Fast Inference for Stationary Time Series

ICIAM 2023, Japan.

Samir Ben Hariz
Based on several joint works with A.Brouste, C.Cai, Y.Esstafa and M.Soltane

Laboratoire Manceau de Mathématiques, Le Mans Université

August 2023.

Summary

1. Introduction
   - Motivations
   - The model
   - Classical inference methods
   - Some limitations

- Time series are commonly used to model time dependencies in econometrics for example,
- A natural way to generalise the classical i.i.d setting.
- Popular models : ARMA, GARCH, ... : weak dependence, summable covariance.
- Long memory time series may appear in certain situations : Gaussian processes, FARIMA...
- Most of the models have a spectral density.
- We study different inference methods for a parametric model of time dependence through the spectral representation of the dependence structure.

Let (X_t) be a stationary Gaussian process with zero mean. We denote by

- r_θ the covariance function of (X_t), $\theta \in \mathbb{R}^d$ is an unknown parameter.
- f_θ the spectral density of (X_t) (Fourier's transform of the covariance)

$$f_\theta(\lambda) = \frac{1}{2\pi} \sum_{j \in \mathbb{Z}} \exp(ij\lambda)\, r_\theta(j)$$

- We have that

$$r_\theta(j) = \int_{-\pi}^{\pi} \exp(ij\lambda)\, f_\theta(\lambda)\, d\lambda.$$

- The likelihood function of $X^{(n)} = (X_1, \ldots, X_n)$ is defined by

$$\mathcal{L}\left(\theta, X^{(n)}\right) = \frac{1}{(2\pi)^{\frac{n}{2}} \det\left(\Sigma_{\theta,n}\right)^{\frac{1}{2}}} \exp\left(-\frac{1}{2} X^{(n)} \Sigma_{\theta,n}^{-1} X^{(n)^*}\right)$$

where $\Sigma_{\theta,n} = \left(r_\theta\left(i-j\right)\right)_{1 \leqslant i,j \leqslant n}$

- The maximum likelihood estimator (MLE) of $\theta \in \Theta$:

$$\widehat{\theta}_n^{MLE} = \arg\max_{\theta \in \Theta} \log \mathcal{L}\left(\theta, X^{(n)}\right).$$

- Under some regularity conditions on the spectral density, see Fox and Taqqu (1986); Dahlhaus (1989); Lieberman et al. (2012)

$$\sqrt{n}\left(\widehat{\theta}_n^{MLE} - \theta\right) \xrightarrow[n \to \infty]{\mathcal{L}} \mathcal{N}\left(0, \mathcal{I}^{-1}\left(\theta\right)\right)$$

where $\mathcal{I}\left(\cdot\right)$ is the Fisher information matrix :

$$\mathcal{I}\left(\theta\right) = \left(\frac{1}{4\pi} \int_{-\pi}^{\pi} \frac{\partial \log f_\theta\left(\lambda\right)}{\partial \theta_i} \frac{\partial \log f_\theta\left(\lambda\right)}{\partial \theta_j} \, d\lambda\right)_{1 \leqslant i,j \leqslant d}.$$

- An alternative to the MLE, the Whittle estimator.
- The Whittle likelihood function (which is an approximation of the exact gaussian log-likelihood) is defined by

$$\mathcal{L}_W\left(\theta, X^{(n)}\right) = \frac{1}{4\pi}\int_{-\pi}^{\pi} \log f_\theta(\lambda) + \frac{I_n(\lambda)}{f_\theta(\lambda)}\, d\lambda$$

where $I_n(\cdot)$ is the periodogram :

$$I_n(\lambda) = \frac{1}{2\pi n}\left|\sum_{j=1}^{n} X_j \exp(-ij\lambda)\right|^2 .$$

- The Whittle estimator of θ (WE) is defined by

$$\widehat{\theta}_n^{WE} = \arg\min_{\theta\in\Theta} \mathcal{L}_W\left(\theta, X^{(n)}\right).$$

- Under the same conditions for the MLE, see Fox and Taqqu (1986); Dahlhaus (1989). $\widehat{\theta}_n^{WE}$ have the same asymptotic properties as $\widehat{\theta}_n^{MLE}$.

