψ by*

$$\hat{\Psi}_{hb}^*(\chi) = \frac{\sum\limits_{i=1}^{n-1} \mathcal{X}_{i+1}^* K(h^{-1}d(\mathcal{X}_i, \chi))}{\sum\limits_{i=1}^{n-1} K(h^{-1}d(\mathcal{X}_i, \chi))}. \tag{7.2}$$

(d) *Calculate the bootstrap predictor*

$$\hat{\mathcal{X}}_{n+1}^* = \hat{\Psi}_{hb}^*(\mathcal{X}_n)$$

and the future bootstrap observation

$$\mathcal{X}_{n+1}^* = \hat{\Psi}_b(\mathcal{X}_n) + \mathcal{E}_{n+1}^*$$

where $\mathcal{E}_{n+1}^*$ is also drawn from $\hat{F}_n$.

(e) *Compute $||\mathcal{X}_{n+1}^* - \hat{\mathcal{X}}_{n+1}^*||$, where $||\cdot||$ is a norm of the practitioner's choice; note that difference choices for the norm will lead to prediction regions of different shape.*

4. *Repeat steps (a)-(e) in the above B times. The B bootstrap replicates $||\mathcal{X}_{n+1}^* - \hat{\mathcal{X}}_{n+1}^*||$ are collected in the form of an empirical distribution whose β-quantile is denoted $q(\beta)$.*

5. *Finally, the $(1-\beta)100\%$ bootstrap predictive region for $\mathcal{X}_{n+1}$ consists of all χ such that*

$$||\chi - \hat{\Psi}_h(\mathcal{X}_n)|| \leq q(\beta).$$

Analogously to Pan and Politis (2016) [25], we also consider using predictive, as opposed to fitted, residuals for the bootstrap. We define the predictive residuals as $\hat{\mathcal{E}}_{t,b}^{(t)} = \mathcal{X}_t - \hat{\mathcal{X}}_t^{(t)}$ where $\hat{\mathcal{X}}_t^{(t)}$ is calculated from the delete-$\mathcal{X}_t$ data set, i.e. the available data for the scatterplot of $\mathcal{X}_k$ vs. $\mathcal{X}_{k-1}$ over which the fitting takes place excludes the single point that corresponds to $k = t$. The forward bootstrap with predictive residuals is similar to Algorithm 7.1 except for Step 2.

Algorithm 7.2. *Bootstrap with predictive residuals*

1. *Same as step 1 of Algorithm 7.1.*
2. *Use the delete-$\mathcal{X}_t$ dataset to compute the delete-one kernel estimator*

$$\hat{\Psi}_b^{(t)}(\chi) = \frac{\sum\limits_{i=1,i\neq t}^{n-1} \mathcal{X}_{i+1} K(b^{-1}d(\mathcal{X}_i, \chi))}{\sum\limits_{i=1,i\neq t}^{n-1} K(b^{-1}d(\mathcal{X}_i, \chi))}, \quad for \ t = 1, \ldots, n-1. \tag{7.3}$$

Then compute the predictive residuals: $\hat{\mathcal{E}}_{t,b}^{(t)} = \mathcal{X}_t - \hat{\Psi}_b^{(t)}(\mathcal{X}_{t-1})$ for $t = 1, \ldots, n-1$, where b is a second smoothing parameter.

3-5. Replace $\hat{\mathcal{E}}_{t,b}$ by $\hat{\mathcal{E}}_{t,b}^{(t)}$ in Algorithm 7.1; the remaining steps are the same.

Remark 7.1. Recall that prediction intervals are asymptotically valid (in a conditional sense) when the probability of coverage of the future value $\mathcal{X}_{n+1}$ conditional on the observed data $\mathcal{X}_1, \ldots, \mathcal{X}_n$ gets close to the nominal one as n increases. The prediction regions constructed in Algorithm 7.1 and 7.2 would indeed be asymptotically valid if the predictive root $\mathcal{X}_{n+1} - \hat{\mathcal{X}}_{n+1}$ and the bootstrap predictive root $\mathcal{X}_{n+1}^* - \hat{\mathcal{X}}_{n+1}^*$ have the same distribution asymptotically. For the predictive root, we have

$$\mathcal{X}_{n+1} - \hat{\mathcal{X}}_{n+1} = \Psi(\mathcal{X}_n) - \hat{\Psi}_h(\mathcal{X}_n) + \mathcal{E}_{n+1} = A + \mathcal{E}_{n+1},$$

where $\mathcal{E}_{n+1}$ is independent of the estimation error A. Similarly, the bootstrap predictive root can be written as

$$\mathcal{X}_{n+1}^* - \hat{\mathcal{X}}_{n+1}^* = \hat{\Psi}_b(\mathcal{X}_n) - \hat{\Psi}_{hb}^*(\mathcal{X}_n) + \mathcal{E}_{n+1}^* = A^* + \mathcal{E}_{n+1}^*,$$

where $\mathcal{E}_{n+1}^*$ is independent of A^*. Consequently, the prediction regions of Algorithm 7.1 and 7.2 would be asymptotically valid provided that the distribution of the true errors $\mathcal{E}_1, \ldots, \mathcal{E}_n$ is captured in the limit by the empirical distribution of the residuals (fitted or predictive).

Corollary 7.1. *Assume the conditions of Theorem 4.4 and 5.2 as well as assumptions strong enough to ensure that $||\hat{\mathcal{E}}_{1,b}|| \xrightarrow{d} ||\mathcal{E}_1||$ as $n \to \infty$ where $|| \cdot ||$ is the norm appearing in Algorithm 7.1 and $\xrightarrow{d}$ denotes convergence in distribution. Then, the prediction region of Algorithm 7.1 is asymptotically valid. If in addition $||\hat{\mathcal{E}}_{1,b}^{(1)}|| \xrightarrow{d} ||\mathcal{E}_1||$, then the prediction region of Algorithm 7.2 is asymptotically valid as well.*

Remark 7.2. The proof of the corollary is immediate since, under the conditions of Theorem 4.4 and 5.2, we have both $A \xrightarrow{p} 0$ and $A^* \xrightarrow{p} 0$. Note that the condition $||\hat{\mathcal{E}}_{1,b}|| \xrightarrow{d} ||\mathcal{E}_1||$ would follow if $\hat{F}_n$, the empirical distribution of the residuals, were shown to converge weakly to the distribution of the true errors as shown e.g. by Franke and Nyarige (2016) [13] in the case of a linear operator Ψ. However, the importance of Corollary 7.1 is limited: as discussed in Politis (2015) [27]—asymptotic validity of prediction regions does not tell the whole story since the variability of estimation—which is asymptotically negligible—does not enter the picture. A prediction region will have good finite sample coverage only if the method employed is able to capture the variability of estimation error; that is why Algorithms 7.1 and 7.2 employ the bootstrap for prediction intervals instead of just relying on the empirical quantiles of $\hat{F}_n$. Indeed, by construction, our model-based bootstrap procedure is capable of approximating the distribution of the re-scaled estimation error $\sqrt{nF_\chi(h)}A$ by that of $\sqrt{nF_\chi(h)}A^*$ due to Theorem 5.2. Hence, our bootstrap prediction regions should have good finite sample coverage which is something that can not be captured by the property of asymptotic validity.

