Nonparametric estimation for i.i.d. Gaussian continuous time moving average models

Fabienne Comte (1), Valentine Genon-Catalot (1)

(1) UMR CNRS 8145, Laboratoire MAP5, Université de Paris

Meeting in Mathematical Statistics 2020 - December 14-18th

CMA

Continuous time Moving Average process

$$X(t) = \int_0^t a(t-s)dW(s) \quad (1)$$

$(W(t))$ Wiener process

$a(t) \in \mathbb{L}^2(\mathbb{R}^+)$

Aim:
Non parametric estimation of unknown function $g = a^2$ from *i.i.d.* $(X_i(t), t \in [0, T], i = 1, \dots, N)$ distributed as $(X(t))$, continuously observed.

Assumption: $g(t) = a^2(t) \in \mathbb{L}^2(\mathbb{R}^+)$

Classical Examples

- For $a(t) = \sigma e^{-\theta t}$, $X(t)$ is the classical Ornstein-Uhlenbeck process.
- CARMA(p, q) processes, see *e.g.* Bergstrom, 1990 *Continuous time econometric modelling*.
- Brockwell, P. J. Continuous-time ARMA processes. Stochastic processes: theory and methods, (2001)
- Comte, F. and Renault, E. (1996): Long memory continuous time models
  In particular, in this paper, the continuous time fractional Brownian motion of order d is defined by:

$$X(t) = \int_0^t \frac{(t-s)^d}{\Gamma(d+1)} \tilde{a}(t-s) dW(s)$$

where $\tilde{a}(t) \in C^1([0, +\infty))$ and $-1/2 < d < 1/2$ $(H = 2d - 1)$

Stationary version

Generally, authors rather deal with the stationary version

$$Y(t) = \int_{-\infty}^{t} a(t-s)dW(s) \quad \text{or} \quad Y(t) = \int_{-\infty}^{t} a(t-s)dL(s)$$

where $(L(s))$ is a Lévy process.

- Schnurr, Woerner (2011) (Well-balanced Lévy driven Ornstein-Uhlenbeck process): properties of the model
- Brockwell, Ferrazzano, Kluppelberg (2013), (High-frequency sampling and kernel estimation for CMA processes): pointwise estimation of $a(t)$ (the kernel) based on one sample path $Y(t)$, with discrete-time obvervation at times $t_i = i\Delta/n$, $i = 1, \dots, n$, in high-frequency framework. Only for the special case of CARMA Gaussian models or CARMA(1,0) Lévy model.
- Belomestny, Panov, Woerner (2016), (Low frequency estimation of continuous-time moving average Lévy processes): estimation of the Lévy triplet of $(L(t))$.

Case of *i.i.d.* observations.

No references (to our knowledge) for the estimation of $a(t)$ based on the observation of $(X_i(t), i = 1, \dots, N)$ distributed as $X(t) = \int_0^t a(t-s)dW(s)$.

Two cases to be distinguished:

1. $a(t) \in C^1(\mathbb{R}^+)$. Then, (see Comte and Renault, 1996)

$$dX(t) = a(0)dW_t + \left[\int_0^t a'(t-s)dW(s)\right] dt.$$

   For $a(0) = 0$, $(X(t))$ is differentiable.

2. $a(t) \notin C^1(\mathbb{R}^+)$, maybe not defined at 0 ($e.g.$ $a(t) = t^d \tilde{a}(t)$, $d < 0$, $\tilde{a}(t) \in C^1(\mathbb{R}^+)$)

$\Rightarrow$ two different estimation methods for cases (1) and (2).

Estimation method

Idea: Projection method

Consider a sequence (S_m) of finite-dimensional subspaces of $\mathbb{L}^2([0, +\infty))$ and estimate g_m the orthogonal projection of g onto S_m.

