Sequential model selection and adaptive robust efficiency for nonparametric autoregression

Workshop Analyse des Séries Temporelles

O. Arkoun & J.-Y. Brua

Laboratoire BIRL, Sup'Biotech and Laboratoire de Mathématiques Raphaël Salem, Université de Rouen Normandie

Joint work with S. Pergamenchtchikov (LMRS)

Valenciennes, June 23-24, 2025

Plan

Problem

We consider the following nonparametric autoregressive model

$$y_k = S(x_k) y_{k-1} + \xi_k \,, \quad 1 \le k \le n \,,$$

where $S(\cdot) \in \mathbf{L}_2[a,b]$ is unknown function, the design points

$$x_k = a + k \frac{b-a}{n} \,,$$

y_0 is a constant and the noise $(\xi_k)_{k \ge 1}$ are i.i.d. unobservable random variables with $\mathbb{E}\xi_1 = 0$ and $\mathbb{E}\xi_1^2 = 1$. We denote by p the distribution of ξ_1.

Goal : Estimate the autoregressive function S at any point.
Assumption : S belongs to the stability set

$$\Theta_{\epsilon,L} = \left\{ S \in C^1[a,b] \quad | \quad \|S\|_\infty \le 1 - \epsilon \quad \text{and} \quad \|\dot{S}\|_\infty \le L \right\}.$$

Previous works about autoregressive models

- [Borisov, V.Z. et Konev, V.V. (1977)] : parametric model, sequential estimation.
- [Belitser (2000)] : nonparametric model, recursive estimator, Lipschitzian class, quadratic risk.
- [Moulines et al. (2005)] : nonparametric heteroscedastic model, recursive estimator, convergence rate for the minimax risk and Hölderian class.
- [Arkoun (2011)] : nonparametric model, sequential estimator, optimal convergence rate for the minimax risk in the adaptive case and Hölderian class.
- [Arkoun and Pergamenchtchikov (2016)] :
  non adaptive case $\rightarrow$ robust efficiency ;
  adaptive case $\rightarrow$ optimal rate of convergence.

Risks

To measure the performance of an estimator $\widehat{S}_n$ of the drift S, we use

- the quadratic risk defined as

$$\mathcal{R}_p(\widehat{S}_n, S) = \mathbb{E}_{p,S}\|\widehat{S}_n - S\|^2\,, \quad \|S\|^2 = \int_a^b S^2(x)\mathrm{d}x\,,$$

  where and $\mathbb{E}_{p,S}$ is the expectation with respect to the distribution law $\mathbb{P}_{p,S}$ of the process $(y_k)_{1\le k\le n}$ given the coefficient S;
- the robust risk defined as

$$\mathcal{R}^*(\widehat{S}_n, S) = \sup_{p\in\mathcal{P}} \mathcal{R}_p(\widehat{S}_n, S)\,,$$

  where $\mathcal{P}$ is a family of distributions;
- and then the minimax risk defined as

$$\inf_{\widehat{S}_n} \sup_{S\in W_r^k} \mathcal{R}^*(\widehat{S}_n, S)$$

  where W_r^k is some Sobolev ball.

Plan

Kernel estimators

For the pointwise estimation, i.e. for the estimation of the $S(z)$ at some fixed point $a < z < b$ usually one uses the least quares estimator :

$$\widehat{S}_n(z) = \frac{\sum_{j=1}^{n} Q(u_j)\, y_{j-1}\, y_j}{\sum_{j=1}^{n} Q(u_j)\, y_{j-1}^2}$$

and

$$u_j = \frac{x_j - z}{h}\,.$$

Here Q is the kernel function $\mathbb{1}_{[-1,1]}$, with the bandwidth $h > 0$ which goes to zero as $n \to \infty$.