As mentioned in the introduction, Aue et al. (2015) [3] proposed a method of predicting FAR(1) using functional principal component analysis (FPCA). In addition, they proposed an algorithm for computing the prediction regions to assess the prediction accuracy. That is the only existing method we could find on the construction of prediction regions for functional time series. In the next subsection, we are going to compare it with the algorithm we proposed.

7.2. Monte Carlo studies

In this subsection, we use the Monte Carlo simulations to evaluate the quality of the bootstrap prediction regions with both fitted and predictive residuals. Simulations are performed on the model from the previous section; see Eq. (6.2). 500 'true' datasets each with sample size n are generated and n varies from 100 to 400. For the ith true dataset, we use one of the bootstrap methods to create $B = 1000$ bootstrap sample paths (step 4 of the algorithms), and construct the prediction region.

To assess the corresponding coverage level (CVR) of the constructed region, we also generate 1000 one-step ahead predictions $Y_{n+1,j} = \hat{\Psi}_i(\mathcal{X}_{n,i})$ for $j = 1, 2, ...1000$ where $\hat{\Psi}_i$ is the estimate from the ith 'true' dataset, $\mathcal{X}_{n,i}$ is the ith dataset's last data and $\mathcal{E}_j$ is randomly drawn from the error process (6.3). Then the empirical coverage level from the ith dataset is given by

$$CVR_i = \frac{1}{1000} \sum_{j=1}^{1000} \mathbb{1}\left\{ ||Y_{n+1,j} - \hat{\Psi}_b(\mathcal{X}_n)|| < q_i(\beta) \right\}$$

where $\mathbb{1}(\cdot)$ is the indicator function and β is the nominal coverage level. The coverage level for bootstrap methods is calculated by the average $\{CVR_i\}$ over the 500 'true' datasets, i.e.

$$CVR = \frac{1}{500} \sum_{i=1}^{500} CVR_i.$$

Note that the last observation $\mathcal{X}_{n,i}$ is different for each dataset; hence the coverage CVR represents an unconditional coverage probability.

Prediction regions are constructed with nominal coverage levels of 95% and 90%. Five different norms in the functional space are selected to measure the proximity between functions. The first three are the common l_p norms in the functional space:

- l_1: $||f||_1 = \int_0^1 |f(t)|dt$
- l_2: $||f||_2 = (\int_0^1 |f(t)|^2 dt)^{1/2}$
- l_∞: $||f||_\infty = sup_{t\in[0,1]}|f(t)|$.

We also consider two pointwise measures:

- 1st coordinate: $||x - y||_{coordinate} = |x(0) - y(0)|$
- 1st component: $||x - y||_{component} = |\langle x - y, v_{1,n}\rangle|$

Comparison between empirical coverage rate and nominal coverage rate using (a) fitted residuals and (b) predictive residuals.

fitted residuals	nominal coverage 95%			nominal coverage 90%		
n	100	200	400	100	200	400
l_1	0.956	0.955	0.953	0.914	0.910	0.907
l_2	0.956	0.955	0.953	0.914	0.910	0.908
l_∞	0.959	0.957	0.956	0.918	0.913	0.910
$1st$ coordinate	0.953	0.954	0.953	0.907	0.908	0.904
$1st$ component	0.954	0.955	0.955	0.907	0.909	0.908

(a) Fitted residuals

predictive residuals	nominal coverage 95%			nominal coverage 90%		
n	100	200	400	100	200	400
l_1	0.957	0.955	0.955	0.913	0.910	0.911
l_2	0.957	0.956	0.956	0.914	0.911	0.911
l_∞	0.960	0.959	0.958	0.918	0.915	0.916
$1st$ coordinate	0.956	0.954	0.955	0.911	0.908	0.909
$1st$ component	0.956	0.957	0.956	0.912	0.913	0.910

(b) Predictive residuals

Empirical coverage rate compared to the nominal coverage rate 95%.

n	100	200	400
$p = 2$	0.859	0.923	0.948
$p = 3$	0.874	0.928	0.945
$p = 4$	0.881	0.931	0.951

where x and y are functions in $L^2[0,1]$ and e_1 is the eigenfunction associated with the largest eigenvalue of the sample covariance operator. Table 1 presents the empirical coverage rate of the prediction region we construct. The results are promising as the empirical coverage rates pretty well match the nominal coverage rate even with quite small sample size. As expected, when using predictive residuals, the coverage rate is a bit higher compared to fitted residuals.

For the comparison purpose, we apply the algorithm of Aue et al. (2015) [3] on the functional data we generate. Table 2 presents the coverage rate of their constructed prediction region. The value of p represents the number of components to be kept in the prediction algorithm; see Aue et al. (2015) [3] for more details. The experiment is also done on three different sample size, 100, 200 and 400. As is shown in the table, their prediction regions fail to achieve the ideal coverage rate when sample size is small–see the case $n = 100$–and results get better when sample size grows. This phenomenon is not surprising, and explainable. Recall that there are two constituents in the expression of predictive root $\mathcal{X}_{n+1} - \hat{\mathcal{X}}_{n+1}$, the estimation error A and the true error $\mathcal{E}$. The bootstrap predictive root in our algorithm is capable of capturing the distribution of both errors by A^* and $\mathcal{E}^*$. However, the method of Aue et al. (2015) [3] does not attempt to capture the variability of estimation error and thus, as expected, it yields coverages less than nominal (undercoverage). Coverage rate improves under larger sample size since the estimation error diminishes as n grows. As

a conclusion, we can say that the prediction regions constructed through our bootstrap procedure can achieve better finite sample coverage.

