HAL
open science

Asymptotic distribution of least square estimators for linear models with dependent errors

Emmanuel Caron

To cite this version:

Emmanuel Caron. Asymptotic distribution of least square estimators for linear models with dependent errors. Statistics, 2019, 53 (4), pp.885-902. 10.1080/02331888.2019.1593987 . hal-01815253v2

HAL Id: hal-01815253

https://hal.science/hal-01815253v2

Submitted on 15 Jun 2019

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Asymptotic distribution of least square estimators for linear models with dependent errors

Emmanuel Caron[a]

[a] Ecole Centrale Nantes, Laboratoire de Mathématiques Jean Leray UMR 6629, 1 Rue de la Noë, 44300 Nantes, France

ARTICLE HISTORY
Compiled June 15, 2019

Abstract
In this paper, we consider the usual linear regression model in the case where the error process is assumed strictly stationary. We use a result from Hannan [11], who proved a Central Limit Theorem for the usual least square estimator under general conditions on the design and on the error process. Whatever the design satisfying Hannan's conditions, we define an estimator of the covariance matrix and we prove its consistency under very mild conditions. As an application, we show how to modify the usual tests on the linear model in this dependent context, in such a way that the type-I error rate remains asymptotically correct, and we illustrate the performance of this procedure through different sets of simulations.

KEYWORDS
Stationary process, Linear regression model, Statistical Tests, Asymptotic normality, Spectral density estimates

1. Introduction

The linear regression model is used in many domains of applied mathematics, and the asymptotic behavior of the least square estimators is well known when the errors are i.i.d. (independent and identically distributed) random variables. Many authors have deepened the research on this subject, we can cite for example [3], [1], [2] and [12] among others. However, many science and engineering data exhibit significant temporal dependence so that the assumption of independence is violated (see for instance [6]). It is observed in astrophysics, geophysics, biostatistics, climatology, among others. Consequently all statistical procedures based on this assumption are not efficient and this can be very problematic for the applications.

In this paper, we propose to study the usual linear regression model in the very general framework of Hannan [11]. Let us consider the equation of the model:

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

The process $(\epsilon_i)_{i\in\mathbb{Z}}$ is assumed to be strictly stationary. The $n \times p$ matrix X is the design and can be random or deterministic. In our framework, we consider the interdependence of the variables of the design. As in Hannan, we assume that the design

Corresponding author. Email: emmanuel.caron@ec-nantes.fr

matrix X is independent of the error process. Such a model can be used for time series regression, but also in a more general context when the residuals seem to derive from a stationary correlated process.

Our work is based on the paper by Hannan [11], who proved a Central Limit Theorem for the usual least square estimator under general conditions on the design and on the error process. Let us quote that most of short-range dependent processes satisfies Hannan's conditions on the error process, for instance the class of linear processes with summable coefficients and square integrable innovations, a large class of functions of linear processes, many processes under various mixing conditions and the 2-strong stable processes introduced by Wu [15]. We refer to our previous paper [7], which presents many classes of short-range dependent processes satisfying Hannan's condition.

The linear regression model with dependent errors has also been studied under more restrictive conditions. For instance, Pagan and Nicholls [14] consider the case where the errors follow a $MA(q)$ process, and Chib and Greenberg [8] the case where the errors are an $ARMA(p, q)$ process. A more general framework is used by Wu [16] for a class of short-range dependent processes. These results are based on the asymptotic theory of stationary processes developed by Wu in [15]. However the class of processes satisfying the so called $\mathbb{L}^2$ "physical dependence measure" is included in the class of processes satisfying Hannan's condition (C1). In the present paper, we consider the very general framework of Hannan in order to obtain the most robust results.

In this paper, we present an estimator of the asymptotic covariance matrix of the normalized least square estimators of the parameters. This estimator is derived from the estimator of the spectral density of the error process introduced in [7]. Once the asymptotic covariance matrix is consistently estimated, it is then possible to obtain confidence regions and test procedures for the unknown parameter β. In particular, we shall use our general results to modify the usual Student and Fisher tests in cases where $(\epsilon_i)_{i\in\mathbb{Z}}$ and the design verify the conditions of Hannan, in order to have always a type-I error rate asymptotically correct (approximately equals to 5%).

The paper is organized as follows. In Section 2, we recall Hannan's Central Limit Theorem for the least square estimator. In Section 3, we focus on the estimation of the covariance matrix under Hannan's conditions. Finally, Section 4 is devoted to the correction of the usual Student and Fisher tests in our dependent context, and some simulations with different models are realized.

2. Hannan's theorem

2.1. Notations and definitions

Let us recall the equation of the linear regression model:

$$Y = X\beta + \epsilon, \tag{1}$$

where X is a design matrix and ϵ is an error process defined on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$. Let us notice that the error process ϵ is independent of the design X. Let $X_{.,j}$ be the column j of the matrix X, and $x_{i,j}$ the real number at the row i and the column j, where j is in $\{1, \ldots, p\}$ and i in $\{1, \ldots, n\}$. The random vectors Y and ϵ belong to $\mathbb{R}^n$ and β is a $p \times 1$ vector of unknown parameters.

Let $\|.\|_2$ be the usual euclidean norm on $\mathbb{R}^n$, and $\|.\|_{\mathbb{L}^p}$ be the $\mathbb{L}^p$-norm on Ω, defined for all random variable Z by: $\|Z\|_{\mathbb{L}^p} = [\mathbb{E}(Z^p)]^{\frac{1}{p}}$. We say that Z is in $\mathbb{L}^p(\Omega)$ if

$[\mathbb{E}(Z^p)]^{\frac{1}{p}} < \infty$.

