Prediction of weakly locally stationary processes by auto-regression

François Roueff

LTCI, Télécom ParisTech, Université Paris Saclay

Sep. 7, 2017

Joint work with Andrés Sánchez Pérez.

Outline

The basic model

Consider a time series X_t, $t \in \mathbb{Z}$.

We wish to compute an online predictor of X_t given its past $X_{t-1}, X_{t-2}, \cdots$

We assume a model able to adapt to a non-stationary context.

The basic model

Consider a time series X_t, $t \in \mathbb{Z}$.

We wish to compute an online predictor of X_t given its past $X_{t-1}, X_{t-2}, \cdots$

We assume a model able to adapt to a non-stationary context.

AR(p) processes :

$$X_t = \sum_{k=1}^{p} \theta_k X_{t-k} + \sigma \epsilon_t .$$

The basic model

Consider a time series X_t, $t \in \mathbb{Z}$.

We wish to compute an online predictor of X_t given its past $X_{t-1}, X_{t-2}, \cdots$

We assume a model able to adapt to a non-stationary context.

time varying AR(p) processes :

$$X_t = \sum_{k=1}^{p} \theta_k(t) X_{t-k} + \sigma(t)\,\epsilon_t\,.$$

The basic model

Consider a time series X_t, $t \in \mathbb{Z}$.

We wish to compute an online predictor of X_t given its past $X_{t-1}, X_{t-2}, \dots$

We assume a model able to adapt to a non-stationary context.

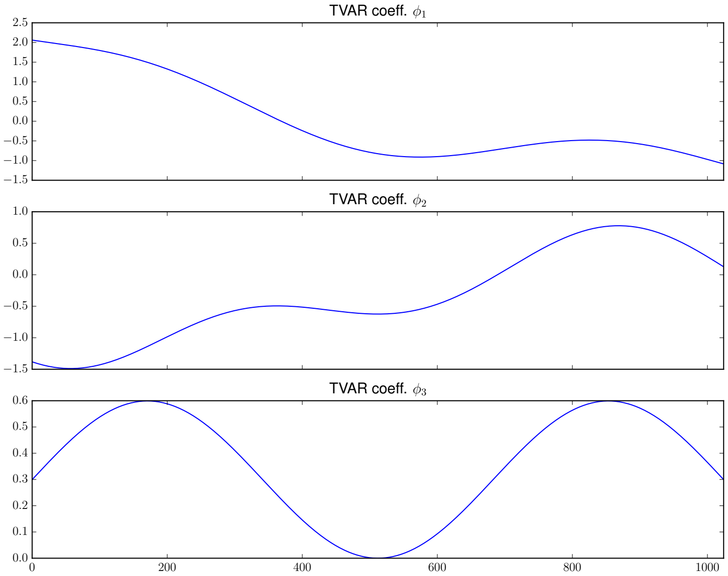
Locally stationary time varying AR(p) processes : (Künsch [1995], Dahlhaus [1996])

$$X_{t,T} = \sum_{k=1}^{p} \theta_k(t/T) X_{t-k,T} + \sigma(t/T)\, \epsilon_t \ ,$$

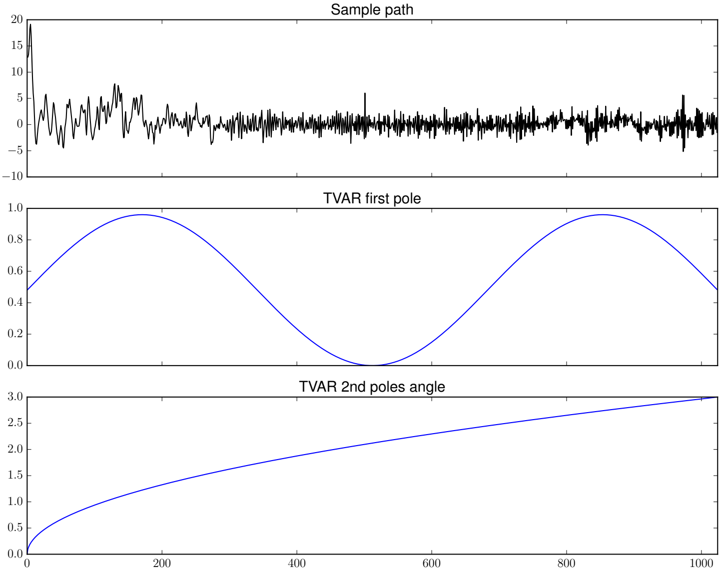
where the unknown parameters $\theta_1, \dots, \theta_p, \sigma^2$ now are functions of T-rescaled time so that as $T \to \infty$ with $t/T \sim u$, $X_{t,T}$ is "closed to" the stationary solution $X_t(u)$ of