Generic notation, the orthonormal basis $(\varphi_j, j \geq 0)$ of $\mathbb{L}^2(\mathbb{R}^+)$:

$$g = \sum_{j \geq 0} \theta_j \varphi_j, \quad \theta_j = \int_0^{+\infty} g(x)\varphi_j(x)dx := \langle g, \varphi_j \rangle.$$

For $S_m = \mathrm{span}(\varphi_0, \ldots, \varphi_{m-1})$, we have

$$g_m = \sum_{j=0}^{m-1} \theta_j \varphi_j$$

and we estimate g_m by defining estimators of the $\theta_j, j = 0, \ldots, m-1$ based on $(X_i(t), t \in [0, T], i = 1, \ldots, N)$.

Two bases

Two settings:

- T is fixed, estimation of $g_T = g\mathbf{1}_{[0,T]}$. We consider the collection $(S_m = S_m^{Trig})$ of finite-dimensional subspaces of $\mathbb{L}^2([0,T])$ where (S_m^{Trig}) has odd dimension and is generated by the **orthonormal trigonometric basis** $(\varphi_{j,T})$ where $\varphi_{0,T}(t) = \sqrt{1/T}\mathbf{1}_{[0,T]}(t)$, and for $j = 1, \dots, (m-1)/2$,

$$\varphi_{2j-1,T}(t) = \sqrt{\frac{2}{T}}\cos(2\pi\frac{jt}{T})\mathbf{1}_{[0,T]}(t),\ \varphi_{2j,T}(t) = \sqrt{\frac{2}{T}}\sin(2\pi\frac{jt}{T})\mathbf{1}_{[0,T]}(t),$$

- T can de large, estimation of g using the collection $(S_m = S_m^{Lag})$ of finite-dimensional subspaces of $\mathbb{L}^2([0,+\infty))$ where $S_m^{Lag} = \mathrm{span}\{\ell_j, j = 0, \dots, m-1\}$ and $(\ell_j)_{j\geq 0}$ is the **Laguerre basis**:

$$\ell_j(t) = \sqrt{2}L_j(2t)e^{-t}\mathbf{1}_{t\geq 0}, \quad j \geq 0, \quad L_j(t) = \sum_{k=0}^{j}(-1)^k\binom{j}{k}\frac{t^k}{k!}.$$

Estimation when $a(t)$ is differentiable.

Lemma

If $a(t)$ is differentiable,

$$\mathbb{E}\left(\int_0^{+\infty}\varphi_j(s)X(s)dX(s)\right)=\frac{1}{2}\left(\theta_j-g(0)\int_0^{+\infty}\varphi_j(s)ds\right),\quad \theta_j=\langle g,\varphi_j\rangle$$

$$\hat{\theta}_j=\hat{\theta}_j(N,T)=2\left[\frac{1}{N}\sum_{i=1}^{N}\left(\int_0^T\varphi_j(s)X_i(s)dX_i(s)\right)\right]+g(0)\int_0^T\varphi_j(s)ds.$$

$$\hat{g}_m=\sum_{j=0}^{m-1}\hat{\theta}_j\varphi_j$$

Remark

$dX(t)=a(0)dW(t)+\left[\int_0^t a'(t-s)dW(s)\right]dt \Rightarrow g(0)=a^2(0)$ *known from continuous observation of* $X_i(t), t\in[0,T]$)

Risk bounds for estimator $\hat{g}_m$ using trigonometric basis

Assumptions.
[H0] The function $g(t) = a^2(t)$ belongs to $\mathbb{L}^1(\mathbb{R}^+) \cap \mathbb{L}^2(\mathbb{R}^+)$.
[H1] The function $a(t)$ belongs to $C^1(\mathbb{R}^+)$, is bounded and $\int_0^{+\infty} [a'(t)]^2 dt < +\infty$.

$\|u\|_T^2 = \int_0^T u^2(s)ds$.