8. Appendix: Proofs

Throughout this section, given some random variable U, P^U stands for the probability measure induced by U. Since the stationarity of the sequence $\{\mathcal{X}_n\}$, we assume $\mathcal{X}_i's$ are identically distributed as $\mathcal{X}$. Also, the pairs $(X_{i+1}, \mathcal{X}_i)$ have the same joint distribution. So are the functions of them. The kernel estimator $\hat{\psi}_h$ will be decomposed as follows:

$$\hat{\psi}_h(\chi) = \frac{\hat{g}_1(\chi)}{\hat{f}(\chi)},$$

where

$$\hat{g}_1(\chi) = \frac{1}{nF_\chi(h)} \sum_{i=1}^n X_{i+1} K\left(\frac{d(\mathcal{X}_i, \chi)}{h}\right)$$

and

$$\hat{f}(\chi) = \frac{1}{nF_\chi(h)} \sum_{i=1}^n K\left(\frac{d(\mathcal{X}_i, \chi)}{h}\right).$$

Similarly, the estimator $\hat{\Psi}_h$ can be decomposed as:

$$\hat{\Psi}_h(\chi) = \frac{\hat{g}_2(\chi)}{\hat{f}(\chi)},$$

where

$$\hat{g}_2(\chi) = \frac{1}{nF_\chi(h)} \sum_{i=1}^n \mathcal{X}_{i+1} K\left(\frac{d(\mathcal{X}_i, \chi)}{h}\right).$$

8.1. *Proof of Theorem 4.1, 4.2 and 4.3*

To prove these theorems, we need the following lemmas:

Lemma 8.1.

$$E[\hat{f}(\chi)] \to K(1) - \int_0^1 K'(s)\tau_{0\chi}(s)ds = M_1, \tag{8.1}$$

$$E[\hat{g}_1(\chi)] \to \psi(\chi)\left(K(1) - \int_0^1 K'(s)\tau_{0\chi}(s)ds\right) = \psi(\chi)M_1. \tag{8.2}$$

Proof. For the first assertion, we have

$$E(\hat{f}(\chi)) = \frac{1}{F_\chi(h)} E\left(K\left(\frac{d(\mathcal{X}, \chi)}{h}\right)\right)$$

$$= \frac{1}{F_\chi(h)} \int K(t) dP^{d(\mathcal{X},\chi)/h}(t)$$

$$= \frac{1}{F_\chi(h)} \left(K(1)F_\chi(h) - \int_0^1 K'(s)F_\chi(hs)ds \right)$$

$$= K(1) - \int_0^1 K'(s)\frac{F_\chi(hs)}{F_\chi(h)}ds$$

$$\to K(1) - \int_0^1 K'(s)\tau_{0\chi}(s)ds = M_1,$$

the third line coming from Eq. (14) in Ferraty et al. (2007) [10] and the fifth line follows from the uniform boundedness of the integrand. The second assertion is proved as follows:

$$E(\hat{g}_1(\chi)) = \frac{1}{F_\chi(h)} E \left(X_{i+1} K \left(\frac{d(\mathcal{X}_i,\chi)}{h} \right) \right)$$

$$= \frac{1}{F_\chi(h)} E \left(E(X_{i+1}|\mathcal{X}_i) K \left(\frac{d(\mathcal{X}_i,\chi)}{h} \right) \right)$$

$$= \frac{1}{F_\chi(h)} E \left(\psi(\mathcal{X}_i) K \left(\frac{d(\mathcal{X}_i,\chi)}{h} \right) \right)$$

$$= \frac{1}{F_\chi(h)} E \left(\psi(\mathcal{X}) K \left(\frac{d(\mathcal{X},\chi)}{h} \right) \right)$$

$$= \frac{1}{F_\chi(h)} E \left((\psi(\chi) + o(1)) K \left(\frac{d(\mathcal{X},\chi)}{h} \right) \right)$$

$$\to \psi(\chi) M_1$$

where the fifth line follows from the continuity of ψ with respect to the semimetric d.   □

Lemma 8.2. *Under the assumption (A1)-(A8), we have*

$$Var(\hat{f}(\chi)) = \frac{M_2}{nF_\chi(h)}(1 + o(1)), \tag{8.3}$$

$$Var(\hat{g}_1(\chi)) = (\sigma_\varepsilon^2 + \psi^2(\chi))\frac{M_2}{nF_\chi(h)}(1 + o(1)), \tag{8.4}$$

$$Cov(\hat{f}(\chi), \hat{g}_1(\chi)) = \psi(\chi)\frac{M_2}{nF_\chi(h)}(1 + o(1)). \tag{8.5}$$

Proof. Since the functional sequence $\{\mathcal{X}_n\}$ is stationary and strong mixing, the n pairs $(\mathcal{X}_i, X_{i+1})_{i=1,\cdots,n}$ are identically distributed and the sequence $(\mathcal{X}_i, X_{i+1})_i$ is also strong mixing. Then the results follows directly from Lemma 2.5 in Delsol (2009) [6].   □

Proof of Theorem 4.1. For the proof of Eq. (4.1), we write the following decomposition:

$$E(\hat{\psi}(\chi)) = \frac{E(\hat{g}_1(\chi))}{E(\hat{f}(\chi))} + \frac{A_1}{(E(\hat{f}(\chi)))^2} + \frac{A_2}{(E(\hat{f}(\chi)))^2}, \tag{8.6}$$

with

$$A_1 = E(\hat{g}_1(\chi)(E(\hat{f}(\chi)) - \hat{f}(\chi))) \tag{8.7}$$

and

$$A_2 = E((\hat{f}(\chi) - E(\hat{f}(\chi)))^2 \hat{\psi}_h(\chi)). \tag{8.8}$$

The last two terms on the right-hand side of Eq. (8.6) are negligible since we have $A_1 = O((nF_\chi(h)^{-1}))$ and $A_2 = O((nF_\chi(h)^{-1}))$, both are direct consequences of Lemma 8.1 and 8.2 (see proof of Lemma 7.5 in Raña et al. (2016) [31]). For the first term on the right-hand side of (8.6), we calculate

