INTRODUCTION (1)

- ARMA models have enjoyed (and still enjoy) a wide popularity.
- Recent developments on ARMA models concern their extension to long-range dependence with fractionally integrated ARMA (FARIMA) models, and multivariate ARMA models.
- However, it was observed early on that some effects such as the *regime* one (different behaviors of residuals for different ranges of signal samples) cannot be dealt with using linear models.

INTRODUCTION (2)

- There are two main possibilities to model these types of effects:
  - Use non-Gaussian probability density functions (pdf)
  - Use non-linear models
- The first approach is usually intractable (it is usually hard to define the appropriate pdf), and most research has been centered on the second approach.

FRAMEWORK FOR MODEL DESCRIPTION (1)

- We will focus now on mimimum mean square error (MMSE) prediction. That is, if the sample x_{n+m}, $m \geq 0$, is to be predicted using the vector $\boldsymbol{X}_{n-1} = [x_{n-1}, x_{n-2}, \ldots, x_1]^{\mathsf{T}}$ of past samples, one looks for a function $f(\boldsymbol{X}_{n-1})$ that minimizes:

$$\mathsf{E}[\{x_{n+m} - f(\boldsymbol{X}_{n-1})\}^2]$$

- The MMSE predictor is the mean of x_{n+m} conditioned on the past, i.e. $f(\boldsymbol{X}_{n-1}) = \mathsf{E}[x_{n+m} | \boldsymbol{X}_{n-1}]$.

FRAMEWORK FOR MODEL DESCRIPTION (2)

- Demonstration: for any predictor function $g(.)$:

$$\mathrm{E}[\{x_{n+m} - g(\boldsymbol{X}_{n\text{-}1})\}^2] = \mathrm{E}[\{x_{n+m} - \mathrm{E}[x_{n+m}|\boldsymbol{X}_{n\text{-}1}]\}^2] + \mathrm{E}[\{\mathrm{E}[x_{n+m}|\boldsymbol{X}_{n\text{-}1}] - g(\boldsymbol{X}_{n\text{-}1})\}^2] + 2C$$

with

$$C = \mathrm{E}[\{x_{n+m} - \mathrm{E}[x_{n+m}|\boldsymbol{X}_{n\text{-}1}]\}\{\mathrm{E}[x_{n+m}|\boldsymbol{X}_{n\text{-}1}] - g(\boldsymbol{X}_{n\text{-}1})\}]$$

$$= \mathrm{E}\Big(\mathrm{E}\Big[\{x_{n+m} - \mathrm{E}[x_{n+m}|\boldsymbol{X}_{n\text{-}1}]\}\{\mathrm{E}[x_{n+m}|\boldsymbol{X}_{n\text{-}1}] - g(\boldsymbol{X}_{n\text{-}1})\}\Big]\Big|\boldsymbol{X}_{n\text{-}1}\Big)$$

FRAMEWORK FOR MODEL DESCRIPTION (3)

$$C = \mathrm{E}\Big(\{\mathrm{E}[x_{n+m}|\boldsymbol{X}_{n\text{-}1}] - g(\boldsymbol{X}_{n\text{-}1})\}\mathrm{E}\{x_{n+m} - \mathrm{E}[x_{n+m}|\boldsymbol{X}_{n\text{-}1}]\big|\boldsymbol{X}_{n\text{-}1}\}\Big)$$

$$= \mathrm{E}\Big(\{\mathrm{E}[x_{n+m}|\boldsymbol{X}_{n\text{-}1}] - g(\boldsymbol{X}_{n\text{-}1})\}\{\mathrm{E}[x_{n+m}|\boldsymbol{X}_{n\text{-}1}] - \mathrm{E}[x_{n+m}|\boldsymbol{X}_{n\text{-}1}]\}\Big)$$

$$= 0.$$

Thus $\mathrm{E}[\{x_{n+m} - g(\boldsymbol{X}_{n\text{-}1})\}^2] \geq \mathrm{E}[\{x_{n+m} - \mathrm{E}[x_{n+m}|\boldsymbol{X}_{n\text{-}1}]\}^2]$

- With $\varepsilon_{n+m} = x_{n+m} - \mathrm{E}[x_{n+m}|\boldsymbol{X}_{n\text{-}1}]$, one can show that:

$$\mathrm{E}[\varepsilon_{n+m}|\boldsymbol{X}_{n\text{-}1}] = 0 \qquad \mathrm{Cov}(\varepsilon_s\,,\ \varepsilon_t) = 0$$

FRAMEWORK FOR MODEL DESCRIPTION (4)

- In the mean square sense, a sample can be written as:

$$x_{n+m} = \mathrm{E}[x_{n+m}|\boldsymbol{X}_{n\text{-}1}]\} + \varepsilon_{n+m}$$

that is, as the sum of the component predictible from the past, and the non-predictible part (innovation).

- It can be proven [1] that, if the signal samples $\{x_n\}$ are Gaussian, then the MMSE predictor is linear.

