Efficiency and consistency of model selection for time series

Joint work with K. Kare (Paris 1) and W. Kengne (Cergy)

Jean-Marc Bardet, SAMM, Université Paris 1, France
bardet@univ-paris1.fr

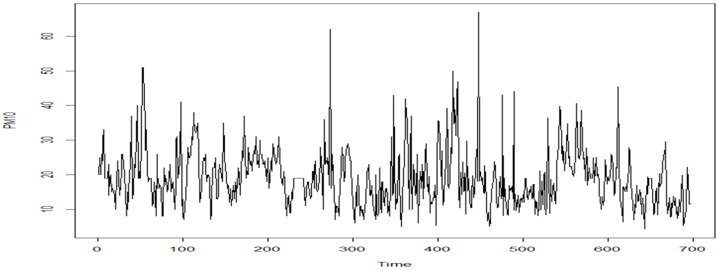
EcoDep 2022 Conference

Paris, June 24

Outline

Example

Observe the daily observations of PM10 at Marseille 01/2018 to 12/2019 :



$\Longrightarrow$ **Aims :** Chosing an "optimal" model for these data from a family $\mathcal{M}$ of models. For instance,

$$\mathcal{M} = \Big\{ \text{ARMA}(p,q) \text{ or GARCH}(p',q'), \quad \text{with } 0 \leq p,p' \leq p_{\max},\ 0 \leq q,q' \leq q_{\max} \Big\}$$

Outline

Two 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 $\forall k \in \mathbb{N}^*$, $\forall (t_1, \ldots, t_k) \in \mathbb{Z}^k$,

$$\big(X_{t_1}, \ldots, X_{t_k}\big) \overset{\mathcal{L}}{\sim} \big(X_{t_1+h}, \ldots, X_{t_k+h}\big) \quad \text{for all } h \in \mathbb{Z}.$$

- 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 $X_t = H\big((\xi_{t-k})_{k\geq 0}\big)$.

Causal AR[∞] and ARCH(∞) models

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

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

  $\Longrightarrow$ 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, $d_q > 0$

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

A common frame for studying time series

A common class of models for AR, ARMA, ARCH and GARCH processes :

Causal affine models : class $\mathcal{CA}(M, f)$

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

- $M(\cdot)$ 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 $\mathbb{E}\big[|\xi_0|^r\big] < \infty$, $r \geq 1$.

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 α_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)

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

Existence and stationarity of causal affine models

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

Our main tool for studying those models :

Assume that $\begin{cases} \dfrac{\partial}{\partial x_i} f\big((x_k)_{k\geq 1}\big) \\ \dfrac{\partial}{\partial x_i} M\big((x_k)_{k\geq 1}\big) \end{cases}$ exist on $\mathbb{R}^\infty$ for any $i \geq 1$.

Proposition (from Doukhan and Wintenberger, 2007)

If $\mathbb{E}\big[|\xi_0|^r\big] < \infty$ with $r \geq 1$, there exists a unique causal solution $(X_t)_{t\in\mathbb{Z}}$ which is stationary, ergodic, such as $\mathbb{E}(|X_0|^r) < \infty$, when

$$\sum_{j=1}^{\infty} \sup_{x\in\mathbb{R}^\infty} \Big| \frac{\partial}{\partial x_j} f\big((x_k)_{k\geq 1}\big)\Big| + \big(\mathbb{E}\big[|\xi_0|^r\big]\big)^{1/r} \sum_{j=1}^{\infty} \sup_{x\in\mathbb{R}^\infty} \Big| \frac{\partial}{\partial x_j} M\big((x_k)_{k\geq 1}\big)\Big| < 1$$

Examples

Conditions on stationarity become :

- Causal AR[∞] :
  $X_t = \sum_{j=0}^{\infty} a_j \xi_{t-j} \Longrightarrow \sum_{j=0}^{\infty} |a_j| < 1$;

- Causal ARCH[∞] :
  $X_t = \xi_t \sqrt{c_0 + \sum_{j=1}^{\infty} c_j X_{t-j}^2} \Longrightarrow \left(\mathbb{E}\left[|\xi_0|^r\right]\right)^{1/r} \sum_{j=1}^{\infty} \sqrt{c_j} < 1$;

