Data driven model selection for same-realization predictions in autoregressive processes

Kamila Kare

Supervisors: Jean-Marc BARDET, William KENGNE

SAMM, Université Paris 1 Panthéon-Sorbonne

5th May 2021, Ecodep Seminar

UNIVERSITÉ PARIS 1 PANTHÉON SORBONNE

Outline

1. AR(∞) Model and some Properties
2. Model Selection
3. Numerical Experiments
4. An Asymptotic Result for Affine causal models

Intuitive Definitions

Let $(X_t)_{t\in\mathbb{Z}}$ be a time series (sequence of r.v. on $(\Omega, \mathcal{A}, \mathbb{P})$)

- $(X_t)_{t\in\mathbb{Z}}$ is a stationary process if $\forall k \in \mathbb{N}^*$, $\forall (t_1, \dots, t_k) \in \mathbb{Z}^k$,

$$\left(X_{t_1}, \dots, X_{t_k}\right) \overset{\mathcal{D}}{\sim} \left(X_{t_1+h}, \dots, X_{t_k+h}\right) \quad \text{for all } h \in \mathbb{Z}.$$

- Assume that $(\xi_t)_{t\in\mathbb{Z}}$ is a white noise (centered i.i.d.r.v.)

  $(X_t)_{t\in\mathbb{Z}}$ causal process if $\exists H : \mathbb{R}^{\mathbb{N}} \to \mathbb{R}$ such as $X_t = H\big((\xi_{t-k})_{k\geq 0}\big)$.

Weak dependence

- (Dedecker and Prieur 2005) mixing coefficients

Let $\mathcal{M}$ a σ-subalgebra of $\mathcal{A}$ and Z a random variable with values in a Banach space $(E, \|.\|_E)$. Assume that $\mathbb{E}|Z| < \infty$ and define

$$\tau^{(p)}(\mathcal{M}, Z) = \Big\| \sup_{f \in \Lambda(E)} \Big\{ \Big| \int f(x) \mathbb{P}_{Z|\mathcal{M}}(dx) - \int f(x) \mathbb{P}_Z(dx) \Big| \Big\} \Big\|_p \qquad (1)$$

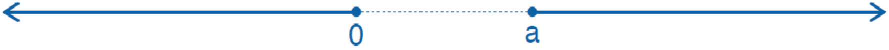
where $\Lambda(E)$ is the set of 1-Lipschitz function, i.e. the functions f from $(E, \|.\|_E)$ to $\mathbb{R}$ such that $|f(x) - f(y)| \le \|x - y\|_E$.

Weak dependence

Let $(Z_t)_{t\in\mathbb{Z}}$ a strictly stationary sequence. Let set by $\mathcal{M}_i = \sigma(Z_t, t \leq i)$ and if $\mathbb{E}(|Z_1|) < \infty$, let

$$\tau_{Z,\infty}^{(p)}(a) = \sup_{l>0} \left\{ \max_{1\leq k\leq l} \frac{1}{k} \sup \left\{ \tau^{(p)}\left(\mathcal{M}_i, (Z_{i_1}, \ldots, Z_{i_k})\right) \ , i + a \leq i_1 < \cdots < i_k \right\} \right\} \quad (2)$$

Finally, the time series $(Z_t)_{t\in\mathbb{Z}}$ is $\tau_{Z,\infty}^{(p)}$-*weakly dependent* when its coefficients $\tau_{Z,\infty}^{(p)}$ tend to 0 as a tends to infinity.



Intuition: at large separations, random vectors are nearly independents.

AR(∞) Model

With $(\xi_t)_{t\in\mathbb{Z}}$ a white noise,

AR(∞) processes $X_t = \sum_{i=1}^{\infty} \theta_i X_{t-i} + \xi_t$

AR(∞) Model

With $(\xi_t)_{t\in\mathbb{Z}}$ a white noise,

AR(∞) processes $X_t = \sum_{i=1}^{\infty} \theta_i X_{t-i} + \xi_t$

$\Longrightarrow$ Invertible ARMA(p,q) processes $X_t + \sum_{i=1}^{p} a_i X_{t-i} = \xi_t + \sum_{i=1}^{q} b_i \xi_{t-i}$.

A special case of Affine causal model

Causal affine models

$$X_t = M(X_{t-1}, X_{t-2}, \ldots)\,\xi_t + f(X_{t-1}, X_{t-2}, \ldots), \quad \forall\, t \in \mathbb{Z}, \text{ a.s..}$$

- $M(\cdot)$ and $f(\cdot)$ are real valued function on $\mathbb{R}^{\mathbb{N}}$;
- $(\xi_t)_{t\in\mathbb{Z}}$ a white noise with $\mathbb{E}(\xi_0) = 0$ and $\mathbb{E}\big(|\xi_0|^r\big) < \infty$, $r \geq 1$.