FRAMEWORK FOR MODEL DESCRIPTION (5)

- In the light of the discussion above, we will focus now on nonlinear predictors/models described by:

$$x_n = g(x_{n\text{-}1},\ldots,x_{n\text{-}p}) + \sigma(x_{n\text{-}1},\ldots,x_{n\text{-}p})\varepsilon_n$$

with $g(.)$ and $\sigma(.)$ well behaved functions and $\{\varepsilon_n\}$ an independent identically distributed sequence with unit variance.

- In almost all cases, $\sigma(.)$ will be constant, except when we consider models with conditional heteroscedasticity.

FRAMEWORK FOR MODEL DESCRIPTION (6)

- The models with $\sigma(.)$ constant are called nonlinear autoregressive models (NAR).
- As for the linear AR models (which constitute a subset of NAR), one tries to have as small an order p as possible.
- Of course, what is desirable is that $g(x_{n-1},\ldots, x_{n-p})$ approximates $\mathsf{E}[x_n|\boldsymbol{X}_{n-1}]\}$ as well as possible. The choice of $g(.)$ may be suggested by some a priori knowledge about the dynamics, or because it is in a set of functions with universal approximation capability.

PARAMETER ESTIMATION (1)

- Minimum variance criterion

Let us examine the case of a single model parameter θ and an estimator $\hat{\theta}$. A natural optimality criterion for the estimator is the mean square error (MSE):

$$\mathrm{mse}(\hat{\theta}) = \mathsf{E}\left[(\hat{\theta}-\theta)^2\right] = \mathsf{E}\left\{\left[\left(\hat{\theta}-\mathsf{E}(\hat{\theta})\right)+\left(\mathsf{E}(\hat{\theta})-\theta\right)\right]^2\right\}$$

$$= \mathrm{var}(\hat{\theta}) + \left[\mathsf{E}(\hat{\theta})-\theta\right]^2 = \mathrm{var}(\hat{\theta}) + b^2(\hat{\theta})$$

with $b(\hat{\theta}) = \mathsf{E}(\hat{\theta}) - \theta$ the estimator bias

PARAMETER ESTIMATION (2)

- Unfortunately, the minimum MSE estimator cannot be obtained in most cases. A feasible approach consists in constraining the bias to be zero and find the estimator with the minimum variance. This estimator is the minimum variance unbiased (MVU) estimator.
- Note that the variance of the MVU estimator should be the smallest for *all possible values* of θ.

PARAMETER ESTIMATION (3)

- There is no general procedure to find the MVU estimator. One possible way to find it is establish the Cramer-Rao lower bound (CRLB).
- The CRLB is the lower bound on the variance of *any* unbiased estimator. If indeed an estimator variance reaches this bound, then it is the MVU one.
- It may happen that no estimator reaches this bound, but that the MVU still exists.

PARAMETER ESTIMATION (4)

- CRLB for a parameter vector $\boldsymbol{\theta}$

The covariance matrix of any unbiased estimator satisfies:

$$\mathbf{C}_{\hat{\boldsymbol{\theta}}} - \mathbf{I}^{-1}(\boldsymbol{\theta}) \geq \mathbf{0}$$

Meaning that the matrix is positive semidefinite. The Fisher information matrix $\mathbf{I}(\boldsymbol{\theta})$ is given by:

$$\mathbf{I}(\boldsymbol{\theta})_{ij} = -\mathsf{E}\left[\frac{\partial^2 \ln p(\boldsymbol{x};\boldsymbol{\theta})}{\partial\boldsymbol{\theta}_i \partial\boldsymbol{\theta}_j}\right]$$

With $p(\boldsymbol{x};\boldsymbol{\theta})$ the probability density function of the data $\boldsymbol{x}$ parameterized by $\boldsymbol{\theta}$.

PARAMETER ESTIMATION (5)

- An unbiased estimator that reaches the CRLB, that is: $\mathbf{C}_{\hat{\boldsymbol{\theta}}} = \mathbf{I}^{-1}(\boldsymbol{\theta})$

can be found if and only if

$$\frac{\partial \ln p(\boldsymbol{x};\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = \mathbf{I}(\boldsymbol{\theta})(h(\boldsymbol{x}) - \boldsymbol{\theta})$$

for some multidimensional function h. In that case the MVU estimator is: $\hat{\boldsymbol{\theta}} = h(\boldsymbol{x})$

Unfortunately, this approach is easy to apply only for linear models and data with Gaussian statistics.

PARAMETER ESTIMATION (6)

- Maximum likelihood (ML) estimation

The ML estimator is the value of $\boldsymbol{\theta}$ maximizing the likelihood $p(\boldsymbol{x};\boldsymbol{\theta})$, where $\boldsymbol{x}$ is now the vector of observed data samples.

It is not optimal in general, but asymptotically is the MVU estimator.