- Causal TARCH[∞] :
  $X_t = \xi_t \left(b_0 + \sum_{j=1}^{\infty} \left[b_j^+ \max(X_{t-j}, 0) - b_j^- \min(X_{t-j}, 0)\right]\right)$
  $\Longrightarrow \left(\mathbb{E}\left[|\xi_0|^r\right]\right)^{1/r} \sum_{j=1}^{\infty} \max\left(b_j^-, b_j^+\right) < 1$;

Additivity of causal affine models

Proposition

Let $\Theta_1 \subset \mathbb{R}^{d_1}$, $\Theta_2 \subset \mathbb{R}^{d_1}$, $M^{(1)}_{\theta_1}, f^{(1)}_{\theta_1}, M^{(2)}_{\theta_2}, f^{(2)}_{\theta_2}$ *for* $\theta_1 \in \Theta_1$, $\theta_2 \in \Theta_2$. *There exist* $\max(d_1, d_2) \leq d \leq d_1 + d_2$, $\Theta \subset \mathbb{R}^d$, *and* M_θ, f_θ *with* $\theta \in \Theta$, *such as for any* $\theta_1 \in \Theta_1 \subset \mathbb{R}^{d_1}$ *and* $\theta_2 \in \Theta_2 \subset \mathbb{R}^{d_2}$,

$$\Big\{\mathcal{CA}\big(M^{(1)}_{\theta_1}, f^{(1)}_{\theta_1}\big) \bigcup \mathcal{CA}\big(M^{(2)}_{\theta_2}, f^{(2)}_{\theta_2}\big)\Big\} \subset \Big\{\mathcal{CA}\big(M_\theta, f_\theta\big)\Big\}.$$

Consequence :

- For any family $\mathcal{M} = \bigcup_{i \in I} \mathcal{CA}\big(M^{(i)}_{\theta_i}, f^{(i)}_{\theta_i}\big)$,

$$\Longrightarrow \quad \exists d \in \mathbb{N}^*,\ M_\theta \text{ and } f_\theta \text{ such as } \mathcal{M} = \bigcup_{i \in I} \Big\{\mathcal{CA}\big(M_\theta, f_\theta\big)\Big\}_{\theta \in \Theta_i \subset \mathbb{R}^d}$$

Finite family of causal affine models

If $\mathcal{M}$ is a finite family of $\mathcal{CA}$ models :

- $\mathcal{M} \sim \{m, \text{ with } m \subset \{1, \dots, d\}\}$;
- for a model $m \in \{1, \dots, d\}$, $\exists \Theta(m) \subset \mathbb{R}^d$ such as

$$X \in \mathcal{CA}\big(M_\theta, f_\theta\big) \text{ with } \theta \in \Theta(m) \subset \big\{(x_1, \dots, x_d) \in \mathbb{R}^d,\ x_i = 0 \text{ if } i \notin m\big\}.$$

Now assume :

$$\Theta(m) \subset \Theta \subset \Big\{\theta \in \mathbb{R}^d,\ \sum_{j=1}^{\infty} \sup_{x \in \mathbb{R}^\infty} \big|\partial_{x_j} f_\theta(x)\big| + \big(\mathbb{E}\big[|\xi_0|^r\big]\big)^{1/r} \sum_{j=1}^{\infty} \sup_{x \in \mathbb{R}^\infty} \big|\partial_{x_j} M_\theta(x)\big| < 1\Big\}$$

$\Longrightarrow$ Semi-parametric model selection...

Outline

Gaussian QMLE of causal affine model

Denote $m^* \in \mathcal{M}$ so-called the true model :

$(X_1, \ldots, X_n)$ observed trajectory of $\mathcal{CA}(M_{\theta^*}, f_{\theta^*})$ with $\theta^* \in \Theta(m^*)$

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

- With $f_\theta^t = f_\theta(X_{t-1}, X_{t-2}, \ldots)$, $M_\theta^t = M_\theta(X_{t-1}, X_{t-2}, \ldots)$,

  Gaussian conditional log-density : $q_t(\theta) = -\frac{1}{2}\Big[\frac{(X_t - f_\theta^t)^2}{(M_\theta^t)^2} + \log\big((M_\theta^t)^2\big)\Big]$

- Let $\widehat{f}_\theta^t = f_\theta(X_{t-1}, \ldots, X_1, 0, \cdots)$ and $\widehat{M}_\theta^t = M_\theta(X_{t-1}, \ldots, X_1, 0, \cdots)$,