Existence and stationarity of causal affine models

We will assume that f and M satisfy Lipschitzian conditions:

$$\begin{cases} |f(x) - f(y)| & \leq \sum_{j=1}^{\infty} \alpha_j(f)|x_j - y_j| \\ |M(x) - M(y)| & \leq \sum_{j=1}^{\infty} \alpha_j(M)|x_j - y_j|. \end{cases}$$

for $x = (x_j)_{j\in\mathbb{N}}$ and $y = (y_j)_{j\in\mathbb{N}}$ two sequences of $\mathbb{R}^\infty$.

Proposition (Doukhan and Wintenberger 2008)

If $\sum_{j=1}^{\infty} \alpha_j(f) + \left(\mathbb{E}(|\xi_0|^r)\right)^{1/r} \sum_{j=1}^{\infty} \alpha_j(M) < 1$, *there exists a unique causal solution* $(X_t)_{t\in\mathbb{Z}}$ *which is stationary,* τ*-mixing, ergodic, such as* $\mathbb{E}(|X_0|^r) < \infty$.

Consequence for AR(∞)

if $(X_t)_{t\in\mathbb{Z}}$ is such that

$$X_t = \sum_{i=1}^{\infty} \theta_i^* X_{t-i} + \xi_t \quad \text{where} \quad \mathbb{E}\big(|\xi_0|^r\big) < \infty, r \geq 1 \tag{3}$$

with

$$\textbf{A1}: \qquad \sum_{i=1}^{\infty} |\theta_i^*| < 1,$$

then $(X_t)_{t\in\mathbb{Z}}$ is stationary, τ-mixing and $\mathbb{E}(|X_0|^r) < \infty$.

Bounding mixing coefficients

Proposition

*Assume **A1** holds and if $|\theta^*_t| = O(t^{-\gamma})$ with $\gamma > 1$, there exists a τ-weakly dependent stationary solution of* (3) *and a constant $C_\tau > 0$ such that for $r > 0$*

$$\tau^{(2)}_{X,\infty}(r) \leq C_\tau \left(\frac{\log r}{r}\right)^{\gamma-1} \quad (4)$$

The blocking trick

Let $(X_1, X_2, \ldots, X_n)$ a trajectory of process (3).

Let denote by $\vec{X}_t$ the random vector $\vec{X}_t := (X_{t-1}, \ldots, X_{t-K_n})^\top$ with K_n an integer. Assume that s_n and q_n are two integers verifying $n = 2\, s_n\, q_n$.

For $k = 0, \ldots, s_n - 1$ let denote by

$$A_k = \left(\vec{X}_{2kq_n+1}, \ldots, \vec{X}_{(2k+1)q_n}\right) \quad \text{and} \qquad B_k = \left(\vec{X}_{(2k+1)q_n+1}, \ldots, \vec{X}_{(2k+2)q_n}\right).$$

Coupling Dedecker and Prieur 2005

Proposition

Let $(X_t)_{t\in\mathbb{Z}}$ be the stationary mixing process (3). Let also s_n, q_n, A_k, B_k defined as above for $k = 0, \ldots, s_n - 1$. There exist random vectors $A_k^* = (\vec{X}^*_{2kq_n+1}, \ldots, \vec{X}^*_{(2k+1)q_n})$, $B_k^* = (\vec{X}^*_{(2k+1)q_n+1}, \ldots, \vec{X}^*_{(2k+2)q_n})$ such that:

1. *For $k = 0, \ldots, s_n - 1$, A_k^* has the same law as A_k, also B_k^* and B_k.*
2. *The random vectors $(A_k^*)_{0\le k\le s_n-1}$ are independent and so are the vectors $(B_k^*)_{0\le k\le s_n-1}$.*

1. AR(∞) Model and some Properties
2. Model Selection
   - Model Selection Approach
   - Assumptions
   - Theoretical Results
3. Numerical Experiments
4. An Asymptotic Result for Affine causal models

Prediction risk

Assume $(X_1, X_2, \ldots, X_n)$ is an observed trajectory of process (3) and we want some way to predict X_{n+1} from this trajectory.

Choose some model S_m (here the set of linear functions from $\mathbb{R}^{D_m}$ to $\mathbb{R}$).
Pick a particular $f_\theta \in S_m$, we want to know how well that f_θ will predict X_{n+1}.

Prediction risk

Assume $(X_1, X_2, \ldots, X_n)$ is an observed trajectory of process (3) and we want some way to predict X_{n+1} from this trajectory.

Choose some model S_m (here the set of linear functions from $\mathbb{R}^{D_m}$ to $\mathbb{R}$).
Pick a particular $f_\theta \in S_m$, we want to know how well that f_θ will predict X_{n+1}.

