Kernel Estimation of Nonparametric Functional Autoregression

Tingyi Zhu UC San Diego email: t8zhu@ucsd.edu Advisor: Dimitris Politis

Functional Time Series

- Funtional data objects collected sequentially over time are characterized as functional time series.
- Each term in series is a function $\mathcal{X}_i(t)$ defined for t taking values in some interval $[a, b]$.
- Long continuous records of temporal sequence segmented into curves over consecutive time intervals.
- Examples: daily price curves of financial transactions, daily patterns of environmental data.

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). $\mathbb{H}$ is endowed with inner product $\langle\cdot,\cdot\rangle$ and corresponding norm $||\cdot||$, and with orthonormal basis $\{e_j : j = 1, \cdots, \infty\}$. We consider the following first-order nonparametric functional autoregression model, namely FAR(1):

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

– Ψ is the autoregressive operator mapping functions from $\mathbb{H}$ to $\mathbb{H}$.
– Innovations $\mathcal{E}_i$'s are i.i.d. $\mathbb{H}$-valued r.v.'s with $E(\mathcal{E}_{i+1}|\mathcal{X}_i) = 0$ and $E(||\mathcal{E}_{i+1}||^2|\mathcal{X}_i) = \sigma^2_{\mathcal{E}}(\mathcal{X}_i) < \infty$.
– Assume the model is homoscedastic: $\sigma_{\mathcal{E}}(\mathcal{X}_i) \equiv \sigma_{\mathcal{E}}$.
– The model is nonparametric in the sense that the operator Ψ is not constrained to be linear.

Estimation of Ψ

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

$$\hat{\Psi}_h(\chi) = \frac{\sum_{i=1}^{n-1} \mathcal{X}_{i+1} K(h^{-1} d(\mathcal{X}_i, \chi))}{\sum_{i=1}^{n-1} K(h^{-1} d(\mathcal{X}_i, \chi))}, \tag{2}$$

where χ is a fixed element in $\mathbb{H}$, K is a kernel function, $d(\cdot,\cdot)$ is a semi-metric defined to measure the proximity between the two elements in $\mathbb{H}$, and h is a bandwidth sequence, tending to zero as n tends to infinity.

Some assumptions are made on the kernel:

- $K(\cdot)$ is supported on $[0, 1]$, has a continuous derivative on $[0, 1)$;
- $K'(s) \le 0$ and $K(1) > 0$.

References

[1] T. Zhu, D. N. Politis, Kernel Estimation of First-order Nonparametric Functional Autoregression and its Bootstrap Approximation, working paper, 2016.

Consistency of estimator $\hat{\Psi}_h$

Theorem 0.2. *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) $\sum_j \alpha(j)^{\frac{\delta}{2+\delta}} < \infty$,

where $\alpha(\cdot)$ is the mixing coefficient of the sequence $\{\mathcal{X}_t, t \in \mathbb{N}\}$. Also assume regularity conditions (C1)-(C4) given in [1]. Then

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

where

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

Remark. The assumptions (i)-(iii) show a trade-off 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.

