Model selection for common time series models

William KENGNE

THEMA, CY Cergy Paris Université

Joint work with J.M. Bardet and K. Kamila (Université Paris 1 Panthéon-Sorbonne)

EcoDep Conference
September 10, 2020

Model selection problem

Assume that the observed trajectory $(X_1, \ldots, X_n)$ is generated from a model m.

m belongs to a collection $\mathcal{M}$: for instance, a set of ARMA(p,q) and GARCH(p',q') processes for $0 \leq p \leq p_{\max}$, $0 \leq q \leq q_{\max}$, $0 \leq p' \leq p'_{\max}$, $0 \leq q' \leq q'_{\max}$

We would like to chose in $\mathcal{M}$, a "best" model for fitting the data $(X_1, \ldots, X_n)$.

For instance, if $p_{\max} = q_{\max} = p'_{\max} = q_{\max} = 9$, in the collection above, there is 200 possible models and we expect to recognize the true process

Outline

1 Introduction

2 The model selection criterion

3 Asymptotic results

4 Some numerical Results

Outline

1 Introduction

2 The model selection criterion

3 Asymptotic results

4 Some numerical Results

Introduction

We consider the class

Class $\mathcal{AC}(M,f)$: A process $X=(X_t)_{t\in\mathbb{Z}}$ belongs to $\mathcal{AC}(M,f)$ if it satisfies:

$$X_t = M\big((X_{t-i})_{i\in\mathbb{N}^*}\big)\,\xi_t + f\big((X_{t-i})_{i\in\mathbb{N}^*}\big) \text{ for any } t\in\mathbb{Z}; \qquad (1)$$

where :

M, f : $\mathbb{R}^\infty \to \mathbb{R}$ are measurable functions

$(\xi)_{t\in\mathbb{Z}}$: sequence of zero-mean i.i.d. random variable satisfying $\mathbb{E}[\xi_0^2]=1$

The existence of stationarity and ergodic solution in such models has been studied by Doukhan and Wintenberger (2008), see also Bardet and Wintenberger (2009).

Introduction

Semiparametric setting :

- The distribution of ξ_0 is unknown
- $(X_1, \ldots, X_n)$ is a trajectory of a process $X = (X_t)_{t\in\mathbb{Z}}$ that belongs to $\mathcal{AC}(M_{\theta^*}, f_{\theta^*})$, with $\theta^* \in \Theta \subset \mathbb{R}^d$
- M_{θ^*}, f_{θ^*} are known up to θ^*
- θ^* corresponds to the true model $m^* \in \mathcal{M}$; where $\mathcal{M}$ is a finite collection of models.

Thus, we will consider in the sequel the class $\mathcal{AC}(M_\theta, f_\theta)$ for $\theta \in \Theta$.

Introduction

The model selection setting :

We will consider several models, which all are particular cases of $\mathcal{AC}(M_\theta, f_\theta)$ with $\theta \in \Theta \subset \mathbb{R}^d$.

- a model m as a subset of $\{1, \ldots, d\}$ and denote $|m| = \#(m)$;
- $\Theta(m) = \big\{(\theta_i)_{1 \leq i \leq d} \in \mathbb{R}^d, \;\theta_i = 0 \text{ if } i \notin m\big\} \cap \Theta$;
- $\mathcal{M}$ as a finite family of models, *i.e.* $\mathcal{M} \subset \mathcal{P}\big(\{1, \ldots, d\}\big)$.

For all $m \in \mathcal{M}$, $m \in \mathcal{AC}(M_\theta, f_\theta)$ when $\theta \in \Theta(m)$.

Our main aim is to select a model $\widehat{m}$ (among the collection $\mathcal{M}$) which is "close" to m^* at least when n is large enough.

Example

$$\begin{cases} AR & \begin{cases} M^{(1)}_{\theta_1}\big((X_{t-i})_{i\in\mathbb{N}^*}\big) = \sigma \\ f^{(1)}_{\theta_1}\big((X_{t-i})_{i\in\mathbb{N}^*}\big) = \phi_1 X_{t-1} + \cdots + \phi_p X_{t-p} \end{cases} \\ \\ ARCH & \begin{cases} M^{(2)}_{\theta_2}\big((X_{t-i})_{i\in\mathbb{N}^*}\big) = \sqrt{a_0 + a_1 X^2_{t-1} + \cdots + a_q X^2_{t-q}} \\ f^{(2)}_{\theta_2}\big((X_{t-i})_{i\in\mathbb{N}^*}\big) = 0 \end{cases} \end{cases}$$

$$\Longrightarrow \begin{cases} M_\theta\big((X_{t-i})_{i\in\mathbb{N}^*}\big) = \sqrt{\theta_0 + \theta_1 X^2_{t-1} + \cdots + \theta_q X^2_{t-q}} \\ f_\theta\big((X_{t-i})_{i\in\mathbb{N}^*}\big) = \theta_{q+1} X_{t-1} + \cdots + \theta_{q+p} X_{t-p} \end{cases} .$$

Example

We could as well consider hierarchical or exhaustive families of models.