Prediction risk

The prediction risk of a function f_θ for predicting X_{n+1} from $(X_1, X_2, \ldots, X_n)$ is

$$R(\theta) = \mathbb{E}\left[(X_{n+1} - f_\theta^{n+1})^2\right],$$

where $f_\theta^n = f_\theta(X_{n-1}, \ldots, X_{n-D_m})$.

Why care about $R(\theta)$

- Measures predictive accuracy on average;
- How much confidence should you have in f_θ's predictions;
- Compare with other models.

Difficulty: Don't know P the joint distribution of $(X_1, \dots, X_{n+1})$

Empirical risk

How do we find $R(\theta)$? One could estimate it with

$$\gamma_n(\theta) = \frac{1}{n}\sum_{t=1}^{n}(X_t - f_\theta^t)^2.$$

From the Uniform LLN

$$\sup_{\theta\in\Theta}\left|\gamma_n(\theta) - R(\theta)\right| \xrightarrow[n\to\infty]{a.s.} 0,$$

we expect to make $R(\theta)$ small despite knowing nothing about P.

Excess loss

The Bayes predictor which minimizes $R(\theta)$ over the set of all predictors is the inaccessible function f_{θ^*}.
Let then introduce the excess loss of the predictor f_θ

$$\ell(\theta, \theta^*) := R(\theta) - R(\theta^*) = \mathbb{E}\big[(f_{\theta^*}^{n+1} - f_{\theta}^{n+1})^2\big] \geq 0.$$

For a model S_m, the best predictor is

$$\theta^*_m = \underset{\theta \in \Theta_m}{\text{argmin}}\ R(\theta)$$

and its empirical version is

$$\widehat{\theta}_m = \underset{\theta \in \Theta_m}{\text{argmin}}\ \gamma_n(\theta). \tag{5}$$

Model selection

Let $\mathcal{M}_n$ a countable collection of hierarchical model S_m and K_n is the dimension of the largest model in $\mathcal{M}_n$ satisfying $|\mathcal{M}_n| \leq K_n < n$.

Let pen: $\mathcal{M}_n \to \mathbb{R}^+$ be a penalty function, and define

$$\widehat{m} = \underset{m \in \mathcal{M}_n}{\operatorname{argmin}} \{C(m)\} \qquad \text{with} \qquad C(m) := \gamma_n(\widehat{\theta}_m) + \text{pen}(S_m). \tag{6}$$

The best possible choice over $\mathcal{M}_n$ is m^* the so-called *oracle* defined as

$$m^* \in \arg \inf_{m \in \mathcal{M}_n} \ell(\widehat{\theta}_m, \theta^*). \tag{7}$$

Oracle Inequality

m^* depends on P and θ^* that are unknowns
$\Longrightarrow m^*$ is unachievable.

Goal Propose a data driven penalty in order to mimic the oracle in terms of excess risk

$$\ell(\widehat{\theta}_{\widehat{m}}, \theta^*) \leq C_1 \inf_{m \in \mathcal{M}_n} \{\ell(\widehat{\theta}_m, \theta^*)\} + \frac{C_2}{n} \tag{8}$$

with the leading constant C_1 close to one and $C_2 > 0$.

Assumptions

A2 X_t is sub-Gaussian with variance proxy $\sigma_0^2 > 0$ i.e.

$$\mathbb{E}[e^{\lambda X_t}] \leq e^{\lambda^2 \sigma_0^2/2} \quad \text{for any } \lambda > 0.$$

Assumptions

A2 X_t is sub-Gaussian with variance proxy $\sigma_0^2 > 0$ i.e.

$$\mathbb{E}[e^{\lambda X_t}] \leq e^{\lambda^2 \sigma_0^2/2} \quad \text{for any } \lambda > 0.$$

A3: For any $f_\theta \in S_m$, $< \alpha, \partial_\theta f_\theta >= 0$ $a.s.$ $\implies \alpha = 0$

Assumptions

A2 X_t is sub-Gaussian with variance proxy $\sigma_0^2 > 0$ i.e.

$$\mathbb{E}[e^{\lambda X_t}] \leq e^{\lambda^2 \sigma_0^2/2} \quad \text{for any } \lambda > 0.$$

A3: For any $f_\theta \in S_m$, $< \alpha, \partial_\theta f_\theta >= 0$ *a.s.* $\implies \alpha = 0$

A4: There exists a constant $a > 0$ such that $\inf_{-\pi \leq \lambda < \pi} g(\lambda) \geq a$, where

$$g(\lambda) = \frac{1}{2\pi} \sum_{h \in \mathbb{Z}} r(h)\, e^{-ih\lambda}$$

Assumptions

A2 X_t is sub-Gaussian with variance proxy $\sigma_0^2 > 0$ i.e.