Simulations

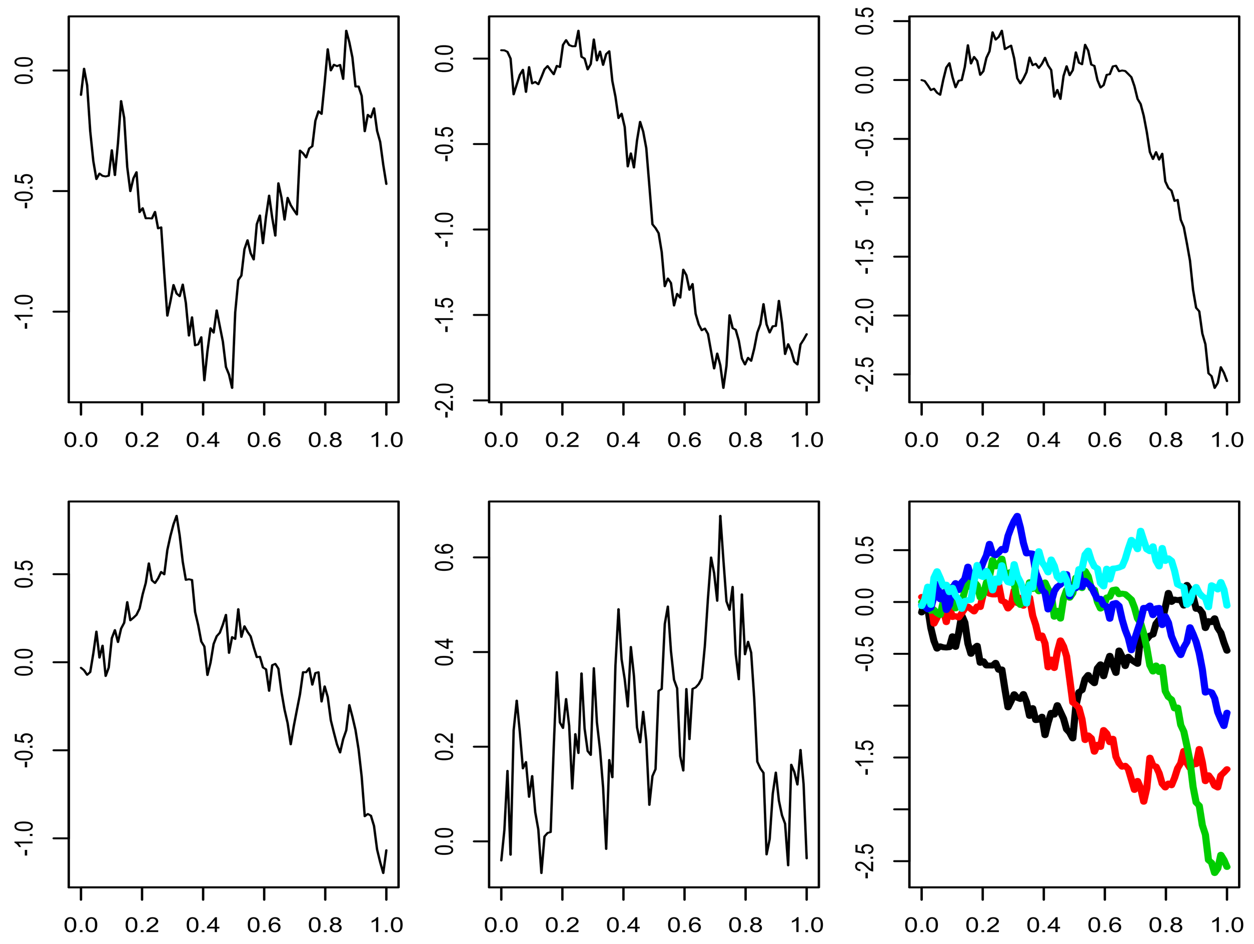

Figure 1: 5 Curves $\mathcal{X}_{101}, \mathcal{X}_{102}, \ldots, \mathcal{X}_{105}$ from the sample.

