Kernel Estimation of Nonparametric Functional Autoregression and its Bootstrap Approximation

Tingyi Zhu

University of California, San Diego

t8zhu@ucsd.edu

Advisor: Dimitris Politis

June 5, 2015
Outline

- Introduction
- The Model
- Assumptions and Notations
- Consistency of the Kernel Estimator
- Componentwise Bootstrap Approximation
- Simulations
Introduction
Motivation: Functional Time Series

- Functional Data Analysis (FDA) has recently grown into an important field of statistical research.

- Functional Data are usually collected in sequential form, exhibiting forms of dependence.

- Curves collected can be characterized as a functional time series ($\mathcal{X}_k: k \in \mathbb{Z}$).

- Each term in the sequence is a function $\mathcal{X}_k(t)$ defined for $t$ taking values in some interval $[a, b]$.

- The most often applied functional time series model (FAR1): $\mathcal{X}_{k+1} = \Psi(\mathcal{X}_k) + \mathcal{E}_{k+1}, k \in \mathbb{Z}$.
Figure: Functional time series
Previous related work

Linear functional autoregression

- Operator $\Psi$ is assumed to be linear
- Bosq (2000):
  - the first to bring up FAR(1) model
  - basic properties and limit theorems
  - estimation of $\Psi$ for linear FAR(1)
- Hormann and Kokoszka (2010)
  - considering the structure of the dependence
- Alexander Aue (2012)
  - methodology of predicting linear FAR(1) process using FPCA
### Nonparametric functional regression

- **Model:** \( Y = \Psi(X) + \varepsilon \)
- **Nonparametric feature:** \( \Psi \) not restricted to be linear

**Ferraty, Mas, Vieu (2007):**
- \( Y \) is scalar, \( X \) is functional
- Kernel estimation and bootstrap approximation

**Ferraty, Van Keilegom, Vieu (2012):**
- Double functional setting, i.e. both \( X \) and \( Y \) are functional

**Delsol (2009):**
- \( Y \) is scalar, \( X \) is functional
- Sequence \( (Y, X)_i \); dependent, strong mixing
### Nonparametric univariate autoregression

- **Model:**  
  \[ X_{i+1} = m(X_i) + \varepsilon_{i+1} \]

- **Variables:**  
  \( X_i \)'s and \( \varepsilon_i \)'s: are scalar; \( m \) is unknown function

- **References:**
  - Robinson (1983) and Masry (1994):
    - Asymptotic study of kernel estimation \( \hat{m} \)
  - Franke, Kreiss and Mammen (2002):
    - Bootstrap method in nonlinear autoregression
    - Bootstrap schemes: autoregression bootstrap, regression bootstrap
Introduction

We focus on the nonlinear functional autoregression model of order one (FAR1):

$$X_{i+1} = \Psi(X_i) + \varepsilon_{i+1}, \quad i \in \mathbb{Z},$$

Main interest:

- Asymptotic study of kernel estimator $\hat{\Psi}$;
- Bootstrap methodology for estimating the distribution of the kernel estimation.
The Model
The functional space

- Let $\mathbb{H}$ be a functional space

- Two topology structures of $\mathbb{H}$:
  - $\mathbb{H}$ is endowed with inner product $\langle \cdot, \cdot \rangle$ and corresponding norm $\| \cdot \|$, and with orthonormal basis $\{ e_j : j = 1, \cdots, \infty \}$.
  - $\mathbb{H}$ is endowed with a semi-metric $d(\cdot, \cdot)$, defining a topology to measure the proximity between two elements in $\mathbb{H}$.

- $(X_i : i = 1, \ldots, n)$ is a stationary and strong mixing functional sequence in $\mathbb{H}$. 
The model

FAR(1)

Consider the first-order nonparametric functional autoregressive model:

\[ X_{i+1} = \Psi(X_i) + \mathcal{E}_{i+1} \quad i = 1, \ldots, n - 1, \]

where \( \Psi \) is the autoregressive operator mapping \( \mathbb{H} \) to \( \mathbb{H} \), and the innovations \( \mathcal{E}_i \)'s are i.i.d. \( \mathbb{H} \)-valued random variables with zero means.
Estimation of $\Psi$