$$\mathbb{E}[e^{\lambda X_t}] \leq e^{\lambda^2 \sigma_0^2/2} \quad \text{for any } \lambda > 0.$$

A3: For any $f_\theta \in S_m$, $< \alpha, \partial_\theta f_\theta >= 0$ *a.s.* $\implies \alpha = 0$

A4: There exists a constant $a > 0$ such that $\inf_{-\pi \leq \lambda < \pi} g(\lambda) \geq a$, where

$$g(\lambda) = \frac{1}{2\pi} \sum_{h \in \mathbb{Z}} r(h)\, e^{-ih\lambda}$$

A5: $\frac{a\, s_n}{2} \min \left\{ \left(\frac{(r \wedge 1)}{2^6\, \sigma_0^2\, K_n} \right)^2, \frac{(r \wedge 1)}{2^7\, \sigma_0^2\, K_n} \right\} \geq 3 \log n.$

Intermediate Results

For $m \in \mathcal{M}_n$, let define $\Gamma_{m,r}$ the set

$$\Gamma_{m,r} = \Big\{ \big\| \widehat{\Sigma}_m^{-1} - \Sigma_m^{-1} \big\|_{\text{op}} \leq r \, \big\| \Sigma_m^{-1} \big\|_{\text{op}} \Big\},$$

where $\mathbf{M}_m = \big[X_{i-1}, \ldots, X_{i-D_m} \big]_{i=1}^{n}$, $\widehat{\Sigma}_m = \mathbf{M}_m^\top \mathbf{M}_m$ and $\Sigma_m = \mathbb{E}\big[\widehat{\Sigma}_m\big]$.

Proposition

Under assumptions $\boldsymbol{A1} - \boldsymbol{A5}$ *and if* $|\theta_t^*| = O(t^{-\gamma})$ *with* $\gamma \geq 8$ *, it holds*

$$\mathbb{P}(\Gamma_{m,r}^c) \leq \frac{c_0}{n^3}, \qquad (9)$$

for some constant $c_0 > 0$.

Oracle Inequality

Theorem

Let consider observations $(X_1, \ldots, X_n)$ arising from a solution of the process (3) satisfying ***A1*** *with $|\theta_t^*| = O(t^{-\gamma})$ where $\gamma \geq 8$ and also verifying* ***A2*** *and* ***A4****. Let $\mathcal{M}_n$ be some countable family of AR models satisfying* ***A3*** *and* ***A5****. For $x \geq 4$, let a penalty function pen: $\mathcal{M}_n \to \mathbb{R}^+$ such that*

$$\mathrm{pen}(S_m) = x\,\sigma^2\,\frac{D_m}{n}. \tag{10}$$

Then with probability at least $1 - c_0\, n^{-2}$, the LSE $\widehat{\theta}_{\widehat{m}}$ with $\widehat{m}$ given in (6), satisfies

$$\ell(\widehat{\theta}_{\widehat{m}}, \theta^*) \leq \inf_{m \in \mathcal{M}_n} \left\{ \ell(\widehat{\theta}_m, \theta^*) \right\} + \frac{2\,x\,\sigma^2}{n}. \tag{11}$$

Theorem

Under the assumptions of the Theorem 1, with the same penalty (10)*, then with probability at least* $1 - c_0\, n^{-2}$*, the LSE* $\widehat{\theta}_{\widehat{m}}$ *with* $\widehat{m}$ *given in* (6)*, satisfies*

$$\ell(\widehat{\theta}_{\widehat{m}}, \theta^*) \leq 2 \inf_{m \in \mathcal{M}_n} \Big\{ \ell(\theta^*_m, \theta^*) + \mathrm{pen}(S_m) \Big\} + \frac{2\, x\, \sigma^2}{n}. \qquad (12)$$

Theorem

Under the assumptions of the Theorem 1, with the same penalty (10), then with probability at least $1 - c_0\, n^{-2}$, the LSE $\widehat{\theta}_{\widehat{m}}$ with $\widehat{m}$ given in (6), satisfies

$$\ell(\widehat{\theta}_{\widehat{m}}, \theta^*) \leq 2 \inf_{m \in \mathcal{M}_n} \Big\{ \ell(\theta^*_m, \theta^*) + \mathrm{pen}(S_m) \Big\} + \frac{2\, x\, \sigma^2}{n}. \qquad (12)$$

Corollary

Under the assumptions of Theorem 1, it holds

$$\frac{\ell(\widehat{\theta}_{\widehat{m}}, \theta^*)}{\inf_{m \in \mathcal{M}_n} \Big\{ \ell(\widehat{\theta}_m, \theta^*) \Big\}} \xrightarrow[n \to \infty]{\mathcal{P}} 1.$$

We generated observations from a causal invertible ARMA$(1,1)$

$$X_t = \phi_0 X_{t-1} + \xi_t + \theta_0 \xi_{t-1}$$

where the ξ_t' s are independent and identically $\mathcal{N}(0,1)$ distributed and

$$(\phi_0, \theta_0) \in \Big\{(a,b) : a \in \{0.9, 0.7, 0.5, -0.9, -0.7, -0.5\} \text{ and } \quad b \in \{0.8, 0.6, -0.8, -0.6\}\Big\}$$

as in Ing, Wei, et al. 2005.