$$X_t(u) = \sum_{k=1}^{p} \theta_k(u) X_{t-k}(u) + \sigma(u) \epsilon_t \ .$$

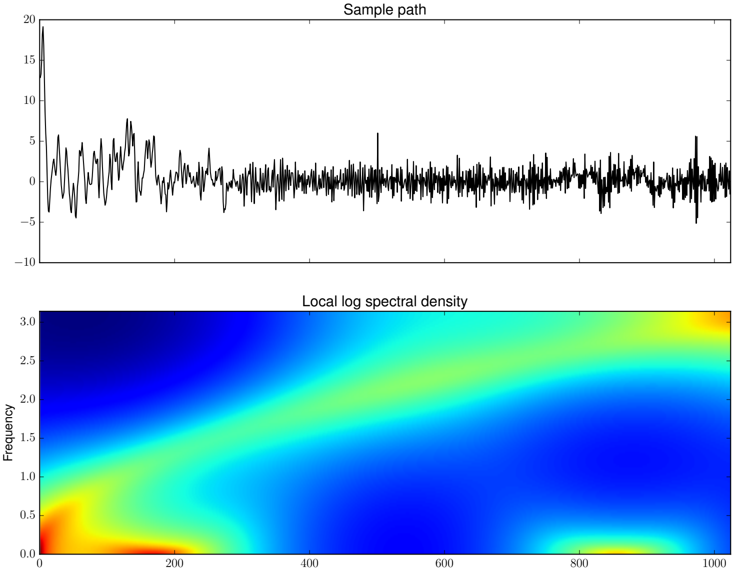
An example with $\sigma^2 = 1$, $p = 3$ and $T = 2^{12}$:



Same example, roots reciprocals :



Same example, local spectral density function :



Linear prediction for L^2 processes

Take a L^2 centered process $(X_{t,T})_{t\in\mathbb{Z},T\geq T_0}$ Denote

$$\gamma^*(t,T,\ell) = \mathrm{cov}\left(X_{t,T}, X_{t-\ell,T}\right) .$$

Then the best linear predictor of order d is $\theta^{*\prime}_{t,T}\mathbf{X}_{t-1,T}$ where $\mathbf{X}_{t-1,T} = [X_{t-1,T} \ \ldots \ X_{t-d,T}]'$ and $\theta^*_{t,T}$ is solution of

$$\Gamma^*_{t,T}\theta^*_{t,T} = \boldsymbol{\gamma}^*_{t,T} ,$$

with

$$\Gamma^*_{t,T} = \mathrm{Cov}\left(\mathbf{X}_{t-1,T}\right) = \left(\gamma^*(t-i,T,j-i);\ i,j=1,\ldots,d\right) ,$$

and

$$\boldsymbol{\gamma}^*_{t,T} = \mathrm{Cov}\left(\mathbf{X}_{t-1,T}, X_{t,T}\right) = [\gamma^*(t,T,1) \ \ldots \ \gamma^*(t,T,d)]' .$$

Linear prediction for weakly locally stationary processes

Let $(X_{t,T})_{t\in\mathbb{Z},T\geq T_0}$ be a centered (β, R)-weakly locally stationary with local spectral density $f(u,\lambda)$ and local autocovariance function $\gamma(u,\ell)$, $u, \lambda \in \mathbb{R}$, $\ell \in \mathbb{Z}$.