The error process $(\epsilon_i)_{i\in\mathbb{Z}}$ is assumed to be strictly stationary with zero mean. Moreover, for all i in $\mathbb{Z}$, ϵ_i is supposed to be in $\mathbb{L}^2(\Omega)$. More precisely, the error process satisfies, for all i in $\mathbb{Z}$:

$$\epsilon_i = \epsilon_0 \circ \mathbb{T}^i,$$

where $\mathbb{T} : \Omega \to \Omega$ is a bijective bimeasurable transformation preserving the probability measure $\mathbb{P}$. Note that any strictly stationary process can be represented in this way.

Let $(\mathcal{F}_i)_{i\in\mathbb{Z}}$ be a non-decreasing filtration built as follows, for all i:

$$\mathcal{F}_i = \mathbb{T}^{-i}(\mathcal{F}_0),$$

where $\mathcal{F}_0$ is a sub-σ-algebra of $\mathcal{F}$ such that $\mathcal{F}_0 \subseteq \mathbb{T}^{-1}(\mathcal{F}_0)$. For instance, one can choose the past σ-algebra before time 0: $\mathcal{F}_0 = \sigma(\epsilon_k, k \leq 0)$, and then $\mathcal{F}_i = \sigma(\epsilon_k, k \leq i)$. In that case, ϵ_0 is $\mathcal{F}_0$-measurable.

As in Hannan, we shall always suppose that $\mathcal{F}_{-\infty} = \bigcap_{i\in\mathbb{Z}} \mathcal{F}_i$ is trivial. Moreover ϵ_0 is assumed $\mathcal{F}_\infty$-measurable. These implie that the ϵ_i's are all regular random variables in the following sense:

Definition 2.1.1 (Regular random variable)**.** *Let Z be a random variable in $L^1(\Omega)$. We say that Z is regular with respect to the filtration $(\mathcal{F}_i)_{i\in\mathbb{Z}}$ if $\mathbb{E}(Z|\mathcal{F}_{-\infty}) = \mathbb{E}(Z)$ almost surely and if Z is $\mathcal{F}_\infty$-measurable.*

Hence there exists a spectral density f for the error process, defined on $[-\pi, \pi]$. The autocovariance function γ of the process ϵ then satisfies:

$$\gamma(k) = \mathrm{Cov}(\epsilon_m, \epsilon_{m+k}) = \mathbb{E}(\epsilon_m \epsilon_{m+k}) = \int_{-\pi}^{\pi} e^{ik\lambda} f(\lambda) d\lambda.$$

Furthermore we denote by Γ_n the covariance matrix of the error process:

$$\Gamma_n = [\gamma(j-l)]_{1\leq j,l\leq n}. \tag{2}$$

2.2. Hannan's Central Limit Theorem

Let $\hat{\beta}$ be the usual least square estimator for the unknown vector β. Given the design X, Hannan [11] has shown a Central Limit Theorem for $\hat{\beta}$ when the error process is stationary. In this section, the conditions for applying this theorem are recalled.

Let $(P_j)_{j\in\mathbb{Z}}$ be a family of projection operators, defined for all j in $\mathbb{Z}$ and for any Z in $\mathbb{L}^2(\Omega)$ by:

$$P_j(Z) = \mathbb{E}(Z|\mathcal{F}_j) - \mathbb{E}(Z|\mathcal{F}_{j-1}).$$

We shall always assume that Hannan's condition on the error process is satisfied:

$$\sum_{i\in\mathbb{Z}} \|P_0(\epsilon_i)\|_{\mathbb{L}^2} < +\infty. \tag{C1}$$

Note that this condition implies that:

$$\sum_{k\in\mathbb{Z}} |\gamma(k)| < \infty, \tag{3}$$

(see for instance [9]).

Hannan's condition provides a very general framework for stationary processes. The hypothesis (C1) is a sharp condition to have a Central Limit Theorem for the partial sum sequence (see the paper of Dedecker, Merlevède and Volný [9] for more details). Notice that the condition (3) implies that the error process is short-range dependent. However, Hannan's condition is satisfied for most short-range dependent stationary processes. The reader can see the paper [7] where some examples checking Hannan's condition are developed.

Let us now recall Hannan's assumptions on the design. Let us introduce:

$$d_j(n) = \|X_{.,j}\|_2 = \sqrt{\sum_{i=1}^{n} x_{i,j}^2},$$

and let $D(n)$ be the diagonal matrix with diagonal term $d_j(n)$ for j in $\{1,\dots,p\}$.

Following Hannan, we also require that the columns of the design X satisfy, almost surely, the following conditions:

$$\forall j \in \{1,\dots,p\}, \qquad \lim_{n\to\infty} d_j(n) = \infty, \tag{C2}$$

and:

$$\forall j \in \{1,\dots,p\}, \qquad \lim_{n\to\infty} \frac{\sup_{1\le i\le n} |x_{i,j}|}{d_j(n)} = 0. \tag{C3}$$

Moreover, we assume that the following limits exist:

$$\forall j,l \in \{1,\dots,p\},\ k \in \{0,\dots,n-1\}, \qquad \rho_{j,l}(k) = \lim_{n\to\infty} \sum_{m=1}^{n-k} \frac{x_{m,j}x_{m+k,l}}{d_j(n)d_l(n)}. \tag{C4}$$

Note that Conditions (C2) and (C3) correspond to the usual Lindeberg's conditions for linear statistics in the i.i.d. case. In the dependent case, we also need Condition (C4).