Proposition

Assume **[H0]-[H1]**. *For* $(\varphi_j = \varphi_{j,T})$ *trigonometric basis,* $a(t)$ *differentiable and* $\int_0^{+\infty} (a'(t))^2 dt < +\infty$,

$$\mathbb{E}(\|\hat{g}_m - g\|_T^2) \leq \|g_m - g\|_T^2 + 8g(0)G(T)\frac{m}{N} + \frac{c(T)}{N}$$

where $G(T) = \int_0^T g(s)ds \leq \int_0^{+\infty} g(s)ds < +\infty$, $g_m = \sum_{j=0}^{m-1} \theta_j \varphi_{j,T}$.
If $g(0) = 0$,

$$\mathbb{E}(\|\hat{g}_m - g\|_T^2) \leq \|g_m - g\|_T^2 + \frac{c(T)}{N}$$

Risk bounds for estimator $\hat{g}_m$ using Laguerre basis

$\|u\|^2 = \int_0^{+\infty} u^2(s)ds.$

Proposition

Assume **[H0]-[H1]**. *If* (φ_j) *is an orthonormal basis of* $\mathbb{L}^2(\mathbb{R}^+)$, *for all* $T \geq 1$, $N \geq 1$, $m \geq 1$, *we have*

$$\mathbb{E}(\|\hat{g}_m - g\|^2) \quad \leq \quad \|g_m - g\|^2 + 8g(0)G(T)\frac{m}{N} + c_G\frac{T}{N} + 2\int_T^{+\infty} g^2(s)ds + \frac{4\|g\|^2}{N}.$$

If in addition $g(0) = 0$,

$$\mathbb{E}(\|\hat{g}_m - g\|^2) \leq \|g_m - g\|^2 + c'_G\frac{T}{N} + \int_T^{+\infty} g^2(s)ds + \frac{2\|g\|^2}{N} \qquad (2)$$

If (φ_j) *is the Laguerre basis of* $\mathbb{L}^2(\mathbb{R}^+)$, g *is bounded and* $T \geq 6m - 3$, *then*

$$\mathbb{E}(\|\hat{g}_m - g\|^2) \leq \|g_m - g\|^2 + C\frac{m^2}{N} + C'\|a\|^2 m \exp(-12\gamma_2 m) \qquad (3)$$

Comments on Risk bounds for estimator $\hat{g}_m$

- Risk bounds contain the usual bias term $\|g_m - g\|^2$, bias of the projection method (small when m large).
- Variance terms $\frac{m}{N}, \frac{m^2}{N}$ which balance the bias terms (large for large m).
- If $g(0) = 0$, m should be taken as large as possible.
- Remainder terms: $O(1/N)$
- More puzzling term:

$$c'_G \frac{T}{N} + \int_T^{+\infty} g^2(s)ds$$

for large T (T should not be too large).

Estimation in the general case.

$a(t)$ may be not differentiable

Lemma

Assume that **[H0]** *holds and* $(\varphi_j)_j$ *is differentiable on* $[0,T]$, *then*

$$\mathbb{E}\left(\int_0^T \varphi_j'(s)X^2(s)ds\right) = \varphi_j(T)G(T) - \int_0^T g(u)\varphi_j(u)du.$$

Therefore, we set

$$\tilde{\theta}_j = -\frac{1}{N}\sum_{i=1}^N\left(\int_0^T \varphi_j'(s)X_i^2(s)ds\right) + \varphi_j(T)\widehat{G}(T) \quad \text{and} \quad \widehat{G}(T) = \frac{1}{N}\sum_{i=1}^N X_i^2(T).$$

$$\tilde{g}_m = \sum_{j=0}^{m-1}\tilde{\theta}_j\varphi_j.$$

Estimation in the general case.