Kernel estimator

Estimation of $\Psi$ is given by the functional version of Nadaraya-Watson estimator of time series:

$$
\hat{\Psi}_h(\chi) = \frac{\sum_{i=1}^{n-1} \chi_{i+1} K(h^{-1}d(\chi_i, \chi))}{\sum_{i=1}^{n-1} K(h^{-1}d(\chi_i, \chi))},
$$

(2)

where $\chi$ is a fixed element in $\mathbb{H}$, $K(\cdot)$ is a kernel function and $h$ is a bandwidth sequence, tending to zero as $n$ tends to infinity.
An auxiliary model

- Consider the orthonormal basis of $\mathbb{H}$, $\{e_j : j = 1, \cdots, \infty\}$, for $j \in \mathbb{Z}^+$.
- Apply $\langle \cdot, e_j \rangle$ on both sides of the equation (1) yields

$$
\langle X_{i+1}, e_j \rangle = \langle \Psi(X_i) + \mathcal{E}_{i+1}, e_j \rangle \\
= \langle \Psi(X_i), e_j \rangle + \langle \mathcal{E}_{i+1}, e_j \rangle \quad i = 1, \ldots, n - 1, \quad (3)
$$

- Let $X_{n,j} = \langle X_n, e_j \rangle$, $\varepsilon_{n,j} = \langle \mathcal{E}_n, e_j \rangle$. Also define another operator $\psi_j$ mapping $\mathbb{H}$ to $\mathbb{R}$ such that

$$
\psi_j(\cdot) = \langle \Psi(\cdot), e_j \rangle. \quad (4)
$$

- Then (3) can be written as

$$
X_{i+1,j} = \psi_j(X_{i,j}) + \varepsilon_{i+1,j} \quad i = 1, \ldots, n - 1. \quad (5)
$$
An auxiliary model, cont.

- We consider the model (5) for a fixed basis $e_j$. So for simplicity, we can drop the index $j$ in (5), such that
  - $X_i$ denotes $X_{i,j}$
  - $\varepsilon_i$ denotes $\varepsilon_{i,j}$
  - $\psi$ denotes $\psi_j$

- Rewrite (5) to form a functional autoregressive model with scalar response

$$X_{i+1} = \psi(X_i) + \varepsilon_{i+1} \quad i = 1, \ldots, n - 1. \quad (6)$$

where $\varepsilon_i$'s are i.i.d. scalar innovations and $\psi$ is an operator mapping $\mathbb{H}$ to $\mathbb{R}$ not constrained to be linear.
An auxiliary model, cont.

- Accordingly, the kernel estimator of model (6) is given by

\[
\hat{\psi}_h(\chi) = \frac{\sum_{i=1}^{n-1} X_{i+1}K(h^{-1}d(X_i, \chi))}{\sum_{i=1}^{n-1} K(h^{-1}d(X_i, \chi))}
\]

(7)

- Connection between \(\hat{\Psi}_h\) and \(\hat{\psi}_h\):

\[
\hat{\psi}_h(\chi) = \langle \hat{\Psi}_h(\chi), e_j \rangle.
\]

(8)
Assumptions and notations
Some notations

In the sequel, $\chi$ is a fixed element in the functional space $\mathbb{H}$, we need the following notations

- We denote

  \[ F_\chi(t) = P(d(\mathcal{X}, \chi) \leq t), \]

  which is CDF of the random variable $d(\mathcal{X}, \chi)$ usually called the small ball probability function in the literature.

- Define for $j \geq 1$:

  \[ \varphi_{\chi,j}(s) = E[\psi(\mathcal{X}) - \psi(\chi)|d(\mathcal{X}, \chi) = s] \]
  
  \[ = E[\langle \Psi(\mathcal{X}) - \Psi(\chi), e_j \rangle |d(\mathcal{X}, \chi) = s], \]
Some notations, cont.

- Also let

\[
\tau_h(s) = \frac{F_\chi(hs)}{F_\chi(h)} = P(d(\mathcal{X}, \chi) \leq hs | d(\mathcal{X}, \chi) \leq h)
\]

and

\[
\tau_0(s) = \lim_{h \downarrow 0} \tau_h(s).
\]

- Technical aspects of the functions \( \varphi_\chi, F_\chi \) and \( \tau_h \) have been discussed in Ferraty, Mas & Vieu (2005).
Some notations, cont.