$$\frac{E(\hat{g}_1(\chi))}{E(\hat{f}(\chi))} - \psi(\chi) = \frac{E\left((X_{i+1} - \psi(\chi))K\left(\frac{d(\mathcal{X}_i,\chi)}{h}\right)\right)}{E\left(K\left(\frac{d(\mathcal{X}_i,\chi)}{h}\right)\right)}, \tag{8.9}$$

which follows from the stationarity of the sequence $\{\mathcal{X}_n\}$. Assume $\mathcal{X}_i \sim \mathcal{X}$ for $i = 1, \cdots, \infty$, then we can write the numerator in (8.9) as follows

$$
\begin{aligned}
E\left((X_{i+1} - \psi(\chi))K\left(\frac{d(\mathcal{X}_i, \chi)}{h}\right)\right) &= E\left((\hat{\psi}(\mathcal{X}_i) - \psi(\chi))K\left(\frac{d(\mathcal{X}_i, \chi)}{h}\right)\right) \\
&= E\left((\hat{\psi}(\mathcal{X}) - \psi(\chi))K\left(\frac{d(\mathcal{X}, \chi)}{h}\right)\right) \\
&= E\left(\varphi_\chi(d(\mathcal{X}, \chi))K\left(\frac{d(\mathcal{X}, \chi)}{h}\right)\right) \\
&= \int \varphi_\chi(t)K(\frac{t}{h})dP^{d(\mathcal{X},\chi)}(t) \\
&= \int \varphi_\chi(ht)K(t)dP^{d(\mathcal{X},\chi)/h}(t) \\
&= h\varphi_\chi{}'(0)\int tK(t)dP^{d(\mathcal{X},\chi)/h}(t) + o(h),
\end{aligned}
$$

here P^U stands for the probability measure induced by U, and the last line comes form the first order Taylor's expansion for φ around 0. For the denominator in (8.9), we have

$$E\left(K\left(\frac{d(\mathcal{X}_i,\chi)}{h}\right)\right) = \int K(t)dP^{d(\mathcal{X},\chi)/h}(t).$$

Consequently,

$$\frac{E(\hat{g}_1(\chi))}{E(\hat{f}(\chi))} - \psi(\chi) = h\varphi_\chi{}'(0)I + o(h), \tag{8.10}$$

where

$$I = \frac{\int tK(t)dP^{d(\mathcal{X},\chi)/h}(t)}{\int K(t)dP^{d(\mathcal{X},\chi)/h}(t)}.$$

By Lemma 2 in Ferraty et al. (2007) [10], $I \to M_0/M_1$ as $n \to +\infty$. Finally, combining (8.6) and (8.10), we obtain

$$E[\hat{\psi}_h(\chi)] - \psi(\chi) = \varphi'_\chi(0)\frac{M_0}{M_1}h + O(\frac{1}{nF_\chi(h)}) + o(h),$$

which completes the proof. □

Proof of Theorem 4.2. We use the following decomposition:

$$
\begin{aligned}
Var(\hat{\psi}_h(\chi)) =& \frac{Var(\hat{g}_1(\chi))}{(E(\hat{f}(\chi)))^2} - 4\frac{E(\hat{g}_1(\chi)Cov(\hat{g}_1(\chi),\hat{f}(\chi)))}{(E(\hat{f}(\chi)))^3} \\
& + 3Var(\hat{f}(\chi))\frac{(E(\hat{g}_1(\chi)))^2}{(E(\hat{f}(\chi)))^4} + o(\frac{1}{nF_\chi(h)}).
\end{aligned}
\tag{8.11}
$$

Then Theorem 4.2 follows from this decomposition together with Lemma 8.1 and 8.2.

Proof of Theorem 4.3. Since the sequence $(\mathcal{X}_i, X_{i+1})_i$ is stationary and strong mixing, this theorem is the same as that obtained in Delsol (2009) [6].

8.2. Proof of Theorem 4.4

Proof. Consider the expression

$$\sqrt{nF_\chi(h)^{1+\delta}}\left[\hat{\Psi}_h(\chi) - \Psi(\chi) - \mathcal{B}_n\right].\tag{8.12}$$

Following similar arguments as in the proof of Theorem 4.1 in Ferraty et al. (2012) [9], we have that the above expression has the same asymptotic distribution as

$$
\begin{aligned}
& \frac{\sqrt{nF_\chi(h)^{1+\delta}}}{M_1}\left[\hat{g}_2(\chi) - E\hat{g}_2(\chi) - (\hat{f}(\chi) - E\hat{f}(\chi))\Psi(\chi)\right] \\
=& \frac{1}{M_1\sqrt{nF_\chi(h)^{1-\delta}}}\sum_{i=1}^n\left[\mathcal{X}_{i+1}K\left(\frac{d(\mathcal{X}_i,\chi)}{h}\right) - E\left\{\Psi(\mathcal{X})K\left(\frac{d(\mathcal{X},\chi)}{h}\right)\right\}\right. \\
& \left. - \Psi(\chi)K\left(\frac{d(\mathcal{X}_i,\chi)}{h}\right) + \Psi(\chi)E\left\{K\left(\frac{d(\mathcal{X},\chi)}{h}\right)\right\}\right] \\
=& \frac{1}{\sqrt{n}}\sum_{i=1}^n(Z_{n,i} - EZ_{n,i}),
\end{aligned}
$$

where for $1 \le i \le n$,

$$Z_{n,i} = \frac{1}{M_1\sqrt{F_\chi(h)^{1-\delta}}}\left[\mathcal{X}_{i+1}K\left(\frac{d(\mathcal{X}_i,\chi)}{h}\right) - \Psi(\chi)K\left(\frac{d(\mathcal{X}_i,\chi)}{h}\right)\right].$$

By assumption (i), we can apply Hölder's inequality to obtain

$$E||Z_{n,i}||^{2+\delta}$$

$$= \frac{1}{M_1^{2+\delta} F_\chi(h)^{\frac{(1-\delta)(2+\delta)}{2}}} E\left(||\mathcal{X}_{i+1} - \Psi(\chi)||^{2+\delta} \left\{ K\left(\frac{d(\mathcal{X}_i, \chi)}{h}\right) \right\}^{2+\delta} \right)$$

$$\leq \frac{1}{M_1^{2+\delta} F_\chi(h)^{\frac{(1-\delta)(2+\delta)}{2}}} \left(E||\mathcal{X}_{i+1} - \Psi(\chi)||^{2+\delta'} \right)^{\frac{2+\delta}{2+\delta'}}$$

$$\times \left\{ E\left[K\left(\frac{d(\mathcal{X}_i, \chi)}{h}\right) \right]^{\frac{2}{1-\delta}} \right\}^{\frac{(1-\delta)(2+\delta)}{2}}.$$