For each pair (ϕ_0, θ_0), we compute an empirical version of

$$ME := \frac{\ell(\widehat{\theta}_{\widehat{m}}, \theta^*)}{\inf_{m \in \mathcal{M}_n} \{\ell(\widehat{\theta}_m, \theta^*)\}}$$

with $\widehat{m}$ selected as in (6) where $\text{pen}(S_m) = \widehat{x}\,\sigma^2\,\frac{D_m}{n}$ and the optimal constant $\widehat{x}$ has been calibrated using the dimension jump algorithm implemented in R capushe package and illustrated in Figure 1.

Calibration of x

ARMA(1,1) model with $\phi_0 = 0.9$ and $\theta_0 = 0.7$ and a sample size of 500

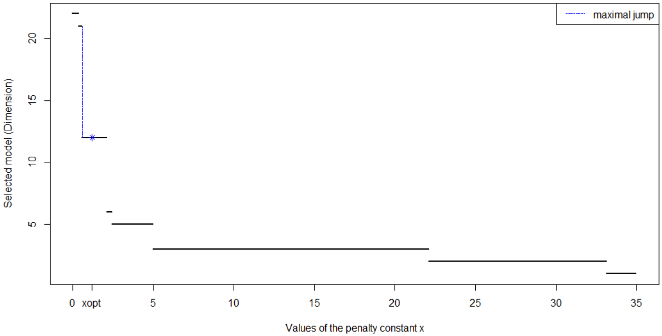


Figure 1: Dimension Jump

		θ_0			
ϕ_0	n/K_n	**0.8**	**0.6**	**-0.8**	**-0.6**
0.9	60/8	1.17	1.24	1.21	1.11
	120/9	1.10	1.19	1.12	1.06
	200/10	1.05	1.16	1.11	1.12
	500/12	1.01	1.04	1.13	1.05
	1000/13	1.01	1.01	1.04	1.06
	2000/15	1.00	1.01	1.03	1.03
0.7	60/8	1.20	1.23	1.15	1.22
	120/9	1.15	1.21	1.09	1.16
	200/10	1.14	1.18	1.18	1.24
	500/12	1.03	1.15	1.09	1.16
	1000/13	1.01	1.07	1.11	1.12
	2000/15	1.01	1.03	1.03	1.12

		θ_0			
ϕ_0	n/K_n	**0.8**	**0.6**	**-0.8**	**-0.6**
0.5	60/8	1.25	1.13	1.12	1.14
	120/9	1.13	1.11	1.06	1.14
	200/10	1.15	1.14	1.07	1.16
	500/12	1.08	1.14	1.05	1.10
	1000/13	1.03	1.08	1.03	1.10
	2000/15	1.01	1.07	1.04	1.11
-0.9	60/8	1.14	1.09	1.17	1.21
	120/9	1.10	1.10	1.07	1.14
	200/10	1.10	1.11	1.04	1.14
	500/12	1.10	1.12	1.02	1.04
	1000/13	1.04	1.08	1.01	1.02
	2000/15	1.07	1.04	1.00	1.01

		θ_0			
ϕ_0	n/K_n	**0.8**	**0.6**	**-0.8**	**-0.6**
-0.7	60/7	1.17	1.21	1.20	1.17
	120/8	1.14	1.21	1.11	1.11
	200/9	1.18	1.12	1.12	1.21
	500/12	1.11	1.12	1.03	1.13
	1000/13	1.06	1.05	1.01	1.07
	2000/15	1.05	1.11	1.00	1.03
-0.5	60/8	1.10	1.22	1.13	1.12
	120/9	1.12	1.13	1.14	1.12
	200/10	1.09	1.08	1.14	1.08
	500/12	1.06	1.17	1.11	1.08
	1000/13	1.05	1.12	1.04	1.06
	2000/15	1.04	1.07	1.01	1.09

1. AR(∞) Model and some Properties
2. Model Selection
3. Numerical Experiments
4. An Asymptotic Result for Affine causal models
   - Examples
   - Quasi-likelihood contrast
   - Asymptotic Results

Examples

- AR(∞) processes $X_t = \sum_{i=1}^{\infty} \theta_i X_{t-i} + \xi_t$

$\Longrightarrow$ Causal ARMA(p,q) processes $X_t + \sum_{i=1}^{p} a_i X_{t-i} = \xi_t + \sum_{i=1}^{q} b_i \xi_{t-i}$.