Sequential version

Key idea [Borisov, V.Z. et Konev, V.V. (1977)]

Replace the random denominator by some non random threshold $H > 0$ through the truncated sequential estimator

$$\widehat{S}_\tau(z) = \frac{\sum_{j=1}^{\tau} Q(u_j)\, y_{j-1}\, y_j}{\sum_{j=1}^{\tau} Q(u_j)\, y_{j-1}^2},$$

where

$$\tau = \inf\{k \geq 1 \,:\, \sum_{j=1}^{k} Q(u_j)\, y_{j-1}^2 \geq H\} \wedge n\,.$$

Sequential version

On the set $\Gamma = \{\tau < n\}$ we can obtain the following decomposition

$$\widehat{S}_\tau(z) = S(z) + B_H(z) + \eta_H(z)\,,$$

where the approximative term

$$B_H(z) = \frac{\sum_{j=1}^{\tau} \Delta_j(z)\, y_{j-1}^2}{\sum_{j=1}^{\tau} Q(u_j)\, y_{j-1}^2}\,, \quad \Delta_j(z) = S(x_j) - S(z)\,,$$

and

$$\eta_H(z) = \frac{\sum_{j=1}^{\tau} Q(u_j)\, y_{j-1}\xi_j}{\sum_{j=1}^{\tau} Q(u_j)\, y_{j-1}^2}\,.$$

Correction version

Similarly to [Arkoun (2011)] on the set $\Gamma = \{\tau < n\}$ we correct the sequential procedure by introducing some special coefficient for the last term, i.e.

$$\widehat{S}_\tau(z) = \frac{\sum_{j=1}^{\tau-1} Q(u_j)\, y_{j-1}\, y_j + \sqrt{\kappa}\, Q(u_\tau)\, y_{\tau-1}\, y_\tau}{H},$$

where the corrected coefficient $\kappa \in]0,1]$ is chosen as

$$\sum_{j=1}^{\tau-1} Q(u_j)\, y_{j-1}^2 + \kappa\, Q(u_\tau)\, y_{\tau-1}^2 = H\,.$$

Correction version

The decomposition for the correction sequential procedure

$$\widehat{S}_\tau(z) = S(z) + B_H(z) + \eta_H(z)\,,$$

where the approximative term

$$B_H(z) = \frac{\sum_{j=1}^{\tau-1} \Delta_j(z) Q(u_j)\, y_{j-1}^2 + \sqrt{\kappa}\, \Delta_j(z) Q(u_\tau)\, y_{\tau-1}^2}{H}$$

and

$$\eta_H(z) = \frac{\sum_{j=1}^{\tau} Q(u_j)\, y_{j-1}\xi_j + \sqrt{\kappa} y_{\tau-1}\xi_\tau}{H}\,.$$

Plan

A new grid on $[a, b]$

Similarly to [Arkoun and Pergamenchtchikov (2016)], in order to estimate in the quadratic metric, we use the following partition $(z_l)_{1 \le l \le d}$ defined as

$$z_l = a + \frac{l}{d}(b - a)$$

where $d = d_n = 2\left[\sqrt{n}/2\right] + 1$.
For each point z_l we construct an estimator of $S(z_l)$ using the truncated sequential kernel estimator defined above :

$$\widehat{S}_{\tau_l}(z_l) = \frac{\sum_{j=1}^{\tau_l - 1} Q(u_j)\, y_{j-1}\, y_j + \sqrt{\kappa_l}\, Q(u_{\tau_l})\, y_{\tau_l - 1}\, y_{\tau_l}}{H_l},$$

where $\tau_l = \tau(H_l)$ and $H_l > 0$ is the threshold depending on z_l.

Choice of the threshold

3 steps :

- We can show that : $\tau_l \approx (1 - S^2(z_l))\, H_l$.
- Separate the sample in two parts and use an auxiliary estimator $\widehat{S}_{\iota_l}$ of $S(z_l)$.
- Project $\widehat{S}_{\iota_l}$ onto $]-1+\varepsilon; 1-\varepsilon[$: $\widetilde{S}_{\iota_l}$

$$H_l = \frac{k_{2,l} - \iota_l}{1 - \widetilde{S}^2_{\iota_l}}.$$

Sequential estimation procedure :

$$\left(\widehat{S}_{\tau(H_l)}(z_l)\,,\ \tau_l\right), \text{with} \quad h = \frac{b-a}{2d}\,.$$