- The MLE and WE have no explicit form.
- It is therefore necessary to perform a numerical optimization which :
    - requires a numerical inversion of the covariance matrix for the MLE,
    - greatly depends on the form of the spectral density for the WE,
    - is often time consuming and can be numerically unstable.
- It is interesting to look for alternative estimation methods that keep the same asymptotic properties as the classical methods.

2 The Gaussian case

- One step estimator
- An asymptotic result
- An example

- Let $\ell\left(\theta, X^{(n)}\right)$ the score function of the Gaussian likelihood.
- $\widetilde{\theta}_n$ some initial estimator θ.
- We peform a Fisher-scoring step on the score at $\widetilde{\theta}_n$, :

$$\widehat{\theta}_n = \widetilde{\theta}_n + \frac{1}{n}\mathcal{I}^{-1}\left(\widetilde{\theta}_n\right)\ell\left(\widetilde{\theta}_n, X^{(n)}\right).$$

- This method called "one-step" allows to build a new estimator of θ in only one step. She was introduced in Le Cam (1956) for variance reduction in $i.i.d.$ models and used fairly recently in Kamatani and Masayuki (2015); Kutoyants and Motrunich (2016) for estimator speed improvements.

Asymptotic efficiency in the Gaussian setting

Theorem (Ben Hariz et al. (2022))

Assume some regularity conditions on the spectral density and

$$n^{\delta}\left(\widetilde{\theta}_n - \theta\right) = O_{\mathbb{P}}(1) \ \text{ for some } \delta > \frac{1}{4}.$$

Then,

$$\sqrt{n}\left(\widehat{\theta}_n - \theta\right) \xrightarrow[n\to\infty]{\mathcal{L}} \mathcal{N}\left(0, \mathcal{I}^{-1}(\theta)\right).$$

We obtain an optimal estimator in speed and variance. The initial estimator can be :

- The MLE or WE on a subsample of size $\left[n^{\beta}\right]$ for some $\frac{1}{2} < \beta \leqslant 1$,
- An estimator from moments methods,
- A semi-parametric estimator (log periodogram, local Whittle...).

- We consider the following AR(1) model as example :

$$X_t = \alpha X_{t-1} + \varepsilon_t^H,$$

- $\left(\varepsilon_t^H\right)$ is an FGN with Hurst exponent H and variance σ_2.
- The covariance function of $\left(\varepsilon_t^H\right)$ is

$$r_{H,\sigma_2}(k) = \frac{\sigma_2}{2}\left(|k+1|^{2H} - 2|k|^{2H} + |k-1|^{2H}\right).$$

- The corresponding spectral density is

$$g_{H,\sigma_2}(\lambda) = C_{H,\sigma_2}\, 2\,(1-\cos\lambda) \sum_{k\in\mathbb{Z}} \frac{1}{|\lambda + 2k\pi|^{2H+1}}$$

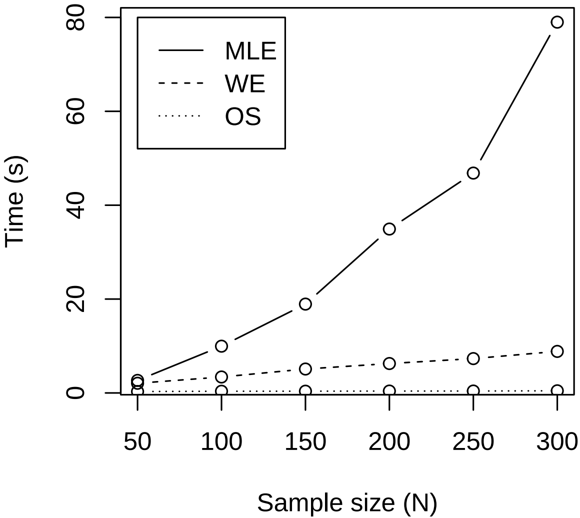
for some constant C_{H,σ_2}.