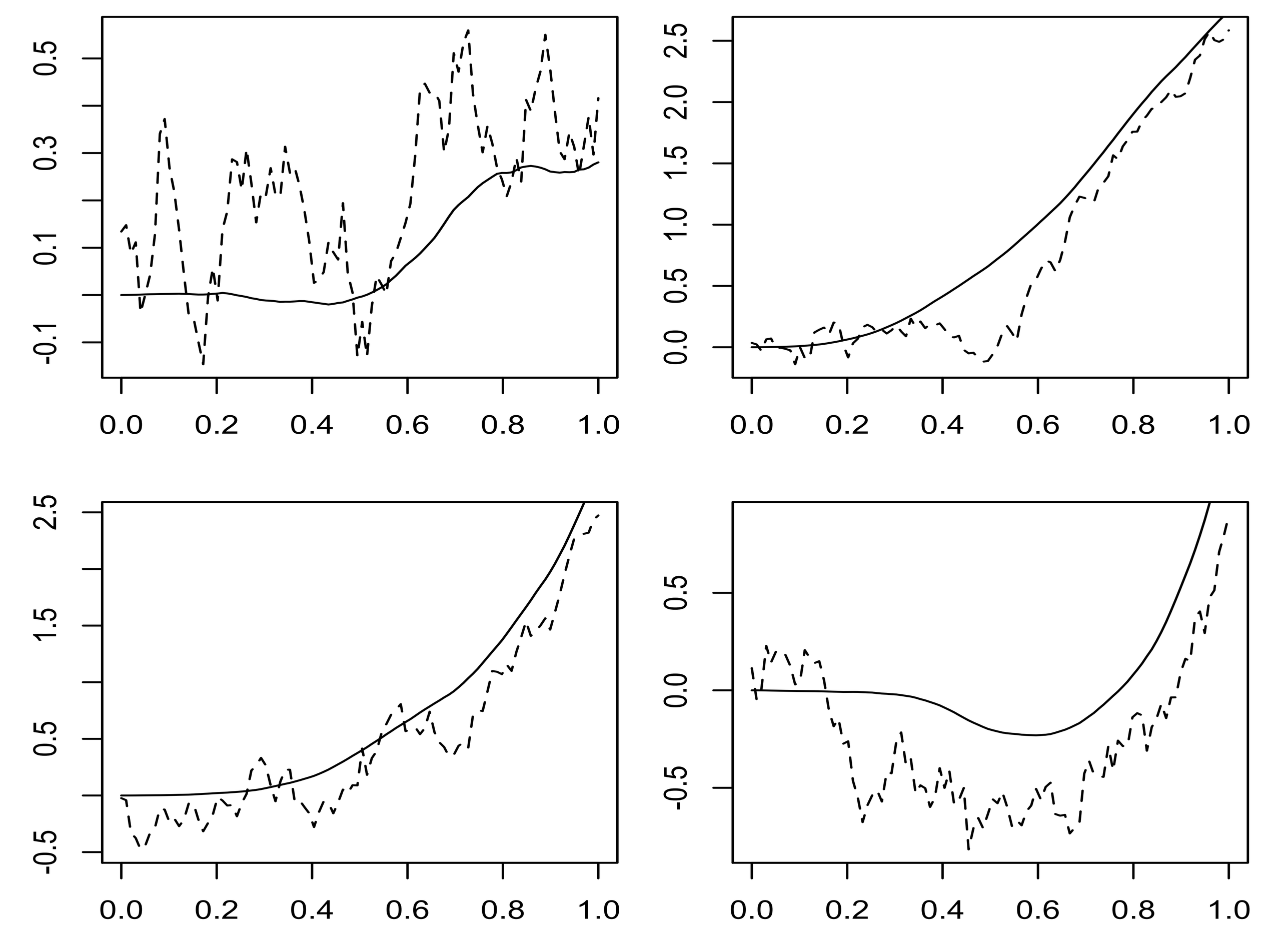

Figure 2: Estimator $\hat{\Psi}_h(\chi)$ (dashed); true operator $\Psi(\chi)$ (solid)

Data Generating Process. For $t \in [0, 1]$, we choose the initial curve $\mathcal{X}_1 = cos(t)$, and generate the FAR(1) series for $n = 1, \ldots, 249$ according to

$$\mathcal{X}_{n+1}(t) = 3\int_0^t s\mathcal{X}_n(s)ds + \mathcal{E}_{n+1}(t), \tag{3}$$

with the error process

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

where $W(\cdot)$ is the standard Weiner process.

Computing Kernel Estimator. With the 250 simulated curves, we use a learning sample (the first 200 curves) to compute the kernel estimator $\hat{\Psi}_h$, and evaluate it among the test sample (the last 50 curves). Parameter h is chosen by a standard cross validation procedure, with a fixed semi-metric $d(\chi_1, \chi_2)$–see reference for details.

Figure 2 compares the kernel estimation (i.e.$\hat{\Psi}_h(\chi)$) with the true operator (i.e.$\Psi(\chi)$) at $\chi = \mathcal{X}_{201}, \ldots, \mathcal{X}_{204}$.

Componentwise Bootstrap Approximation

We propose a bootstrap procedure to approximates the distribution of $\hat{\Psi}_h(\chi) - \Psi(\chi)$, which consists of the following steps:

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
$$\hat{\Psi}^*_{hb}(\chi) = \frac{\sum_{i=1}^{n-1} \mathcal{X}^*_{i+1} K(h^{-1} d(\mathcal{X}_i, \chi))}{\sum_{i=1}^{n-1} K(h^{-1} d(\mathcal{X}_i, \chi))}. \tag{4}$$

This procedure can be applied to construct the bootstrap prediction region. Its validity is shown in the following theorem.

Theorem 0.1. *For any $k = 1, 2, \ldots$, and any bandwidth h and b, let $\hat{\Psi}_{k,h}(\chi) = \langle\hat{\Psi}_h(\chi), e_k\rangle$ and $\hat{\Psi}^*_{k,hb}(\chi) = \langle\hat{\Psi}^*_{hb}(\chi), e_k\rangle$. Under some regularity conditions, we have*

$$\sup_{y\in\mathbb{R}} \left| P^*\left(\sqrt{nF_\chi(h)}\{\hat{\Psi}^*_{k,hb}(\chi) - \hat{\Psi}_{k,b}(\chi)\} \le y\right) - P\left(\sqrt{nF_\chi(h)}\{\hat{\Psi}_{k,h}(\chi) - \Psi_k(\chi)\} \le y\right)\right| \xrightarrow{a.s.} 0, \tag{5}$$

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