Transition to the regression model

Setting $\Gamma = \bigcap_{l=1}^{d} \{\tau_l < k_{2,l}\}$ and

$$Y_l = \widehat{S}_{\tau_l}(z_l)\mathbb{1}_\Gamma,$$

we obtain the following regression model on the set Γ :

$$Y_l = S(z_l) + \zeta_l$$

where

$$\zeta_l = B_{H_l}(z_l) + \eta_{H_l}(z_l) := B_l + \eta_l \,.$$

Asymptotic properties

Proposition

For any $\gamma > 0$

$$\lim_{n\to\infty} n^{1-\gamma} \sup_{p\in\mathcal{P}} \sup_{S\in\Theta_{\epsilon,L}} \mathbb{E}_{p,S} \max_{1\le l\le n} B_l^2 = 0\,.$$

Proposition

For any $b > 0$

$$\lim_{n\to\infty} n^{b} \sup_{p\in\mathcal{P}} \sup_{S\in\Theta_{\epsilon,L}} \mathbb{P}_{p,S}\left(\Gamma^c\right) = 0\,.$$

Plan

We use the model selection procedure proposed by [Galtchouk and Pergamenchtchikov (2011)] for heteroscedastic regression.
Let $(\phi_j)_{1\leq j\leq d}$ be an orthonormal family with respect to the empirical inner product, i.e.

$$(\phi_i\,,\,\phi_j)_d = \frac{b-a}{d}\sum_{l=1}^{d}\phi_i(z_l)\phi_j(z_l) = \mathbf{1}_{\{i=j\}}\,.$$

For example, we can take the trigonometric basis of $\mathbf{L}_2[a,b]$. For this basis we obtain

$$S(z_l) = \sum_{j=1}^{d}\theta_{j,d}\,\phi_j(z_l) \quad \text{and} \quad \theta_{j,d} = \Big(S,\phi_j\Big)_d\,.$$

Model selection procedure

To estimate the function S we have to estimate the Fourier coefficients $(\theta_{j,d})_{1\le j\le d}$.

$$\widehat{\theta}_{j,d} = \Big(Y, \phi_j\Big)_d\,, \quad Y = (Y_1, \dots, Y_d)'\,.$$

We obtain immediately the following regression scheme

$$\widehat{\theta}_{j,d} = \theta_{j,d} + \widetilde{\zeta}_{j,d}\,, \quad \text{where} \quad \widetilde{\zeta}_{j,d} = (\zeta\,, \phi_j)_d = \widetilde{B}_{j,d} + \sqrt{\frac{b-a}{d}}\,\widetilde{\eta}_{j,d}$$

with

$$\widetilde{B}_{j,d} = \frac{b-a}{d}\sum_{l=1}^{d} B_l\,\phi_j(z_l) \quad \text{and} \quad \widetilde{\eta}_{j,d} = \sqrt{\frac{b-a}{d}}\sum_{l=1}^{d}\eta_l\phi_j(z_l)\,.$$

Note that by the Bounyakovskii-Cauchy-Schwarz inequality we obtain that

$$|\widetilde{B}_{j,d}| \le \|B\|_d\,\|\phi_j\|_d = \|B\|_d\,.$$

Model selection procedure

We estimate the function S on the sieve by the weighted least squares estimator

$$\widehat{S}_\lambda(z_l) = \sum_{j=1}^{d} \lambda(j)\, \widehat{\theta}_{j,d}\, \phi_j(z_l)\, \mathbf{1}_\Gamma\,, \quad 1 \le l \le d\,,$$

where the weight vector $\lambda = (\lambda(1), \ldots, \lambda(d))'$ belongs to some finite set $\Lambda \subset [0, 1]^d$.

We set for any $a \le t \le b$

$$\widehat{S}_\lambda(t) = \sum_{l=1}^{d} \widehat{S}_\lambda(z_l) \mathbf{1}_{\{z_{l-1} < t \le z_l\}}\,.$$

Model selection procedure

In order to obtain a good estimator, we have to write a rule to choose a specific weight vector $\lambda \in \Lambda$.