Proposition

Assume that **[H0]** *holds.*

- *If* $(\varphi_j = \varphi_{j,T})$ *the trigonometric basis, then*

$$\mathbb{E}(\|\tilde{g}_m - g\|_T^2) \leq \|g_m - g\|_T^2 + c_G\left(\frac{m^2}{NT} + \frac{m}{NT}\right).$$

- *Let* $(\varphi_j = \ell_j)$ *be the Laguerre basis,*
  - *if* $\int_0^1 \frac{G^2(s)}{s} ds = c_0 < +\infty$,

$$\mathbb{E}(\|\tilde{g}_m - g\|^2) \leq \|g_m - g\|^2 + c'_G \log(T)\frac{m}{N} + c'_G\frac{T}{N} + \int_T^\infty g^2(s)ds$$

  - *If* $T \geq 6m - 3$

$$\mathbb{E}(\|\tilde{g}_m - g\|^2) \leq \|g_m - g\|^2 + c_1\frac{m^3}{N} + c_2 m \exp(-12\gamma_2 m)$$

where c_1, c_2, γ_2 *are constants depending on the basis only.*

Comments.

Tools.

- $\mathbb{E}(X^2(t)) = G(t) = \int_0^T g(s)ds$
  $\Rightarrow$ Moments of $X(t)$ provide information on $g = a^2$.
  Contrary to stationary $Y(t) = \int_{-\infty}^t a(t-s)dW(s)$ where
  $\mathbb{E}(Y^2(t)) = \int_0^{+\infty} g(s)ds$
- Properties of Laguerre functions (especially Askey & Wainger, 1965)

Consequences.

Rates of convergence for the $\mathbb{L}^2$-risks may be deduced form the previous risks bounds for g belonging to a specified function space (to assess the bias rate).
However, we know nothing about optimal rates for this problem.

Adaptive procedure. General case.

For $(B) = (Lag)$ (Laguerre basis) or $(B) = (Trig)$ (trigonometric basis), we define a data-driven choice of m (mimicks the squared bias-variance compromise) by

$$\widetilde{m}^{(B)} := \arg \min_{m \in \mathcal{M}_N^{(B)}} \left(-\|\tilde{g}_m\|^2 + \mathrm{pen}^{(B)}(m) \right),$$

where **Laguerre case**:

$$\mathcal{M}_N^{(Lag)} = \{m \in \mathbb{N}, m \leq \frac{N}{\log(T)}\}, \quad \mathrm{pen}^{(Lag)}(m) = \kappa \log(N) \frac{m}{N} \left(G^2(T) + \int_0^T \frac{G^2(u)}{u} du \right),$$

Trigonometric case:

$$\mathcal{M}_N^{(Trig)} = \{m \in \mathbb{N}, m^2 \leq N\}, \quad \mathrm{pen}^{(Trig)}(m) = \kappa \log(N) \frac{m^2}{NT} G^2(T),$$

- $-\|\tilde{g}_m\|^2$ provides an estimation of the squared bias, up to a constant.
- $\mathrm{pen}^{(B)}(m)$ has the variance order, up to the $\log(N)$ factor.
- κ a numerical constant.

Risk bounds for the adaptive estimator.

Theorem

There exists a numerical value $\kappa_0^{(B)} > 0$ *such that* $\forall \kappa \geq \kappa_0^{(B)}$,

$$\mathbb{E}(\|\tilde{g}_{\widetilde{m}^{(B)}} - g\|^2) \leq 4 \inf_{m \in \mathcal{M}_N^{(B)}} \left(\|g - g_m\|^2 + \mathrm{pen}^{(B)}(m) \right) + C^{(B)}(T, N),$$

where $C^{(B)}(T, N)$ *remainder terms:*

$$C^{(Lag)}(T, N) = 32 G^2(T) \frac{T \log(N)}{N} + \int_T^{+\infty} g^2(s) ds + \frac{C}{N} \left(\frac{T G^2(T)}{N} + \int_0^T \frac{G^2(u)}{u} du \right)$$

$$C^{(Trig)}(T, N) = \frac{C}{N} \left(\frac{1}{T} + \frac{G^2(T)}{T} \frac{1}{N^{1/2}} \right)$$

- $\tilde{g}_{\widetilde{m}^{(B)}}$ is adaptive in the sense that its $\mathbb{L}^2$-risk automatically achieves the best compromise between squared bias and variance terms, up to remainder terms.
- **Tool**: Deviation inequality for χ^2- variables.