Under some conditions, asymptotically:

$$\hat{\theta} \approx N(\boldsymbol{\theta}, \mathbf{I}^{-1}(\boldsymbol{\theta}))$$

PARAMETER ESTIMATION (7)

- Least squares (LS) estimation

For NAR models, and *N* samples available, the LS estimator is the value of minimizing:

$$\mathrm{lse} = \sum_{n=1}^{N} [x_n - g(x_{n-1}, \cdots, x_{n-p}; \boldsymbol{\theta})]^2$$

- The estimation performance depends on the distribution of the modeling errors. If this distribution is Gaussian and the errors are uncorrelated, then LS is equvalent to ML.

WHEN IS A NONLINEAR MODEL STABLE? (1)

- Conditions have been derived in the literature [2] on the ergodicity of (vector) Markov chains $\{\boldsymbol{Y}_n\}$.
- A sufficient condition for ergodicity is the existence of positive constants α, β, and γ such that:

$$\mathrm{E}\Big(\|\boldsymbol{Y}_n\| - \|\boldsymbol{Y}_{n\text{-}1}\| \;\Big|\; \boldsymbol{Y}_{n\text{-}1} = \boldsymbol{y} \Big) \leq -\beta, \;\; \|\boldsymbol{y}\| > \alpha \qquad (1)$$

$$\mathrm{E}\Big(\|\boldsymbol{Y}_n\| - \|\boldsymbol{Y}_{n\text{-}1}\| \;\Big|\; \boldsymbol{Y}_{n\text{-}1} = \boldsymbol{y} \Big) \leq \gamma, \;\; \|\boldsymbol{y}\| \leq \alpha \qquad (2)$$

WHEN IS A NONLINEAR MODEL STABLE? (2)

- Condition (2) is not very stringent. It just imposes that the increase in norm inside a ball of radius α is bounded.
- Condition (1) expresses that there must be some mechanism for *drift back to the center*, i.e. that if the norm of the instance of the Markov chain becomes large, then the norms of successive instances must decrease. In this way, ergodicity (existence of an equilibrium pdf) is guaranteed.

WHEN IS A NONLINEAR MODEL STABLE? (3)

- A stronger condition for geometrical ergodicity (exponentially fast convergence to the equilibrium pdf starting from arbitrary initial conditions) is obtained by replacing condition (1) with:

$$\mathrm{E}\left(r\|\boldsymbol{Y}_n\| - \|\boldsymbol{Y}_{n-1}\| \,\Big|\, \boldsymbol{Y}_{n-1} = \boldsymbol{y}\right) \leq -\beta, \quad \|\boldsymbol{y}\| > \alpha \qquad (1b)$$

with r a constant > 1.

WHEN IS A NONLINEAR MODEL STABLE? (4)

- Now $\{x_n\}$ in the NAR model is not Markovian. A Markov chain is obtained by using a state space representation:

$$\boldsymbol{X}_n = G(\boldsymbol{X}_{n\text{-}1}) + \boldsymbol{E}_n$$

with

$$\boldsymbol{X}_{n\text{-}1} = [x_{n\text{-}1}, x_{n\text{-}2}, \ldots, x_{n\text{-}p}]^{\mathsf{T}}$$

$$G(\boldsymbol{X}_{n\text{-}1}) = [g(x_{n\text{-}1}, \ldots, x_{n\text{-}p}), x_{n\text{-}1}, \ldots, \mathrm{x}_{\mathrm{n\text{-}p+1}}]^{\mathsf{T}}$$

$$\boldsymbol{E}_n = [\sigma(x_{n\text{-}1}, \ldots, x_{n\text{-}p})\boldsymbol{\varepsilon}_n, 0, \ldots, 0]^{\mathsf{T}}$$

SOME PRE-PROCESSING (1)