From the previous example, we can consider:

- a family $\mathcal{M}_1$ such as $\mathcal{M}_1 = \big\{\{1\}, \{1,2\}, \ldots, \{1, \ldots, q+1\}\big\}$: this family is the hierarchical one of ARCH processes with orders varying from 0 to q
- a family $\mathcal{M}_2$ such as $\mathcal{M}_2 = \mathcal{P}\big(\{1, \ldots, p+q+1\}\big)$: this family is the exhaustive one and contains as well the AR(2) process $X_t = \phi_2 X_{t-2} + \sigma\,\xi_t$ as the process $X_t = \phi_1 X_{t-1} + \phi_3 X_{t-3} + \xi_t\,\sqrt{\theta_0 + a_2 X_{t-2}^2}$.

Outline

1 Introduction

2 The model selection criterion

3 Asymptotic results

4 Some numerical Results

The QMLE

Consider a model $m \in \mathcal{M}$ and the class $\mathcal{AC}(M_\theta, f_\theta)$ for $\theta \in \Theta(m) \subset \Theta \subset \mathbb{R}^d$.

Assume that $(X_1, \cdots, X_n)$ is observed.

Consider the approximated quasi (log)likelihood given (up to a constant) for all $\theta \in \Theta(m)$ by

$$\widehat{L}_n(\theta) := -\frac{1}{2} \sum_{t=1}^{n} \widehat{q}_t(\theta) \ , \ \text{with } \widehat{q}_t(\theta) := \frac{(X_t - \widehat{f}_\theta^t)^2}{\widehat{H}_\theta^t} + \log(\widehat{H}_\theta^t) \quad (2)$$

where $\widehat{f}_\theta^t = f_\theta(X_{t-1}, X_{t-2}, \cdots, X_1, 0, \cdots)$, $\widehat{M}_\theta^t = M_\theta(X_{t-1}, X_{t-2}, \cdots, X_1, 0, \cdots)$, $\widehat{H}_\theta^t = (\widehat{M}_\theta^t)^2$.

The QMLE

The "best" parameter associated to the model m if defined by,

$$\theta^*(m) = \underset{\theta \in \Theta(m)}{\text{argmin}} \, \mathbb{E}[q_0(\theta)].$$

According to Bardet *et al.* (2020), $\theta^*(m)$ exists and it is unique under some identifiability assumptions.

When $m = m^*$, we have $\theta^*(m^*) = \theta^*$.

For any $m \in \mathcal{M}$, the QMLE of $\theta^*(m)$ is given by

$$\widehat{\theta}(m) = \underset{\theta \in \Theta(m)}{\text{argmax}} \, \widehat{L}_n(\theta). \tag{3}$$

The criterion

The selection of the "best" model $\widehat{m}$ among the collection $\mathcal{M}$ is performed by minimizing the penalized contrast

$$\widehat{C}(m) = -2\widehat{L}_n\big(\widehat{\theta}(m)\big) + |m|\kappa_n,$$

that is

$$\widehat{m} = \underset{m \in \mathcal{M}}{\text{argmin}} \ \widehat{C}(m),$$

where

- $(\kappa_n)_n$ is a sequence of a regularization parameter
- $|m|$ denotes the dimension of the model m, *i.e.* the cardinal of m, subset of $\{1, \ldots, d\}$, which is also the number of estimated components of θ (the others are fixed to zero).