Risk bounds for the adaptive estimator (continued)

Penalties contain unknown quantities:

$$G^2(T) + \int_0^T \frac{G^2(u)}{u} du, \quad G^2(T)$$

The term $G(T) = \int_0^T g(s)ds$ is unknown. In practical implementation, we set

$$\widehat{G^2(T)} = \frac{1}{3N} \sum_{i=1}^{N} X_i^4(T), \quad \int_0^T \frac{G^2(u)}{u} du \to \log T$$

(unbiased estimator, order)
For the implementation of the procedure, we have to fix the constants κ in the penalties . The choice of κ in the penalties is standardly calibrated by preliminary simulations.

Simulation study

To simulate an exact discrete sampling of $(X_i(t), i = 1, \ldots, N$ with small sampling interval Δ, with $T = n\Delta$, we use:

Vectors $(X_i(k\Delta), k = 1, \ldots, n)'$ are *i.i.d.* centered Gaussian vectors with covariance matrix $A = (A_{j,k})$ where for $1 \leq j \leq k$,

$$A_{j,k} = \int_0^{j\Delta} a(j\Delta - u)a(k\Delta - u)du = \int_0^{j\Delta} a(v)a((k-j)\Delta + v)dv.$$

Integrals in the estimators formulae are discretized.
Settings for (N, T): $T = n\Delta = 10$, $n = 400, \Delta = 0.1/4$ with $N = 500, 8000$.
Functions $a(.)$:

1. $a_3(t) = (\frac{1}{2}\beta(3, 3, t/3) + \frac{1}{2}\beta(3, 3, t/3 - 2))^{1/2}$.
2. $a_6(t) = t^{0.25}e^{-t/3}$.
3. $a_7(t) = t^{-0.125}e^{-t/5}$.

Simulation study (continued)

We implemented method 1 (less general) and method 2 for the two bases

- Trigonometric
- Laguerre

Results for Laguerre basis are better then for trigonometric basis for most tested functions (except trigonometric ones).

For method 1 ($\hat{g}_m$), dimension was selected as follows: for $(B) = (Lag), (Trig)$,

$$\hat{m}^{(B)} = \arg \min_{1 \le m \le D_{\max}} \left\{ -\|\hat{g}_m^{(B)}\|^2 + \kappa_1^{(B)} g^{\dagger}(0) \widehat{G(T)} \frac{m}{N} \right\}.$$

We compute $(\hat{g}_m^{(Trig)})_{1 \le m \le D_{\max}}$ and $(\hat{g}_m^{(Lag)})_{1 \le m \le D_{\max}}$ $\widehat{G}(T) = (1/N) \sum_{i=1}^N X_i^2(T)$,
$g^{\dagger}(0)$ is computed using the quadratic variation
No theoretical study.

For method 2: $(\tilde{g}_m^{(Trig)})_{1\leq m\leq D_{\max}}$ and $(\tilde{g}_m^{(Lag)})_{1\leq m\leq D_{\max}}$ analogously,

$$\tilde{m}^{(Trig)} = \arg\min_m \left\{ -\|\tilde{g}_m^{(Trig)}\|^2 + \kappa_2^{(Trig)} \widehat{G^2(T)} \log(N) \frac{m^2}{NT} \right\}$$

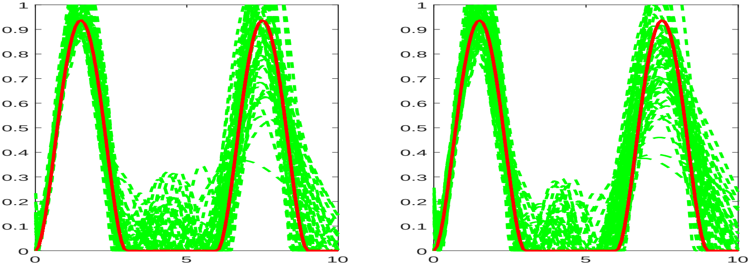
and

$$\tilde{m}^{(Lag)} = \arg\min_m \left\{ -\|\tilde{g}_m^{(Lag)}\|^2 + \kappa_2^{(Lag)} \widehat{G^2(T)} \log(T) \log(N) \frac{m}{N} \right\}$$