In the above expression, $\left(E||\mathcal{X}_{i+1} - \Psi(\chi)||^{2+\delta'} \right)^{\frac{2+\delta}{2+\delta'}}$ is finite because of assumption (ii). For the last item, we note that

$$K^{\frac{2}{1-\delta}}(t) = K^{\frac{2}{1-\delta}}(1) - \int_t^1 (K^{\frac{2}{1-\delta}}(s))' ds.$$

Applying Fubini's Theorem, we obtain

$$E\left[K\left(\frac{d(\mathcal{X}_i, \chi)}{h}\right) \right]^{\frac{2}{1-\delta}} = \int_0^1 K^{\frac{2}{1-\delta}}(t) dP^{d(\mathcal{X}, \chi)/h}(t)$$

$$= K^{\frac{2}{1-\delta}}(1) F_\chi(h) - \int_0^1 \left(\int_t^1 (K^{\frac{2}{1-\delta}}(s))' ds \right) dP^{d(\mathcal{X}, \chi)/h}(t)$$

$$= K^{\frac{2}{1-\delta}}(1) F_\chi(h) - \int_0^1 (K^{\frac{2}{1-\delta}}(s))' F_\chi(hs) ds$$

$$= F_\chi(h) \left(K^{\frac{2}{1-\delta}}(1) - \int_0^1 (K^{\frac{2}{1-\delta}}(s))' \tau_{h\chi}(s) ds \right).$$

As the factor of $F_\chi(h)$ converges to $M_{\frac{2}{1-\delta}}$ for $h \to 0$ where

$$M_{\frac{2}{1-\delta}} = K^{\frac{2}{1-\delta}}(1) - \int_0^1 (K^{\frac{2}{1-\delta}}(s))' \tau_{0\chi}(s) ds$$

is a constant depending on $K(\cdot)$, the right-hand side is bounded by $C F_\chi(h)$ for a suitable C and all small enough h.

Now we have, $E||Z_{n,i}||^{2+\delta} \leq C < \infty$ for all n. Combining this with assumption (iii), it follows from Theorem 2.3(i) in Politis and Romano (1994) [28] that $\frac{1}{\sqrt{n}} \sum_{i=1}^n (Z_{n,i} - EZ_{n,i})$ converge weakly to a Gaussian measure with mean 0 in $\mathbb{H}$. Hence, (8.12) converges weakly to the same measure and

$$\hat{\Psi}_h(\chi) = \Psi(\chi) - \mathcal{B}_n + O_p\left(\frac{1}{\sqrt{nF_\chi(h)^{1+\delta}}}\right). \tag{8.13}$$

8.3. Proof of Theorem 5.1

Let us first state some additional assumptions needed:

(D1) For some $\beta > 0$, all $0 \le s \le \beta$ and all χ_1 in a neighborhood of χ with respect to the semi-metric d, $\varphi_{\chi_1}(0) = 0$, $\varphi'_{\chi_1}(s)$ exists, $\varphi'_{\chi_1}(0) \ne 0$, and $\varphi'_{\chi_1}(s)$ is uniformly Hölder continuous of order $0 < \alpha \le 1$ at 0, i.e. there exists a constant $0 < L < \infty$ such that

$$|\varphi'_{\chi_1}(s) - \varphi'_{\chi_1}(0)| \le Ls^{\alpha}, \quad \text{for all } 0 \le s \le \beta,$$

uniformly in χ_1 in a neighborhood of χ.

(D2) For all $\chi_1 \in \mathbb{H}$, $F_{\chi_1}(0) = 0$ and $F_{\chi_1}(t)/F_{\chi}(t)$ is Hölder continuous of order α in χ_1, uniformly in t in a neighborhood of 0 (with α defined in (D1)), i.e. for some $\beta > 0$, there exists a constant $0 < M < \infty$ such that

$$|(F_{\chi_1}(t) - F_{\chi_2}(t))/F_{\chi}(t)| \le Md(\chi_1, \chi_2)^{\alpha}, \quad \text{for all } \chi_1, \chi_2 \in \mathbb{H},$$

uniformly for all $0 < t \le \beta$.

(D3) For all $\chi_1 \in \mathbb{H}$ and all $0 \le s \le 1$, $\tau_{0\chi_1}(s)$ exists, $\sup_{\chi_1 \in \mathbb{H}, \, 0 \le s \le 1}|\tau_{h\chi_1}(s) - \tau_{0\chi_1}(s)| = o(1)$, $M_{0\chi} > 0$, $M_{2\chi} > 0$, $\inf_{d(\chi_1,\chi)<\varepsilon}M_{1\chi_1} > 0$ for some $\varepsilon > 0$, and $M_{k\chi_1}$ is Hölder continuous of order α for $k = 0, 1, 2$, (with α defined in (D1)), i.e. for $k = 0, 1, 2$, there exists a $0 < N_k < \infty$ such that

$$|M_{k\chi_1} - M_{k\chi_2}| \le N_k d(\chi_1, \chi_2)^{\alpha}, \quad \text{for all } \chi_1, \chi_2 \in \mathbb{H}.$$

(Here the quantities M_k, $k = 0, 1, 2$ are indexed with an additional index χ, which is necessary to the above continuity conditions of M_k as a function of χ.)

(D4) $E(|X_{i+1}||\mathcal{X}_i = \cdot)$ is continuous in a neighborhood of χ with respect to the semi-metric d, and $\sup_{d(\chi_1,\chi)<\epsilon}E(|X_{i+1}|^m|\mathcal{X}_i = \chi_1) < \infty$ for some $\epsilon > 0$ and all $m \ge 1$.

(D5) $b \to 0$, $h/b \to 0$, $h(nF_{\chi}(h))^{1/2} = O(1)$, $n^{1/p}F_{\chi}(h)^{1/2}log(n) = o(1)$, $b^{1+\alpha}(nF_{\chi}(h))^{1/2} = o(1)$, $F_{\chi}(b+h)/F_{\chi}(b) \to 1$, $[F_{\chi}(h)/F_{\chi}(b)]log(n) = o(1)$, $F_{\chi}(b)^{-1}h/b = O(1)$, and $bh^{\alpha-1} = O(1)$ (with p and α defined in (A6) and (D1), respectively).

