Introduction
Some definitions
Hannan's theorem
Estimation of the covariance matrix
Tests
Gaussian non-parametric regression

The regression models with dependent errors

Emmanuel Caron [1]

Université Lumière Lyon 2, Laboratoire ERIC EA3083, 86 Rue Pasteur, 69007 Lyon

2 Mars 2020

[1]E-mail: emmanuel.caron@univ-lyon2.fr, Page web: http://ecaron.perso.math.cnrs.fr

Introduction
Some definitions
Hannan's theorem
Estimation of the covariance matrix
Tests
Gaussian non-parametric regression

1 **Introduction**

2 Some definitions

3 Hannan's theorem

4 Estimation of the covariance matrix

5 Tests

6 Gaussian non-parametric regression

Introduction
Some definitions
Hannan's theorem
Estimation of the covariance matrix
Tests
Gaussian non-parametric regression

Introduction
Time Series: CO2

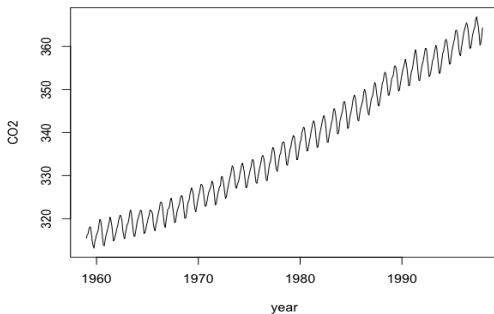A typical example to show the correlations between the observations:

Figure: CO2 rate as a function of time.

Introduction
Some definitions
Hannan's theorem
Estimation of the covariance matrix
Tests
Gaussian non-parametric regression

From Time Series to Linear Regression Model

$$Y_t = \underbrace{trend + seasonality}_{deterministic} + \underbrace{errors}_{random}.$$

Then:

$$Y = X\beta + \epsilon,$$

with:

$$X = \begin{pmatrix} 1 & 1^2 & 1^3 & \cos(\frac{2\pi}{3}) & \sin(\frac{2\pi}{3}) & \ldots & \cos(\frac{2\pi}{12}) & \sin(\frac{2\pi}{12}) \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ t & t^2 & t^3 & \cos(\frac{2\pi t}{3}) & \sin(\frac{2\pi t}{3}) & \ldots & \cos(\frac{2\pi t}{12}) & \sin(\frac{2\pi t}{12}) \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ n & n^2 & n^3 & \cos(\frac{2\pi n}{3}) & \sin(\frac{2\pi n}{3}) & \ldots & \cos(\frac{2\pi n}{12}) & \sin(\frac{2\pi n}{12}) \end{pmatrix}$$

Introduction
Some definitions
Hannan's theorem
Estimation of the covariance matrix
Tests
Gaussian non-parametric regression

ACF of the residuals

$\hat{\beta} = (X^t X)^{-1} X^t Y$: Least Squares Estimators, $\hat{\epsilon} = Y - \hat{Y}$: residuals.
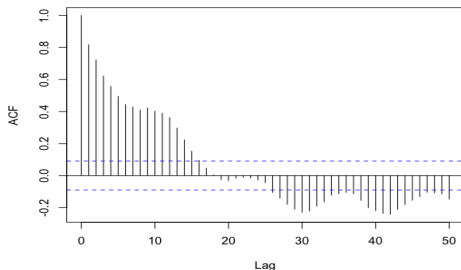

Figure: Autocorrelation of the residuals.

This is important for the applications to consider the dependency of the error process.

Introduction
Some definitions
Hannan's theorem
Estimation of the covariance matrix
Tests
Gaussian non-parametric regression

Goals

1. Investigate the linear regression model in the case where the errors are dependent

2. Modification of the usual results (confidence intervals, tests, ...). Focus on the "Fisher's test" and its calibration

3. Study the non-parametric regression model in the case where the errors are Gaussian and dependent, via a model selection approach.

Introduction
Some definitions
Hannan's theorem
Estimation of the covariance matrix
Tests
Gaussian non-parametric regression

Summary

1. Some definitions
2. Presentation of Hannan's Theorem (1973) [10]: convergence of the LSE in the stationary case under very mild conditions
3. Estimation of the asymptotic covariance matrix
4. Application: modification and calibration of the "Fisher's tests"
5. Gaussian non-parametric regression.

Introduction
Some definitions
Hannan's theorem
Estimation of the covariance matrix
Tests
Gaussian non-parametric regression

1. Introduction

2. Some definitions

3. Hannan's theorem

4. Estimation of the covariance matrix

5. Tests

6. Gaussian non-parametric regression

Introduction
Some definitions
Hannan's theorem
Estimation of the covariance matrix
Tests
Gaussian non-parametric regression

Some definitions
Linear model

Let us consider the regression linear model:

$$Y = X\beta + \epsilon,$$

where:

- X is a random or deterministic design, matrix of size $[n \times p]$
- Y is a n random vector of observations
- β is the p vector of unknown parameters
- ϵ are the errors and $\epsilon \in \mathbb{R}^n$. In the following, the error process is independent of the design X.

Introduction
Some definitions
Hannan's theorem
Estimation of the covariance matrix
Tests
Gaussian non-parametric regression

Least Squares Estimators

Let us recall the definition of the Least Squares Estimators (LSE):

$$\hat{\beta} = \mathrm{argmin}_{\beta \in \mathbb{R}^p} \|Y - X\beta\|_2^2 = (X^t X)^{-1} X^t Y,$$

($\|.\|_2$ = euclidean norm).