- We want to estimate $\theta = (\alpha, H, \sigma_2)$ from the observations $X^{(n)}$.
- The spectral density of (X_t) is given by

$$f_{\vartheta}(\lambda) = \frac{g_{H,\sigma_2}(\lambda)}{1 - 2\alpha\cos(\lambda) + \alpha^2}$$

- The covariance function of (X_t) is not in closed form.
- We cannot use an initial estimator like moments methods or quadratic generalized variations as in Istas and Lang (1997).
- Since the spectral density is in closed form, we can use an initial semi parametric estimator in order to estimate H.

- We construct an initial estimator of θ via the following steps :
  - We estimate H via an adapted method of log-periodogram regression, that is the GPH estimator (see Hurvich et al. (1998) for more details).
  - Then we estimate α via a generalized least square estimator involving the estimation of H.
  - Finally we estimate σ_2 via the residual process.
- The initial estimator $\widetilde{\theta}_n$ satisfy the condition of the Theorem with $\frac{1}{4} < \delta < \frac{1}{3}$ hence the one-step procedure apply.
- In the next slide, the parameter is fixed to $\theta = (0.2; 0.6; 0.4)$ and we peform 20 Monte-Carlo simulations in order to evaluate the computation time for each method for different sample size.


MLE
WE
OS
Time (s)
0
20
40
60
80
50
100
150
200
250
300
Sample size (N)

n	B IE	B OS	SD IE	SD OS	RMSE IE	RMSE OS
500	0.0195	0.0088	0.1017	0.0494	0.1036	0.0502
1000	0.0116	0.0036	0.0720	0.0335	0.0729	0.0336
1500	0.0082	0.0006	0.0641	0.0274	0.0646	0.0274

Table: Bias SD and RMSE for α, where $\theta =$(-0.6 ; 0.8 ; 1)

n	B IE	B OS	SD IE	SD OS	RMSE IE	RMSE OS
500	-0.0010	-0.0038	0.1167	0.0473	0.1167	0.0474
1000	-0.0011	-0.0012	0.0909	0.0332	0.0909	0.0333
1500	0.0007	0.0020	0.0783	0.0252	0.0783	0.0253

Table: Bias SD and RMSE for H, where $\theta =$(-0.6; 0.8; 1)

n	B IE	B OS	SD IE	SD OS	RMSE IE	RMSE OS
500	0.0348	0.0047	0.1223	0.0501	0.1272	0.0503
1000	0.0250	0.0011	0.1011	0.0322	0.1042	0.0322
1500	0.0239	-0.0029	0.0855	0.0269	0.0888	0.0271

Table: Bias SD and RMSE for α, where $\theta =$(-0.6 ; 0.3 ; 1)

n	B IE	B OS	SD IE	SD OS	RMSE IE	RMSE OS
500	-0.0078	-0.0023	0.1068	0.0381	0.1071	0.0382
1000	-0.0058	-0.0009	0.0896	0.0247	0.0898	0.0247
1500	-0.0108	0.0015	0.0801	0.0195	0.0808	0.0196

Table: Bias SD and RMSE for H, where $\theta =$(-0.6 ; 0.3 ; 1)

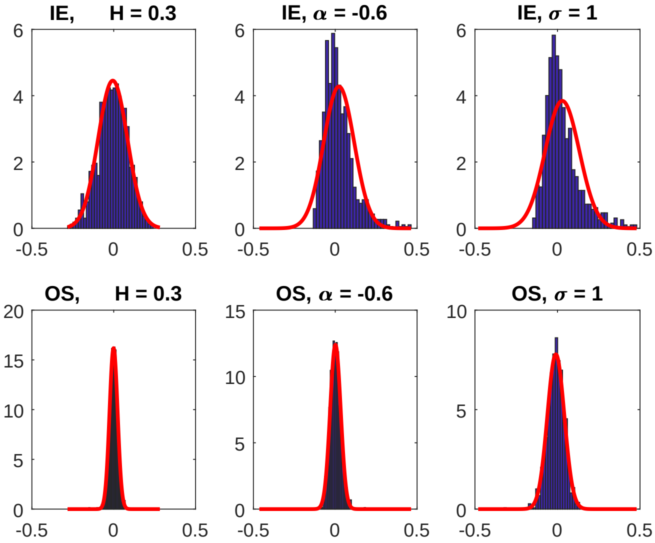


Figure: Statistical error of the initial estimator and the one-step method where $\theta = (-0.6; 0.3; 1)$ and $n = 1500$.