(D6) For each n, there exist $r_n \ge 1$, $l_n > 0$ and curves $t_{1n}, \ldots, t_{r_n n}$ such that $B(\chi, h) \subset \cup_{k=1}^{r_n} B(t_{kn}, l_n)$, with $r_n = O(n^{b/h})$ and $l_n = o(b(nF_{\chi}(h))^{-1/2})$, where $B(\chi, t) = \{\chi \in \mathbb{H} : d(\chi_1, \chi) \le t\}$ is the ball in $\mathbb{H}$ with center χ and radius t.

(D7) $a > 4.5$ (with a defined in (A8)).

Note that (D1)-(D3) are regularity conditions related to the smoothness of the functions ψ, φ_{χ}, F_{χ} and $\tau_{0\chi}$, which have been used in the context of bootstrap in functional regression. Assumption (D7), along with $n^{1/p}F_{\chi}(h)^{1/2}log(n) = o(1)$ and $F_{\chi}(b)^{-1}h/b = O(1)$ in (D5) are additional assumptions made in the dependent case in Raña et al. (2016) [31] which facilitates the proof of Lemma 8.4. See more details on these assumptions in Ferraty et al. (2007) [10], (2010) [8].

Proof. The expression between the absolute values of (5.2) can be decomposed as

$$P^* \left(\sqrt{nF_\chi(h)} \{\hat{\psi}_{hb}^*(\chi) - \hat{\psi}_b(\chi)\} \le y \right) - P \left(\sqrt{nF_\chi(h)} \{\hat{\psi}_h(\chi) - \psi(\chi)\} \le y \right)$$
$$= T_1(y) + T_2(y) + T_3(y)$$

where

$$T_1(y) = P^* \left(\sqrt{nF_\chi(h)} \{\hat{\psi}_{hb}^*(\chi) - \hat{\psi}_b(\chi)\} \le y \right)$$
$$- \Phi \left(\frac{y - \sqrt{nF_\chi(h)} \{E^* \hat{\psi}_{hb}^*(\chi) - \hat{\psi}_b(\chi)\}}{\sqrt{nF_\chi(h) var^*(\hat{\psi}_{hb}^*(\chi))}} \right),$$

$$T_2(y) = \Phi \left(\frac{y - \sqrt{nF_\chi(h)} \{E^* \hat{\psi}_{hb}^*(\chi) - \hat{\psi}_b(\chi)\}}{\sqrt{nF_\chi(h) var^*(\hat{\psi}_{hb}^*(\chi))}} \right)$$
$$- \Phi \left(\frac{y - \sqrt{nF_\chi(h)} \{E \hat{\psi}_h(\chi) - \psi(\chi)\}}{\sqrt{nF_\chi(h) var(\hat{\psi}_h(\chi))}} \right)$$

and

$$T_3(y) = \Phi \left(\frac{y - \sqrt{nF_\chi(h)} \{E \hat{\psi}_h(\chi) - \psi(\chi)\}}{\sqrt{nF_\chi(h) var(\hat{\psi}_h(\chi))}} \right)$$
$$- P \left(\sqrt{nF_\chi(h)} \{\hat{\psi}_h(\chi) - \psi(\chi)\} \le y \right).$$

By the asymptotic normality of $\hat{\psi}_h$ given in Theorem 4.3, we have $T_3(y) \to 0$. The a.s. convergence to 0 of $T_1(y)$ is given by the asymptotic normality of $\hat{\psi}_{hb}^*$ conditioning on $\mathcal{X}_i, i = 1, ..., n$, which is proved below.

We decompose $\hat{\psi}_{hb}^*$ as follows

$$\hat{\psi}_{hb}^*(\chi) = \frac{\sum_{i=1}^{n-1} X_{i+1}^* K(h^{-1} d(\mathcal{X}_i, \chi))}{\sum_{i=1}^{n-1} K(h^{-1} d(\mathcal{X}_i, \chi))} = \frac{\hat{g}^*(\chi)}{\hat{f}(\chi)},$$

where

$$\hat{g}^*(\chi) = \frac{1}{nF_\chi(h)} \sum_{i=1}^{n-1} X_{i+1}^* K(h^{-1} d(\mathcal{X}_i, \chi)),$$

$$\hat{f}(\chi) = \frac{1}{nF_\chi(h)} \sum_{i=1}^{n-1} K(h^{-1} d(\mathcal{X}_i, \chi)).$$

Then we have

$$\hat{g}^*(\chi) = \frac{1}{nF_\chi(h)} \sum_{i=1}^{n-1} (\hat{\psi}_b(\mathcal{X}_i) + \varepsilon^*_{i+1}) K(h^{-1}d(\mathcal{X}_i, \chi)),$$

$$E^*(\hat{g}^*(\chi)) = \frac{1}{nF_\chi(h)} \sum_{i=1}^{n-1} (\hat{\psi}_b(\mathcal{X}_i) + E^*\varepsilon^*_{i+1}) K(h^{-1}d(\mathcal{X}_i, \chi)).$$

Therefore,

$$\frac{\hat{\psi}^*_{hb}(\chi) - E^*(\hat{\psi}^*_{hb}(\chi))}{\sqrt{var^*(\hat{\psi}^*_{hb}(\chi))}} = \frac{\frac{\hat{g}^*(\chi)}{\hat{f}(\chi)} - E^*(\frac{\hat{g}^*(\chi)}{\hat{f}(\chi)})}{\sqrt{var^*(\frac{\hat{g}^*(\chi)}{\hat{f}(\chi)})}}$$

$$= \frac{\hat{g}^*(\chi) - E^*(\hat{g}^*(\chi))}{\sqrt{var^*(\hat{g}^*(\chi))}}$$

$$= \frac{\hat{h}^*(\chi) - E^*(\hat{h}^*(\chi))}{\sqrt{var^*(\hat{h}^*(\chi))}}$$

where

$$\hat{h}^*(\chi) = \frac{1}{nF_\chi(h)} \sum_{i=1}^{n-1} \varepsilon^*_{i+1} K(h^{-1}d(\mathcal{X}_i, \chi)).$$

$\hat{h}^*(\chi)$ is a sum of a mixing sequence and its asymptotic normality follows from the similar arguments in the proof of Theorem 3 in Delsol (2009) [6].