We have:

- $\hat{Y} = X\hat{\beta}$: Orthogonal Projection of Y on $\mathcal{M}_X = Vect\{X_{.,1}, \ldots, X_{.,p}\}$
- Residual vector: $\hat{\epsilon} = Y - \hat{Y} = Y - X\hat{\beta} \in \mathcal{M}_X^{\perp}$
- $\hat{\sigma}^2 = \frac{\|\hat{\epsilon}\|_2^2}{n-p}$.

Introduction
Some definitions
Hannan's theorem
Estimation of the covariance matrix
Tests
Gaussian non-parametric regression

Strict Stationarity

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space.

The error process $(\epsilon_i)_{i \in \mathbb{Z}}$ is defined on $(\Omega, \mathcal{F}, \mathbb{P})$, supposed strictly stationary with zero mean, and $\epsilon_0 \in \mathbb{L}^2(\Omega)$.

Definition: Strict Stationarity

A stochastic process $(\epsilon_i)_{i \in \mathbb{Z}}$ is said to be strictly stationary if the joint distributions of $(\epsilon_{t_1}, \ldots, \epsilon_{t_k})$ and $(\epsilon_{t_1+h}, \ldots, \epsilon_{t_k+h})$ are the same for all positive integers k and for all $t_1, \ldots, t_k, h \in \mathbb{Z}$.

Let $(\mathcal{F}_i)_{i \in \mathbb{Z}}$ be a non-decreasing filtration on $(\Omega, \mathcal{F}, \mathbb{P})$ defined as follows: $\mathcal{F}_i = \sigma(\epsilon_k, k \leq i)$.

Introduction
Some definitions
Hannan's theorem
Estimation of the covariance matrix
Tests
Gaussian non-parametric regression

Spectral density

Let us define the autocovariance function of the error process:

$$\gamma(k) = \mathrm{Cov}(\epsilon_m, \epsilon_{m+k}) = \mathbb{E}(\epsilon_m \epsilon_{m+k}),$$

and we denote by Γ_n the toeplitz covariance matrix:
$\Gamma_n = [\gamma(j - l)]_{1 \leq j, l \leq n}$.

Let f be the associated spectral density, that is the positive function on $[-\pi, \pi]$ such that:

$$\gamma(k) = \int_{-\pi}^{\pi} e^{ik\lambda} f(\lambda) d\lambda.$$

Introduction
Some definitions
Hannan's theorem
Estimation of the covariance matrix
Tests
Gaussian non-parametric regression

1. **Introduction**

2. **Some definitions**

3. **Hannan's theorem**

4. **Estimation of the covariance matrix**

5. **Tests**

6. **Gaussian non-parametric regression**

Introduction
Some definitions
Hannan's theorem
Estimation of the covariance matrix
Tests
Gaussian non-parametric regression

Hannan's theorem
Condition on the error process

In the following, we work conditionally at design X. Given X, Hannan (1973) [10] has proved a Central Limit Theorem in the stationary case for the usual LSE $\hat{\beta}$ under very mild conditions.

- Let $(P_j)_{j \in \mathbb{Z}}$ be a family of projection operators: $\forall j \in \mathbb{Z}$ and $\forall Z \in \mathbb{L}^2(\Omega)$: $P_j(Z) = \mathbb{E}(Z|\mathcal{F}_j) - \mathbb{E}(Z|\mathcal{F}_{j-1})$.
- **Hannan's condition** on the error process:

$$\sum_{i \geq 0} \|P_0(\epsilon_i)\|_{\mathbb{L}^2} < +\infty.$$

This implies the short memory: $\sum_k |\gamma(k)| < +\infty$.
Hannan's condition is satisfied for most of short-range dependent processes. (Linear Processes, Functions of linear processes (Dedecker, Merlevède, Volný (2007) [7]), Weakly dependent sequences (Dedecker and Prieur (2005) [8], Caron and Dede (2018) [5]), ...).

Introduction
Some definitions
Hannan's theorem
Estimation of the covariance matrix
Tests
Gaussian non-parametric regression

Hannan's conditions on the design

Let $X_{\cdot,j}$ be the column j of the matrix X, $j \in \{1, \ldots, p\}$, and $d_j(n)$ the euclidean norm of $X_{\cdot,j}$: $d_j(n) = \|X_{\cdot,j}\|_2 = \sqrt{\sum_{i=1}^n x_{i,j}^2}$.

Let $D(n)$ be the diagonal normalization matrix with diagonal term $d_j(n)$.

Conditions on the design:

- $\forall j \in \{1, \ldots, p\}, \qquad \lim_{n \to \infty} d_j(n) = \infty \qquad a.s.$

- $\forall j \in \{1, \ldots, p\}, \qquad \lim_{n \to \infty} \frac{\sup_{1 \leq i \leq n} |x_{i,j}|}{d_j(n)} = 0 \qquad a.s.,$

  and the following limits exist, $\forall j, l \in \{1, \ldots, p\}, k \in \{0, \ldots, n-1\}$:

- $\rho_{j,l}(k) = \lim_{n \to \infty} \sum_{m=1}^{n-k} \frac{x_{m,j} x_{m+k,l}}{d_j(n) d_l(n)} \qquad a.s.$

Introduction
Some definitions
Hannan's theorem
Estimation of the covariance matrix
Tests
Gaussian non-parametric regression

Hannan's theorem

Theorem (Hannan (1973) [10])

Under the previous conditions, for all bounded continuous function f:

$$\mathbb{E}\left(f\left(D(n)(\hat{\beta}-\beta)\right)\Big|X\right)\xrightarrow[n\to\infty]{a.s.}\mathbb{E}\left(f(Z)\Big|X\right),$$

where the distribution of Z given X is: $\mathcal{N}(0,C)$. Furthermore we have the convergence of second order moment:

$$\mathbb{E}\left(D(n)(\hat{\beta}-\beta)(\hat{\beta}-\beta)^{t}D(n)^{t}\Big|X\right)\xrightarrow[n\to\infty]{a.s.}C.$$