Linear prediction for weakly locally stationary processes

Let $(X_{t,T})_{t\in\mathbb{Z},T\geq T_0}$ be a centered (β, R)-weakly locally stationary with local spectral density $f(u,\lambda)$ and local autocovariance function $\gamma(u,\ell)$, $u,\lambda\in\mathbb{R}$, $\ell\in\mathbb{Z}$.

Suppose that $f(u,\lambda)\geq f_- > 0$ for all u,λ,

Linear prediction for weakly locally stationary processes

Let $(X_{t,T})_{t\in\mathbb{Z},T\geq T_0}$ be a centered (β, R)-weakly locally stationary with local spectral density $f(u,\lambda)$ and local autocovariance function $\gamma(u,\ell)$, $u,\lambda\in\mathbb{R}$, $\ell\in\mathbb{Z}$.

Suppose that $f(u,\lambda)\geq f_- > 0$ for all u,λ,

Let

$$\theta_u = \Gamma_u^{-1}\boldsymbol{\gamma}_u$$

where Γ_u and $\boldsymbol{\gamma}_u$ are the analogues of $\Gamma^*_{t,T}$ and $\boldsymbol{\gamma}^*_{t,T}$, with $\gamma(u,\ell)$ replacing $\gamma^*(t-i, T, \ell)$.

Linear prediction for weakly locally stationary processes

Let $(X_{t,T})_{t\in\mathbb{Z},T\geq T_0}$ be a centered (β, R)-weakly locally stationary with local spectral density $f(u,\lambda)$ and local autocovariance function $\gamma(u,\ell)$, $u,\lambda\in\mathbb{R}$, $\ell\in\mathbb{Z}$.

Suppose that $f(u,\lambda)\geq f_- > 0$ for all u,λ,

Let

$$\theta_u = \Gamma_u^{-1}\boldsymbol{\gamma}_u$$

where Γ_u and $\boldsymbol{\gamma}_u$ are the analogues of $\Gamma^*_{t,T}$ and $\boldsymbol{\gamma}^*_{t,T}$, with $\gamma(u,\ell)$ replacing $\gamma^*(t-i, T, \ell)$.

Then, we have

$$\left\|\theta^*_{t,T} - \theta_{t/T}\right\| \leq C_1\, T^{-\min(1,\beta)}\ ,$$

where C_1 only depends on β, d, R and f_-.

Estimation of autoregressive coefficients for weakly locally stationary

We use a h-tapered estimator of $\gamma(u, \ell)$, (following Dahlhaus and Giraitis [1998])

$$\widehat{\gamma}_{T,M}(u, \ell) = \frac{1}{H_M} \sum_t h\left(\frac{t}{M}\right) h\left(\frac{t-\ell}{M}\right) X_{\lfloor uT \rfloor + t - M/2, T} X_{\lfloor uT \rfloor + t - \ell - M/2, T} ,$$

where h is supported on $[0, 1]$ and

$$H_M = \sum_{k=1}^{M} h^2(k/M) \sim M \int_0^1 h^2(x)\mathrm{d}x = M .$$

Estimation of autoregressive coefficients for weakly locally stationary

We use a h-tapered estimator of $\gamma(u,\ell)$, (following Dahlhaus and Giraitis [1998])

$$\widehat{\gamma}_{T,M}(u,\ell) = \frac{1}{H_M}\sum_t h\left(\frac{t}{M}\right) h\left(\frac{t-\ell}{M}\right) X_{\lfloor uT\rfloor+t-M/2,T} X_{\lfloor uT\rfloor+t-\ell-M/2,T} ,$$

where h is supported on $[0,1]$ and

$$H_M = \sum_{k=1}^{M} h^2(k/M) \sim M \int_0^1 h^2(x)\mathrm{d}x = M .$$

Hence the estimator $\widehat{\theta}_{t,T}(M)$ defined by

$$\widehat{\Gamma}_{t,T,M}\widehat{\theta}_{t,T}(M) = \widehat{\gamma}_{t,T,M} ,$$

where $\widehat{\Gamma}_{t,T,M}$ and $\widehat{\gamma}_{t,T,M}$ are obtained from $\widehat{\gamma}_{T,M}(t/T,\ell)$.

Estimation of autoregressive coefficients for weakly locally stationary (cont.)