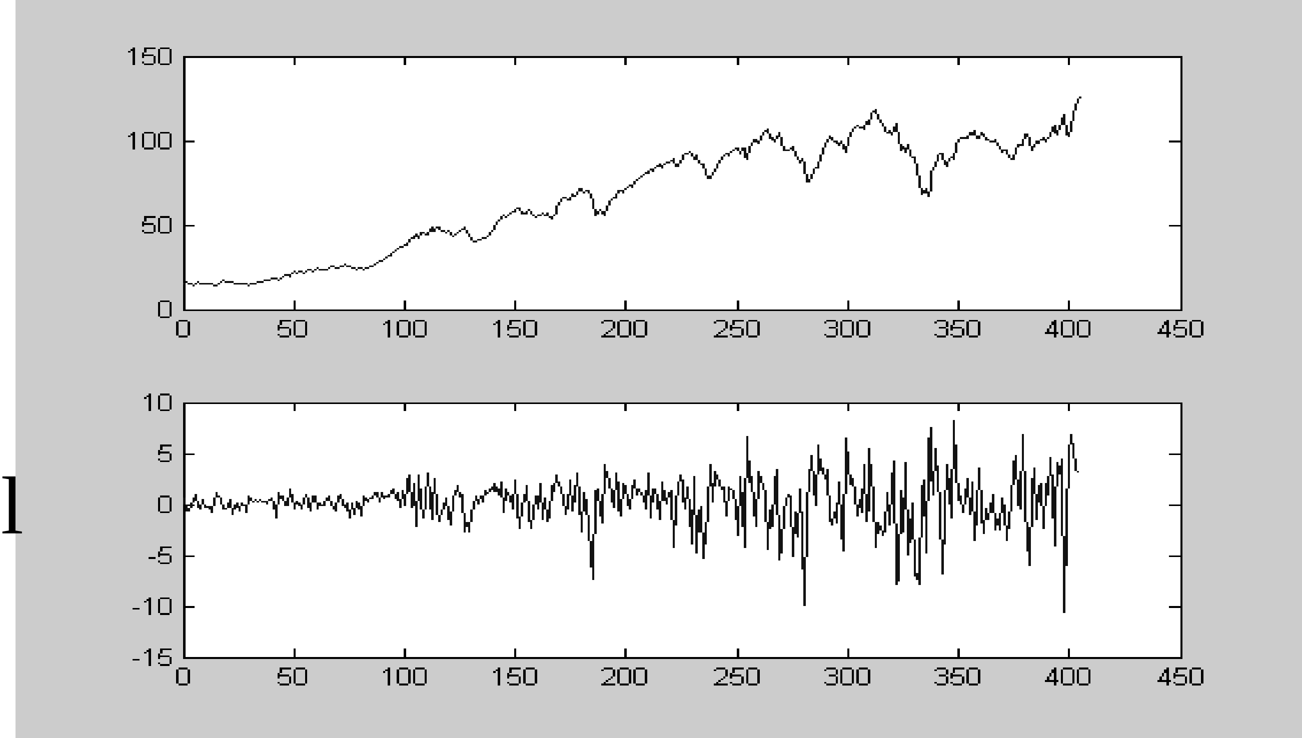
- If the mean does not carry information by itself, it is often judicious to remove it from the data. If it is not removed, one should include a constant term in the NAR model to take it into account.
- Also, deterministic trends (ramp) and oscillations (for instance seasonal effects) should be removed.
- Stochastic trends (unit root) should also be dealt with. Statistical tests (such as the Augmented Dickey-Fuller one) test for the presence of such a trend.

SOME PRE-PROCESSING (2)

By all means, the presence of a unit root is usually visible in the signal, and differencing $y_n = x_n - x_{n-1}$, usually solves this problem.

Standard & Poor
500 Index
(1947 – 1983)

Differenced signal



TESTING FOR NONLINEARITY (1)

- Obviously, the first test is to check whether the signal samples are Gaussian or not, using the Kolmogorov-Smirnov or the Chi-square test for instance.
- Likelihood-ratio (LR) tests have been developed to test the significance of a nonlinear model with respect to a linear one. Unfortunately, LR tests have to be tailored to the nonlinear models used, and the test statistics may be hard to evaluate [3].

TESTING FOR NONLINEARITY (2)

- A test for nonlinearity based on higher-order statistics has also been proposed in [4]. It uses 3rd and 4th order cumulants and test statistics are derived.
- A test based on time irreversibility (linear signals are time reversible) has also been proposed. The measure of time irreversibility is given by:

$$\phi_{rev}(\tau) = \frac{1}{N-\tau}\sum_{n=\tau+1}^{N}(x_n - x_{n-\tau})^3$$

For a suitable lag τ.

TESTING FOR NONLINEARITY (3)

- It is also possible to select the best polynomial model on the signal with respect to a model selection criterion such as MDL.

The linear AR models constitute a subset of the set of polynomial models. If only linear terms ar retained, then there is non nonlinearity in the signal.

TESTING FOR NONLINEARITY (4)

- In order to assess the significance of a test, a powerful approach, *surrogate analysis*, has recently been introduced [5].

The idea is the following: suppose one has measured some feature *m* on the signal at hand. One generates synthetic (surrogate) signals sharing some properties of the original signal (sample pdf and 2nd-order statistics), *but not the hypothetized nonlinear relationship between samples*.

TESTING FOR NONLINEARITY (5)

Then one can compute a significance *S*:

$$S = \frac{\left| m - \langle m \rangle_{surr} \right|}{\sigma_{surr}}$$

Where $<m>_{surr}$ is the mean of the distribution of the feature for the surrogates and σ_{surr} its standard deviation.

Assuming *m* is Gaussian a value $S = 2.6$ corresponds to a significance level of 0.01 for the value of *m* obtained on the original signal.