The semi-metric $d$ will act on the asymptotic behavior of the estimator through $\varphi_\chi$, $F_\chi$ and $\tau_h$, and the following quantities:

\[ M_0 = K(1) - \int_0^1 (sK(s))' \tau_0(s) ds, \]
\[ M_1 = K(1) - \int_0^1 K'(s) \tau_0(s) ds, \]
\[ M_2 = K^2(1) - \int_0^1 (K^2)'(s) \tau_0(s) ds. \]
Assumptions

We consider the following assumptions:

(A1) $\psi$ and $\sigma^2$ are continuous in a neighborhood of $\chi$, and $F_\chi(0) = 0$.
(A2) $\varphi'(0)$ exists.
(A3) $h \to 0$ and $nF_\chi(h) \to \infty$, as $n \to 0$.
(A4) The kernel function $K$ is supported on $[0, 1]$ and has a continuous derivative with $K'(s) \leq 0$ and $K(1) > 0$.
(A5) For $s \in [0, 1]$, $\tau_h(s) \to \tau_0(s)$ as $h \to 0$.

(A6) $\exists \delta > 2, E(|\epsilon|^{2+\delta}|\mathcal{X}) < \infty$.
(A7) $\max(E(|X_{i+1}X_{j+1}||\mathcal{X}_i, \mathcal{X}_j), E(|X_{i+1}||\mathcal{X}_i, \mathcal{X}_j)) < \infty$.
(A8) Assumption (H1) in Delsol (2009).
(A9) Assumption (H2) in Delsol (2009).
Consistency of the Kernel Estimator
Consistency of $\hat{\psi}_h$

First, we have the following asymptotic results:

**Theorem 1**
Assume (A1)-(A6), then

$$E[\hat{\psi}_h(\chi)] - \psi(\chi) = \varphi'(0) \frac{M_0}{M_1} h + O\left(\frac{1}{nF_\chi(h)}\right) + o(h).$$  \hspace{1cm} (9)

**Theorem 2**
Assume (A1)-(A8), then

$$\text{Var}(\hat{\psi}_h(\chi)) = \frac{\sigma_\epsilon^2}{M_1^2} \frac{M_2}{nF_\chi(h)} + o\left(\frac{1}{nF_\chi(h)}\right).$$  \hspace{1cm} (10)

**Corollary 1**
Assume (A1)-(A8), then

$$\hat{\psi}_h(\chi) \xrightarrow{p} \psi(\chi).$$  \hspace{1cm} (11)
Asymptotic normality of $\hat{\psi}_h$

**Theorem 3**

Assume (A1)-(A9), then

$$
\sqrt{n\hat{F}_\chi(h)} \left( \hat{\psi}_h(\chi) - \psi(\chi) - B_n \right) \frac{M_1}{\sqrt{\sigma_\epsilon^2 M_2}} \overset{d}{\to} N(0, 1), \quad (12)
$$

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

$$
\hat{F}_\chi(h) = \frac{\#(i : d(\chi_i, \chi) \leq h)}{n}
$$
Asymptotic normality of $\hat{\psi}_h$

The bias term in (12) can be cancelled with the following additional assumption:

$$(A10) \lim_{n \to \infty} h \sqrt{n F_X(h)} = 0.$$ 

**Corollary 2**

Assume (A1)-(A10), then

$$\sqrt{n F_X(h)} \left( \hat{\psi}_h(\chi) - \psi(\chi) \right) \xrightarrow{d} \frac{M_1}{\sqrt{\sigma^2 \epsilon M_2}} \overset{d}{\to} N(0, 1).$$ (13)
Consistency of $\hat{\Psi}_h$

By the structures of the $\hat{\Psi}_h$ and $\hat{\psi}_h$ in (2) (7), we have

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

Noting that $\psi_j(\cdot) = \langle \Psi(\cdot), e_j \rangle$, corollary 1 indicates

$$\langle \hat{\Psi}_h(\chi) - \Psi(\chi), e_j \rangle \overset{p}{\to} 0. \quad j = 1, \ldots, \infty \quad (14)$$

(14) does not guarantee the consistency of estimator $\hat{\Psi}_h$ in an infinite-dimensional space.
Consistency of $\hat{\Psi}_h$

To provide a limit theorem for $\hat{\Psi}_h(\chi)$, we need to make additional assumptions on the mixing coefficient and the moment condition as follows:

Assume $\exists \delta' > \delta > 0$ such that

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

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

(iii) $\sum_j \alpha(j) \frac{\delta}{2+\delta} < \infty$.