Let $(X_{t,T})$ be as above, with $\beta = k + \alpha$, $\alpha \in (0,1]$. Suppose that $\widehat{\gamma}_{T,M}(u,\ell) - \mathbb{E}\left[\widehat{\gamma}_{T,M}(u,\ell)\right]$ is decreasing at $M^{-1/2}$-rate in L^p-norm for all $p \geq 2$. Then we have (Roueff and Sánchez-Pérez [2017])

$$\widehat{\theta}_{t,T}(M) = \theta_{t/T} + \sum_{j=1}^{k} c_j \left(\frac{M}{T}\right)^j + O_{L^p}\left(\left(\frac{M}{T}\right)^\beta + M^{-1/2}\right) ,$$

with $c_1 = 0$ if $h(x) = h(1-x)$.

Estimation of autoregressive coefficients for weakly locally stationary (cont.)

Let $(X_{t,T})$ be as above, with $\beta = k + \alpha$, $\alpha \in (0,1]$. Suppose that $\widehat{\gamma}_{T,M}(u,\ell) - \mathbb{E}\left[\widehat{\gamma}_{T,M}(u,\ell)\right]$ is decreasing at $M^{-1/2}$-rate in L^p-norm for all $p \geq 2$. Then we have (Roueff and Sánchez-Pérez [2017])

$$\widehat{\theta}_{t,T}(M) = \theta_{t/T} + \sum_{j=1}^{k} c_j \left(\frac{M}{T}\right)^j + O_{L^p}\left(\left(\frac{M}{T}\right)^\beta + M^{-1/2}\right),$$

with $c_1 = 0$ if $h(x) = h(1-x)$.

We then define, using appropriate ω_ℓ's,

$$\widetilde{\theta}_{u,T}(M) = \theta_{t/T} + \sum_{\ell=0}^{k} \omega_\ell \widehat{\theta}_{t,T}\left(M 2^\ell\right)$$

so that

$$\widetilde{\theta}_{u,T}(M) = \theta_{t/T} + O_{L^p}\left(\left(\frac{M}{T}\right)^\beta + M^{-1/2}\right).$$

Estimation of autoregressive coefficients for weakly locally stationary (cont.)

Let $(X_{t,T})$ be as above, with $\beta = k + \alpha$, $\alpha \in (0,1]$. Suppose that $\widehat{\gamma}_{T,M}(u,\ell) - \mathbb{E}\left[\widehat{\gamma}_{T,M}(u,\ell)\right]$ is decreasing at $M^{-1/2}$-rate in L^p-norm for all $p \geq 2$. Then we have (Roueff and Sánchez-Pérez [2017])

$$\widehat{\theta}_{t,T}(M) = \theta_{t/T} + \sum_{j=1}^{k} c_j \left(\frac{M}{T}\right)^j + O_{L^p}\left(\left(\frac{M}{T}\right)^\beta + M^{-1/2}\right),$$

with $c_1 = 0$ if $h(x) = h(1-x)$.

We then define, using appropriate ω_ℓ's,

$$\widetilde{\theta}_{u,T}(M) = \theta_{t/T} + \sum_{\ell=0}^{k} \omega_\ell \widehat{\theta}_{t,T}\left(M2^\ell\right)$$

so that

$$\widetilde{\theta}_{u,T}(M) = \theta^*_{t,T} + O_{L^p}\left(\left(\frac{M}{T}\right)^\beta + M^{-1/2}\right).$$

Simulation setting

- Choose a model:
  - we set $p = 3$,
  - and pick random functions $\theta_1, \dots, \theta_p, \sigma^2$ on $[0, 1]$.
- Monte Carlo loops:
  - Generate $X_{t,T}$ for $t = 1, \dots, T$ and $T = 2^{10}, 2^{12}, \dots, 2^{30}$.
  - Compute $\widehat{\theta}_{T/2,T}(M)$ and $\widetilde{\theta}_{T/2,T}(M)$ for $M = 2^{11}, \dots, 2^{18}$.
  - Compute the estimation quadratic errors