We define the empirical squared risk as

$$\mathrm{Err}_d(\lambda) = \frac{d}{b-a} \|\widehat{S}_\lambda - S\|_d^2 \,.$$

We can rewrite this risk as

$$\mathrm{Err}_d(\lambda) = \sum_{j=1}^{d} \lambda^2(j) \widehat{\theta}_{j,d}^2 - 2 \sum_{j=1}^{d} \lambda(j) \widehat{\theta}_{j,d} \, \theta_{j,d} + \sum_{j=1}^{d} \theta_{j,d}^2 \,.$$

Model selection procedure

Since the coefficient $\theta_{j,d}$ is unknown, we need to replace the term $\widehat{\theta}_{j,d}\,\theta_{j,d}$ by one of its estimators which we choose as

$$\widetilde{\theta}_{j,d} = \widehat{\theta}^2_{j,d} - \frac{b-a}{d}\,\sigma^2_{j,d} \quad \text{with} \quad \sigma^2_{j,d} = \frac{b-a}{d}\sum_{l=1}^{d}\frac{1}{H_l}\,\phi_j^2(z_l)\,.$$

Note that

$$\sigma^2_{j,d} \le (b-a)\max_{1\le l\le d}\frac{1}{H_l} := \sigma_* \xrightarrow[n\to\infty]{} 0\,.$$

Cost function

Finally, we define the cost function of the form

$$J_d(\lambda) = \sum_{j=1}^{d} \lambda^2(j)\widehat{\theta}^2_{j,d} - 2\sum_{j=1}^{d} \lambda(j)\,\widetilde{\theta}_{j,d} + \delta P_d(\lambda)\,,$$

where the penalty term is defined as

$$P_d(\lambda) = \frac{b-a}{d}\sum_{j=1}^{d} \lambda^2(j)\sigma^2_{j,d}\,.$$

Minimizing the cost function

$$\widehat{\lambda} = \operatorname*{argmin}_{\lambda\in\Lambda} J_d(\lambda)\,,$$

yields the model selection procedure

$$\widehat{S}_* = \widehat{S}_{\widehat{\lambda}}\,.$$

Plan

Oracle inequalities

Theorem (A., B., Pergamenchtchikov (2019))

For any function S, *for any* $n \geq 3$, $p \in \mathcal{P}$ *and* $0 < \delta \leq 1/12$

$$\mathbb{E}_{p,S} \| \widehat{S}_* - S \|_d^2 \leq \frac{1+4\delta}{1-6\delta} \min_{\lambda \in \Lambda} \mathbb{E}_{p,S} \| \widehat{S}_\lambda - S \|_d^2 + \nu \frac{U_n(S)}{\delta n},$$

where $\nu = \#\Lambda$ *and* $U_n(S)$ *is such that for any* $\gamma > 0$, $\lim_{n \to \infty} \frac{U_n(S)}{n^\gamma} = 0$.

Theorem (A., B., Pergamenchtchikov (2019))

For any function $S \in \Theta_{\epsilon,L}$, *any* $n \geq 3$ *and* $0 < \delta \leq 1/12$

$$\mathcal{R}^*(\widehat{S}_*, S) \leq (1 + G(\delta)) \min_{\lambda \in \Lambda} \mathcal{R}^*(\widehat{S}_\lambda, S) + \nu \frac{\widetilde{U}_n(S)}{\delta n},$$

where $\lim_{\delta \to 0} G(\delta) = 0$ *and* $\widetilde{U}_n(S)$ *is such that for any* $\gamma > 0$, $\lim_{n \to \infty} \frac{\widetilde{U}_n(S)}{n^\gamma} = 0$.