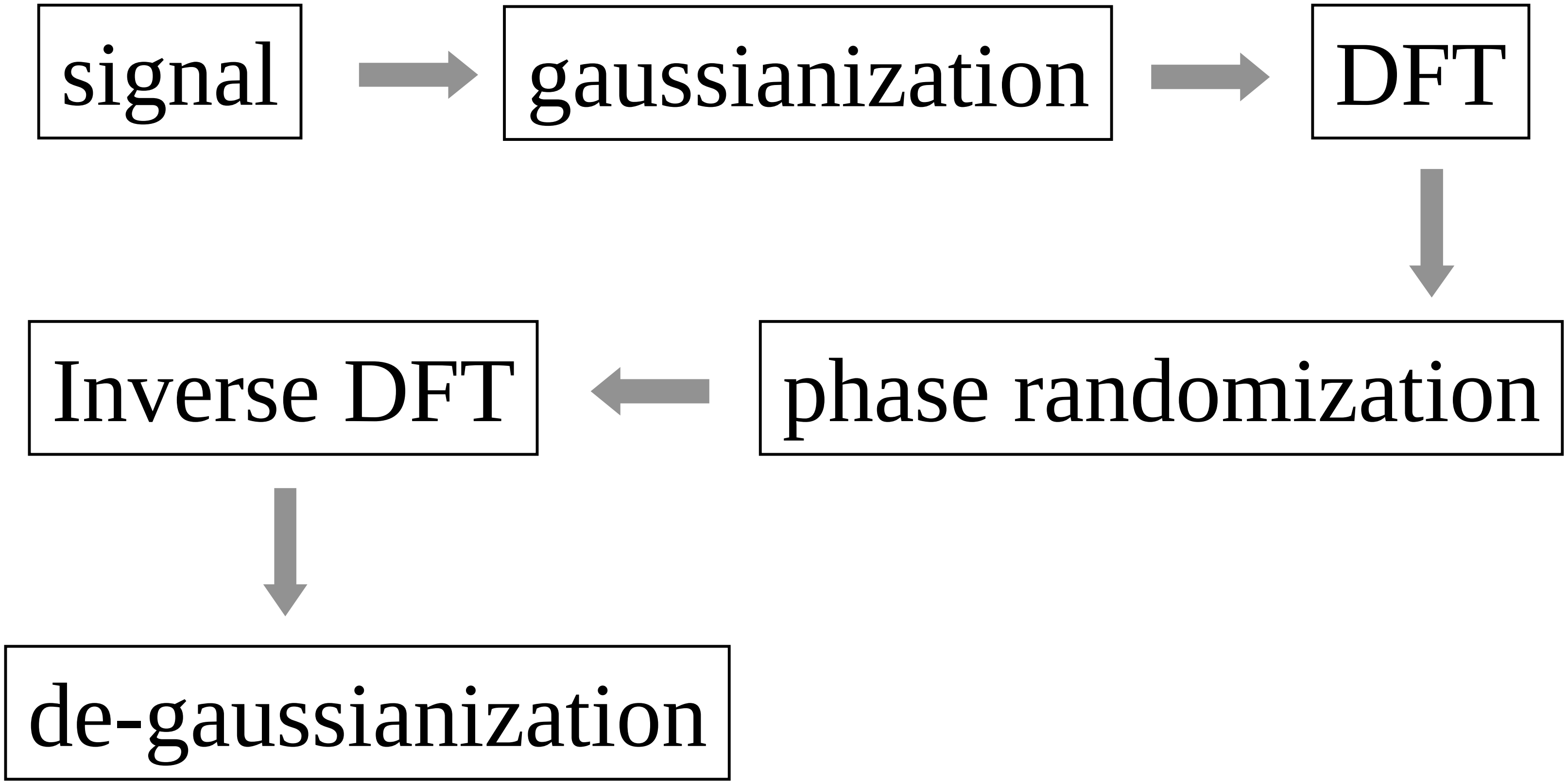
TESTING FOR NONLINEARITY (6)

Since the assumption that *m* is Gaussian may be bold, one can also use a rank-order test [5].

- To build these surrogates, one uses the fact that linear relationships between samples imply only 2nd-order statistics, i.e. the autocorrelation function, which is even and doest not carry any phase information.

TESTING FOR NONLINEARITY (7)

- Principle of surrogate generation:



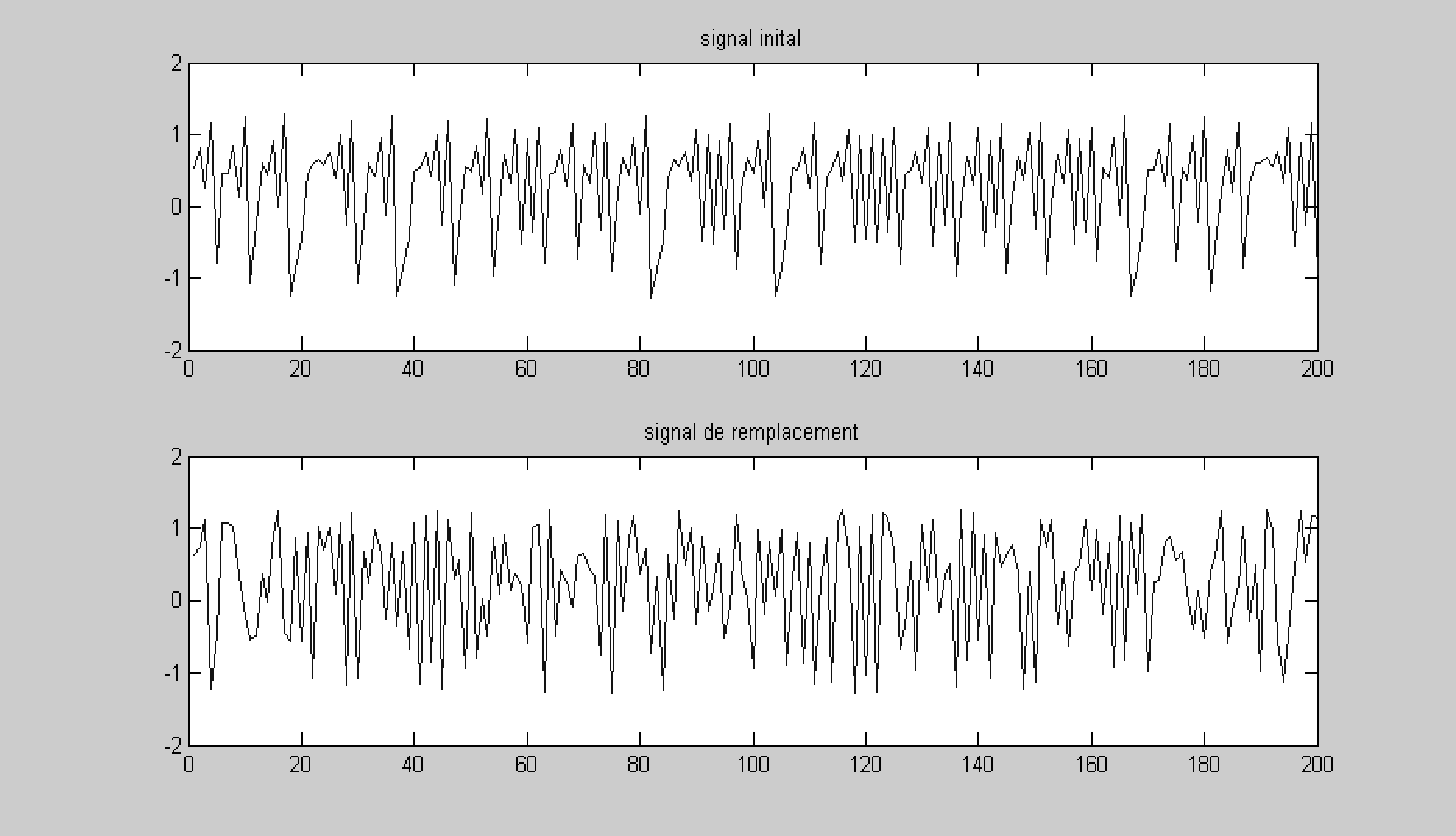
signal → gaussianization → DFT → phase randomization → Inverse DFT → de-gaussianization

TESTING FOR NONLINEARITY (8)

- To “Gaussianize“ the samples, one feeds them through an instantaneous nonlinearity which is the distribution of the samples.
- Phase randomization on the discrete Fourier transform (phases uniformly drawn between 0 and 2π), destroys any potential nonlinear structure.
- De-Gaussianization consist in applying the inverse of the instantaneous linear transform.