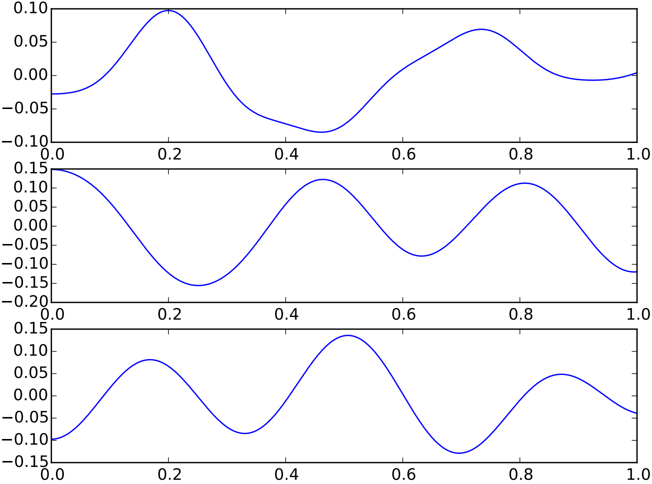
$$\left\| \widehat{\theta}_{T/2,T}(M) - \theta(1/2) \right\|^2$$

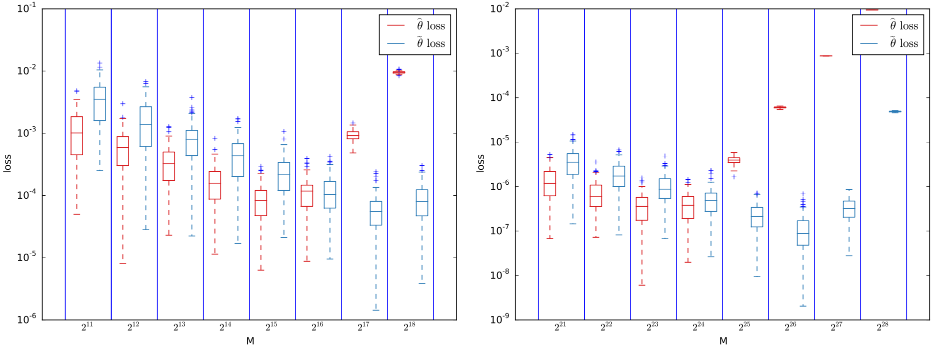
and

$$\left\| \widetilde{\theta}_{T/2,T}(M) - \theta(1/2) \right\|^2$$

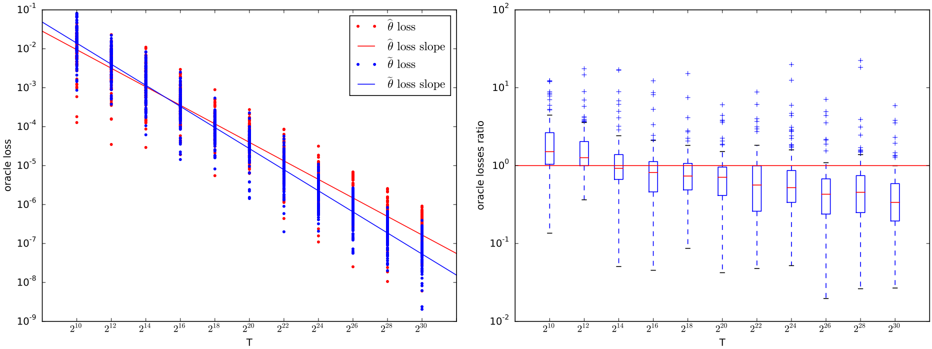
over the Monte-Carlo runs.

Model parameters :





$T = 2^{20}$ (left) and $T = 2^{30}$ (right)



Optimal M (oracle estimators).

Aggregation of N predictors: a simple example

Let $(\widehat{X}^{(i)}_t)_{t=1\dots,T}$, $i = 1, \dots, N$, be N predictors of X_t.
Compute for some given $\eta > 0$ the *exponential* weights, for all $t = 1, \dots, T$,

$$\widehat{\alpha}^{(i)}_t \propto \mathrm{e}^{-\eta \sum_{s=1}^{t-1} (\widehat{X}^{(i)}_s - X_s)^2} \qquad \text{summing to 1 over } i = 1, \dots, N.$$

The aggregated predictor is defined by $\widehat{X}_t = \sum_{i=1}^{N} \widehat{\alpha}^{(i)}_t \widehat{X}^{(i)}_t, \quad t = 1, \dots, T$.