The $p \times p$ matrix formed by the coefficients $\rho_{j,l}(k)$ is called $R(k)$:

$$R(k) = [\rho_{j,l}(k)] = \int_{-\pi}^{\pi} e^{ik\lambda} F_X(d\lambda), \quad a.s. \tag{4}$$

where F_X is the spectral measure associated with the matrix $R(k)$. The matrix $R(0)$ is supposed to be positive definite:

$$R(0) > 0, \quad a.s. \tag{C5}$$

Let then F and G be the matrices:

$$F = \frac{1}{2\pi}\int_{-\pi}^{\pi} F_X(d\lambda), \quad a.s.$$

$$G = \frac{1}{2\pi}\int_{-\pi}^{\pi} F_X(d\lambda)\otimes f(\lambda), \quad a.s.$$

The Central Limit Theorem for the regression parameter, due to Hannan [11], can be stated as follows:

Theorem 2.1. *Let $(\epsilon_i)_{i\in\mathbb{Z}}$ be a stationary process with zero mean. Assume that $\mathcal{F}_{-\infty}$ is trivial, ϵ_0 is $\mathcal{F}_\infty$-measurable, and that the sequence $(\epsilon_i)_{i\in\mathbb{Z}}$ satisfies Hannan's condition (C1). Assume that the design X satisfies, almost surely, the conditions (C2), (C3), (C4) and (C5). Then, for all bounded continuous function f:*

$$\mathbb{E}\left(f\left(D(n)(\hat{\beta}-\beta)\right)\Big|X\right) \xrightarrow[n\to\infty]{} \mathbb{E}\left(f(Z)\Big|X\right), \quad a.s. \tag{5}$$

where the distribution of Z given X is a gaussian distribution, with mean zero and covariance matrix equal to $F^{-1}GF^{-1}$. Furthermore, there is the convergence of the second order moment: [1]

$$\mathbb{E}\left(D(n)(\hat{\beta}-\beta)(\hat{\beta}-\beta)^t D(n)^t\Big|X\right) \xrightarrow[n\to\infty]{} F^{-1}GF^{-1}, \quad a.s. \tag{6}$$

Remark 2.1. *Let us notice that, by the dominated convergence theorem, the property (5) implies that for any bounded continuous function f,*

$$\mathbb{E}\left(f\left(D(n)(\hat{\beta}-\beta)\right)\right) \xrightarrow[n\to\infty]{} \mathbb{E}\left(f(Z)\right).$$

Remark 2.2. *In this remark, for the sake of clarity, we give a direct proof of (6). We shall see that, in fact, (6) holds under (3) and (C4) - (C5) (Hannan's condition (C1), which implies (3), is needed for (5) only). Moreover, this proof will serve as a preliminary to the proof of Theorem 3.1. We start from the exact expression of the second order moment:*

$$\begin{aligned}\mathbb{E}&\left(D(n)(\hat{\beta}-\beta)(\hat{\beta}-\beta)^t D(n)^t\Big|X\right)\\ &= D(n)(X^tX)^{-1}X^t\Gamma_n X(X^tX)^{-1}D(n)\\ &= \hat{R}(0)^{-1}\left(D(n)^{-1}X^t\Gamma_n X D(n)^{-1}\right)\hat{R}(0)^{-1},\end{aligned}$$

with $\hat{R}(0) = D(n)^{-1}X^tXD(n)^{-1}$. The $n\times n$ covariance matrix Γ_n is a symmetric Toeplitz matrix and is equal to:

$$\Gamma_n = \sum_{k=-n+1}^{n-1}\gamma(k)J_n^{(k)},$$

[1] The transpose of a matrix X is denoted by X^t.

where $J^{(k)}$ is the matrix with some 1 on the k-th diagonal and 0 elsewhere.

Hence, we deduce that:

$$\mathbb{E}\left(D(n)(\hat{\beta}-\beta)(\hat{\beta}-\beta)^t D(n)^t \middle| X\right) = \hat{R}(0)^{-1}\left(\sum_{k=-n+1}^{n-1}\gamma(k)B_{k,n}\right)\hat{R}(0)^{-1},$$

with:

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

For all k in $\{-n+1,\dots,n-1\}$, the matrices $B_{k,n}$ are equal to:

$$B_{k,n} = [\hat{\rho}_{j,l}(k)] \quad \text{if } k \geq 0, \qquad B_{k,n} = [\hat{\rho}_{j,l}(-k)] \quad \text{if } k \leq -1, \tag{7}$$

where $\hat{\rho}_{j,l}(k) = \sum_{m=1}^{n-k}\frac{x_{m,j}x_{m+k,l}}{d_j(n)d_l(n)}$. Under (C4), $\hat{\rho}_{j,l}(k)$ converges almost surely to $\rho_{j,l}(k)$.

By the dominated convergence theorem, since every term of $B_{k,n}$ is in $[-1,1]$, we deduce that:

$$\sum_{k=-n+1}^{n-1}\gamma(k)B_{k,n} \xrightarrow[n\to\infty]{} \sum_{k=-\infty}^{\infty}\gamma(k)B_k,$$

where $B_k = [\rho_{j,l}(k)]$ if $k \geq 0$ and $B_k = [\rho_{j,l}(-k)]$ if $k \leq -1$.

