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

1. TVAR processes
2. Prediction set-up
3. Numerical experiments (bias reduction in action)
4. Sequential aggregation of experts
5. Numerical experiments (aggregation in action)
6. References
1. TVAR processes
2. Prediction set-up
3. Numerical experiments (bias reduction in action)
4. Sequential aggregation of experts
5. Numerical experiments (aggregation in action)
6. References
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 $X_{t-1}, X_{t-2}, \ldots$

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 $X_{t-1}, X_{t-2}, \ldots$

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 $X_{t-1}, X_{t-2}, \ldots$

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 $X_{t-1}, X_{t-2}, \ldots$

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 $\theta_1, \ldots, \theta_p, \sigma^2$ 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 $\sigma^2 \equiv 1$, $p = 3$ and $T = 2^{12}$:
Same example, roots reciprocals:

Sample path

TVAR first pole

TVAR 2nd poles angle
Same example, local spectral density function:
TVAR processes

Prediction set-up

Numerical experiments (bias reduction in action)

Sequential aggregation of experts

Numerical experiments (aggregation in action)

References
Linear prediction for $L^2$ processes

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

$$
\gamma^*(t, T, \ell) = \text{cov}(X_t, T, X_{t-\ell}, T).
$$

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

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

with

$$
\Gamma^*_{t, T} = \text{Cov}(X_{t-1, T}) = (\gamma^*(t - i, T, j - i); i, j = 1, \ldots, d),
$$

and

$$
\gamma^*_{t, T} = \text{Cov}(X_{t-1, T}, X_{t, T}) = [\gamma^*(t, T, 1) \ldots \gamma^*(t, T, d)]'.
$$
Linear prediction for weakly locally stationary processes

Let \((X_t, T)_{t \in \mathbb{Z}, T \geq T_0}\) be a centered \((\beta, R)\)-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 \((X_t, \tau)_t \in \mathbb{Z}, \tau \geq T_0\) be a centered \((\beta, R)\)-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 \(u, \lambda\),
Linear prediction for weakly locally stationary processes

Let \((X_t, T)_{t \in \mathbb{Z}, T \geq T_0}\) be a centered \((\beta, R)\)-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 \(u, \lambda\),

Let

\[\theta_u = \Gamma_u^{-1} \gamma_u\]

where \(\Gamma_u\) and \(\gamma_u\) are the analogues of \(\Gamma_{t, T}^*\) and \(\gamma_{t, T}^*\), 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 \((\beta, R)\)-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 \(u, \lambda\),

Let

\[ \theta_u = \Gamma_u^{-1} \gamma_u \]

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

Then, we have

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

where \(C_1\) only depends on \(\beta, d, R\) 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])

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

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

$$H_M = \sum_{k=1}^{M} h^2(k/M) \sim M \int_0^{1} h^2(x) dx = M .$$
Estimation of autoregressive coefficients for weakly locally stationary

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

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

where \( h \) is supported on \([0, 1]\) and

\[
H_M = \sum_{k=1}^{M} h^2(k/M) \sim M \int_0^1 h^2(x)dx = M.
\]

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

\[
\hat{\Gamma}_{t,T,M} \hat{\theta}_{t,T}(M) = \hat{\gamma}_{t,T,M},
\]

where \( \hat{\Gamma}_{t,T,M} \) and \( \hat{\gamma}_{t,T,M} \) are obtained from \( \hat{\gamma}_{T,M}(t/T, \ell) \).
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 
\[\hat{\gamma}_{T,M}(u, \ell) - \mathbb{E}[\hat{\gamma}_{T,M}(u, \ell)]\] is decreasing at \(M^{-1/2}\)-rate in \(L^p\)-norm for all \(p \geq 2\). Then we have (Roueff and Sánchez-Pérez [2017])

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

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 
\(\hat{\gamma}_{T,M}(u, \ell) - \mathbb{E}[\hat{\gamma}_{T,M}(u, \ell)]\) is decreasing at \(M^{-1/2}\)-rate in \(L^p\)-norm for all \(p \geq 2\). Then we have (Roueff and Sánchez-Pérez [2017])

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

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

We then define, using appropriate \(\omega_\ell\)'s,

\[
\tilde{\theta}_{u,T}(M) = \frac{\theta_{t/T}}{T} + \sum_{\ell=0}^{k} \omega_\ell \hat{\theta}_{t,T}(M2^{\ell})
\]

so that

\[
\tilde{\theta}_{u,T}(M) = \frac{\theta_{t/T}}{T} + O_{L^p} \left( \left( \frac{M}{T} \right)^{\beta} + M^{-1/2} \right).
\]
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 
\[
\hat{\gamma}_{T,M}(u, \ell) - \mathbb{E} [\hat{\gamma}_{T,M}(u, \ell)]
\]
is decreasing at \(M^{-1/2}\)-rate in \(L^p\)-norm for all \(p \geq 2\). Then we have (Roueff and Sánchez-Pérez [2017])

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

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

We then define, using appropriate \(\omega_\ell\)'s,

\[
\tilde{\theta}_{u,T}(M) = \theta_{t/T} + \sum_{\ell=0}^{k} \omega_\ell \hat{\theta}_{t,T}(M2^\ell)
\]

so that

\[
\tilde{\theta}_{u,T}(M) = \theta_{t,T}^* + O_{L^p} \left( \left( \frac{M}{T} \right)^\beta + M^{-1/2} \right).
\]
1 TVAR processes

2 Prediction set-up

3 **Numerical experiments (bias reduction in action)**

4 Sequential aggregation of experts

5 **Numerical experiments (aggregation in action)**

6 References
Simulation setting

▷ Choose a model:
  ▷ we set \( p = 3 \),
  ▷ and pick random functions \( \theta_1, \ldots, \theta_p, \sigma^2 \) on \([0, 1]\).