  Quasi conditional log-density : $\widehat{q}_t(\theta) = -\frac{1}{2}\Big[\frac{(X_t - \widehat{f}_\theta^t)^2}{(\widehat{M}_\theta^t)^2} + \log\big((\widehat{M}_\theta^t)^2\big)\Big]$.

$$\Longrightarrow \text{Gaussian QMLE : } \widehat{\theta}_n = \operatorname*{argmax}_{\theta \in \Theta} \widehat{L}_n(\theta) \text{ with } \widehat{L}_n(\theta) = \sum_{t=1}^{n} \widehat{q}_t(\theta).$$

A risk for a family of causal affine models

Let $X \in \mathcal{CA}\big(M_\theta, f_\theta\big)$ and $\theta \in \Theta \subset \mathbb{R}^d$, define its risk by :

$$R(\theta) = -2 \times \mathbb{E}[q_1(\theta)] = \mathbb{E}[\gamma(\theta, X_1)], \quad \gamma(\theta, X_t) = \frac{(X_t - f_\theta^t)^2}{(M_\theta^t)^2} + \log\big((M_\theta^t)^2\big)$$

$$\text{where } f_\theta^t = f_\theta(X_{t-1}, X_{t-2}, \ldots),\ M_\theta^t = M_\theta(X_{t-1}, X_{t-2}, \ldots).$$

Assumption A1 : for $\theta, \theta' \in \Theta$, $\big(f_\theta^0 = f_{\theta'}^0 \textit{ and } M_\theta^0 = M_{\theta'}^0\big)$ *a.s.* $\implies \theta = \theta'$.

From Assumption **A1**, for $m \in \mathcal{M}$, θ_m^* exists and is unique with

$$\theta_m^* = \operatorname*{argmin}_{\theta \in \Theta(m)} R(\theta)$$

$$\implies \quad \theta_{m^*}^* = \theta^* \quad \text{and if } m^* \subset m, \quad \theta_m^* = \theta^*.$$

Empirical risk and computable empirical risk

Define the empirical risk :

$$R_n(\theta) = \frac{1}{n} \sum_{t=1}^{n} \gamma(\theta, X_t).$$

Not computable ! Define the computable empirical risk :

$$\widehat{R}_n(\theta) = \frac{1}{n} \sum_{t=1}^{n} \widehat{\gamma}(\theta, X_t) \quad \text{with } \widehat{\gamma}(\theta, X_t) := \frac{(X_t - \widehat{f}_\theta^t)^2}{(\widehat{M}_\theta^t)^2} + \log\left((\widehat{M}_\theta^t)^2\right)$$

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

Finally, for $m \in \mathcal{M}$, the QMLE $\widehat{\theta}_m$ is

$$\widehat{\theta}_m = \underset{\theta \in \Theta(m)}{\text{argmin}} \, \widehat{R}_n(\theta).$$

Model selection procedure

Define a penalty function $m \in \mathcal{M} \mapsto \text{pen}(m) \in \mathbb{R}^+$, possibly data-dependent, such as $\text{pen}(m_1) \leq \text{pen}(m_2)$ when $m_1 \subset m_2$.

Then define the penalized contrast and the model selected by it :

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

Natural aim : find $\widehat{m}_{id} = \underset{m \in \mathcal{M}}{\text{argmin}}\, R(\widehat{\theta}_m)$.

$\Longrightarrow$ Let the ideal penalty be defined by

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

Outline

Assumptions

- **A0** : Θ for $r > 8$, where $\mathbb{E}[\xi_0^2] = 1$;
- **A1** : for $\theta, \theta' \in \Theta$, $(f_\theta^0 = f_{\theta'}^0$ *and* $M_\theta^0 = M_{\theta'}^0)$ *a.s.* $\Longrightarrow$ $\theta = \theta'$.
- **A2** : $\exists \underline{M} > 0$ such that $M_\theta(x) \geq \underline{M}$ for all $\theta \in \Theta$, $x \in \mathbb{R}^{\mathbb{N}}$.
- **A3** : For any $m \in \mathcal{M}$, θ_m^* belongs to the interior of $\Theta(m)$.
- **A4** : For any $\theta \in \Theta$, $x \in \mathbb{R}^\infty$, $\partial_{x_k}\partial^2_{\theta^2} f_\theta(x)$ and $\partial_{x_k}\partial^2_{\theta^2} M_\theta(x)$ exist and