Remark

Let us notice that, by the dominated convergence theorem, we have for any bounded continuous function f:
$$\mathbb{E}\left(f\left(D(n)(\hat{\beta}-\beta)\right)\right)\xrightarrow[n\to\infty]{}\mathbb{E}\left(f(Z)\right).$$

Introduction
Some definitions
Hannan's theorem
Estimation of the covariance matrix
Tests
Gaussian non-parametric regression

1. **Introduction**

2. **Some definitions**

3. **Hannan's theorem**

4. **Estimation of the covariance matrix**

5. **Tests**

6. **Gaussian non-parametric regression**

Introduction
Some definitions
Hannan's theorem
Estimation of the covariance matrix
Tests
Gaussian non-parametric regression

Estimation of the covariance matrix

To obtain confidence regions or test procedures, one needs to estimate the limiting covariance matrix C. By Hannan, we have:

$$\mathbb{E}\left(D(n)(\hat{\beta}-\beta)(\hat{\beta}-\beta)^t D(n)^t \Big| X\right) \xrightarrow[n\to\infty]{a.s.} C,$$

and:

$$\mathbb{E}\left(D(n)(\hat{\beta}-\beta)(\hat{\beta}-\beta)^t D(n)^t \Big| X\right) = D(n)(X^t X)^{-1} X^t \Gamma_n X (X^t X)^{-1} D(n),$$

with $\Gamma_n = [\gamma(j-l)]_{1\leq j,l\leq n}$ (covariance matrix of the error process).

Consequently, we only need an estimator of Γ_n.

Introduction
Some definitions
Hannan's theorem
Estimation of the covariance matrix
Tests
Gaussian non-parametric regression

Residual-based Kernel estimator

Let us consider the following estimator of Γ_n:

$$\widehat{\Gamma}^*_{n,h_n} = \left[K\left(\frac{j-l}{h_n}\right) \hat{\gamma}^*_{j-l} \right]_{1 \leq j,l \leq n},$$

with[2] : $\hat{\gamma}^*_k = \frac{1}{n} \sum_{j=1}^{n-|k|} \hat{\epsilon}_j \hat{\epsilon}_{j+|k|}, \qquad 0 \leq |k| \leq (n-1).$

The function K is a kernel such that:

- K is nonnegative, symmetric, and $K(0) = 1$
- K has compact support
- the fourier transform of K is integrable.

The sequence of positive reals h_n is such that $h_n \to \infty$ and $\frac{h_n}{n} \to 0$ as $n \to \infty$.

---

[2]In our context, $(\epsilon_i)_{i \in \{1,\ldots,n\}}$ is not observed. Only the residuals $\hat{\epsilon}_i = Y_i - (x_i)^t \hat{\beta}$ are available, because only the data Y and the design X are observed.

Introduction
Some definitions
Hannan's theorem
Estimation of the covariance matrix
Tests
Gaussian non-parametric regression

Covariance matrix estimator

To estimate the asymptotic covariance matrix C, we use the estimator:

$$C_n = D(n)(X^tX)^{-1}X^t\widehat{\Gamma}^*_{n,h_n}X(X^tX)^{-1}D(n).$$

The coefficients of the matrices C_n and C are respectively denoted by $c_{n,(j,l)}$ and $c_{j,l}$, for all j,l in $\{1,...,p\}$.

Introduction
Some definitions
Hannan's theorem
Estimation of the covariance matrix
Tests
Gaussian non-parametric regression

Consistency

The following theorem proves, under mild conditions, the $\mathbb{L}^1$-norm consistency given X of the covariance matrix estimator:

Theorem (C. (2019) [4])

Let h_n be a sequence of positive reals such that $h_n \to \infty$ as $n \to \infty$, and:

$$h_n \mathbb{E}\left(|\epsilon_0|^2 \left(1 \wedge \frac{h_n}{n}|\epsilon_0|^2\right)\right) \xrightarrow[n \to \infty]{} 0.$$

Then, under the assumptions of Hannan's Theorem, the estimator C_n is consistent, that is for all j, l in $\{1, ..., p\}$:

$$\mathbb{E}\left(\left|c_{n,(j,l)} - c_{j,l}\right| \,\Big|\, X\right) \xrightarrow[n \to \infty]{} 0.$$

Introduction
Some definitions
Hannan's theorem
Estimation of the covariance matrix
Tests
Gaussian non-parametric regression

h_n condition

Corollary

Under the same conditions, the estimator C_n converges in probability to C as n tends to infinity.

The condition:

$$h_n \mathbb{E}\left(|\epsilon_0|^2 \left(1 \wedge \frac{h_n}{n}|\epsilon_0|^2\right)\right) \xrightarrow[n\to\infty]{} 0. \tag{1}$$

is a very general condition.

Remark

If $\epsilon_0 \in \mathbb{L}^2$, then there exists $h_n \to \infty$ such that (1) holds.
In particular, if ϵ_0 has a fourth order moment, then the condition is verified if $\frac{h_n}{\sqrt{n}} \to 0$.