Assume that $|\widehat{X}^{(i)}_t - X_t| \leq C$ for all t and all i. Then, if $2C^2\eta \leq 1$, we have

$$\frac{1}{T} \sum_{t=1}^{T} \left(\widehat{X}_t - X_t\right)^2 \leq \inf_{1 \leq i \leq N} \frac{1}{T} \sum_{t=1}^{T} \left(\widehat{X}^{(i)}_t - X_t\right)^2 + \frac{\ln N}{\eta T} .$$

Setting the experts and aggregate them

- Compute a finite number of autoregression estimators $\widehat{\theta}^{(i)}_{t,T}$, $i = 1, \ldots, N$ which are minimax rate for well chosen smoothness indices $\beta_1, \ldots, \beta_N$.

Setting the experts and aggregate them

- Compute a finite number of autoregression estimators $\widehat{\theta}^{(i)}_{t,T}$, $i = 1, \dots, N$ which are minimax rate for well chosen smoothness indices $\beta_1, \dots, \beta_N$.
- Compute the corresponding predictors

$$\widehat{X}^{(i)}_{t,T} = \widehat{\theta}^{(i)\prime}_{t,T} \mathbf{X}_{t-1,T} \ .$$

They will be our experts.

Setting the experts and aggregate them

- Compute a finite number of autoregression estimators $\widehat{\theta}^{(i)}_{t,T}$, $i = 1, \dots, N$ which are minimax rate for well chosen smoothness indices $\beta_1, \dots, \beta_N$.
- Compute the corresponding predictors

$$\widehat{X}^{(i)}_{t,T} = \widehat{\theta}^{(i)\prime}_{t,T} \mathbf{X}_{t-1,T} .$$

They will be our experts.

- Use online aggregation to find a new adaptive predictor, based on these experts.

$$\widehat{X}_{t,T}(A) = \sum_{i=1}^{N} \alpha_{i,t,T} \, \widehat{X}^{(i)}_{t,T} .$$

Setting the experts and aggregate them

- Compute a finite number of autoregression estimators $\widehat{\theta}^{(i)}_{t,T}$, $i = 1, \ldots, N$ which are minimax rate for well chosen smoothness indices $\beta_1, \ldots, \beta_N$.
- Compute the corresponding predictors

$$\widehat{X}^{(i)}_{t,T} = \widehat{\theta}^{(i)\prime}_{t,T} \mathbf{X}_{t-1,T} \; .$$

They will be our experts.

- Use online aggregation to find a new adaptive predictor, based on these experts.

$$\widehat{X}_{t,T}(A) = \sum_{i=1}^{N} \alpha_{i,t,T} \, \widehat{X}^{(i)}_{t,T} \; .$$

- What can be said about the obtained prediction error ?

Theoretical results

(See Giraud, Roueff, and Sanchez-Perez [2015])

- Upper bound: one can choose the individual predictors to get

$$\frac{1}{T}\sum_{t=1}^{T}\left(\mathbb{E}\left[\left(\widehat{X}_{t,T}(A) - X_{t,T}\right)^2\right] - \sigma^2_{t,T}\right) = O\left(T^{-2\beta/(1+2\beta)}\right) \ .$$

Theoretical results

(See Giraud, Roueff, and Sanchez-Perez [2015])

- ▷ Upper bound: one can choose the individual predictors to get

$$\frac{1}{T}\sum_{t=1}^{T}\left(\mathbb{E}\left[\left(\widehat{X}_{t,T}(A) - X_{t,T}\right)^2\right] - \sigma_{t,T}^2\right) = O\left(T^{-2\beta/(1+2\beta)}\right) .$$

- ▷ Lower bound: this is the minimax prediction rate over $\theta \in s_p(\delta) \cap \Lambda_p(\beta, L)$ and σ valued in $[\sigma_-, \sigma_+]$.