Based on preliminary simulations, the constants are calibrated once and for all to the following values

$$\kappa_1^{(Lag)} = 27, \quad \kappa_1^{(Trig)} = 6, \quad \kappa_2^{(Lag)} = 0.11 \text{ and } \kappa_2^{(Trig)} = 0.6.$$

Simulation study (continued)

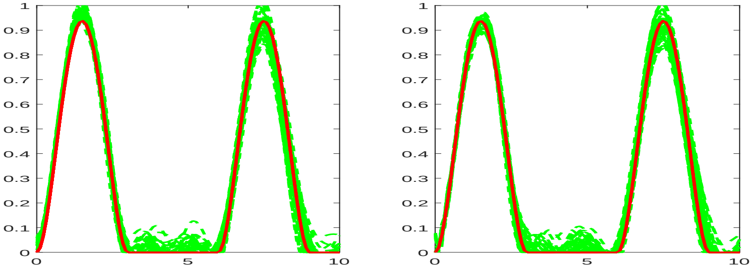


$N = 500$

MISE $= 0.15$, dim $= 25$ MISE $= 0.19$, dim $= 16$

Figure: Examples of 50 estimated curves (in green) using the Laguerre basis, with the two methods (method 1 left, method 2 right) for functions 3 , for $N = 500$. The bold curve is the true function. Under each plot, the MISE over the 50 paths, and the mean of the selected dimensions.

Simulation study (continued)

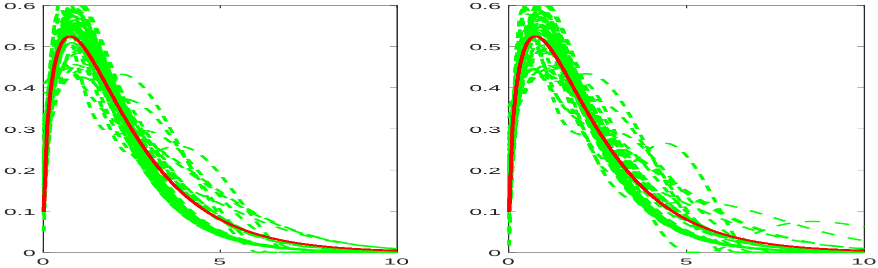


$N = 8000$

MISE $= 0.015$, dim $=39$ MISE$=0.013$, dim $= 31$

Figure: Examples of 50 estimated curves (in green) using the Laguerre basis, with the two methods (method 1 left, method 2 right) for functions 3 , for $N = 8000$. The bold curve is the true function. Under each plot, the MISE over the 50 paths, and the mean of the selected dimensions.

Simulation study (continued)

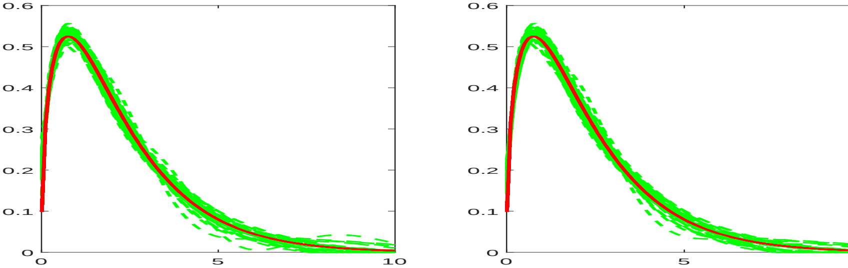


$N = 500$

MISE $= 0.012$, dim=2.5   MISE=0.013, dim= 2.8

Figure: Examples of 50 estimated curves (in green) using the Laguerre basis, with the two methods (method 1 left, method 2 right) for function 6, for $N = 500$. The bold curve is the true function. Under each plot, the MISE over the 50 paths, and the mean of the selected dimensions.