- ARCH(∞) processes, (Robinson, 1991), with $b_0 > 0$ and $b_j \geq 0$

$$\begin{cases} X_t &= \sigma_t \xi_t, \\ \sigma_t^2 &= \phi_0 + \sum_{j=1}^{\infty} \phi_j X_{t-j}^2. \end{cases}$$

$\Longrightarrow$ GARCH(p,q) processes, with $c_0 > 0$, c_j, $d_j \geq 0$, c_p, $d_q > 0$

$$\begin{cases} X_t &= \sigma_t \xi_t, \\ \sigma_t^2 &= c_0 + \sum_{j=1}^{p} c_j X_{t-j}^2 + \sum_{j=1}^{q} d_j \sigma_{t-j}^2 \end{cases}$$

Extensions of univariate ARCH models

- TGARCH(∞) processes, (Zakoïan, 1994), with $b_0,\ b_j^+,\ b_j^- \geq 0$

$$\begin{cases} X_t &= \sigma_t\, \xi_t, \\ \sigma_t &= b_0 + \sum_{j=1}^{\infty} \left[b_j^+ \max(X_{t-j},0) - b_j^- \min(X_{t-j},0) \right] \end{cases} .$$

- APARCH(δ, p, q) processes, (Ding *et al.*, 1993)

$$\begin{cases} X_t &= \sigma_t\, \zeta_t, \\ \sigma_t^\delta &= \omega + \sum_{j=1}^{p} \alpha_i (|X_{t-i}| - \gamma_i X_{t-i})^\delta + \sum_{j=1}^{q} \beta_j \sigma_{t-j}^\delta, \end{cases}$$

with $\delta \geq 1$, $\omega > 0$, $-1 < \gamma_i < 1$ and $\alpha_i,\ \beta_j \geq 0$.

Combinations of models

- ARMA-GARCH processes, (Ding *et al.*, 1993, Ling and McAleer, 2003)

$$\begin{cases} X_t &= \sum_{i=1}^{p} a_i X_{t-i} + \varepsilon_t + \sum_{j=1}^{q} b_j \varepsilon_{t-j}, \\ \varepsilon_t &= \sigma_t \zeta_t, \quad \text{with} \quad \sigma_t^2 = c_0 + \sum_{i=1}^{p'} c_i \varepsilon_{t-i}^2 + \sum_{j=1}^{q'} d_j \sigma_{t-j}^2 \end{cases}$$

- ARMA-APARCH processes, (Ding *et al.*, 1993)

$$\begin{cases} X_t &= \sum_{i=1}^{p} a_i X_{t-i} + \varepsilon_t + \sum_{j=1}^{q} b_j \varepsilon_{t-j}, \\ \varepsilon_t &= \sigma_t \zeta_t, \quad \text{with} \quad \sigma_t^\delta = \omega + \sum_{j=1}^{p'} \alpha_i (|X_{t-i}| - \gamma_i X_{t-i})^\delta + \sum_{j=1}^{q'} \beta_j \sigma_{t-j}^\delta \end{cases}$$

Quasi-likelihood contrast

Let $(X_1, \ldots, X_n)$ an observed trajectory of an $\mathcal{CA}(M_{\theta^*}, f_{\theta^*})$

$$X_t = M_{\theta^*}(X_{t-1}, X_{t-2}, \ldots)\,\xi_t + f_{\theta^*}(X_{t-1}, X_{t-2}, \ldots), \quad \forall\, t \in \mathbb{Z}.$$

Goal: Prediction of X_{n+1}.

Given a predictor θ, we measure its quality by the risk defined as

$$R(\theta) := \mathbb{P}\gamma(\theta) = \mathbb{E}[\gamma(\theta, X_1)]$$

$$\text{with} \quad \gamma(\theta, X_t) := \frac{(X_t - f_\theta^t)^2}{H_\theta^t} + \log(H_\theta^t), \text{ and } \begin{cases} f_\theta^t & := & f_\theta\big((X_{t-k})_{k\geq 1}\big) \\ M_\theta^t & := & M_\theta\big((X_{t-k})_{k\geq 1}\big) \\ H_\theta^t & := & \big(M_\theta^t\big)^2 \end{cases} \quad (13)$$

Quasi-likelihood contrast

Let set by γ_n the associated empirical criterion

$$\gamma_n(\theta) := \mathbb{P}_n \gamma(\theta, .) = \frac{1}{n} \sum_{t=1}^{n} \gamma(\theta, X_t).$$

Problem γ_n is not computable since it depends on $(X_t)_{t \leq 0}$ that are unobserved.