Plan

Minimax framework

For S belonging to some space W, we consider the minimax risk :

$$\inf_{\widetilde{S}_n} \sup_{S \in W} \nu(S)\ \mathcal{R}^*(\widetilde{S}_n, S)$$

Goal 1 : Find the convergence rate φ_n and a constant $c > 0$ s.t.

$$\liminf_{n\to\infty} \varphi_n \inf_{\widetilde{S}_n} \sup_{S \in W} \nu(S)\ \mathcal{R}^*(\widetilde{S}_n, S) \geq c.$$

Goal 2 : Find an asymptotically efficient estimator $\widehat{S}$ s.t.

$$\limsup_{n\to\infty} \varphi_n \sup_{S \in W} \nu(S)\ \mathcal{R}^*(\widehat{S}, S) \leq c.$$

The Sobolev ball

We consider the Sobolev ball

$$W = W_{k,r} = \left\{ S \in \Theta_{\epsilon,L} \mid \sum_{j=0}^{k} \| S^{(j)} \|_2^2 \leq r \right\},$$

which can be represented as

$$W_{k,r} = \left\{ S \in \Theta_{\epsilon,L} \mid \sum_{j=1}^{\infty} a_j \, \theta_j^2 \leq r \right\}$$

where

$$a_j = \sum_{i=0}^{k} \| \phi_j^{(i)} \|_2^2,$$

and

$$\theta_j = (S, \phi_j) = \int_a^b S(x) \, \phi_j(x) \mathrm{d}x.$$

Lower bound

We set the normalizing coefficient

$$\nu(S) = \eta_*(S)^{-2k/(2k+1)}, \quad \text{where} \quad \eta_*(S) = \int_a^b \left(1 - S^2(u)\right) \mathrm{d}u.$$

Theorem (A., B., Pergamenchtchikov (2024))

The robust minimax risk normalized by the coefficient is bounded from below as

$$\liminf_{n\to\infty} n^{2k/(2k+1)} \inf_{\widetilde{S}_n} \sup_{S\in W_{k,r}} \nu(S)\, \mathcal{R}^*(\widetilde{S}_n, S) \geq c_{k,r},$$

where

$$c_{k,r} = ((2k+1)r)^{1/(2k+1)} \left(\frac{k(b-a)}{\pi(k+1)}\right)^{2k/(2k+1)}.$$

Specification of weights

To obtain the efficiency property we take the weights proposed in [Pinsker (1980)]. For some fixed $k^* \geq 1$ and $\varepsilon > 0$ we defined the following set

$$\mathcal{A} = \{1, \ldots, k^*\} \times \{r_1, \ldots, r_m\} ,$$

where $r_i = i\varepsilon$ and $m = [\varepsilon^{-2}]$. For a couple $\alpha = (\ell, r_i) \in \mathcal{A}$, we introduce the weight sequence

$$\lambda_\alpha = (\lambda_\alpha(j))_{1 \leq j \leq d}$$

with the elements

$$\lambda_\alpha(j) = \mathbf{1}_{\{1 \leq j < j_*\}} + \left(1 - (j/\omega_\alpha)^\ell\right) \mathbf{1}_{\{j_* \leq j \leq \omega_\alpha\}} .$$

Now we define the set Λ as

$$\Lambda = \{\lambda_\alpha \, , \, \alpha \in \mathcal{A}\} .$$

Optimal weights

In this case $\nu = \text{card}(\Lambda) = k^* m$, i.e. ν is the function of n such that

$$\lim_{n\to\infty} \frac{\nu}{n^\gamma} = 0 \quad \text{for any} \quad \gamma > 0\,.$$

Theorem

For any function $S \in \Theta_{\epsilon,L}$, $n \geq 3$ *and* $0 < \delta \leq 1/12$

$$\mathcal{R}^*(\widehat{S}_*, S) \leq (1 + G(\delta)) \min_{\lambda\in\Lambda} \mathcal{R}^*(\widehat{S}_\lambda, S) + \nu \frac{\widetilde{U}_n(S)}{\delta n}\,,$$

where $G(\delta) \to 0$ *as* $\delta \to 0$ *and the term* $\widetilde{U}_n(S)$ *is such that*

$$\lim_{n\to\infty} \sup_{S\in W_r^k} \frac{\widetilde{U}_n(S)}{n^\gamma} = 0 \quad \textit{for any} \quad \gamma > 0\,.$$