Introduction
Some definitions
Hannan's theorem
Estimation of the covariance matrix
Tests
Gaussian non-parametric regression

Sketch of the proof

Let $V(X)$ be the matrix $\mathbb{E}\left(D(n)(\hat{\beta} - \beta)(\hat{\beta} - \beta)^t D(n)^t \Big| X\right)$, and let $v_{j,l}$ be its coefficients. By the triangle inequality, $\forall j, l \in \{1, \ldots, p\}$:

$$\left| c_{n,(j,l)} - c_{j,l} \right| \leq \left| v_{j,l} - c_{j,l} \right| + \left| c_{n,(j,l)} - v_{j,l} \right|.$$

Thanks to Hannan's Theorem:

$$\lim_{n \to \infty} \mathbb{E}\left(\left| v_{j,l} - c_{j,l} \right| \Big| X \right) = 0, \quad a.s.$$

Then it remains to prove that:

$$\lim_{n \to \infty} \mathbb{E}\left(\left| c_{n,(j,l)} - v_{j,l} \right| \Big| X \right) = 0, \quad a.s.$$

We have:

$$V(X) = D(n)(X^t X)^{-1} X^t \Gamma_n X (X^t X)^{-1} D(n)$$

$$C_n = D(n)(X^t X)^{-1} X^t \widehat{\Gamma}_{n,h_n}^* X (X^t X)^{-1} D(n).$$

Introduction
Some definitions
Hannan's theorem
Estimation of the covariance matrix
Tests
Gaussian non-parametric regression

Thanks to the convergence of $D_n(X^tX)^{-1}D_n$ (Hannan's conditions), it is sufficient to consider the matrices:

$$V' = D_n^{-1}X^t\Gamma_n X D_n^{-1}, \qquad C_n' = D_n^{-1}X^t\widehat{\Gamma}^*_{n,h_n}X D_n^{-1}.$$

We know that $\Gamma_n = \sum_{k=-n+1}^{n-1}\gamma(k)J_n^{(k)}$, where $J_n^{(k)}$ is a matrix with some 1 on the k-th diagonal. Thus we have:

$$D(n)^{-1}X^t\Gamma_n X D(n)^{-1} = \sum_{k=-n+1}^{n-1}\gamma(k)B_{k,n}$$

$$D(n)^{-1}X^t\widehat{\Gamma}^*_{n,h_n}X D(n)^{-1} = \sum_{k=-n+1}^{n-1}K\left(\frac{k}{h_n}\right)\hat{\gamma}^*_k B_{k,n},$$

with $B_{k,n} = D(n)^{-1}X^tJ_n^{(k)}X D(n)^{-1}$.

Introduction
Some definitions
Hannan's theorem
Estimation of the covariance matrix
Tests
Gaussian non-parametric regression

$$\left| c'_{n,(j,l)} - v'_{j,l} \right| = \left| \sum_{k=-n+1}^{n-1} \left(K\left(\frac{k}{h_n} \right) \hat{\gamma}^*_k - \gamma(k) \right) b^{k,n}_{j,l} \right|,$$

where $b^{k,n}_{j,l}$ is the coefficient (j,l) of the $B_{k,n}$ matrix.

We recall that:

$$f(\lambda) = \frac{1}{2\pi} \sum_{k=-\infty}^{\infty} \gamma(k) e^{ik\lambda}, \qquad \gamma(k) = \int_{-\pi}^{\pi} e^{ik\lambda} f(\lambda) d\lambda,$$

and:

$$f^*_n(\lambda) = \frac{1}{2\pi} \sum_{k=-n+1}^{n-1} K\left(\frac{|k|}{h_n} \right) \hat{\gamma}^*_k e^{ik\lambda}, \qquad K\left(\frac{|k|}{h_n} \right) \hat{\gamma}^*_k = \int_{-\pi}^{\pi} e^{ik\lambda} f^*_n(\lambda) d\lambda.$$

Introduction
Some definitions
Hannan's theorem
Estimation of the covariance matrix
Tests
Gaussian non-parametric regression

Then:

$$\sum_{k=-n+1}^{n-1} \left(K\left(\frac{k}{h_n}\right) \hat{\gamma}_k^* - \gamma(k) \right) B_{k,n} = \int_{-\pi}^{\pi} (f_n^*(\lambda) - f(\lambda)) \, g_n(\lambda)(d\lambda),$$

with:

$$g_n(\lambda) = \frac{1}{2\pi} \sum_{k=-(n-1)}^{n-1} e^{ik\lambda} B_{k,n},$$

in such a way that the matrices $B_{k,n}$ are the Fourier coefficients of the function $g_n(\lambda)$:

$$B_{k,n} = \int_{-\pi}^{\pi} e^{ik\lambda} g_n(\lambda) d\lambda.$$

Thus it remains to prove that, for all j, l in $\{1, \ldots, p\}$:

$$\lim_{n \to \infty} \mathbb{E} \left(\left| \int_{-\pi}^{\pi} (f_n^*(\lambda) - f(\lambda)) \, [g_n(\lambda)]_{j,l} d\lambda \right| \, \Big| \, X \right) = 0, \quad a.s.$$

Introduction
Some definitions
Hannan's theorem
Estimation of the covariance matrix
Tests
Gaussian non-parametric regression

We have:
$$\mathbb{E}\left(\left|\int_{-\pi}^{\pi}\left(f_n^*(\lambda)-f(\lambda)\right)[g_n(\lambda)]_{j,l}d\lambda\right|\,\bigg|\,X\right)$$

$$\leq \sup_{\lambda\in[-\pi,\pi]}\mathbb{E}\left(\left|f_n^*(\lambda)-f(\lambda)\right|\,\bigg|\,X\right)\int_{-\pi}^{\pi}\left|[g_n(\lambda)]_{j,l}\right|d\lambda,$$

because $[g_n(\lambda)]_{j,l}$ is measurable with respect to the σ-algebra generated by the design X.