$$a/ \ \sup_{\theta\in\Theta}\Big(\sup_{x\in\mathbb{R}^\infty}\big|\partial_{x_k} f_\theta(x)\big| + \sup_{x\in\mathbb{R}^\infty}\big|\partial_{x_k} M_\theta(x)\big| + \sup_{x\in\mathbb{R}^\infty}\big\|\partial_{x_k}\partial_\theta f_\theta(x)\big\| + \sup_{x\in\mathbb{R}^\infty}\big\|\partial_{x_k}\partial_\theta M_\theta(x)\big\|\Big) = O(k^{-\delta}) \text{ with } \delta > 7/2$$

$$b/ \ \sup_{\theta\in\Theta}\Big(\sum_{k=1}^\infty \sup_{x\in\mathbb{R}^\infty}\big\|\partial_{x_k}\partial^2_{\theta^2} f_\theta(x)\big\| + \sup_{x\in\mathbb{R}^\infty}\big\|\partial_{x_k}\partial^2_{\theta^2} M_\theta(x)\big\|\Big) < \infty$$

Asymptotic normality of the estimator

Théorème

Under Assumptions **A0-A4**, *for any* $m \in \mathcal{M}$,

$$\sqrt{n}\,\big((\widehat{\theta}_m)_i - (\theta^*_m)_i\big)_{i\in m} \xrightarrow[n\to\infty]{\mathcal{L}} \mathcal{N}\Big(0\,,\,\big(F_m(\theta^*_m)\big)^{-1} G_m(\theta^*_m)\,\big(F_m(\theta^*_m)\big)^{-1}\Big),$$

with G_m *and* F_m *defined by*

- $$G_m(\theta) = \frac{1}{4}\Big(\sum_{t\in\mathbb{Z}} cov\big(\partial_{\theta_i}\gamma(\theta,X_0)\,,\,\partial_{\theta_j}\gamma(\theta,X_t)\big)\Big)_{i,j\in m}$$
  $$\Longrightarrow\; G_m(\theta^*) = \frac{1}{4}\Big(cov\big(\partial_{\theta_i}\gamma(\theta^*,X_0)\,,\,\partial_{\theta_j}\gamma(\theta^*,X_0)\big)\Big)_{i,j\in m} \quad \text{if } m^* \subset m$$
- $$F_m(\theta) = -\frac{1}{2}\Big(\mathbb{E}\big[\partial^2_{\theta_i\,\theta_j}\gamma(\theta,X_0)\big]\Big)_{i,j\in m}.$$

- Could be applied to all cited processes ARMA, ARCH, APARCH,...

Consequences of asymptotic normality

Proposition

Under Assumptions **A0-A4**, *there exists* $N_0 \in \mathbb{N}$ *such as for any* $n \geq N_0$,

$$\underset{m \in \mathcal{M}}{\text{argmin}}\ \mathbb{E}\big[R(\widehat{\theta}_m)\big] = m^*.$$

Proposition

Under Assumptions **A0-A4** *and for any* $m \in \mathcal{M}$, $\exists$ *a bounded sequence* $(v_n^*)_{n \in \mathbb{N}^*}$, *not depending on* m *when* $m^* \subset m$, *satisfying*

$$\mathbb{E}\big[pen_{id}(m)\big] \underset{n \to \infty}{\sim} -\frac{2}{n}\, Trace\Big(\big(F_m(\theta_m^*)\big)^{-1} G_m(\theta_m^*)\Big) + \frac{v_n^*}{n}.$$

Rem : $-2\text{Trace}\Big(\big(F_m(\theta_m^*)\big)^{-1} G_m(\theta_m^*)\Big) = \begin{cases} 2\,|m| & \text{Gaussian process} \\ 2\,|m| & \text{ARMA process} \\ (\mu_4 - 1)\,|m| & \text{GARCH process} \end{cases}$

Efficiency

Théorème

Under Assumptions **A0-A4**, *and if for any* $\varepsilon > 0$, $\exists K_\varepsilon > 0$ *such as*

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

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

$$\mathbb{P}\Big(R\big(\widehat{\theta}_{\widehat{m}_{\mathrm{pen}}}\big) \leq \min_{m\in\mathcal{M}}\big\{R\big(\widehat{\theta}_m\big)\big\} + \frac{M_\varepsilon}{n}\Big) \geq 1-\varepsilon.$$

Example : Satisfied for $\mathrm{pen}(m) = \mathbb{E}\big[\mathrm{pen}_{id}(m)\big]$, not for $\mathrm{pen}(m) = \frac{\log n}{n}$.