Since moreover $\hat{R}(0)$ converges almost surely to $R(0)$ (which is positively definite, see (C5)) as n tends to infinity, we conclude that:

$$\mathbb{E}\left(D(n)(\hat{\beta}-\beta)(\hat{\beta}-\beta)^t D(n)^t \middle| X\right) \xrightarrow[n\to\infty]{} R(0)^{-1}\left(\sum_{k=-\infty}^{\infty}\gamma(k)B_k\right)R(0)^{-1}.$$

Note that $R(0) = \int_{-\pi}^{\pi} F_X(d\lambda) = 2\pi F$ and $\sum_{k=-\infty}^{\infty}\gamma(k)B_k = 4\pi^2 G$, which is consistent with (6).

3. Estimation of the covariance matrix

To obtain confidence regions or test procedures from Theorem 2.1, one needs to estimate the limiting covariance matrix $F^{-1}GF^{-1}$. In this section, we propose an estimator of this covariance matrix, and we show its consistency under Hannan's conditions.

Let us first consider a preliminary random matrix defined as follows:

$$\widehat{\Gamma}_{n,h_n} = \left[K\left(\frac{j-l}{h_n}\right)\hat{\gamma}_{j-l}\right]_{1\leq j,l\leq n}, \tag{8}$$

with:

$$\hat{\gamma}_k = \frac{1}{n}\sum_{j=1}^{n-|k|}\epsilon_j\epsilon_{j+|k|}, \quad 0 \leq |k| \leq n-1.$$

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 tends to infinity and $\frac{h_n}{n}$ tends to 0 when n tends to infinity.

In our context, the errors $(\epsilon_i)_{1\leq i\leq n}$ are not observed. Only the residuals are available:

$$\hat{\epsilon}_j = Y_j - (x_j)^t\hat{\beta},$$

because only the data Y and the design X are observed. Consequently, we consider the following estimator of Γ_n:

$$\widehat{\Gamma}^*_{n,h_n} = \left[K\left(\frac{j-l}{h_n}\right)\hat{\gamma}^*_{j-l}\right]_{1\leq j,l\leq n}, \tag{9}$$

with:

$$\hat{\gamma}^*_k = \frac{1}{n}\sum_{j=1}^{n-|k|}\hat{\epsilon}_j\hat{\epsilon}_{j+|k|}, \quad 0 \leq |k| \leq n-1.$$

This estimator is a truncated version of the full matrix $\widehat{\Gamma}^*_n = \left[\hat{\gamma}^*_{j-l}\right]_{1\leq j,l\leq n}$, preserving the diagonal and some sub-diagonals. Following [4], $\widehat{\Gamma}^*_{n,h_n}$ is called the tapered covariance matrix estimator. The motivation for tapering comes from the fact that, for a large k, either $\gamma(k)$ is close to zero or $\hat{\gamma}^*_k$ is an unreliable estimate of $\gamma(k)$. Thus, prudent use of tapering may bring considerable computational economy in the former case, and statistical efficiency in the simulations, by keeping small or unreliable $\hat{\gamma}^*_k$ out of the calculations.

To estimate the asymptotic covariance matrix $F^{-1}GF^{-1}$, we use the estimator:

$$C_n = D(n)(X^tX)^{-1}X^t\widehat{\Gamma}^*_{n,h_n}X(X^tX)^{-1}D(n). \tag{10}$$

Let us denote by C the matrix $F^{-1}GF^{-1}$ and 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 $\{1, \ldots, p\}$. Our first result is the following:

Theorem 3.1. *Let h_n be a sequence of positive reals such that $h_n \to \infty$ as n tends to infinity, and:*

$$h_n\mathbb{E}\left(|\epsilon_0|^2\left(1 \wedge \frac{h_n}{n}|\epsilon_0|^2\right)\right) \xrightarrow[n\to\infty]{} 0. \tag{11}$$

Then, under the assumptions of Theorem 2.1, the estimator C_n is consistent, that is for all j, l in $\{1, \ldots, p\}$:

$$\mathbb{E}\left(\left|c_{n,(j,l)} - c_{j,l}\right| \middle| X\right) \xrightarrow[n\to\infty]{} 0, \quad a.s. \tag{12}$$

Remark 3.1. *If ϵ_0 is square integrable, then there exists $h_n \to \infty$ such that (11) holds.*

Furthermore if $\mathbb{E}\left(|\epsilon_0|^{\delta+2}\right) < \infty$ *with* δ *in* $]0,2]$*, then:*

$$h_n\mathbb{E}\left(|\epsilon_0|^2\left(1 \wedge \frac{h_n}{n}|\epsilon_0|^2\right)\right) \leq h_n\mathbb{E}\left(|\epsilon_0|^2\frac{h_n^{\delta/2}}{n^{\delta/2}}|\epsilon_0|^{\delta}\right) \leq \frac{h_n^{1+\delta/2}}{n^{\delta/2}}\mathbb{E}\left(|\epsilon_0|^{\delta+2}\right).$$

Thus, if h_n *satisfies* $\frac{h_n^{1+\delta/2}}{n^{\delta/2}} \xrightarrow[n\to\infty]{} 0$*, then* (11) *holds. In particular, if the random variable* ϵ_0 *has a fourth order moment, then the condition on* h_n *is* $\frac{h_n^2}{n} \xrightarrow[n\to\infty]{} 0$*.*