Also assume,

(C1) For each $k \geq 1$, $\psi_k$ is continuous in a neighborhood of $\chi$, and $F_\chi(0) = 0$. 

(C2) For some $\beta > 0$, all $0 \leq s \leq \beta$ and all $k \geq 1$, $\varphi_{\chi,k}(0) = 0$, $\varphi'_{\chi,k}(s)$ exists, and $\varphi'_{\chi,k}(s)$ is uniformly Lipschitz continuous of order $0 < \alpha \leq 1$, i.e. there exists a $0 < L_k < \infty$ such that $|\varphi'_{\chi,k}(s) - \varphi'_{\chi,k}(0)| \leq L_k s^\alpha$ uniformly for all $0 \leq s \leq \beta$. 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 $(nF_{\chi}(h))^{1/2} h^{1+\alpha} = o(1)$.

(C4) The kernel function $K$ is supported on $[0, 1]$ and has a continuous derivative on $[0, 1)$, with $K'(s) \leq 0$ for $0 \leq s < 1$ and $K(1) > 0$. 
Consistency of $\hat{\psi}_h$

**Theorem 4**

Assume (i), (ii), (iii) and (C1)-(C4), we have

$$\hat{\psi}_h(\chi) = \psi(\chi) - B_n + O_p\left(\frac{1}{\sqrt{nF_\chi(h)^{1+\delta}}}\right)$$  \hspace{1cm} (15)

where

$$B_n = h\frac{M_0}{M_1} \sum_{k=1}^{\infty} \varphi'_k(0)e_k.$$
Proof of Theorem 4

Consider the expression

\[ \sqrt{nF_\chi(h)^{1+\delta}} \left[ \hat{\psi}_h(\chi) - \psi(\chi) - B_n \right]. \]  

(16)

Following the similar arguments as in the proof of theorem 4.1 in Ferraty (2012), (16) has the same asymptotic distribution as

\[ \frac{1}{\sqrt{n}} \sum_{i=1}^{n} (Z_{n,i} - EZ_{n,i}), \]

where for 1 ≤ i ≤ n,

\[ Z_{n,i} = \frac{1}{M_1 \sqrt{F_\chi(h)^{1-\delta}}} \left[ \chi_{i+1} K \left( \frac{d(\chi_i, \chi)}{h} \right) - \psi(\chi) K \left( \frac{d(\chi_i, \chi)}{h} \right) \right]. \]
Theorem (Politis and Romano, 1992)

Assume $X_1, X_2, \ldots$ is a stationary sequence of $H$-valued random variables with mean $m$ and mixing sequence $\alpha_X(\cdot)$. If $E(||X_1||^{2+\delta}) < \infty$ for some $\delta > 0$ and $\sum_j [\alpha_X(j)]^{\delta/(2+\delta)} < \infty$, then $Z_n = n^{-1/2} \sum_{i=1}^n (X_i - m)$ converge weakly to a Gaussian measure with mean 0 and covariance operator $S$.

By assumption (i), we can apply Holder’s inequality to obtain

$$E||Z_{n,i}||^{2+\delta} = \frac{1}{M_1^{2+\delta} F_{\chi}(h) \frac{(1-\delta)(2+\delta)}{2}} E \left(||X_{i+1} - \psi(\chi)||^{2+\delta} \left\{ K \left( \frac{d(X_i, \chi)}{h} \right) \right\}^{2+\delta} \right)$$

$$\leq \frac{1}{M_1^{2+\delta} F_{\chi}(h) \frac{(1-\delta)(2+\delta)}{2}} \left( E||X_{i+1} - \psi(\chi)||^{2+\delta'} \right)^{\frac{2+\delta}{2+\delta'}} \left\{ E K \left( \frac{d(X_i, \chi)}{h} \right)^{1-\delta} \right\}^{\frac{(1-\delta)(2+\delta)}{2}}.$$
In the above expression, \( \left( E\|\mathcal{X}_{i+1} - \Psi(\chi)\|^2 + \delta' \right)^{2+\delta} \) 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 get

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

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

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

\[
= F_{\chi}(h) M^{\frac{2}{1-\delta}}.
\]
Hence we have, $E\|Z_{n,i}\|^{2+\delta} \leq C < \infty$ for all $n$. Along with assumption (iii), we can conclude that $\frac{1}{\sqrt{n}} \sum_{i=1}^{n}(Z_{n,i} - EZ_{n,i})$ converges weakly to a Gaussian measure with mean 0 in $H$.