Then, we can prove that:
$$\int_{-\pi}^{\pi}\left|[g_n(\lambda)]_{j,l}\right|d\lambda \leq 1.$$

Consequently:
$$\sup_{\lambda\in[-\pi,\pi]}\mathbb{E}\left(\left|f_n^*(\lambda)-f(\lambda)\right|\,\bigg|\,X\right)\int_{-\pi}^{\pi}\left|[g_n(\lambda)]_{j,l}\right|d\lambda$$

$$\leq \sup_{\lambda\in[-\pi,\pi]}\mathbb{E}\left(\left|f_n^*(\lambda)-f(\lambda)\right|\,\bigg|\,X\right).$$

Introduction
Some definitions
Hannan's theorem
Estimation of the covariance matrix
Tests
Gaussian non-parametric regression

Proof: Spectral density estimate

Let us consider the following estimator of the spectral density, for
$\lambda \in [-\pi, \pi]$: $f_n^*(\lambda) = \frac{1}{2\pi} \sum_{|k| \leq n-1} K\left(\frac{|k|}{h_n}\right) \hat{\gamma}_k^* e^{ik\lambda}$, where:
$\hat{\gamma}_k^* = \frac{1}{n} \sum_{j=1}^{n-|k|} \hat{\epsilon}_j \hat{\epsilon}_{j+|k|}, \qquad 0 \leq |k| \leq (n-1)$.

Theorem (C. and Dede (2018) [5])

Under the same assumptions of the consistency Theorem:

$$\sup_{\lambda \in [-\pi, \pi]} \|f_n^*(\lambda) - f(\lambda)\|_{\mathbb{L}^1} \xrightarrow[n \to \infty]{} 0.$$

This theorem has been proved for a fixed design X, but it remains true
with a random design:

$$\lim_{n \to \infty} \sup_{\lambda \in [-\pi, \pi]} \mathbb{E}\left(|f_n^*(\lambda) - f(\lambda)| \,\Big|\, X\right) = 0, \quad a.s.$$

The proof is complete.

Introduction
Some definitions
Hannan's theorem
Estimation of the covariance matrix
Tests
Gaussian non-parametric regression

Corollary (Hannan's theorem + Consistency theorem)

Corollary

Under the assumptions of Hannan's Theorem and the consistency theorem (Consistency of C_n), we get:

$$C_n^{-\frac{1}{2}} \left(D(n)(\hat{\beta} - \beta) \right) \xrightarrow[n \to \infty]{\mathcal{L}} \mathcal{N}(0, I_p),$$

where I_p is the $p \times p$ identity matrix.

Consequently, we can obtain confidence regions and tests for β in this dependent context.

Introduction
Some definitions
Hannan's theorem
Estimation of the covariance matrix
Tests
Gaussian non-parametric regression

Introduction
Some definitions
Hannan's theorem
Estimation of the covariance matrix
Tests
Gaussian non-parametric regression

Tests

- We are interested in test procedures on the linear model, particularly the "Fisher's" tests
- Thanks to the previous corollary, we can establish a new test statistic, so that the tests on the linear model always have an asymptotically good level, even when the underlying error process is dependent
- The level of a test (denoted by α) is the probabilty to choose H_1 hypothesis while H_0 is true.

Introduction
Some definitions
Hannan's theorem
Estimation of the covariance matrix
Tests
Gaussian non-parametric regression

"Fisher's" test: Dependent case

$H_0 : \beta_{j_1} = \ldots = \beta_{j_{p_0}} = 0$, against $H_1 : \exists j_z \in \{j_1, \ldots, j_{p_0}\}$ such that $\beta_{j_z} \neq 0$. If the error process is strictly stationary, we have:

$$C_{n_{p_0}}^{-1/2} \begin{pmatrix} d_{j_1}(n)(\hat{\beta}_{j_1} - \beta_{j_1}) \\ \vdots \\ d_{j_{p_0}}(n)(\hat{\beta}_{j_{p_0}} - \beta_{j_{p_0}}) \end{pmatrix} \xrightarrow[n \to \infty]{\mathcal{L}} \mathcal{N}(0_{p_0 \times 1}, I_{p_0}).$$

Then under H_0-hypothesis:

$$\begin{pmatrix} Z_{1,n} \\ \vdots \\ Z_{p_0,n} \end{pmatrix} = C_{n_{p_0}}^{-1/2} \begin{pmatrix} d_{j_1}(n)\hat{\beta}_{j_1} \\ \vdots \\ d_{j_{p_0}}(n)\hat{\beta}_{j_{p_0}} \end{pmatrix} \xrightarrow[n \to \infty]{\mathcal{L}} \mathcal{N}(0_{p_0 \times 1}, I_{p_0}),$$

and we define the following test statistic: $\Xi = Z_{1,n}^2 + \cdots + Z_{p_0,n}^2$. Under the H_0-hypothesis, $\Xi \xrightarrow[n \to \infty]{\mathcal{L}} \chi_{p_0}^2$.
In the same way, we can define an univariate test.

Introduction
Some definitions
Hannan's theorem
Estimation of the covariance matrix
Tests
Gaussian non-parametric regression

Bandwidth calibration

We have defined test procedures with a level asymptotically equal to α
(α to be determined, typically 5%).

Question: With a finite value of observations, how to choose the
bandwidth h_n in order to have well-calibrated tests and a non-asymptotic
level as close as possible to the wanted level α ?

Two main difficulties in our context:

1. Our target is the level of a test, which differs from classical
   approaches where the risk of an estimator is considered
2. We are not only in a context of dependent variables, but also in the
   very general framework of Hannan whose theorem applies for most
   stationary short-memory processes.

Consequently we can not use directly the classical methods of adaptive
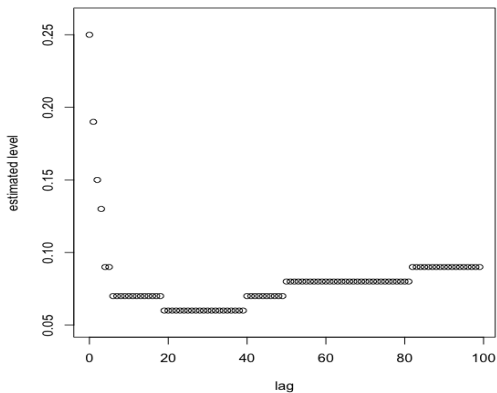statistics in our framework.