Simulation setting

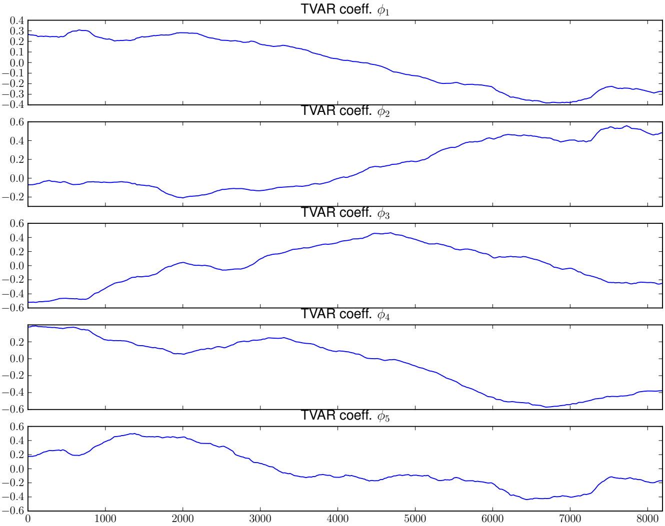
- Choose a model:
  - we set $T = 2^{13}$ and $p = 5$,
  - and pick random functions $\theta_1, \dots, \theta_p, \sigma^2$ on $[0, 1]$.
- Monte Carlo loops:
  - Generate $X_{t,T}$ for $t = 1, \dots, T$.
  - Compute $\widehat{\theta}^{(i)}_{t,T}$, $\widehat{X}^{(i)}_{t,T}$, $n = 1, \dots, N$ and $\widehat{X}_{t,T}(A)$ for $t = 1, \dots, T$.
  - Compute the regret

$$\frac{1}{T} \sum_{t=1}^{T} \left(\widehat{X}^{(i)}_{t,T} - X_{t,T} \right)^2 , \quad i = 1, \dots, N ,$$

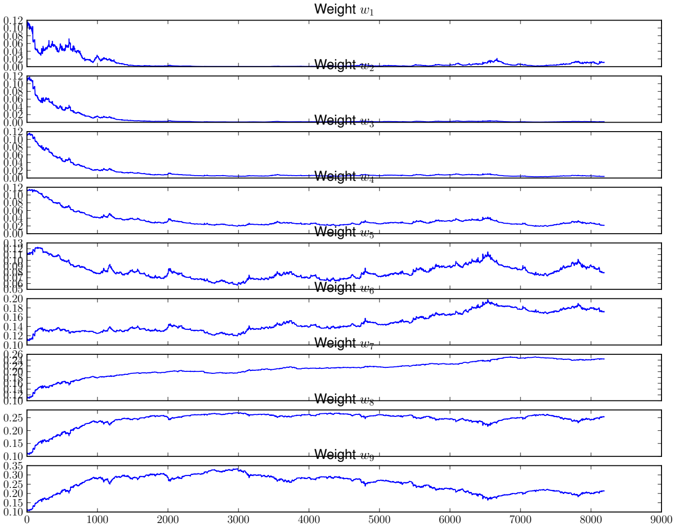
and

$$\frac{1}{T} \sum_{t=1}^{T} \left(\widehat{X}_{t,T}(A) - X_{t,T} \right)^2 .$$

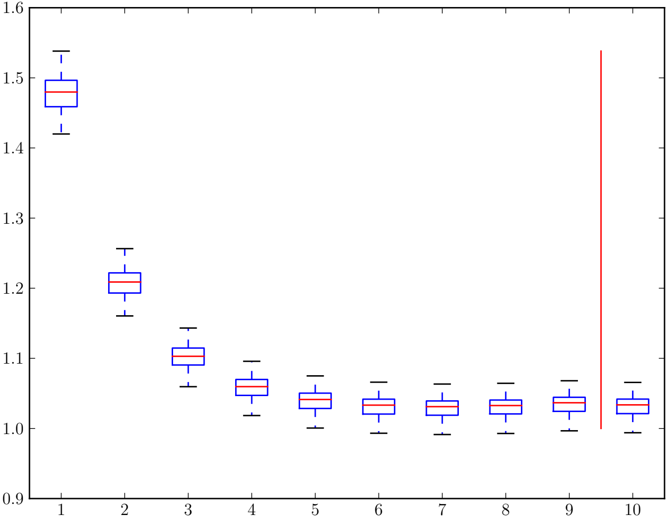
Model parameters :