Therefore,

$$\hat{\Psi}_h(\chi) = \Psi(\chi) - B_n + O_p\left(\frac{1}{\sqrt{nF}\chi(h)^{1+\delta}}\right). \quad (17)$$
Componentwise Bootstrap Approximation
Bootstrap procedure for $\hat{\psi}_h$

A bootstrap procedure for $\hat{\psi}_h$ is proposed as follows:

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 $(\hat{\varepsilon}_{1,b} - \bar{\varepsilon}_b, \ldots, \hat{\varepsilon}_{n,b} - \bar{\varepsilon}_b)$ where $\bar{\varepsilon}_b = n^{-1} \sum_{i=1}^n \hat{\varepsilon}_{i,b}$.

3. For $i = 1, \ldots, n - 1$, let $X_{i+1}^* = \hat{\psi}_b(\mathcal{X}_i) + \varepsilon^*_{i+1}$.

4. Define

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

(18)
Validity of bootstrap

**Theorem 5**

If conditions of Theorem 3 holds, and assume (C1)-(C7) in Ferraty (2010), we have

\[
\sup_{y \in \mathbb{R}} P^* \left( \sqrt{nF_x(h)} \{\hat{\psi}_{hb}(\chi) - \hat{\psi}_b(\chi)\} \leq y \right)
\]

\[
- P \left( \sqrt{nF_x(h)} \{\hat{\psi}_h(\chi) - \psi(\chi)\} \leq y \right) \xrightarrow{a.s.} 0,
\]

where \( P^* \) denotes probability conditioned on the sample \( \{X_1, \ldots, X_n\} \).
Proof of Theorem 5

The expression between absolute values can be written as

$$P^*(\sqrt{nF_X(h)\{\hat{\psi}_{hb}^*(\chi) - \hat{\psi}_b(\chi)\}} \leq y) - \Phi \left( \frac{y - \sqrt{nF_X(h)\{E^*\hat{\psi}_{hb}^*(\chi) - \hat{\psi}_b(\chi)\}}}{\sqrt{nF_X(h)\operatorname{var}^*(\hat{\psi}_{hb}^*(\chi))}} \right)$$

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

$$+ \Phi \left( \frac{y - \sqrt{nF_X(h)\{E\hat{\psi}_h(\chi) - \psi(\chi)\}}}{\sqrt{nF_X(h)\operatorname{var}(\hat{\psi}_h(\chi))}} \right) - P(\sqrt{nF_X(h)\{\hat{\psi}_h(\chi) - \psi(\chi)\}} \leq y)$$

$$= T_1(y) + T_2(y) + T_3(y)$$

By the asymptotic normality of $\hat{\psi}_h$ given in Theorem 3, $T_3(y) \to 0$ a.s. The a.s. convergence to 0 of $T_1(y)$ is given by the asymptotic normality of $\hat{\psi}_{hb}^*$ 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 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,

\[
\frac{\hat{\psi}_{hb}(\chi) - E^*(\hat{\psi}_{hb}(\chi))}{\sqrt{\text{var}^*(\hat{\psi}_{hb}(\chi))}} = \frac{\hat{g}^*(\chi) - E^*(\hat{g}^*(\chi))}{\sqrt{\text{var}^*(\hat{g}^*(\chi))}} = \frac{\hat{g}^*(\chi) - E^*(\hat{g}^*(\chi))}{\sqrt{\text{var}^*(\hat{g}^*(\chi))}}
\]

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

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 (see Delsol 2009).