▷ Monte Carlo loops:
  ▷ Generate \( X_{t,T} \) for \( t = 1, \ldots, T \) and \( T = 2^{10}, 2^{12}, \ldots, 2^{30} \).
  ▷ Compute \( \hat{\theta}_{T/2,T}(M) \) and \( \tilde{\theta}_{T/2,T}(M) \) for \( M = 2^{11}, \ldots, 2^{18} \).
  ▷ Compute the estimation quadratic errors

\[
\left\| \hat{\theta}_{T/2,T}(M) - \theta(1/2) \right\|^2
\]

and

\[
\left\| \tilde{\theta}_{T/2,T}(M) - \theta(1/2) \right\|^2
\]

over the Monte-Carlo runs.
Model parameters:
\[ T = 2^{20} \text{ (left)} \text{ and } T = 2^{30} \text{ (right)} \]
Optimal $M$ (oracle estimators).
1 TVAR processes
2 Prediction set-up
3 Numerical experiments (bias reduction in action)
4 **Sequential aggregation of experts**
5 Numerical experiments (aggregation in action)
6 References
Aggregation of $N$ predictors: a simple example

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

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

The aggregated predictor is defined by

$$\hat{X}_t = \sum_{i=1}^{N} \hat{\alpha}^{(i)}_t \hat{X}^{(i)}_t, \quad t = 1, \ldots, T.$$ 

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

$$\frac{1}{T} \sum_{t=1}^{T} (\hat{X}_t - X_t)^2 \leq \inf_{1 \leq i \leq N} \frac{1}{T} \sum_{t=1}^{T} (\hat{X}^{(i)}_t - X_t)^2 + \frac{\ln N}{\eta T}.$$
Setting the experts and aggregate them

Compute a finite number of autoregression estimators $\hat{\theta}_t, T_i$, $i = 1, \ldots, N$ which are minimax rate for well chosen smoothness indices $\beta_1, \ldots, \beta_N$.

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

What can be said about the obtained prediction error?
Setting the experts and aggregate them

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

$$\hat{X}^{(i)}_{t,T} = \hat{\theta}_{t,T}^{(i)} x_{t-1,T}.$$  

They will be our experts.
Setting the experts and aggregate them

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

- Compute the corresponding predictors

$$\hat{X}_{t,T}^{(i)} = \hat{\theta}_{t,T}^{(i)}' \mathbf{x}_{t-1,T}.$$ 

They will be our experts.

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

$$\hat{X}_{t,T}(A) = \sum_{i=1}^{N} \alpha_{i,t,T} \hat{X}_{t,T}^{(i)}.$$
Setting the experts and aggregate them

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

- Compute the corresponding predictors

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

They will be our experts.

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

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

- 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

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

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

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

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

- Lower bound: this is the minimax prediction rate over \( \theta \in s_p(\delta) \cap \Lambda_p(\beta, L) \) and \( \sigma \) valued in \([\sigma_-, \sigma_+]\).
1. TVAR processes
2. Prediction set-up
3. Numerical experiments (bias reduction in action)
4. Sequential aggregation of experts
5. Numerical experiments (aggregation in action)
6. References
Simulation setting

▷ Choose a model:
 ▷ we set $T = 2^{13}$ and $p = 5$,
 ▷ and pick random functions $\theta_1, \ldots, \theta_p, \sigma^2$ on $[0, 1]$.

▷ Monte Carlo loops:
 ▷ Generate $X_{t,T}$ for $t = 1, \ldots, T$.
 ▷ Compute $\hat{\theta}^{(i)}_{t,T}, \hat{X}^{(i)}_{t,T}, n = 1, \ldots, N$ and $\hat{X}_{t,T}(A)$ for $t = 1, \ldots, T$.
 ▷ Compute the regret

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

and

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

TVAR coeff. $\phi_1$

TVAR coeff. $\phi_2$

TVAR coeff. $\phi_3$

TVAR coeff. $\phi_4$

TVAR coeff. $\phi_5$
An example of computed weights:

<table>
<thead>
<tr>
<th>Weight ( w_i )</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.00 0.02 0.04 0.06 0.08 0.10 0.12</td>
</tr>
<tr>
<td>0.05 0.06 0.07 0.08 0.09 0.10 0.12</td>
</tr>
<tr>
<td>0.10 0.12 0.14 0.16 0.18 0.20 0.22</td>
</tr>
<tr>
<td>0.10 0.15 0.20 0.25 0.30 0.35 0.40</td>
</tr>
</tbody>
</table>

The graphs above illustrate the computed weights for each weight \( w_i \). Each graph shows the variation of the weight over different values, indicating how the weight changes with respect to the input data.
Obtained regrets:

![Box plot showing obtained regrets](image-url)
An example of computed weights (not as many experts):

For weights $w_1$ to $w_7$, the graphs show trends over a range of values, indicating how each weight might change or remain constant with respect to an input parameter or other factors.
Obtained regrets (not as many experts):

1.00 1.05 1.10 1.15 1.20 1.25 1.30 1.35 1.40
1. TVAR processes

2. Prediction set-up

3. Numerical experiments (bias reduction in action)

4. Sequential aggregation of experts

5. Numerical experiments (aggregation in action)

6. References