Efficiency (2)

Théorème

Assume that there exists $g : \mathcal{M} \to [0, \infty[$ *such as* $\text{pen}(m) = \frac{g(m)}{n}$ *for any* $m \in \mathcal{M}$. *Then, under Assumptions* **A0-A4**,

$$\liminf_{n\to\infty} \ \mathbb{P}\big(\widehat{m}_{\text{pen}} \text{ overfits}\big) > 0.$$

and there exists $M > 0$ *such as for* n *large enough,*

$$\mathbb{E}\big[R(\widehat{\theta}_{\widehat{m}_{\text{pen}}})\big] \geq \min_{m\in\mathcal{M}} \mathbb{E}\big[R(\widehat{\theta}_m)\big] + \frac{M}{n}.$$

Example : Satisfied for $\text{pen}(m) = \mathbb{E}\big[\text{pen}_{id}(m)\big]$, not for $\text{pen}(m) = \frac{\log n}{n}$.

Efficiency and consistency

Théorème

Under Assumptions **A0-A4** *and if for any* $\varepsilon > 0$,

$$n\,\mathbb{P}\big(\mathrm{pen}(m) \geq \varepsilon\big) \underset{n\to+\infty}{\longrightarrow} 0 \quad \text{for any } m \in \mathcal{M}.$$

Then,

$$n\,\mathbb{P}\big(m^* \not\subset \widehat{m}_{\mathrm{pen}}\big) \underset{n\to+\infty}{\longrightarrow} 0.$$

$\Longrightarrow$ if the penalty decreases to 0 (in proba), $\widehat{C}_{\mathrm{pen}}$ does not select a misspecified model asymptotically.

Efficiency and consistency (2)

Théorème

Under Assumptions **A0-A4**, *if the penalty pen satisfies* (25), *and if for* $m^* \subset m$, $e_n(m) = pen(m) - pen(m^*) > 0$ *satisfies*

$$n\,\mathbb{E}[e_n(m)] \underset{n\to+\infty}{\longrightarrow} \infty \qquad \text{and} \qquad n\,\mathbb{E}\Big[\big|e_n(m) - \mathbb{E}[e_n(m)]\big|\Big] \underset{n\to+\infty}{\longrightarrow} 0,$$

$$\text{then} \quad \mathbb{P}\big(\widehat{m}_{\text{pen}} = m^*\big) \underset{n\to+\infty}{\longrightarrow} 1.$$

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

$$\begin{cases} \mathbb{P}\Big(R\big(\widehat{\theta}_{\widehat{m}_{\text{pen}}}\big) \leq R\big(\widehat{\theta}_{m^*}\big)\big) + \frac{\eta}{n}\Big) \geq 1 - \varepsilon \\ \mathbb{E}\Big[R\big(\widehat{\theta}_{\widehat{m}_{\text{pen}}}\big)\Big] \leq \min_{m\in\mathcal{M}} \mathbb{E}\Big[R\big(\widehat{\theta}_m\big)\Big] + \frac{\eta}{n}. \end{cases}$$

$\Longrightarrow$ Results valid for instance for BIC penalty $\text{pen}(m) = \frac{\log n}{n}$.

A new consistent criterion

Théorème (Laplace approximation)

Under Assumptions **A0-A4**, *and for any* $x \in R^\infty$, *the functions* $\theta \to M_\theta$ *and* $\theta \to f_\theta$ *are* $\mathcal{C}^6(\Theta)$, *then*

$$-2\times\log\left(\mathbb{P}\left((X_1,\ldots,X_n)\,|\,m\right)\right) = -2\times\widehat{L}_n(\widehat{\theta}_m)+\log(n)\,|m|-2\,\log\left(b_m(\widehat{\theta}_m)\right) - \log(2\pi)\,|m| + \log\left(\det\left(-\widehat{F}_n(m)\right)\right) + 2\,\log(|\mathcal{M}|) + O(n^{-1}) \quad a.s.$$

where $\widehat{F}_n(m) := \left(\partial^2_{\theta_i\theta_j}\widehat{R}_n(\widehat{\theta}_m)\right)_{i,j\in m}$ *and* b_m *a bounded function on* Θ.