Choice of δ

We choose now in the model selection procedure the parameter $\delta = \delta_n$ as a slowly decreasing function of n, i.e.

$$\lim_{n\to\infty} \delta_n = 0 \quad \text{and} \quad \lim_{n\to\infty} n^{\gamma}\,\delta_n = +\infty \quad \text{for any} \quad \gamma > 0\,.$$

For example, $\delta_n = (12 + \ln(n))^{-1}$.

For any $k \geq 1$ the oracle inequality implies

$$n^{\frac{2k}{2k+1}}\,\mathcal{R}^*(\widehat{S}_*, S) \leq \,(1 + \mathrm{o}(1))\; n^{\frac{2k}{2k+1}}\,\min_{\lambda\in\Lambda} \mathcal{R}^*(\widehat{S}_\lambda, S) + \,\mathrm{o}(1)\,.$$

Upper bound

Theorem (A., B., Pergamenchtchikov (2024))

The model selection procedure estimator $\widehat{S}_*$ *constructed with the previous choices of* δ *and* Λ *satisfies*

$$\limsup_{n\to\infty} n^{2k/(2k+1)} \sup_{S\in W_{k,r}} \upsilon(S)\, \mathcal{R}^*(\widehat{S}_*, S) \le c_{k,r}.$$

Finally, $\widehat{S}_*$ is an asymptotically efficient estimator of the autoregressive function S :

- for the robust integral quadratic risk,
- and over the Sobolev ball $W_{k,r}$.

Plan

Perspectives

Consider the p-order autoregressive model for $p > 1$,

$$y_k = S_1(x_k)y_{k-1} + S_2(x_k)y_{k-2} + \cdots + S_p(x_k)y_{k-p} + \xi_k, \quad k = 1, \ldots, n$$

where $x_k = \frac{k}{n}$, $S : [0,1] \to \mathbb{R}^p$ and ξ_k is the noise.

- Pointwise estimation, absolute error, Höderian classes
- Adaptive case
- Quadratic risk, Sobolev classes

Thank you for your attention !

References

Arkoun, O. (2011) : Sequential adaptive estimators in nonparametric autoregressive models. *Sequential Anal.* **30** (2), 228-246.

Arkoun, O., Brua J.-Y. and Pergamenchtchikov, S. (2019) : Sequential model selection method for nonparametric autoregression. *Sequential Anal.* **38** (4), 437-460.

Arkoun, O., Brua, J.-Y. and Pergamenchtchikov, S. (2024) : Adaptive efficient robust sequential analysis for autoregressive big data models. *Sequential Anal.* **43** (4), 432-460.

Arkoun, O. and Pergamenchtchikov, S. (2016) : Sequential robust estimation for nonparametric autoregressive models. *Sequential Anal.* **35** (4), 489-515.

Belitser, E. (2000) : Recursive estimation of a drifted autoregressive parameter. *Ann. Statist.* **28** (3), 860-870.

Borisov, V.Z. and Konev, V.V. (1977) : Sequential Estimation of Parameters of Discrete Processes. *Automat. and Remote control* **10**, 58-64.

Dahlhaus, R. (1996) : Maximum Likelihood Estimation and Model Selection for Locally Stationary Processes. *J. Nonparametr. Stat.* **6**, 171-191.

Galthouk, L. and Pergamenshchikov, S. M. (2011) : Adaptive sequential estimation for ergodic diffusion processes in quadratic metric. *J. Nonparametr. Stat.* **23** (2), 255-285.

Moulines, E., Priouret P. and Roueff, F. (2005) : On recursive estimation for time varying autoregressive processes. *Ann. Statist.* **33** (6), 2610-2654.

Pinsker, M. S. (1980) : Optimal Filtering of Square-Integrable Signals in Gaussian Noise. *Problemy Peredachi Informatsii* **16** (2), 52-68.