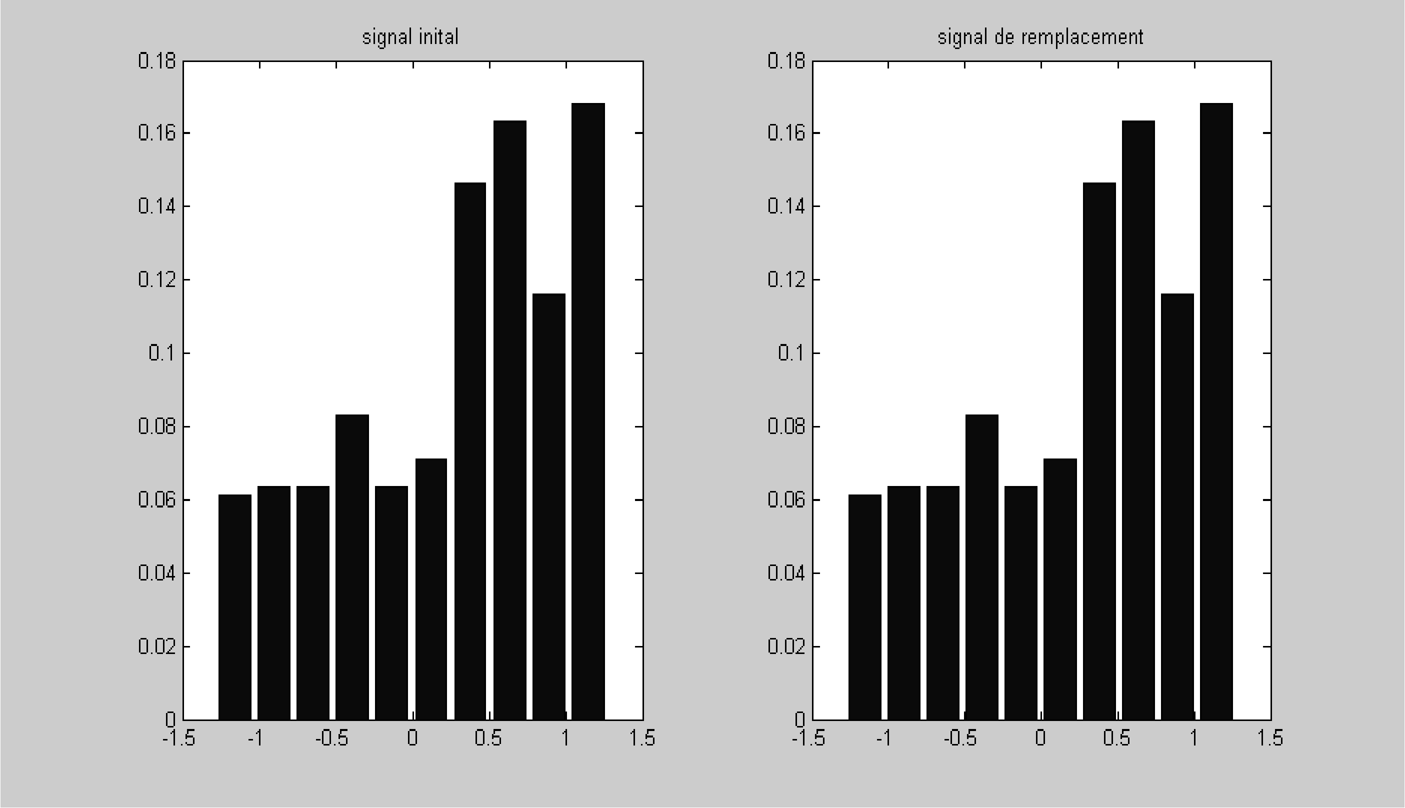
SURROGATE DATA (4)