A special case is when $K(\cdot) = \mathbb{1}_{[0,1]}(\cdot)$, under which

$$\hat{h}^*(\chi) = \frac{1}{\#\{i : d(\mathcal{X}_i, \chi) \leq h\}} \sum_{i:d(\mathcal{X}_i,\chi)\leq h} \varepsilon^*_{i+1},$$

so that $\hat{h}^*(\chi)$ is a sum of independent random variable given $\mathcal{X}_i, i = 1, \ldots, n$. and the asymptotic normality follows directly.

It remains to consider $T_2(y)$, and its a.s convergence to 0 follows from Lemma 8.3 and 8.4 that follow:

Lemma 8.3. *Assume (A1), (A4) and (D1)-(D5). Then*

$$\frac{var^*[\hat{\psi}^*_{hb}(\chi)]}{var[\hat{\psi}_h(\chi)]} \to 1 \quad a.s.$$

Proof. Define $\hat{\sigma}^2_\varepsilon = n^{-1} \sum_{i=1}^{n} (\hat{\varepsilon}_{i,b} - \bar{\hat{\varepsilon}}_b)^2$. Then

$$var^*[\hat{\psi}^*_{hb}(\chi)] = var^* \left[\frac{\sum_{i=1}^{n-1} (\hat{\psi}_b(\mathcal{X}_i) + \varepsilon^*_{i+1}) K(h^{-1}d(\mathcal{X}_i, \chi))}{\sum_{i=1}^{n-1} K(h^{-1}d(\mathcal{X}_i, \chi))} \right]$$

$$= var^* \left[\frac{\sum_{i=1}^{n-1} \varepsilon_{i+1}^* K(h^{-1}d(\mathcal{X}_i, \chi))}{\sum_{i=1}^{n-1} K(h^{-1}d(\mathcal{X}_i, \chi))} \right]$$

$$= \frac{\sum_{i=1}^{n-1} K^2(h^{-1}d(\mathcal{X}_i, \chi)) var^*(\varepsilon_{i+1}^*)}{\left(\sum_{i=1}^{n-1} K(h^{-1}d(\mathcal{X}_i, \chi)) \right)^2}$$

$$= \frac{\hat{\sigma}_\varepsilon^2}{\hat{f}(\chi)^2} (nF_\chi(h))^{-2} \sum_{i=1}^{n-1} K^2(h^{-1}d(\mathcal{X}_i, \chi))$$

$$= \frac{\sigma_\varepsilon^2}{E[\hat{f}(\chi)]^2} (nF_\chi^2(h))^{-1} \cdot E[K^2(h^{-1}d(\mathcal{X}_i, \chi))] \cdot (1 + o(1))$$

$$= \frac{\sigma_\varepsilon^2}{M_1^2} \frac{M_2}{nF_\chi(h)} (1 + o(1))$$

$$= var[\hat{\psi}_h(\chi)] + o((nF_\chi(h))^{-1}).$$

Since $var[\hat{\psi}_h(\chi)] = O((nF_\chi(h))^{-1})$ by Theorem 4.2, the result follows by dividing $var[\hat{\psi}_h(\chi)]$ on both sides. $\square$

Lemma 8.4. *Assume (A1)-(A11) and (D1)-(D7). Then*

$$\sqrt{nF_\chi(h)}\{E^*[\hat{\psi}_{hb}^*(\chi)] - \hat{\psi}_b(\chi) - E[\hat{\psi}_h(\chi)] + \psi(\chi)\} \xrightarrow{a.s.} 0.$$

Proof. Write

$$E^*[\hat{\psi}_{hb}^*(\chi)] - \hat{\psi}_b(\chi)$$

$$= E^* \left[\frac{\sum_{i=1}^{n-1} (\hat{\psi}_b(\mathcal{X}_i) + \varepsilon_{i+1}^*) K(h^{-1}d(\mathcal{X}_i, \chi))}{\sum_{i=1}^{n-1} K(h^{-1}d(\mathcal{X}_i, \chi))} \right] - \hat{\psi}_b(\chi)$$

$$= \frac{\sum_{i=1}^{n-1} \hat{\psi}_b(\mathcal{X}_i) K(h^{-1}d(\mathcal{X}_i, \chi))}{\sum_{i=1}^{n-1} K(h^{-1}d(\mathcal{X}_i, \chi))} - \hat{\psi}_b(\chi)$$

$$= \frac{(nF_\chi(h))^{-1}}{\hat{f}_h(\chi)} \sum_{i=1}^{n-1} \{\hat{\psi}_b(\mathcal{X}_i) - \hat{\psi}_b(\chi)\} K(h^{-1}d(\mathcal{X}_i, \chi))$$

$$= \frac{(nF_\chi(h))^{-1}}{\hat{f}_h(\chi)} \sum_{i=1}^{n-1} \{\hat{\psi}_b(\mathcal{X}_i) - \hat{\psi}_b(\chi) - E[\hat{\psi}_b(\mathcal{X}_i)] + E[\hat{\psi}_b(\chi)]\} K(h^{-1}d(\mathcal{X}_i, \chi))$$

$$+ \frac{(nF_\chi(h))^{-1}}{\hat{f}_h(\chi)} \sum_{i=1}^{n-1} \{E[\hat{\psi}_b(\mathcal{X}_i)] - E[\hat{\psi}_b(\chi)] - \psi(\mathcal{X}_i) + \psi(\chi)\} K(h^{-1}d(\mathcal{X}_i, \chi))$$

$$+ \frac{(nF_\chi(h))^{-1}}{\hat{f}_h(\chi)} \sum_{i=1}^{n-1} \{\psi(\mathcal{X}_i) - \psi(\chi)\} K(h^{-1}d(\mathcal{X}_i, \chi))$$
$$= U_1 + U_2 + U_3$$

Using our Lemma 8.2 instead of Theorem 1 in Ferraty et al. (2007) [10], we can follow the lines of the proof of Lemma 5 in Ferraty et al. (2010) [8] to obtain $U_1 = o((nF_\chi(h))^{-1/2})$ a.s. We can also obtain $U_2 = o((nF_\chi(h))^{-1/2})$ a.s. by following the lines of the proof of (8.11) in Raña et al. (2016) [31], which is the extension of the Lemma 6 in Ferraty et al. (2010) [8] to the dependent data case. Finally, it is easy to see from Theorem 4.3 that $U_3 = E[\hat{\psi}_h(\chi)] - \psi(\chi) + o((nF_\chi(h))^{-1/2})$, which completes the proof of the lemma. $\square$