3 The general case

- The Whittle method
- One-step Whittle
- An example

- We assume that (X_t) admit the following Wold representation

$$X_t = \sum_{j \geqslant 0} a_{j,\theta} \varepsilon_{t-j}$$

where (ε_t) is a weak white noise of variance σ_2. The sequence (ε_t) is uncorrelated but not necessarily independent.

- From the Wold representation, we have that

$$f_\theta(\lambda) = \frac{\sigma_2}{2\pi} \left| \sum_{j=0}^{\infty} \exp(ij\lambda) a_{j,\theta} \right|^2 .$$

- The discrete Whittle-likelihood is given by

$$\mathcal{L}_W^D\left(\theta, X^{(n)}\right) = \frac{1}{2n} \sum_{j=1}^{n-1} \log f_\theta(\lambda_j) + \frac{I_n(\lambda_j)}{f_\theta(\lambda_j)}$$

where $\lambda_j = \frac{2\pi j}{n}$. In this setting the Whittle estimator is defined by

$$\overline{\theta}_n^{WE} = \arg\min_{\theta \in \Theta} \mathcal{L}_W^D\left(\theta, X^{(n)}\right).$$

- Let $\widetilde{I}_n(\lambda) = \frac{1}{2\pi n}\left|\sum_{j=1}^{n} \varepsilon_j \exp(-ij\lambda)\right|^2$, we have (see Lemma A.7 of Shao (2010)) that

$$\frac{1}{\sqrt{2n}} \sum_{j=1}^{n-1} \frac{\partial \log f_\theta(\lambda_j)}{\partial \theta_k} \left(\frac{I_n(\lambda_j)}{f_\theta(\lambda_j)} - \frac{2\pi \widetilde{I}_n(\lambda_j)}{\sigma_2} \right) \xrightarrow[n\to\infty]{\mathbb{P}} 0,$$

- The asymptotic normality of the score follows therefore from the asymptotic normality of the random vector

$$\frac{1}{\sqrt{2n}} \sum_{j=1}^{n-1} \frac{\partial \log f_\theta(\lambda_j)}{\partial \theta_k} \times \frac{2\pi \widetilde{I}_n(\lambda_j)}{\sigma_2},$$

- We consider the following function

$$f_4\left(\underline{\lambda}\right)=\frac{1}{(2\pi)^3}\sum_{\underline{k}=(k_1,k_3,k_3)\in\mathbb{Z}^3}\text{cum}\left(\varepsilon_0,\varepsilon_{k_1},\varepsilon_{k_2},\varepsilon_{k_3}\right)\exp\left(-i\left\langle\underline{\lambda},\underline{k}\right\rangle\right),$$

where $\underline{\lambda}\in[-\pi;\pi]^3$. This function is the fourth order cumulant spectral density of the noise (ε_t).

- From Lemma A.8 of Shao (2010),

$$\begin{aligned}\mathbb{C}ov\left(\widetilde{I}_n\left(\lambda_j\right),\widetilde{I}_n\left(\lambda_k\right)\right)=&\mathbb{1}_{j=k}\left(\frac{\sigma_2}{2\pi}+o\left(1\right)\right)\\&+\mathbb{1}_{j\neq k}\left(\frac{2\pi}{n}f_4\left(\lambda_j,-\lambda_k,\lambda_k\right)+o\left(\frac{1}{n}\right)\right),\end{aligned}$$

- The functional f_4 (which depends on an unobservable process) therefore contributes to the asymptotic distribution of the score.