Consequences : Using this approximation :

- $\widehat{m}_{BIC} = \underset{m\in\mathcal{M}}{\text{argmin}}\left\{-2\,\widehat{L}_n(\widehat{\theta}_m) + \log(n)\,|m|\right\}$, (Schwarz, 1978)
- $\widehat{m}_{KC} = \underset{m\in\mathcal{M}}{\text{argmin}}\left\{-2\,\widehat{L}_n(\widehat{\theta}_m) + \log(n)\,|m| + \log\left(\det\left(-\widehat{F}_n(m)\right)\right)\right\}$, (Kashyap, 1982).

A new consistent criterion (2)

$\Longrightarrow$ By taking more terms in the Laplace approximation, define :

$$\widehat{KC'}(m) = BIC(m) - \log(2\pi)\,|m| + \log\Big(\det\big(-\widehat{F}_n(m)\big)\Big) + 2\log\big(|m|\big)$$

$$\text{and} \quad \widehat{m}_{KC'} = \operatorname*{argmin}_{m\in\mathcal{M}} \big\{\widehat{KC}'(m)\big\}$$

Corollary

The criteria BIC, KC and KC' are consistent model selection criteria and satisfy $o(1/n)$ efficiency.

Outline

Simulation results for classical models

Consider the following test bench :

DGP I	AR(2)	$X_t = 0.4\,X_{t-1} + 0.4\,X_{t-2} + \xi_t,$
DGP II	ARMA(1,1)	$X_t - 0.5\,X_{t-1} = \xi_t + 0.6\,\xi_{t-1},$
DGP III	GARCH(1,1)	$X_t = \sigma_t\,\xi_t \quad \text{with} \quad \sigma_t^2 = 1 + 0.35X_{t-1}^2 + 0.4\sigma_{t-1}^2,$
DGP IV	GARCH(1,1)	$X_t = \sigma_t\,\xi_t \quad \text{with} \quad \sigma_t^2 = 0.01 + 0.15X_{t-1}^2 + 0.75\sigma_{t-1}^2,$
DGP V	GARCH(1,2)	$X_t = \sigma_t\,\xi_t \quad \text{with} \quad \sigma_t^2 = 0.01 + 0.2X_{t-1}^2 + 0.2\sigma_{t-1}^2 + 0.5\sigma_{t-2}^2,$
DGP VI	GARCH(2,2)	$X_t = \sigma_t\,\xi_t \quad \text{with} \quad \sigma_t^2 = 0.01 + 0.05X_{t-1}^2 + 0.2X_{t-2}^2 + 0.3\sigma_{t-1}^2$

where $(\xi_t)_t$ is a Gaussian white noise with variance unity.

Simulation results (2)

	n	200			500			1000			2000		
		AIC	BIC	KC'	AIC	BIC	KC'	AIC	BIC	KC'	AIC	BIC	KC'
DGP I	True	17.2	36.2	35.6	30.4	73.2	78.2	36.4	87.4	92.2	32.4	96.2	98.4
	Wrong	82.8	63.8	64.4	69.6	26.8	21.8	63.6	13.6	7.8	67.6	03.8	01.6
DGP II	True	27.8	80.8	92.0	30.6	88.4	96.6	31.0	89.1	97.5	33.3	95.2	99.9
	Wrong	72.2	19.2	08.0	69.7	11.6	03.4	69.0	10.9	02.5	66.7	04.8	00.1
DGP III	True	00.4	10.8	14.8	01.4	32.2	55.8	01.0	54.8	82.0	02.0	75.8	93.8
	Wrong	99.6	89.2	85.2	98.6	67.8	44.2	99.0	45.2	18.0	98.0	24.2	06.2

Table – Percentage of "true" selected models for DGP I-III.

n	200			500			1000			2000		
	AIC	BIC	KC'	AIC	BIC	KC'	AIC	BIC	KC'	AIC	BIC	KC'
DGP I	4.91	2.59	5.35	3.46	1.11	1.18	3.08	0.98	0.75	3.05	0.38	0.29
DGP II	3.66	0.87	0.54	3.37	0.42	0.11	2.62	0.15	0.05	2.5	0.10	0.04
DGP III	2.39	4.63	13.16	2.53	4.08	9.54	2.69	2.96	2.52	3.21	2.06	0.76

Table – $\widehat{ME} = n\left(\overline{\widetilde{R}(\widehat{\theta}_{\widehat{m}})} - \overline{\widetilde{R}(\widehat{\theta}_{m^*})}\right)$ for DGP I-III.