Quasi-likelihood contrast

Introduce an approximation of the empirical risk:

$$\widehat{\gamma}_n(\theta) := \frac{1}{n}\sum_{t=1}^{n}\widehat{\gamma}(\theta, X_t)$$

with $\widehat{\gamma}(\theta, X_t) := \frac{(X_t - \widehat{f}_\theta^t)^2}{\widehat{H}_\theta^t} + \log(\widehat{H}_\theta^t)$ and $\begin{cases} \widehat{f}_\theta^t & := & f_\theta(X_{t-1}, X_{t-2}, \cdots, X_1, u) \\ \widehat{M}_\theta^t & := & M_\theta(X_{t-1}, X_{t-2}, \cdots, X_1, u) \\ \widehat{H}_\theta^t & := & (\widehat{M}_\theta^t)^2 \end{cases}$ (14)

for any deterministic sequence $u = (u_n)_{n\in\mathbb{N}}$ with finitely many non-zero values (we will use $u = 0$ without loss of generality).

Penalization procedure

Given a finite family of affine models $\mathcal{M}$, the goal is to come up with the model that verifies

$$\inf_{m \in \mathcal{M}} \ell(\widehat{\theta}_m, \theta^*). \tag{15}$$

where $\widehat{\theta}_m = \underset{\theta \in \Theta_m}{\operatorname{argmin}} \, \widehat{\gamma}_n(\theta)$.

Then define for some penalty pen: $m \in \mathcal{M} \mapsto \text{pen}(m) \in \mathbb{R}^+$,

$$\widehat{m}_{\text{pen}} = \underset{m \in \mathcal{M}}{\operatorname{argmin}} \left\{ \widehat{C}_{\text{pen}}(m) \right\} \quad \text{with} \quad \widehat{C}_{\text{pen}}(m) := \widehat{\gamma}_n\big(\widehat{\theta}_m\big) + \text{pen}(m). \tag{16}$$

In order to achieve (15), the *ideal penalty* to consider in (16) is

$$\text{pen}_{id}(m) = R(\widehat{\theta}_m) - \widehat{\gamma}_n(\widehat{\theta}_m). \tag{17}$$

Computation of the ideal penalty?

Obviously, pen_{id} is not accessible.

Computation of the ideal penalty?

Obviously, pen_{id} is not accessible.

$\Longrightarrow$ Asymptotic approximation?

Extra assumptions

Assumptions

Assumption A0: *The process* $X \in \mathcal{AC}(M_{\theta^*}, f_{\theta^*})$ *verifies:*

- *the white noise* $(\xi_t)_t$ *is such as* $\|\xi_0\|_r < \infty$ *with* $8 < r$;
- *for any* $x \in \mathbb{R}^\infty$, *the functions* $\theta \to M_\theta$ *and* $\theta \to f_\theta$ *are* $\mathcal{C}^2(\Theta)$ *functions:*
- $\Theta \in \mathbb{R}^d$ *is a compact set such as*

$$\Theta \subset \Big\{\theta \in \mathbb{R}^d,\ A(f_\theta, \{\theta\}) \text{ and } A(M_\theta, \{\theta\}) \text{ hold with } \sum_{k=1}^{\infty} \alpha_k(f_\theta, \{\theta\}) + \|\xi_0\|_r \sum_{k=1}^{\infty} \alpha_k(M_\theta, \{\theta\}) < 1\Big\}. \quad (18)$$

Assumptions

- **A1**: *For all* $\theta, \theta' \in \Theta$, $(f^0_\theta = f^0_{\theta'}$ *and* $M^0_\theta = M^0_{\theta'})$ *a.s.* $\implies \theta = \theta'$.

Assumptions

- **A1**: *For all* θ, $\theta' \in \Theta$, $(f^0_\theta = f^0_{\theta'}$ *and* $M^0_\theta = M^0_{\theta'})$ *a.s.* $\implies \theta = \theta'$.
- **A2**: $< \alpha, \partial_\theta f^0_\theta >= 0 \implies \alpha = 0$ *a.s. or* $< \alpha, \partial_\theta H^0_\theta >= 0 \implies \alpha = 0$ *a.s.*

Assumptions

- **A1**: *For all* $\theta,\ \theta' \in \Theta$, $(f^0_\theta = f^0_{\theta'}$ *and* $M^0_\theta = M^0_{\theta'})$ *a.s.* $\implies \theta = \theta'$.
- **A2**: $<\alpha, \partial_\theta f^0_\theta>= 0 \implies \alpha = 0$ *a.s. or* $<\alpha, \partial_\theta H^0_\theta>= 0 \implies \alpha = 0$ *a.s.*
- **A3**: $\exists \underline{h} > 0$ *such that* $\inf_{\theta\in\Theta}(H_\theta(x)) \geq \underline{h}$ *for all* $x \in \mathbb{R}^\infty$.