- Example: surrogate for a chaotic signal

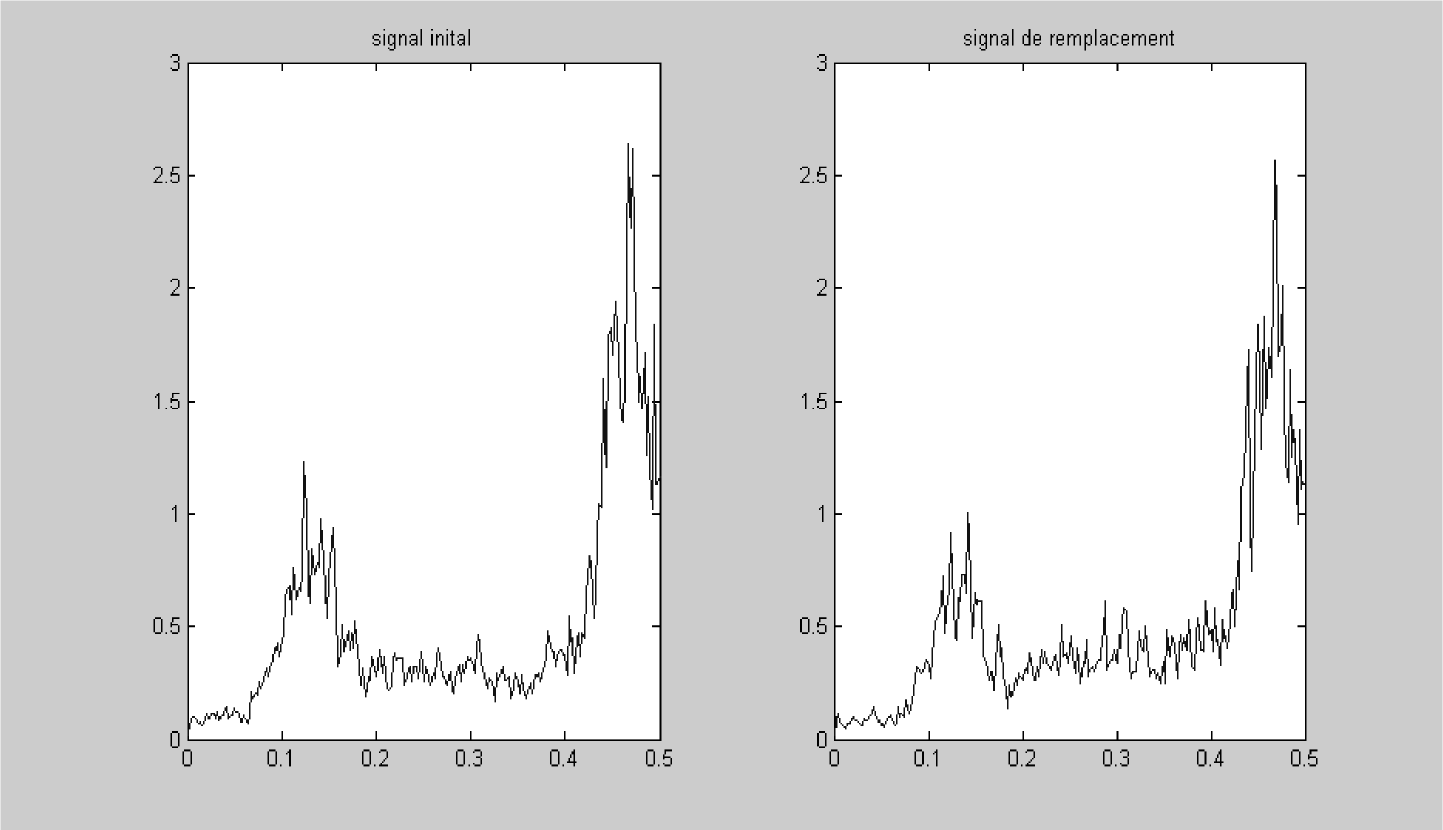


SURROGATE DATA (5)

Estimated probability density functions:

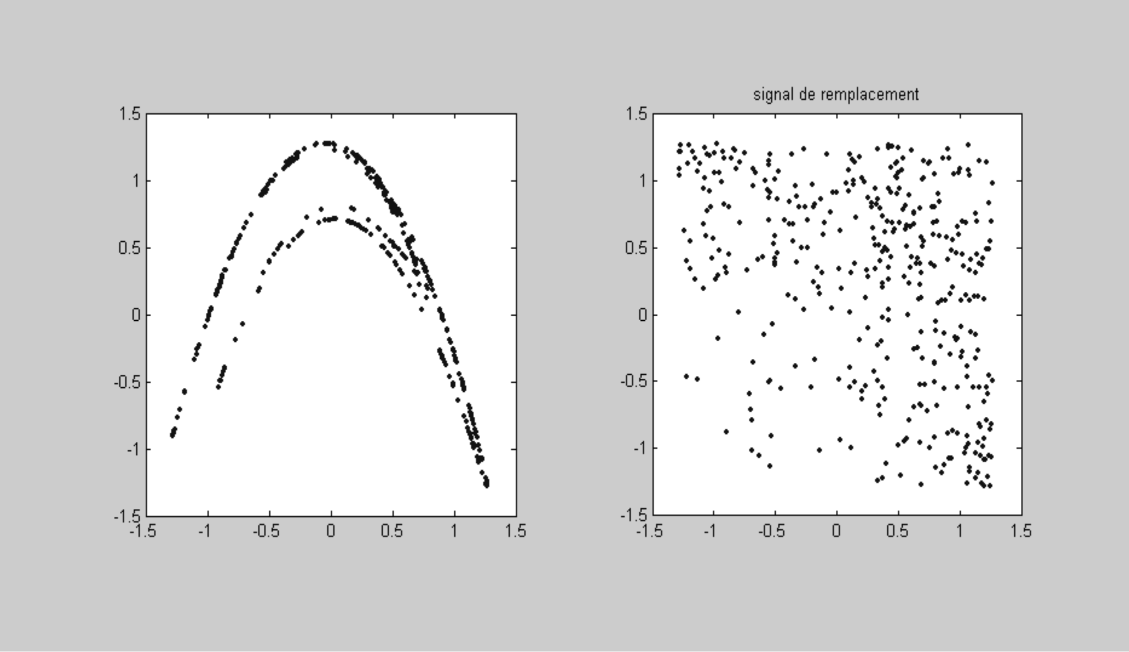


Estimated power spectra



SURROGATE DATA (7)

But in the state space…



The structure present in the initial signal has been destroyed.

REFERENCES

1. M. Pourahmadi, *Foundations of Time Series Analysis and Prediction Theory*, Wiley, NY, 2001.
2. S. P. Meyn and R. L. Tweedie, *Markov Chains and Stochastic Stability*, Springer-Verlag, London, 1993.
3. P. H. Franses and D. van Dijk, *Non-Linear Time Series Models in Empirical Finance*, Cambridge Univ. Press, 2000.
4. G. B. Giannakis and M. K. Tsatsanis, “Time-domain tests for Gaussianity and time-reversibility,“ *IEEE Trans. Sig. Proc.*, vol. 42, no. 12, pp. 3460-3472, Dec. 1994.
5. T. Schreiber and A. Schmitz, “Surrogate time series,“ *Physica D*, vol. 142, pp. 346-382, 2000.