Introduction
Some definitions
Hannan's theorem
Estimation of the covariance matrix
Tests
Gaussian non-parametric regression

Figure: Level curve: estimated level as a function of the lag; $n = 1000$ observations, $T = 100$ simulations.

Introduction
Some definitions
Hannan's theorem
Estimation of the covariance matrix
Tests
Gaussian non-parametric regression

Empirical methods

- It is of first importance to provide hypothesis tests with correct significance levels
- We need data driven methods for the applications
- We partially answered to this problem by constructing empirical methods based on the data
- We propose a "plug-in" approach which consists in replacing the estimator of Γ_n. So we introduce the following estimator:

$$\widehat{C} = \widehat{C}(\widehat{\Gamma}_n) := D(n)(X^t X)^{-1} X^t \widehat{\Gamma}_n X (X^t X)^{-1} D(n),$$

and we use $\widehat{C}$ to compute the usual statistics of the linear model.

Introduction
Some definitions
Hannan's theorem
Estimation of the covariance matrix
Tests
Gaussian non-parametric regression

We have defined different ways to obtain the $\widehat{\Gamma}_n$ matrix:

1. by adapting an autoregressive process on the residual process and computing the theoretical covariances of the obtained AR(p) process. The order of the AR process is chosen by an AIC criterion

2. using the kernel estimator defined in Caron [4] with a bootstrap method to choose the value of the window (Wu and Pourahmadi (2009) [12])

3. by using an alternative choice of the window for the rectangular kernel (Efromovich (1998) [9])

4. in using an adaptive estimator of the spectral density via a histogram base (Comte (2001) [6]), with the slope heuristic algorithm to choose the dimension.

All these methods have been programmed on **R** in the *slm* package available on the CRAN.

Introduction
Some definitions
Hannan's theorem
Estimation of the covariance matrix
Tests
Gaussian non-parametric regression

Simulations

Let us define the three following processes:

1. AR(1) process (called "AR1"): $\epsilon_i - 0.7\epsilon_{i-1} = W_i$, where $W_i \sim \mathcal{N}(0, 1)$

2. MA(12) process (called "MA12"):
   $\epsilon_i = W_i + 0.5W_{i-2} + 0.3W_{i-3} + 0.2W_{i-12}$, where the (W_i)'s are i.i.d. random variables following Student's distribution with 10 degrees of freedom

3. A dynamical system (called "Sysdyn"): for $\gamma \in]0, 1[$, the intermittent map $\theta_\gamma : [0, 1] \mapsto [0, 1]$ introduced by Liverani, Saussol and Vaienti [11] is defined by

$$\theta_\gamma(x) = \begin{cases} x(1 + 2^\gamma x^\gamma) & \text{if} \quad x \in [0, 1/2[\\ 2x - 1 & \text{if} \quad x \in [1/2, 1]. \end{cases}$$

The Sysdyn process is then defined by $\varepsilon_i = \theta_\gamma^i$ (For the simulations, $\gamma = 1/4$). It is a non-mixing process (in the sense of Rosenblatt), with arithmetic decay of the correlations ($\sim \frac{1}{k^3}$ if $\gamma = 1/4$).

Introduction
Some definitions
Hannan's theorem
Estimation of the covariance matrix
Tests
Gaussian non-parametric regression

- Let us define the following linear regression model, for all i in $\{1, \ldots, n\}$ ($\beta_1 = 3$ and Z_i is a gaussian AR(1) process):

$$Y_i = \beta_1 + \beta_2(\log(i) + \sin(i) + Z_i) + \beta_3 i + \varepsilon_i$$

- We simulate a n-error process according to the AR1, the MA12 or the Sysdyn processes (small samples ($n = 200$) and larger ($n = 1000, 5000$))

- We simulate realizations of the linear regression model under the null hypothesis: $H_0 : \beta_2 = \beta_3 = 0$

- We make the test like described above

- The simulations are repeated 1000 times.

Introduction
Some definitions
Hannan's theorem
Estimation of the covariance matrix
Tests
Gaussian non-parametric regression

n	Method / Process	Fisher i.i.d.	fitAR	spectral proj	efromo vich	kernel
200	AR1 process	0.465	**0.097**	0.14	0.135	0.149
	Sysdyn process	0.385	**0.105**	0.118	0.124	0.162
	MA12 process	0.228	**0.113**	**0.113**	0.116	0.15
1000	AR1 process	0.418	0.043	**0.049**	**0.049**	0.086
	Sysdyn process	0.393	**0.073**	0.077	0.079	0.074
	MA12 process	0.209	0.064	0.066	0.069	**0.063**
5000	AR1 process	0.439	0.044	**0.047**	**0.047**	**0.047**
	Sysdyn process	0.381	0.058	0.061	**0.057**	0.064
	MA12 process	0.242	0.044	**0.048**	0.043	0.057

Table: Estimated levels.

Introduction
Some definitions
Hannan's theorem
Estimation of the covariance matrix
Tests
Gaussian non-parametric regression

1. **Introduction**

2. **Some definitions**

3. **Hannan's theorem**

4. **Estimation of the covariance matrix**

5. **Tests**

6. **Gaussian non-parametric regression**

Introduction
Some definitions
Hannan's theorem
Estimation of the covariance matrix
Tests
Gaussian non-parametric regression

Gaussian non-parametric regression

- Estimation of a non-random vector $f^* \in \mathbb{R}^n$ in the model:

$$Y = f^* + \epsilon,$$

  where $\epsilon \sim \mathcal{N}(0_{n \times 1}, \Sigma_{n \times n})$

- Study the regression model in the non-parametric case via a model selection approach

- Develop a model selection theory with penalization in the framework of Gaussian dependent variables

- Establish an oracle inequality for the minimal risk estimator among a collection of models.