Under regularity conditions on the spectral density and the noise, Shao (2010)

$$\sqrt{n}\left(\overline{\theta}_n^{WE} - \theta\right) \xrightarrow[n\to\infty]{\mathcal{L}} \mathcal{N}\left(0, \mathcal{I}^{-1}(\theta)\,\mathcal{W}(\theta)\,\mathcal{I}^{-1}(\theta)\right)$$

where $\mathcal{W}(\theta)$ is some matrix whose coefficients are expressed in the form to an integration between the spectral density and f_4.

- If the noise is Gaussian then $\mathcal{W}(\theta) = \mathcal{I}(\theta)$ and we recover the known result,
- It is necessary to estimate $\mathcal{I}^{-1}(\theta)\,\mathcal{W}(\theta)\,\mathcal{I}^{-1}(\theta)$ in order to build confidence bound.

Let $\ell_W^D\left(\theta, X^{(n)}\right)$ the Whittle score and $\widetilde{\theta}_n$ some initial estimator θ.
We perform an adapted Fisher-scoring step on the Whittle score at $\widetilde{\theta}_n$,

$$\widehat{\theta}_n = \widetilde{\theta}_n - \mathcal{I}^{-1}\left(\widetilde{\theta}_n\right)\ell_W^D\left(\widetilde{\theta}_n, X^{(n)}\right).$$

We proove a similar result to the Gaussian case in Ben Hariz et al. (2023) :

Theorem

We assume that certain regularity conditions on the spectral density and the noise are satisfied. We also assume that

$$n^{\delta}\left(\widetilde{\theta}_n - \theta\right) = O_{\mathbb{P}}(1) \text{ for some } \delta > \frac{1}{4}.$$

Then,

$$\sqrt{n}\left(\widehat{\theta}_n - \theta\right) \xrightarrow[n\to\infty]{\mathcal{L}} \mathcal{N}\left(0, \mathcal{I}^{-1}(\theta)\,\mathcal{W}(\theta)\,\mathcal{I}^{-1}(\theta)\right).$$

- We consider a FARIMA$(1, d, 1)$ model of the form

$$(1-L)^d (1-aL) X_t = (1-bL) \varepsilon_t$$

for different type of noise (ε_t) where L is the backward operator.

- We denote $\theta = (a, b, d)$. a is the autoregressive parameter, b the moving average parameter and d the fractional parameter of the filter. We want to estimate θ from the observations $X^{(n)}$.
- The spectral density of the process is therefore

$$f_\theta(\lambda) = \frac{\sigma_2}{2\pi} |1-\exp(i\lambda)|^{-2d} |1-a\exp(i\lambda)|^{-2} |1-b\exp(i\lambda)|^2 ,$$

- The initial estimator is the Whittle estimator on a sub-sample.

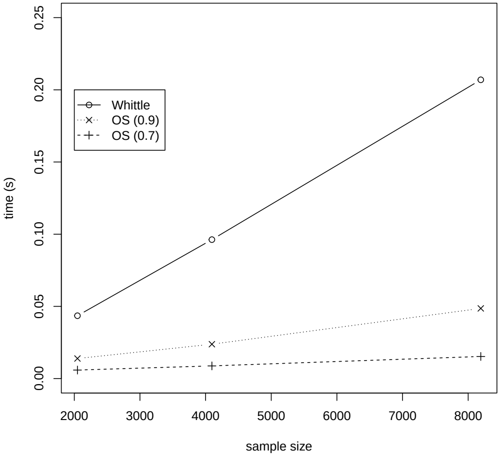


Figure: Evolution of the computation times (in seconds), in a FARIMA model of parameter $\vartheta_0 = (0,2; 0,5; 0,3)$.

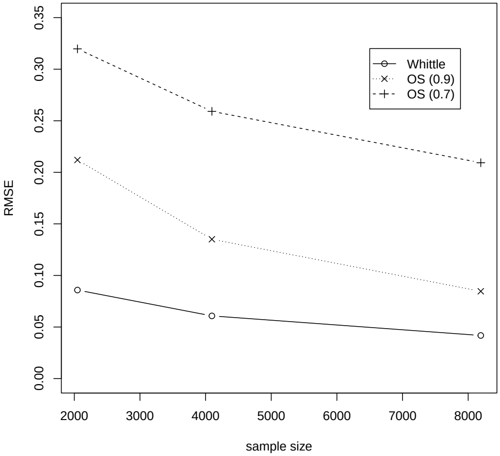


Figure: Evolution of the RMSE for the autoregressive parameter in a FARIMA model of parameter $\theta_0 = (0,2;0,5;0,3)$.