From this theorem, we get the non-conditional convergence in probability:

Corollary 3.1. *Let* h_n *be a sequence satisfying* (11)*. Then the estimator* C_n *converges in probability to* C *as* n *tends to infinity.*

Remark 3.2. *Since* $F^{-1}GF^{-1}$ *is assumed to be positive definite, our estimator* C_n *is also asymptotically positive definite. But it has no reason to be positive definite for any kernel and for any* n*. To overcome this problem, one can consider the estimator* $\tilde{C}_n$ *which is built as* C_n *but with a positive definite kernel, like for instance the triangular kernel.*

Indeed, following Wu [17], we can define:

$$\widehat{\Gamma}^*_{n,h_n} = \widehat{\Gamma}^*_n \star W_n,$$

where $\star$ *is the Hadamard (or Schur) product, which is formed by element-wise multiplication of matrices, and* W_n *is the kernel's matrix equal to* $\left[K\left(\frac{j-l}{h_n}\right)\right]_{1\leq j,l\leq p}$*. Let us notice that the full matrix* $\widehat{\Gamma}^*_n$ *is positive definite if and only if* $\hat{\gamma}^*_0 > 0$ *(see [6]). Consequently, by the Schur Product Theorem in matrix theory [13], since* $\hat{\Gamma}^*_n$ *and* W_n *are both positive definite, their Schur product* $\widehat{\Gamma}^*_{n,h_n}$ *is also positive definite.*

Let us recall that $C_n = \Psi\widehat{\Gamma}^*_{n,h_n}\Psi^t$ *with* $\Psi = D(n)(X^tX)^{-1}X^t$*. Then the estimator* C_n *is positive definite if for all* $x \neq 0$*,* $x^t\Psi\widehat{\Gamma}^*_{n,h_n}\Psi^t x$ *is strictly greater than* 0*. It is true if* $\hat{\gamma}^*_0 > 0$ *and if the design* X *is a rank* p *matrix.*

Combining Theorem 2.1 and Theorem 3.1, we obtain the following corollary, which is the main result of our paper:

Corollary 3.2. *Under the assumptions of Theorem 2.1 and Theorem 3.1, we get:*

$$C_n^{-\frac{1}{2}}\left(D(n)(\hat{\beta}-\beta)\right) \xrightarrow[n\to\infty]{\mathcal{L}} \mathcal{N}(0, I_p), \tag{13}$$

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

4. Tests and simulations

As an application of this main result, we show how to modify the usual tests on the linear regression model.

4.1. Tests

Let us recall the assumptions. We consider the linear regression model (1), and we assume that Hannan's condition (C1) as well as the conditions (C2) to (C5) on the design are satisfied. We also assume that ϵ_0 is $\mathcal{F}_\infty$-measurable and that $\mathcal{F}_{-\infty}$ is trivial. With these conditions, the usual tests can be modified and adapted to the case where the errors are short-range dependent and for any design verifying Hannan's conditions.

As usual, the null hypothesis H_0 means that the parameter β belongs to a vector space with dimension strictly smaller than p, and we denote by H_1 the alternative hypothesis (meaning that H_0 is not true, but (1) holds).

In order to test $H_0 : \beta_j = 0$ against $H_1 : \beta_j \neq 0$, for j in $\{1, \ldots, p\}$, under the H_0-hypothesis and according to Corollary 3.2 we have:

$$d_j(n)\hat{\beta}_j \xrightarrow[n\to\infty]{} \mathcal{N}(0, c_{j,j}).$$

We introduce the following univariate test statistic:

$$T_{j,n} = \frac{d_j(n)\hat{\beta}_j}{\sqrt{c_{n,(j,j)}}}. \tag{14}$$

Under the H_0-hypothesis, the distribution of $T_{j,n}$ converges to a standard normal distribution when n tends to infinity.

Now we want test 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 $\beta_{j_z} \neq 0$. By Corollary 3.2, it follows that:

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

where $C_{n_{p_0}}$ is the covariance matrix C_n built with removing the rows and the columns which do not belong to the discrete set $\{j_1, \ldots, j_{p_0}\}$. The $p_0 \times p_0$ identity matrix is denoted by I_{p_0} and $0_{p_0\times 1}$ is a p_0 vector of zeros.

Then under H_0-hypothesis, we have:

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

and we define the following test statistic:

$$\Xi = Z_{1,n}^2 + \cdots + Z_{p_0,n}^2. \tag{15}$$

Under the H_0-hypothesis, the distribution Ξ converges to a χ^2-distribution with parameter p_0.

For the simulations, we shall use for the estimator C_n the kernel K defined by:

$$\begin{cases} K(x) &= 1 & if\ |x| < 0.8 \\ K(x) &= 5 - 5|x| & if\ 0.8 \leq |x| \leq 1 \\ K(x) &= 0 & if\ |x| > 1. \end{cases} \tag{16}$$

This kernel verifies the conditions defined at the beginning of Section 3, and it is close to the rectangular kernel (whose Fourier transform is not integrable). Hence, the parameter h_n can be understood as the number of covariance terms that are necessary to obtain a good approximation of Γ_n. To choose its values, we shall use the graph of the empirical autocovariance of the residuals.