Introduction
Some definitions
Hannan's theorem
Estimation of the covariance matrix
Tests
Gaussian non-parametric regression

Framework

- Estimation of a non-random vector $f^* \in \mathbb{R}^n$ in the model:

$$Y = f^* + \epsilon,$$

where $\epsilon \sim \mathcal{N}(0_{n \times 1}, \Sigma_{n \times n})$

- Let $\{S_m, m \in \mathcal{M}\}$ be a collection of finite-dimensional spaces. For all $m \in \mathcal{M}$:
  - $d_m = dim(S_m)$
  - $\hat{f}_m = Proj_{S_m}^{\perp} Y$: least squares estimator of f^* on S_m

- $\hat{f}_m$ minimizes the contrast function defined $\forall t \in S_m$ ($\|.\|_2$: euclidean norm):

$$\gamma(t) = \|Y - t\|_2^2$$

- Risk of an estimator $\hat{f}_m$:

$$R(\hat{f}_m) = \mathbb{E} \left[\left\| \hat{f}_m - f^* \right\|_2^2 \right].$$

Introduction
Some definitions
Hannan's theorem
Estimation of the covariance matrix
Tests
Gaussian non-parametric regression

- We want to select the oracle estimator $\hat{f}_{m_0}$:

$$m_0 \in \text{argmin}_{m \in \mathcal{M}}\{R(\hat{f}_m)\}$$

- Usually, risk $=$ bias^2 $+$ variance. Bias and variance have opposite behaviors according to the dimension
- The goal is to find the dimension that balances bias and variance
- We can not compute the risks directly because only the data are available. Definition of the empirical risk:

$$\widehat{R}(\hat{f}_m) = \frac{1}{n}\left\|Y - \hat{f}_m\right\|_2^2$$

- This typically leads to overfitting, then we have to penalize the larger models
- Thus we want to find $\hat{f}_{\hat{m}}$ such that:

$$\hat{m} \in \text{argmin}_{m \in \mathcal{M}}\left\{\left\|Y - \hat{f}_m\right\|_2^2 + pen(m)\right\},$$

where $pen : \mathcal{M} \to \mathbb{R}^+$ is a penalty function.

Introduction
Some definitions
Hannan's theorem
Estimation of the covariance matrix
Tests
Gaussian non-parametric regression

Penalty and oracle inequality

- Let $\pi = \{\pi_m, m \in \mathcal{M}\}$ be a distribution of probability on $\mathcal{M}$ associated with the collection of models $\{S_m, m \in \mathcal{M}\}$, such that $\sum_{m \in \mathcal{M}} e^{-\pi_m} < +\infty$
- $(\lambda_i)_{\{1 \leq i \leq n\}}$: eigenvalues of the covariance matrix Σ
- $\rho(\Sigma)$ its spectral radius $\rho(\Sigma) = \max_{1 \leq i \leq n} \lambda_i$.

According to the penalized criterion, for a constant $K > 1$ and the penalty:

$$pen(m) = K\rho(\Sigma) \left(\sqrt{d_m} + \sqrt{2 \log \left(\frac{1}{\pi_m} \right)} \right)^2,$$

we have the following theorem:

Introduction
Some definitions
Hannan's theorem
Estimation of the covariance matrix
Tests
Gaussian non-parametric regression

Theorem (Oracle inequality)

For the estimator $\hat{f}_{\hat{m}}$, there exists a constant $C_K > 1$, dependent only on $K > 1$, such that:

$$
\mathbb{E}\left[\left\|\hat{f}_{\hat{m}} - f^*\right\|_2^2\right] \leq C_K \min_{m \in \mathcal{M}} \left\{ \mathbb{E}\left[\left\|\hat{f}_m - f^*\right\|_2^2\right] \right.
$$
$$
\left. + \rho(\Sigma)\left(1 + \left(\sqrt{d_m} + \sqrt{2\log\left(\frac{1}{\pi_m}\right)}\right)^2\right)\right\}.
$$

Remark

If the Gaussian process ϵ is stationary and the spectral density is bounded, then the spectral radius is bounded and therefore the penalty is the same as in the i.i.d. case, up to a constant (cf. Birgé and Massart [1], [2]).

σ^2 has been replaced by $\rho(\Sigma)$.

Introduction
Some definitions
Hannan's theorem
Estimation of the covariance matrix
Tests
Gaussian non-parametric regression

Slope heuristic

- If the collection of model is not too large, the penalty is:

$$pen(m) = K'\rho(\Sigma)d_m$$

- For the applications, we calibrate the constant K' of the penalty from data

- Given the penalty, it is reasonable to use the slope heuristic method (Birgé and Massart [1], [2], [3])

- We use here the algorithm of dimension jump. Let $\hat{m}(K')$ be the selected model by the penalized criterion for a choice of K':

$$\hat{m}(K') \in \mathrm{argmin}_{m \in \mathcal{M}} \left\{ \left\| Y - \hat{f}_m \right\|_2^2 + K'\rho(\Sigma)d_m \right\}.$$

Introduction
Some definitions
Hannan's theorem
Estimation of the covariance matrix
Tests
Gaussian non-parametric regression

Simulations

- Let ϵ be the following ARMA(2,1) gaussian process
  $(W_i \sim \mathcal{N}(0, 0.5))$: $\epsilon_i - 0.4\epsilon_{i-1} - 0.2\epsilon_{i-2} = W_i + 0.3W_{i-1}$
- For all t in $[0, 1]$, $f^* = 3 - 0.1t + 0.5t^2 - t^3 + \sin(8t)$
- We generate a sample of size $n = 1000$, defined for all i in $\{1, \ldots, n\}$:

$$Y_i = f^* \left(\frac{i}{n} \right) + \epsilon_i$$

- The goal is to adapt a regressogram and choose the best regular partition to approach the f^* function.