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 an independent sum and asymptotic normality follows directly.
It remains to consider $T_2(y)$. Its a.s convergence to 0 follows from the following lemma:

**Lemma**

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

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

\[
\text{var}^*[\hat{\psi}_{hb}^*(\chi)] = \text{var}^*
\left[
\frac{\sum_{i=1}^{n-1} (\hat{\psi}_b(\chi_i) + \varepsilon_{i+1}^*) K(h^{-1}d(\chi_i, \chi))}{\sum_{i=1}^{n-1} K(h^{-1}d(\chi_i, \chi))}
\right]
\]

\[
= \text{var}^*
\left[
\frac{\sum_{i=1}^{n-1} \varepsilon_{i+1}^* K(h^{-1}d(\chi_i, \chi))}{\sum_{i=1}^{n-1} K(h^{-1}d(\chi_i, \chi))}
\right]
\]
\[
\begin{align*}
\sum_{i=1}^{n-1} K^2(h^{-1}d(X_i, \chi))\text{var}^* (\varepsilon^*_{i+1}) &= \frac{\left(\sum_{i=1}^{n-1} K(h^{-1}d(X_i, \chi))\right)^2}{\left(\sum_{i=1}^{n-1} K(h^{-1}d(X_i, \chi))\right)^2} \\
&= \frac{\hat{\sigma}_\varepsilon^2}{\hat{f}(\chi)^2} (nF_\chi(h))^{-2} \sum_{i=1}^{n-1} K^2(h^{-1}d(X_i, \chi)) \\
&= \frac{\sigma_\varepsilon^2}{E[\hat{f}(\chi)]^2} (nF_\chi^2(h))^{-1} \cdot E[K^2(h^{-1}d(X_i, \chi))] \cdot (1 + o(1)) \\
&= \frac{\sigma_\varepsilon^2}{E[\hat{f}(\chi)]^2} \frac{M_2}{M_1^2 nF_\chi(h)} (1 + o(1)) \\
&= \text{var}[\hat{\psi}_h(\chi)] + o((nF_\chi(h))^{-1}).
\end{align*}
\]

Since \(\text{var}[\hat{\psi}_h(\chi)] = O((nF_\chi(h))^{-1})\) by Theorem 2, the result follows by deviding \(\text{var}[\hat{\psi}_h(\chi)]\) on both sides. That completes the proof.
Bootstrap procedure for $\hat{\Psi}_h$

A bootstrap procedure for $\hat{\Psi}_h$ is proposed as follows:

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

$$\hat{\Psi}_{hb}^*(\chi) = \frac{n^{-1} \sum_{i=1}^{n-1} \mathcal{X}_{i+1}^* K(h^{-1}d(\mathcal{X}_i, \chi))}{n^{-1} \sum_{i=1}^{n-1} K(h^{-1}d(\mathcal{X}_i, \chi))}.$$  (19)
Validity of bootstrap

**Theorem 6**

For any \( k = 1, 2, \ldots \), and any bandwidth \( h \) and \( b \), let \( \hat{\Psi}_{k,h}(\chi) = \langle \hat{\Psi}(\chi), e_k \rangle \) and \( \hat{\Psi}_{k,hb}(\chi) = \langle \hat{\Psi}_{hb}(\chi), e_k \rangle \). If, in addition to (C1), (C2) and (C4), (i)-(v) in Ferraty (2012) hold, then for any \( k = 1, 2, \ldots \), we have

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

where \( \mathbb{P}^* \) denotes probability conditioned on the sample \( \{ \chi_1, \ldots, \chi_n \} \).

**Remark:** This theorem is a direct consequence of Theorem 5, since the problem is a one-dimension response problem for a fixed \( k \).
Simulations
Data generating process

- Simulated realization of linear FAR(1) series, Diderickson (2011)
- Functional space: $L^2[0, 1]$
- Linear operator: $\Psi(X) = \int_0^1 \psi(s, t)X(s)ds$
- Error Process: $\mathcal{E}_i$’s

$\Psi$ is acted on functions $X_i$’s, and the functional series are generated according to

$$X_{n+1}(t) = \int_0^1 \psi(t, s)X_n(s)ds + \mathcal{E}_{n+1}(t).$$  \hspace{1cm} (20)
Choice of $\psi$

- We use the kernel

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

- Then

$$\mathcal{X}_{n+1}(t) = C \int_0^t s \mathcal{X}_n(s) ds + \varepsilon_{n+1}(t).$$  \hspace{1cm} (21)

- $C$ 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).

- Choose $C = 3$, such that $||\psi|| = 0.5$. 
We use the error process introduced by Didericksen (2011):

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

- \( W(\cdot) \) is the standard Weiner process

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

- \( Z_k \)'s are independent standard normal and \( Z_0 = 0 \).
Equally partition the interval $[0, 1]$ such that $0 = t_1 < t_2 < \cdots < t_{99} < t_p = 1$ with $p = 100$.