Simulation study (continued)

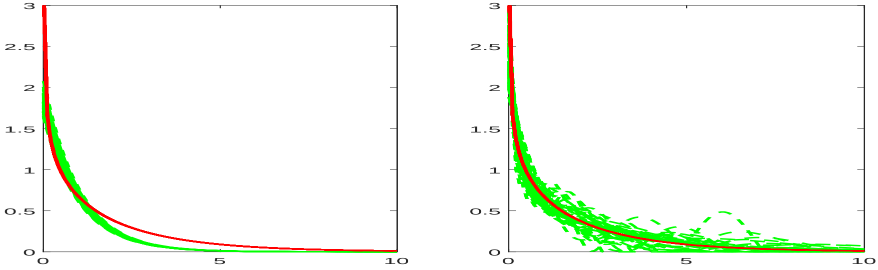


$N = 8000$

MISE $= 0.0024$ dim$= 4.4$ MISE$=0.0018$, dim$=4.4$

Figure: Examples of 50 estimated curves (in green) using the Laguerre basis, with the two methods (method 1 left, method 2 right) for function 6 and $N = 8000$. The bold curve is the true function. Under each plot, the MISE over the 50 paths, and the mean of the selected dimensions.

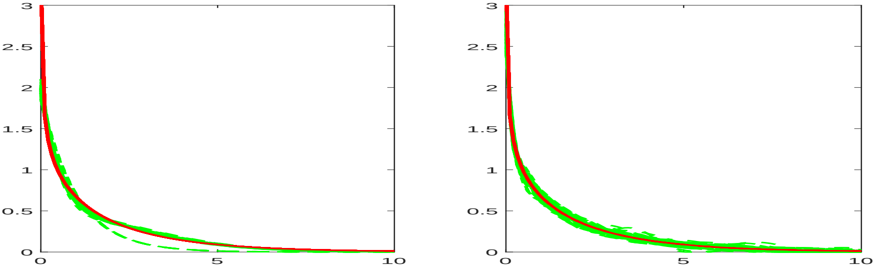
Simulation study (continued)



$N = 500$

MISE =0.25, dim= 1 MISE=0.13, dim=5

Figure: Examples of 50 estimated curves (in green) using the Laguerre basis, with the two methods (method 1 left, method 2 right) for function 7, for $N = 500$. The bold curve is the true function. Under each plot, the MISE over the 50 paths, and the mean of the selected dimensions.



$N = 8000$

MISE $=0.16$, dim $= 2.8$ MISE $= 0.035$, dim $=11$

Figure: Examples of 50 estimated curves (in green) using the Laguerre basis, with the two methods (method 1 left, method 2 right) for function 7, for $N = 8000$. The bold curve is the true function. Under each plot, the MISE over the 50 paths, and the mean of the selected dimensions.

Références

[1] Askey, R. and Wainger, S. (1965). Mean convergence of expansions in Laguerre and Hermite series. *Amer. J. Math.* **87**, 695-708.

[2] Belomestny, D., Panov, V. and Woerner, J. H. C. (2019) Low-frequency estimation of continuous-time moving average Lévy processes. *Bernoulli* **25**, 902-931.

[3] Brockwell, P. J. (2001) Continuous-time ARMA processes. Stochastic processes: theory and methods, 249-276. *In Handbook of Statist.*, **19**, C.R. Rao and D.N. Shanbhag (eds), North-Holland, Amsterdam.

[4] Brockwell, P. J., Ferrazzano, V. and Klüppelberg, C. (2013) High-frequency sampling and kernel estimation for continuous-time moving average processes. *J. Time Series Anal.* **34**, 385-404.

[5] Comte, F. and Renault, E. (1996). Long memory continuous time models. *Journal of Econometrics*, **73**, 101-149.

[6] Schnurr, A. and Woerner, J. H. C. (2011) Well-balanced Lévy driven Ornstein-Uhlenbeck processes. *Stat. Risk Model.* **28**, 343-357.

Thank you for your attention!