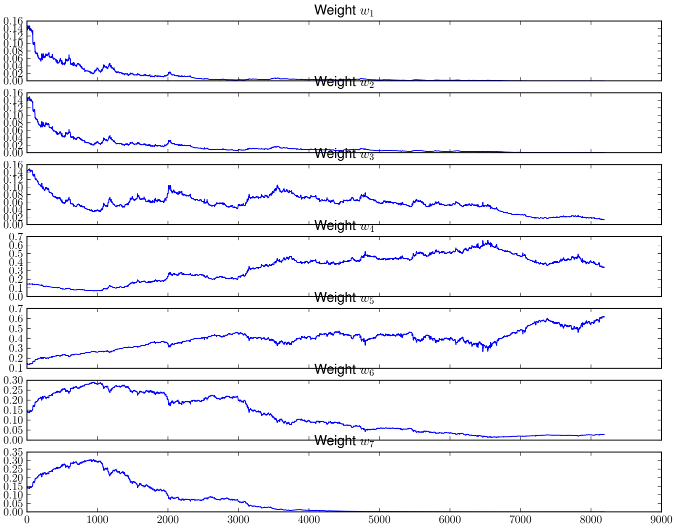
An example of computed weights :



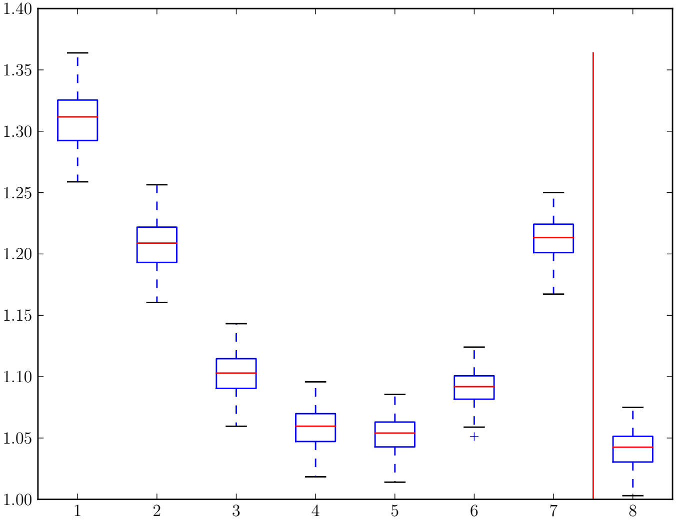
Obtained regrets :



An example of computed weights (not as many experts) :



Obtained regrets (not as many experts) :



References I

R. Dahlhaus. On the Kullback-Leibler information divergence of locally stationary processes. *Stochastic Process. Appl.*, 62(1):139–168, 1996. ISSN 0304-4149. doi: 10.1016/0304-4149(95)00090-9. URL http://dx.doi.org/10.1016/0304-4149(95)00090-9.

Rainer Dahlhaus and Liudas Giraitis. On the optimal segment length for parameter estimates for locally stationary time series. *J. Time Ser. Anal.*, 19(6):629–655, 1998. ISSN 0143-9782. doi: 10.1111/1467-9892.00114. URL http://dx.doi.org/10.1111/1467-9892.00114.

Christophe Giraud, François Roueff, and Andres Sanchez-Perez. Aggregation of predictors for nonstationary sub-linear processes and online adaptive forecasting of time varying autoregressive processes. *Ann. Statist.*, 43(6):2412–2450, 2015. ISSN 0090-5364. doi: 10.1214/15-AOS1345. URL http://dx.doi.org/10.1214/15-AOS1345.

References II

Hans Rudolf Künsch. A note on causal solutions for locally stationary ar-processes. Preprint ETH Zürich, 1995.

F. Roueff and A. Sánchez-Pérez. Prediction of weakly locally stationary processes by auto-regression. Technical report, LTCI, 2017. preprint available at [HAL].