The procedure is consistent when

$$\mathbb{P}(\widehat{m} = m^*) \underset{n \to \infty}{\longrightarrow} 1.$$

Outline

1 Introduction

2 The model selection criterion

3 Asymptotic results

4 Some numerical Results

Some Lipschitz-type conditions

For any compact subset $\mathcal{K} \subset \Theta$ and for $\Psi = f$ or M and $k = 0, 1, 2$, define

Assumption $(A_k(\Psi, \mathcal{K}))$: $\|\frac{\partial^k \Psi_\theta(0)}{\partial \theta^k}\|_{\mathcal{K}} < \infty$ and there exists a non-negative sequence $(\alpha_i^{(k)}(\Psi, \mathcal{K}))_i$ such that $\forall x, y \in \mathbb{R}^{\mathbb{N}}$,

$$\|\frac{\partial^k \Psi_\theta(x)}{\partial \theta^k} - \frac{\partial^k \Psi_\theta(y)}{\partial \theta^k}\|_{\mathcal{K}} \leq \sum_{i=1}^{\infty} \alpha_i^{(k)}(\Psi, \mathcal{K})|x_i - y_i|$$

with $\sum_{i=1}^{\infty} \alpha_i^{(0)}(\Psi, \mathcal{K}) < 1$ and $\sum_{i=1}^{\infty} \alpha_i^{(k)}(\Psi, \mathcal{K}) < \infty$ for $k = 1, 2$.

The results

In the sequel, we assume that : $\kappa_n \underset{n\to\infty}{\longrightarrow} \infty$, $\kappa_n/n \underset{n\to\infty}{\longrightarrow} 0$, $\mathbb{E}|\xi_0|^r < \infty$ for some $r \geq 3$ and

$$\sum_{k\geq 1} \frac{1}{\kappa_k} \sum_{j\geq k} \alpha_j^{(0)}(f_\theta, \Theta) + \alpha_j^{(0)}(M_\theta, \Theta) + \alpha_j^{(1)}(f_\theta, \Theta) + \alpha_j^{(1)}(M_\theta, \Theta) < \infty.$$

Consistency:

Theorem

Under some regular conditions, including the above Lipschitz-type conditions on f_θ and M_θ :

1. $P(\widehat{m} = m^*) \underset{n\to\infty}{\longrightarrow} 1$
2. $\widehat{\theta}(\widehat{m}) \xrightarrow[n\to\infty]{\mathcal{P}} \theta^*$

The results

Asymptotic normality of the selected model:

Theorem

Under some regular conditions, including the above Lipschitz-type conditions on f_θ and M_θ :

$$\sqrt{n}\left(\left(\widehat{\theta}(\widehat{m})\right)_i - (\theta^*)_i\right)_{i\in m^*} \xrightarrow[n\to\infty]{\mathcal{D}} \mathcal{N}_{|m^*|}\left(0, F(\theta^*, m^*)^{-1} G(\theta^*, m^*) F(\theta^*, m^*)^{-1}\right)$$

where

$$\left(F(\theta^*, m^*)\right)_{i,j} = \mathbb{E}\left[\frac{\partial^2 q_0(\theta^*)}{\partial\theta_i\partial\theta_j}\right]$$ and

$$\left(G(\theta^*, m^*)\right)_{i,j} = \mathbb{E}\left[\frac{\partial q_0(\theta^*)}{\partial\theta_i}\frac{\partial q_0(\theta^*)}{\partial\theta_j}\right]$$ for $i, j \in m^*$.

The results

The conditions on the Lipschitz-type coefficients show that :

the BIC procedure (with $\kappa_n = \log n$) is consistent in the case of AR(p), ARMA, ARCH(q), GARCH, ARMA-GARCH, $\cdots$

Outline

1 Introduction

2 The model selection criterion

3 Asymptotic results

4 Some numerical Results

Simulation results

We consider the following DGP as "true" models m^*:

1. Model 1, AR(2): $X_t = 0.4X_{t-1} + 0.4X_{t-2} + \xi_t$;
2. Model 2, ARMA(1,1): $X_t = 0.3X_{t-1} + \xi_t + 0.5\xi_{t-1}$;
3. Model 3, ARCH(2): $X_t = \xi_t \sqrt{0.2 + 0.4X_{t-1}^2 + 0.2X_{t-2}^2}$;
4. Model 4, GARCH(1,1): $X_t = \sigma_t \, \xi_t$, with $\sigma_t^2 = 0.2 + 0.3X_{t-1}^2 + 0.5\sigma_{t-1}^2$.

The collection of the competing models is :
$\mathcal{M} = \big\{ARMA(p,q) \ or \ GARCH(p',q') \ with \ 0 \leq p,q,p' \leq 5, \ 1 \leq q' \leq 5\big\}$

There are 66 candidate models.