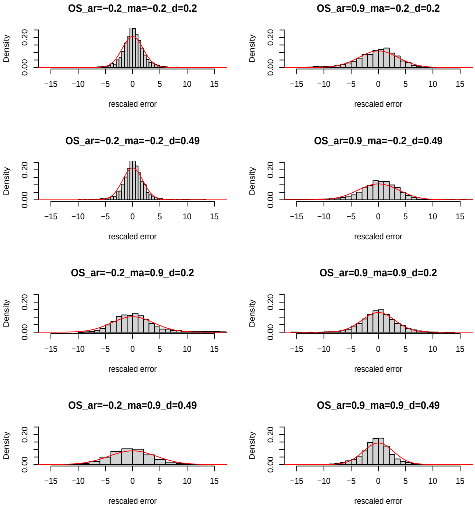


Figure: Statistical error of the one-step estimator of d in a FARIMA model.

4 Comments and Perspectives

Comments and Perspectives

- The one-step procedure makes it possible to obtain an estimator whose asymptotic properties are the same as the Whittle estimator with reduced computation time.
- The asymptotic efficiency is even reached via this method in the Gaussian case.
- We can obtain the same variance as in Gaussian case by reducing the number of 's frequencies. Less speed but simpler vaiance!!
- It would be interesting, apart from the Gaussian case, to estimate the covariance matrix $\mathcal{W}(\theta)$ in order to build confidence bound. We are currently developing an approach in Ben Hariz et al. (2023) that draws on the work of Taniguchi (1982); Keenan (1987).

Ben Hariz, S., Brouste, A., Cai, C., and Soltane, M. (2022). Fast and Asymptotically-efficient estimation in a Fractional autoregressive process. working paper or preprint.

Ben Hariz, S., Brouste, A., Esstafa, Y., and Soltane, M. (2023). Fast inference for stationary time series. *In preparation*.

Dahlhaus, R. (1989). Efficient parameter estimation for self-similar processes. *The Annals of Statistics*, 17(4):1749–1766.

Fox, R. and Taqqu, M. (1986). Large-sample properties of parameter estimates for strongly dependent stationary gaussian time series. *The Annals of Statistics*, 14(2):517–532.

Hurvich, C. M., Deo, R., and Brodsky, J. (1998). The mean squared error of Geweke and Porter-Hudak's estimator of the memory parameter of a long-memory time series. *Journal of Time Series Analysis*, 19(1):19–46.

Istas, J. and Lang, G. (1997). Quadratic variations and estimation of the local Hölder index of a Gaussian process. *Annales de l'I.H.P. section B*, 33(4):407–436.

Kamatani, K. A. and Masayuki, Y. (2015). Hybrid multi-step estimators for stochastic differential equations based on sampled data. *Statistical Inference for Stochastic Processes*, 18:177–204.

Keenan, D. M. (1987). Limiting Behavior of Functionals of Higher-Order Sample Cumulant Spectra. *The Annals of Statistics*, 15(1):134 – 151.

Kutoyants, Y. A. and Motrunich, A. (2016). On multi-step mle-process for markov sequences. *Metrika*, 79:705–724.

Le Cam, L. (1956). On the asymptotic theory of estimation and testing hypothesis. *Proceedings of the 3rd Berkeley Symposium*, 1:355–368.

Lieberman, O., Rosemarin, R., and Rousseau, J. (2012). Asymptotic theory for maximum likelihood estimation of the memory parameter in stationary gaussian processes. *Econometric Theory*, 28(2):457–470.

Shao, X. (2010). Nonstationarity-extended whittle estimation. *Econometric Theory*, 26(4):1060–1087.

Taniguchi, M. (1982). On estimation of the integrals of the fourth order cumulant spectral density. *Biometrika*, 69(1):117–122.