Choose the initial curve $X_1 = \cos(t)$.

Build the series $X_1, \ldots, X_n$ with $n = 250$, for $j = 1, \ldots, 100$:

$$X_1(t_j) = \cos(t_j),$$

$$X_i(t_j) = 3 \int_0^{t_j} s X_{i-1}(s) ds + \varepsilon_i(t_j), \quad i = 2, \ldots, 250.$$
Sample curves

Figure: 5 Curves $x_{101}, x_{102}, \ldots, x_{105}$ from the sample.
Computing kernel estimator

- 250 Curves generated
- Learning sample: \( X_1, \ldots, X_{200} \)
- Testing sample: \( X_{201}, \ldots, X_{250} \)
- Use learning sample to compute kernel estimator \( \hat{\Psi}_h \)
- Compare the kernel estimation (i.e. \( \hat{\Psi}_h(\chi) \)) and the true operator (i.e. \( \Psi(\chi) \)) and \( \chi \) is taken from the testing sample.
h, b and semi-metric \( d(\cdot, \cdot) \)

- Semi-metric \( d \):

\[
d(x_1, x_2) = \sqrt{\sum_{j=1}^{J} \langle x_1 - x_2, v_{j,n} \rangle^2},
\]

where \( v_{1,n}, v_{2,n}, \ldots \) are eigenfunctions associated with the largest eigenvalues of the empirical covariance operator of the learning sample:

\[
C(\cdot) = \frac{1}{200} \sum_{i=1}^{200} \langle x_i, \cdot \rangle x_i.
\]

- \( h \) is chosen by a cross validation procedure, see Ferraty (2012).

- \( b = h \)
Comparison between $\hat{\Psi}_h$ and $\Psi$

Figure: Kernel estimations $\hat{\Psi}_h(\chi)$ (dashed lines); true operator $\Psi(\chi)$ (solid lines), for $\chi = \chi_{201}, \chi_{202}, \chi_{203}, \chi_{204}$. 
To demonstrate the bootstrap method, we compare

- the density function $f_{k,\chi}^*$ of the componentwise bootstrapped error

\[ \langle \hat{\Psi}_{hb}(\chi) - \hat{\Psi}_b(\chi), e_k \rangle \]

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

\[ \langle \hat{\Psi}_h(\chi) - \Psi(\chi), e_k \rangle. \]

- $\{e_1, e_2, \ldots \}$ is the basis derived from the sample covariance operator.
Estimation of $f^*_{k,\chi}$

- compute $\hat{\Psi}_b(\chi)$ over the learning sample $\mathcal{X}_1, \ldots, \mathcal{X}_{200}$,

- repeat 200 times the bootstrap algorithm introduced in previous section to obtain

  \[
  \hat{\Psi}_{hb}^1(\chi), \ldots, \hat{\Psi}_{hb}^{200}(\chi),
  \]

- estimate the density $f^*_{k,\chi}$ over the 200 values

  \[
  \langle \hat{\Psi}_{hb}^1(\chi) - \hat{\Psi}_b(\chi), e_k \rangle, \ldots, \langle \hat{\Psi}_{hb}^{200}(\chi) - \hat{\Psi}_b(\chi), e_k \rangle.
  \]
Estimation of $f^{true}_{k,\chi}$ (Monte-Carlo scheme)

- build 200 samples $\{\mathcal{X}^s_1, \ldots, \mathcal{X}^s_{200}\}_{s=1,\ldots,200}$,
- for the $s$th sample $\{\mathcal{X}^s_1, \ldots, \mathcal{X}^s_{200}\}$, compute $\hat{\Psi}^s_h$ to obtain
  $$\hat{\Psi}^1_h(\chi), \ldots, \hat{\Psi}^{200}_h(\chi),$$
- estimate the density $f^{true}_{k,\chi}$ over the 200 values
  $$\langle \hat{\Psi}^1_h(\chi) - \Psi(\chi), e_k \rangle, \ldots, \langle \hat{\Psi}^{200}_h(\chi) - \Psi(\chi), e_k \rangle.$$
Comparison between $f_{k,\chi}^*$ and $f_{k,\chi}^{true}$

Figure: Solid line: true error, dashed line: bootstrap error.
Thank you!