Acknowledgements

References

[1] ANTONIADIS, A., PAPAROTIDIS, E. AND SAPATINAS, T. (2006). A functional wavelet-kernel approach for time series prediction. *Journal of the Royal Statistical Society, Series B.* **68** 837–857. MR2301297

[2] ANTONIADIS, A. AND SAPATINAS, T. (2003). Wavelet methods for continuous time prediction using Hilbert-valued autoregressive processes. *Journal of Multivariate Analysis.* **87** 133–158. MR2007265

[3] AUE, A., NORINHO, D. D. AND HÖRMANN, S. (2015). On the prediction of stationary functional time series. *Journal of the American Statistical Association.* **110** 378–392. MR3338510

[4] BOSQ, D. (2000). *Linear processes in function space.* Springer, New York. MR1783138

[5] BOSQ, D. (2007). General linear processes in Hilbert spaces and prediction. *Journal of Statistical Planning and Inference.* **137** 879–894. MR2301723

[6] DELSOL, L. (2009) Advances on asymptotic normality in non-parametric functional time series analysis. *Statistics: A Journal of Theoretical and Applied Statistics.* **43(1)** 13–33. MR2499359

[7] DIDERICKSEN, D., KOKOSZKA, P. AND ZHANG, X. (2012) Empirical properties of forecasts with the functional autoregressive model. *Comput. Stat.* **27(2)** 285–298. MR2923229

[8] FERRATY, F., KEILEGOM, I. V. AND VIEU, P. (2010). On the validity of the bootstrap in non-parametric functional regression. *Scandinavian Journal of Statistics.* **37** 286–306. MR2682301

[9] FERRATY, F., KEILEGOM, I. V. AND VIEU, P. (2012). Regression when both response and predictor are functions. *Journal of Multivariate Analysis.* **109** 10–28. MR2922850

[10] FERRATY, F., MAS, A. AND VIEU, P. (2007). Nonparametric regression on functional data: inference and practical aspects. *Aust. N. Z. J. Stat.* **49** 267–286. MR2396496

[11] FERRATY, F. AND VIEU, P. (2006). *Nonparametric functional data analysis, Theory and Practice.* Springer, New York. MR2229687

[12] FRANKE, J., KREISS, J.-P. AND MAMMEN, E. (2002). Bootstrap of kernel smoothing in nonlinear time series. *Bernoulli.* **8(1)** 1–37. MR1884156

[13] FRANKE, J. AND NYARIGE, E. (2016). On the residual-based bootstrap for functional autoregression. *working paper, Univ. of Kaiserslautern, Germany.*

[14] GABRYS, R., HORVÁTH, L. AND KOKOSZKA, P. (2010). Tests for error correlation in the functional linear model. *Journal of American Statistical Association.* **105** 1113–1125. MR2752607

[15] HÖRMANN, S., KIDZIŃSKI, L. AND HALLIN, M. (2015). Dynamic functional principal components. *Journal of the Royal Statistical Society, Series B.* **77(2)** 319–348. MR3310529

[16] HÖRMANN, S. AND KOKOSZKA, P. (2010). Weakly dependent functional data. *The Annals of Statistics.* **38(3)** 1845–1884. MR2662361

[17] HORVÁTH, L. AND KOKOSZKA, P. (2011). *Inference for functional data with applications.* Springer Series in Statistics, Springer, New York. MR2920735

[18] KARGIN, V. AND ONATSKI, A. (2008). Curve forecasting by functional autoregression. *Journal of Multivariate Analysis.* **99** 2508–2526. MR2463404

[19] KLEPSCH, J. AND KLÜPPELBERG, C. (2017). An innovations algorithm for the prediction of functional linear processes. *Journal of Multivariate Analysis.* **155** 252–271. MR3607894

[20] KLEPSCH, J., KLÜPPELBERG, C. AND WEI, T. (2017). Prediction of functional ARMA processes with an application to traffic data. Econometrics and Statistics. **1** 128–149.

[21] KREISS, J.-P. AND LAHIRI, S. N. (2012). Bootstrap methods for time series. *Handbook of Statistics: Time Series Analysis: Methods and Applications.* **30(1)**.

[22] MASRY, E. (1996). Multivariate regression estimation: Local polynomial fitting for time series. *Stochastic Process. Appl.* **65** 81–101. MR1422881

[23] MASRY, E. (2005). Nonparametric regression estimation for dependent functional data: asymptotic normality. *Stochastic Process. Appl.* **115(1)** 155–177. MR2105373

[24] NEUMANN, M. H. AND KREISS, J.-P. (1998). Regression-type inference in nonparametric autoregression. *The Annals of Statistics.* **26**, 1570–1613. MR1647701

[25] PAN, L. AND POLITIS, D. N. (2016) Bootstrap prediction intervals for linear, nonlinear and nonparametric autoregression (with discussion). *Journal of Statistical Planning and Inference.* **177** 1–27. MR3511318

[26] POLITIS, D. N. (2013). Model-free model fitting and predictive distribution (with discussion). *Test.* **22(2)** 183–250. MR3062250

[27] Politis, D. N. (2015). *Model-free prediction and regression: A transformation-based approach to inference.* Springer International Publishing. MR3442999

[28] Politis, D. N. and Romano, J. (1994). Limit theorems for weakly dependent Hilbert space valued random variables with application to the stationary bootstrap. *Statistica Sinica.* **4** 461–476. MR1309424

[29] Ramsay, J. and Silverman, B. W. (1997). *Functional Data Analysis.* Springer-Verlag, New York. MR2168993

[30] Ramsay, J. and Silverman, B. W. (2002). *Applied Functional Data Analysis: Methods and Case Studies.* Springer-Verlag, New York. MR1910407

[31] Raña, P., Aneiros, G., Vilar, J. and Vieu, P. (2016). Bootstrap confidence intervals in functional nonparametric regression under dependence. *Electronic Journal of Statistics.* **10** 1973–1999. MR3522666

[32] Robinson, P. M. (1983). Nonparametric estimators for time series. *J. Time Ser. Anal.* **4** 185–207. MR0732897

[33] Rosenblatt, M. (1956). A central limit theorem and a strong mixing condition. *Proc. Nat. Acad. Sci U.S.A.* **42** 43–47. MR0074711