Simulation results

	n	100			500			1000			2000		
		$\log n$	$\sqrt{n}$	$\widehat{\kappa}_n$	$\log n$	$\sqrt{n}$	$\widehat{\kappa}_n$	$\log n$	$\sqrt{n}$	$\widehat{\kappa}_n$	$\log n$	$\sqrt{n}$	$\widehat{\kappa}_n$
M1	W	21.4	32.3	18.4	1.7	0.8	0.9	0.8	0.1	0.1	0.2	0	0
	T	74.2	67.6	79.7	97.2	99.2	99.1	98.2	99.9	99.9	99.2	100	100
	O	4.4	0.1	1.9	1.1	0	0	1.0	0	0	0.6	0	0
M2	W	30.4	57.7	28.0	4.8	4.2	4.0	0.7	0.3	0.3	0.4	0	0
	T	64.1	42.1	67.3	93.6	95.8	95.8	98.2	99.7	99.6	99.2	100	100
	O	5.5	0.2	4.7	1.6	0	0.2	1.1	0	0.1	0.4	0	0
M3	W	76.1	90.8	53.5	27.3	67.1	18.0	14.0	41.5	13.3	4.6	12.0	4.6
	T	23.8	9.2	39.8	72.7	32.9	79.9	85.9	58.5	86.7	95.4	88.0	95.4
	O	0.1	0	6.7	0	0	2.1	0.1	0	0	0	0	0
M4	W	83.8	94.3	73.4	22.1	61.5	20.4	5.8	31.3	5.7	1.8	6.2	0.7
	T	15.9	5.7	21.6	77.5	38.5	75.9	93.2	68.7	92.6	98.0	93.8	99.3
	O	0.3	0	5.0	0.4	0	3.7	1.0	0	1.7	0.2	0	0

Real data example : Air quality analysis

Air quality : the level of cleanliness of the air

Air pollution is probably one of the first environmental (ecological) concerns of this century

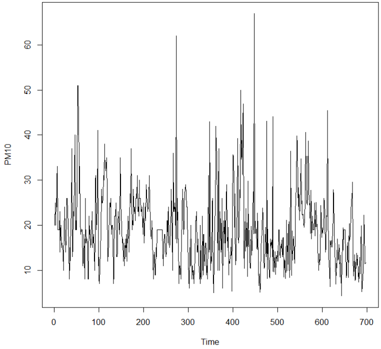
Due to the human activities, the air is degraded by a wide variety of pollutants, including PM.

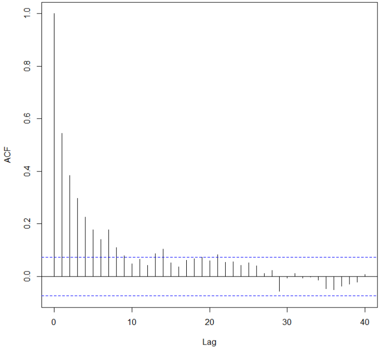
PM stands for particulate matter : the term for a mixture of solid particles and liquid droplets found in the air

We consider daily observations of PM10 (with diameters ≤ 10 micrometers) at Marseille Kaddouz station from January 1, 2018 to November 30, 2019.

Real data example : Air quality analysis

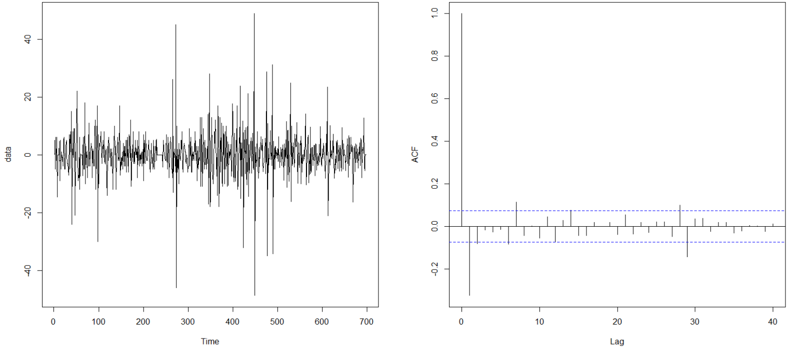
The data





Real data example : Air quality analysis

The differenced data



Real data example : Air quality analysis

The selection results:

	$\kappa_n = \log(n)$	$\kappa_n = \sqrt{n}$	$\kappa_n = \widehat{\kappa}_n$
$\widehat{m}$	ARMA$(1,2)$	ARMA$(1,1)$	ARMA$(1,1)$

The Portmanteau test applied to these models shows that ARMA(1,2) is more suitable.

THANK YOU
FOR YOUR
ATTENTION.