Simulation results (3)

	n	AIC	500 BIC	KC'	AIC	1000 BIC	KC'	AIC	2000 BIC	KC'	AIC	5000 BIC	KC
DGP IV	True	81.8	88.4	63.0	86.8	98.2	87.8	87.2	98.4	94.4	88.8	100	100
	Wrong	18.2	11.6	37.0	13.2	1.8	12.2	12.8	1.6	5.6	11.2	0	0
DGP V	True	29.8	9.6	24.0	51.2	22.6	54.6	76.4	49.8	84.2	83.8	95.8	97.6
	Wrong	70.2	90.4	76.0	48.8	77.4	45.4	23.6	50.2	15.8	16.2	4.2	2.4
DGP VI	True	25.4	3.4	22.0	44.8	8.6	48.8	68.8	24.4	75.0	84.8	71.2	94.0
	Wrong	74.6	96.6	78.0	55.2	91.4	51.2	31.2	75.6	25.0	15.2	28.8	6.0

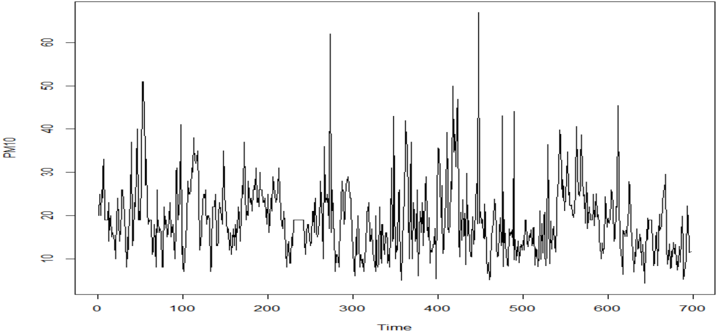
Table – Percentage of "true" selected models for DGP IV-VI.

n	AIC	500 BIC	KC'	AIC	1000 BIC	KC'	AIC	2000 BIC	KC'	AIC	5000 BIC	KC'
DGP IV	0.52	4.23	12.17	0.95	0.13	1.71	0.67	0.08	0.24	0.43	0	0
DGP V	3.58	8.23	15.12	2.16	4.3	4.45	1.43	4.45	1.05	0.65	0.84	0.15
DGP VI	2.42	10.63	21.16	2.30	5.27	2.65	1.26	5.14	1.24	0.90	3.08	0.46

Table – $\widehat{ME} = n\left(\overline{\widetilde{R}(\widehat{\theta}_{\widehat{m}})} - \overline{\widetilde{R}(\widehat{\theta}_{m^*})}\right)$ for DGP IV-VI.

Example

For the daily observations of PM10 at Marseille 01/2018 to 12/2019 :



$$\Longrightarrow \quad \mathrm{ARMA}(1,2)$$

References

- Bardet, Kare and Kengne (2020). Consistent model selection criteria and goodness-of-fit test for common time series models. *Elec. Journ. Statist.*
- Bardet and Wintenberger (2009). Asymptotic normality of the Quasi-Maximum Likelihood Estimator for multidimensional causal processes. *Ann. Statist.*
- Doukhan and Wintenberger (2008). Weakly dependent chains with infinite memory. *Stoch. Proc. and their Appli.*
- Francq and Zakoïan (2010). *GARCH Models*. Wiley.
- Hannan (1980). The Estimation of the Order of an ARMA Process. *Ann. Statist.*
- Hsu, Ing and Tong (2019). On model selection from a finite family of possibly misspecified time series models. *Ann. of Statist.*
- Kashyap (1982). Optimal choice of AR and MA parts in autoregressive moving average models. *IEEE Trans Pattern Anal. Machine Intel.*