4.2. Simulations

We first simulate $(Z_1, \ldots, Z_n)$ according to the $AR(1)$ equation $Z_{k+1} = \frac{1}{2}(Z_k + \eta_{k+1})$, where Z_1 is uniformly distributed over $[0, 1]$ and $(\eta_i)_{i\geq 2}$ is a sequence of i.i.d. random variables with distribution $\mathcal{B}(1/2)$, independent of Z_1. The transition kernel of the chain $(Z_i)_{i\geq 1}$ is:

$$K(f)(x) = \frac{1}{2}\left(f\left(\frac{x}{2}\right) + f\left(\frac{x+1}{2}\right)\right),$$

and the uniform distribution on $[0, 1]$ is the unique invariant distribution by K. Hence, the chain $(Z_i)_{i\geq 1}$ is strictly stationary. Furthermore, it is not α-mixing in the sense of Rosenblatt [5], but it is $\tilde{\phi}$-dependent in the sense of Dedecker-Prieur [10] (see also Caron-Dede [7], Section 4.4). Indeed, one can prove that the coefficients $\tilde{\phi}(k)$ of the chain $(Z_i)_{i\geq 1}$ decrease geometrically [10]: $\tilde{\phi}(k) \leq 2^{-k}$. Let now Q_{0,σ^2} be the inverse of the cumulative distribution function of the law $\mathcal{N}(0, \sigma^2)$. Let then:

$$\epsilon_i = Q_{0,\sigma^2}(Z_i).$$

The sequence $(\epsilon_i)_{i\geq 1}$ is also a stationary Markov chain (as an invertible function of a stationary Markov chain), and one can easily check that its $\tilde{\phi}(k)$ coefficients are exactly equal to those of the sequence $(Z_i)_{i\geq 1}$ (hence, (ϵ_i) satisfies Hannan's condition (C1), see Section 5.1 in [7]). By construction, ϵ_i is $\mathcal{N}(0, \sigma^2)$-distributed, but the sequence $(\epsilon_i)_{i\geq 1}$ is not a Gaussian process (otherwise it would be mixing in the sense of Rosenblatt). Consequently Hannan's conditions are satisfied and the tests can be corrected as indicated above. For the simulations, let us notice that the variance σ^2 is chosen equal to 25.

The first model simulated with this error process is the following linear regression model, for all i in $\{1, \ldots, n\}$:

$$Y_i = \beta_0 + \beta_1(i^2 + X_i) + \epsilon_i, \tag{17}$$

with $(X_i)_{i\geq 1}$ a gaussian $AR(1)$ process (the variance is equal to 9), independent of the Markov chain $(\epsilon_i)_{i\geq 1}$. The coefficient β_0 is chosen equal to 3.

We test the hypothesis H_0: $\beta_1 = 0$, against the hypothesis H_1: $\beta_1 \neq 0$, thanks to the statistic $T_{j,n}$ defined above (14). The estimated level of the test will be studied for different choices of n and h_n, which is linked to the number of covariance terms considered. Under the hypothesis H_0, the same test is carried out 2000 times. Then

we look at the frequency of rejection of the test when we are under H_0, that is to say the estimated level of the test. Let us specify that we want an estimated level close to 5%.

- Case $\beta_1 = 0$ and $h_n = 1$ (no correction):

n	200	400	600	800	1000
Estimated level	0.203	0.195	0.183	0.205	0.202

Here, since $h_n = 1$, we do not estimate any of the covariance terms. The result is that the estimated levels are too large. This means that the test will reject the null hypothesis too often.

The parameter h_n may be chosen by analyzing the graph of the empirical autocovariances, Figure 1. For this example, the shape of the empirical autocovariance suggests to keep only 4 terms. This leads to choose $h_n = 5$.

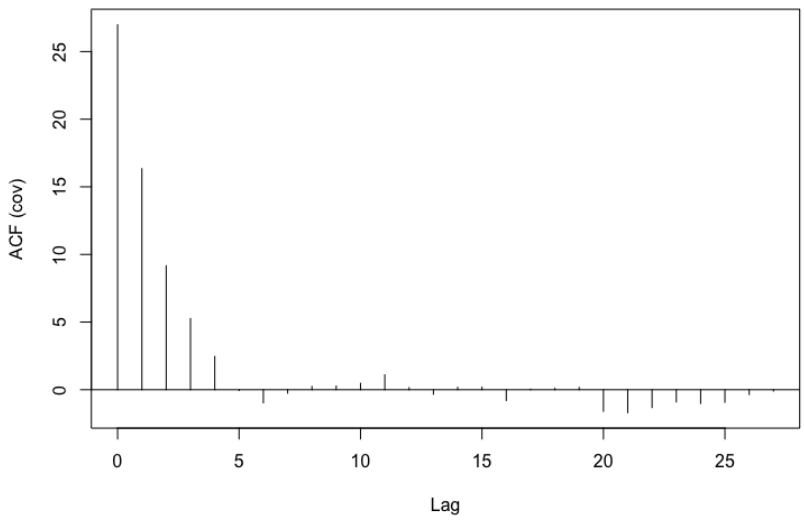

Figure 1. Empirical autocovariance of the residuals of Model (17).

- Case $\beta_1 = 0$, $h_n = 5$:

n	200	400	600	800	1000
Estimated level	0.0845	0.065	0.0595	0.054	0.053

As suggested by the graph of the empirical autocovariances, the choice $h_n = 5$ gives better estimated levels than $h_n = 1$. If one increases the size of the samples n, we are getting closer to the estimated level 5%. If $n = 2000$, the estimated level is around 0.05.