Assumptions

- **A1**: *For all* θ, $\theta' \in \Theta$, $(f^0_\theta = f^0_{\theta'}$ *and* $M^0_\theta = M^0_{\theta'})$ *a.s.* $\implies \theta = \theta'$.
- **A2**: $< \alpha, \partial_\theta f^0_\theta >= 0 \implies \alpha = 0$ *a.s. or* $< \alpha, \partial_\theta H^0_\theta >= 0 \implies \alpha = 0$ *a.s.*
- **A3**: $\exists \underline{h} > 0$ *such that* $\inf_{\theta \in \Theta} (H_\theta(x)) \geq \underline{h}$ *for all* $x \in \mathbb{R}^\infty$.
- **A4**: *For every* $m \in \mathcal{M}$, *if* $(\overline{\theta}_{m,n})$ *is a sequence of* Θ_m *satisfying* $\overline{\theta}_{m,n} \xrightarrow[n\to\infty]{a.s.} \theta^*_m$, *then*

$$\limsup_{n\to\infty} \left\{ \mathbb{E}\left[\left(\left\| \left(\partial^2_{\theta_i \theta_j} \gamma_n(\overline{\theta}_m) \right)_{i,j \in m} \right) ^{-1} \right\|^8 \right] \right\} < \infty. \quad (19)$$

Asymptotic expansion of the ideal penalty

Proposition

Under assumptions **A0-A5** *and for any $m \in \mathcal{M}$, there exists a bounded sequence $(v_n^*)_{n\in\mathbb{N}^*}$ not depending on m satisfying*

$$\mathbb{E}\left[pen_{id}(m)\right] \underset{n\to\infty}{\sim} \left(\frac{1}{n}\, Trace\Big(\big(F(\theta_m^*)\big)^{-1} G(\theta_m^*)\Big)\right) + \frac{v_n^*}{n}, \qquad (20)$$

with

$$F(\theta_m^*) := \left(\mathbb{E}\left[\partial^2_{\theta_i\,\theta_j}\gamma(\theta_m^*, X_0)\right]\right)_{i,j\in m}$$

and

$$G(\theta_m^*) := \frac{1}{4}\left(\sum_{t\in\mathbb{Z}} Cov\big(\partial_{\theta_i}\gamma(\theta_m^*, X_0)\,,\, \partial_{\theta_j}\gamma(\theta_m^*, X_t)\big)\right)_{i,j\in m}$$

Asymptotic efficiency

Theorem

Under assumptions **A0-A5**, *then*

1. *Assume that for any* $\varepsilon > 0$ *there exists* $K_\varepsilon > 0$ *such as*

$$\mathbb{P}\Big(\limsup_{n\to\infty} \max_{m\in\mathcal{M}} \big(n\,\mathrm{pen}(m) \big) \geq K_\varepsilon \Big) \leq \varepsilon. \tag{21}$$

Then for any $\varepsilon > 0$, *there exists* $M_\varepsilon > 0$ *and* $N_\varepsilon \in \mathbb{N}^*$ *such as for any* $n \geq N_\varepsilon$,

$$\mathbb{P}\Big(\ell(\widehat{\theta}_{\widehat{m}_{\mathrm{pen}}}, \theta^*) \leq \inf_{m\in\mathcal{M}} \big\{ \ell(\widehat{\theta}_m, \theta^*) \big\} + \frac{M_\varepsilon}{n} \Big) \geq 1 - \varepsilon. \tag{22}$$

Asymptotic efficiency

Theorem

2. *Define for $m \in \mathcal{M}$,*

$$\widetilde{\text{pen}}(m) := \frac{1}{n}\, Trace\Big(\big(F(\theta^*_m)\big)^{-1} G(\theta^*_m)\Big). \tag{23}$$

Then for any $\eta > 0$, there exists $N_\eta \in \mathbb{N}^*$ such as for any $n \geq N_\eta$,

$$\mathbb{E}\big[\ell(\widehat{\theta}_{\widehat{m}_{\widetilde{\text{pen}}}}, \theta^*)\big] \leq \mathbb{E}\big[\min_{m\in\mathcal{M}} \big\{\ell(\widehat{\theta}_m, \theta^*)\big\}\big] + \frac{\eta}{n}. \tag{24}$$

Dedecker, Jérôme and Clémentine Prieur (2005). "New dependence coefficients. Examples and applications to statistics". In: *Probability Theory and Related Fields* 132.2, pp. 203–236.

Doukhan, P. and O. Wintenberger (2008). "Weakly dependent chains with infinite memory". In: *Stochastic Processes and their Applications* 118.11, pp. 1997–2013.

Ing, Ching-Kang, Ching-Zong Wei, et al. (2005). "Order selection for same-realization predictions in autoregressive processes". In: *The Annals of Statistics* 33.5, pp. 2423–2474.

Thank you