For a dimension m, from 1 to 50, we split the interval $[0, 1]$ into m intervals and the estimator $\hat{f}_m$ is a piecewise constant function, equal to the average of Y_i on each interval.

Introduction
Some definitions
Hannan's theorem
Estimation of the covariance matrix
Tests
Gaussian non-parametric regression

This simulation is repeated 100 times and we obtain the following mean risk curve $\left(\left\| \hat{f}_m - f^* \right\|_2^2 \right)$
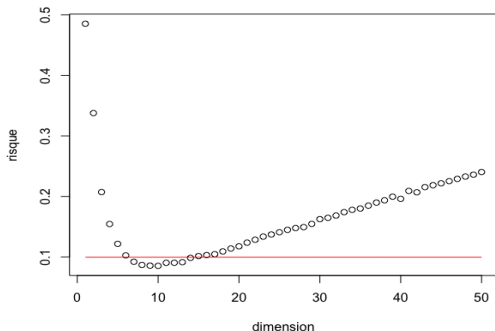
Figure: Mean risk curve for 100 simulations, and total mean risk of the method with slope heuristic (red line).

Introduction
Some definitions
Hannan's theorem
Estimation of the covariance matrix
Tests
Gaussian non-parametric regression

- Evaluation of the performance of the dimension jump algorithm
- We compute the risk $\left\| \hat{f}_m - Y \right\|_2^2$
- Then we use the slope heuristic method to choose the dimension (again the simulation is repeated 100 times).
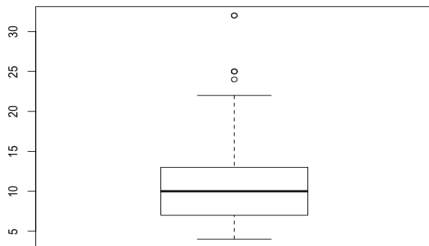
Figure: Boxplot with the dimensions selected by the dimension jump algorithm.

Introduction
Some definitions
Hannan's theorem
Estimation of the covariance matrix
Tests
Gaussian non-parametric regression

This represents the function f^* and its estimator with a regressogram of dimension 10 (dimension with the minimum average theoretical risk).
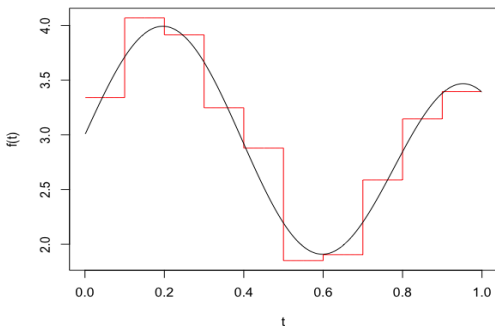
Figure: Function f^* (black) and the regressogram with dimension 10 (red).

Introduction
Some definitions
Hannan's theorem
Estimation of the covariance matrix
Tests
Gaussian non-parametric regression

Perspectives

- Gaussian non-parametric regression: deepen the obtained results in the gaussian short memory case. Then find the shape of the penalty in the gaussian long memory case
- Non-parametric regression: generalize the previous results to the non-Gaussian case
- Dependent variables in Statistical Learning

Introduction
Some definitions
Hannan's theorem
Estimation of the covariance matrix
Tests
Gaussian non-parametric regression

Thank you !

Introduction
Some definitions
Hannan's theorem
Estimation of the covariance matrix
Tests
Gaussian non-parametric regression

📄 L. Birgé and P. Massart.
Gaussian model selection.
Journal of the European Mathematical Society, 3(3):203–268, 2001.

📄 L. Birgé and P. Massart.
A generalized cp criterion for gaussian model selection.
2001.

📄 L. Birgé and P. Massart.
Minimal penalties for gaussian model selection.
Probability theory and related fields, 138(1-2):33–73, 2007.

📄 E. Caron.
Asymptotic distribution of least square estimators for linear models
with dependent errors.
Statistics, 53(4):885–902, 2019.

📄 E. Caron and S. Dede.
Asymptotic distribution of least squares estimators for linear models
with dependent errors: Regular designs.

Introduction
Some definitions
Hannan's theorem
Estimation of the covariance matrix
Tests
Gaussian non-parametric regression

Mathematical Methods of Statistics, 27(4):268–293, 2018.

📄 F. Comte.
Adaptive estimation of the spectrum of a stationary gaussian sequence.
Bernoulli, 7(2):267–298, 2001.

📄 J. Dedecker, F. Merlevède, and D. Volný.
On the weak invariance principle for non-adapted sequences under projective criteria.
Journal of Theoretical Probability, 20(4):971–1004, 2007.

📄 J. Dedecker and C. Prieur.
New dependence coefficients. examples and applications to statistics.

Probability Theory and Related Fields, 132(2):203–236, 2005.

📄 S. Efromovich.
Data-driven efficient estimation of the spectral density.

Introduction
Some definitions
Hannan's theorem
Estimation of the covariance matrix
Tests
Gaussian non-parametric regression

Journal of the American Statistical Association, 93(442):762–769, 1998.

📄 E. J. Hannan.
Central limit theorems for time series regression.
Probability theory and related fields, 26(2):157–170, 1973.

📄 C. Liverani, B. Saussol, and S. Vaienti.
A probabilistic approach to intermittency.
Ergodic theory and dynamical systems, 19(03):671–685, 1999.

📄 W. B. Wu and M. Pourahmadi.
Banding sample autocovariance matrices of stationary processes.
Statistica Sinica, pages 1755–1768, 2009.