Let us notice that even for n moderately large (n approximately 200), it is much better to correct the test than not to do it. The estimated level goes from 20% to 8.5%.

- Case $\beta_1 = 0.00001$, $h_n = 5$:

In this example, H_0 is not satisfied. We choose β_1 equal to 0.00001, and we perform the same tests as above ($N = 2000$) to estimate the power of the test.

n	200	400	600	800	1000
Estimated power	0.1025	0.301	0.887	1	1

As one can see, the estimated power is always greater than 0.05, as expected. Still as expected, the estimated power increases with the size of the samples. For $n = 200$, the power of the test is around 0.10, and for $n = 800$, the power is around 1. As soon as $n = 800$, the test always rejects the H_0-hypothesis.

The second model considered is the following linear regression model, for all i in $\{1, \ldots, n\}$:

$$Y_i = \beta_0 + \beta_1(\log(i) + \sin(i) + X_i) + \beta_2 i + \epsilon_i \tag{18}$$

Here, we test the hypothesis H_0: $\beta_1 = \beta_2 = 0$ against H_1: $\beta_1 \neq 0$ or $\beta_2 \neq 0$, thanks to the statistic Ξ (15). The coefficient β_0 is equal to 3, and we use the same simulation scheme as above.

- Case $\beta_1 = \beta_2 = 0$ and $h_n = 1$ (no correction):

n	200	400	600	800	1000
Estimated level	0.348	0.334	0.324	0.3295	0.3285

As for the first simulation, if $h_n = 1$ the test will reject the null hypothesis too often.

As suggested by the graph of the estimated autocovariances Figure 2, it suggests to keep only 5 terms of covariances. Given the kernel (16), if we want to keep 5 terms of covariances, we must choose a bandwidth equal to $h_n = 6.25$ (because $\frac{5}{0.8} = 6.25$).

- Case $\beta_1 = \beta_2 = 0$, $h_n = 6.25$:

n	200	400	600	800	1000
Estimated level	0.09	0.078	0.066	0.0625	0.0595

Here, we see that the choice $h_n = 6.25$ works well. For $n = 1000$, the estimated level is around 0.06. If $n = 2000$ and $h_n = 6.25$, the estimated level is around 0.05.

- Case $\beta_1 = 0.2$, $\beta_2 = 0$, $h_n = 6.25$:

Now, we study the estimated power of the test. The coefficient β_1 is chosen equal to 0.2 and β_2 is equal to 0.

n	200	400	600	800	1000
Estimated power	0.33	0.5	0.6515	0.776	0.884

As expected, the estimated power increases with the size of the samples, and it is around 0.9 when $n = 1000$.

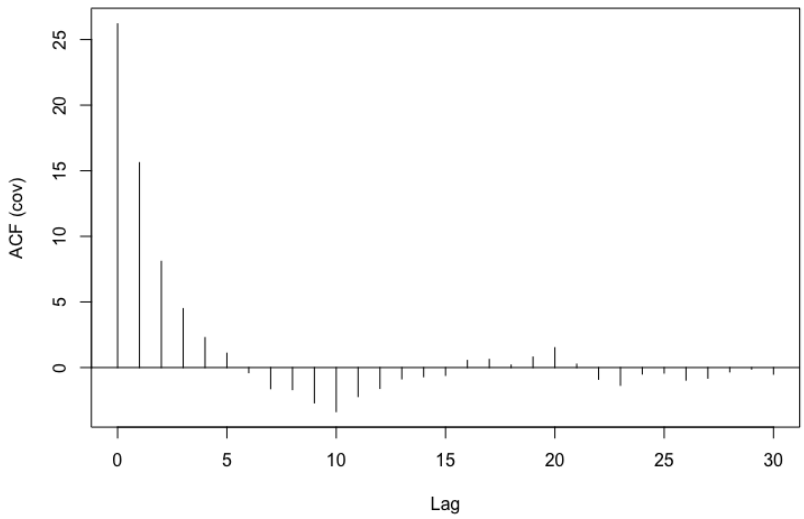

Figure 2. Empirical autocovariance of the residuals of Model (18).

5. Proofs

5.1. Theorem 3.1

Proof. In this proof, we use the notations introduced in Section 2 and Section 3. We denote by $V(X)$ the matrix equal to $\mathbb{E}\left(D(n)(\hat{\beta}-\beta)(\hat{\beta}-\beta)^t D(n)^t \middle| X\right)$ and by $v_{j,l}$ its coefficients.

By the triangle inequality, we have for all j, l in $\{1, \ldots, p\}$:

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

Thanks to Hannan's Theorem 2.1, we already know that:

$$\lim_{n\to\infty} \mathbb{E}\left(|v_{j,l} - c_{j,l}| \,\middle| X\right) = 0, \quad a.s. \tag{19}$$

Then it remains to prove that:

$$\lim_{n\to\infty} \mathbb{E}\left(\left|c_{n,(j,l)} - v_{j,l}\right| \,\middle| X\right) = 0, \quad a.s. \tag{20}$$

The matrix $V(X)$ is equal to:

$$D(n)(X^tX)^{-1}X^t\Gamma_n X(X^tX)^{-1}D(n),$$

with Γ_n defined in (2), and the estimator C_n:

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

with $\widehat{\Gamma}^*_{n,h_n}$ defined in (9). Thanks to the convergence of $D_n(X^tX)^{-1}D_n$ to $R(0)^{-1}$, it is sufficient to consider the matrices:

$$V' = D_n^{-1}X^t\Gamma_n X D_n^{-1},$$

and:

$$C'_n = D_n^{-1}X^t\widehat{\Gamma}^*_{n,h_n}XD_n^{-1}.$$

We know that $\Gamma_n = \sum_{k=-n+1}^{n-1}\gamma(k)J_n^{(k)}$ (see Remark 2.2 for the definition of $J_n^{(k)}$). Thus we have for V' and C'_n the following decomposition:

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

and:

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

with:

$$B_{0,n} = D(n)^{-1}X^tXD(n)^{-1}$$

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

and:

$$\hat{\gamma}^*_k = \frac{1}{n}\sum_{j=1}^{n-|k|}\hat{\epsilon}_j\hat{\epsilon}_{j+|k|}.$$

Let us compute:

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

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

$$\lim_{n\to\infty}\mathbb{E}\left(\left|\sum_{k=-n+1}^{n-1}\left(K\left(\frac{k}{h_n}\right)\hat{\gamma}^*_k - \gamma(k)\right)b^{k,n}_{j,l}\right|\Big|X\right) = 0, \quad a.s.$$

We recall that:

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

where the coefficients $\gamma(k)$ are the Fourier coefficients of the spectral density $f(\lambda)$. We have:

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

and the coefficients $K\left(\frac{k}{h_n}\right) \hat{\gamma}_k^*$ are the Fourier coefficients of the spectral density's estimator $f_n^*(\lambda)$. Let us define:

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

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

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

Consequently we can deduce that:

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

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

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

We have:

$$\begin{aligned} \mathbb{E}\left(\left| \int_{-\pi}^{\pi} \left(f_n^*(\lambda) - f(\lambda)\right) [g_n(\lambda)]_{j,l} d\lambda \right| \Big| X \right) &\leq \mathbb{E}\left(\int_{-\pi}^{\pi} \left|f_n^*(\lambda) - f(\lambda)\right| \left|[g_n(\lambda)]_{j,l}\right| d\lambda \Big| X \right) \\ &\leq \int_{-\pi}^{\pi} \left|[g_n(\lambda)]_{j,l}\right| \mathbb{E}\left(\left|f_n^*(\lambda) - f(\lambda)\right| \Big| X \right) d\lambda, \end{aligned}$$

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

$$\begin{aligned} &\int_{-\pi}^{\pi} \left|[g_n(\lambda)]_{j,l}\right| \mathbb{E}\left(\left|f_n^*(\lambda) - f(\lambda)\right| \Big| X \right) d\lambda \\ &\qquad \leq \sup_{\lambda\in[-\pi,\pi]} \mathbb{E}\left(\left|f_n^*(\lambda) - f(\lambda)\right| \Big| X \right) \int_{-\pi}^{\pi} \left|[g_n(\lambda)]_{j,l}\right| d\lambda. \end{aligned}$$

Theorem 3.1 of our paper [7] states that:

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

for a fixed design X and for the particular kernel defined by: $K(x) = \mathbb{1}_{\{|x|\leq 1\}} + (2 - |x|)\mathbb{1}_{\{1\leq|x|\leq 2\}}$. But a quick look to the proof of this theorem suffices to see that this result is available for any design X, conditionally to X:

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

Furthermore, this result is still available for all kernel K verifying the conditions at the beginning of Section 3.

Thus it remains to find a bound for:

$$\int_{-\pi}^{\pi} |[g_n(\lambda)]_{j,l}| \, d\lambda.$$

Let us recall (see (7)) that the matrices $B_{k,n}$ are equal to, for all k in $\{-n+1, \ldots, n-1\}$:

$$B_{k,n} = [\hat{\rho}_{j,l}(k)], \quad \text{if } k \geq 0, \qquad B_{k,n} = [\hat{\rho}_{j,l}(-k)], \quad \text{if } k \leq -1.$$

By definition we have:

$$\hat{\rho}_{j,l}(k) = \frac{\hat{\gamma}_{j,l}(k)}{\sqrt{\hat{\gamma}_{j,j}(0)\hat{\gamma}_{l,l}(0)}}. \tag{21}$$

For a multivariate time series, let us recall that the cross-periodogram is defined by, for all j, l in $\{1, \ldots, p\}$ [6]:

$$[I_n(\lambda)]_{j,l} = \frac{1}{2\pi} \sum_{k=-n+1}^{n-1} \hat{\gamma}_{j,l}(k) e^{ik\lambda}. \tag{22}$$

Combining (21) and (22), the function $g_n(\lambda)$ is equal to, for all j, l in $\{1, \ldots, p\}$:

$$[g_n(\lambda)]_{j,l} = \frac{[I_n(\lambda)]_{j,l}}{\sqrt{\hat{\gamma}_{j,j}(0)\hat{\gamma}_{l,l}(0)}} = \frac{1}{2\pi} \sum_{k=-n+1}^{n-1} b_{j,l}^{k,n} e^{ik\lambda}.$$

Then using the definition of the coherence [6], we get:

$$\begin{aligned} |[g_n(\lambda)]_{j,l}| = \frac{|[I_n(\lambda)]_{j,l}|}{\sqrt{\hat{\gamma}_{j,j}(0)\hat{\gamma}_{l,l}(0)}} &\leq \sqrt{\frac{[I_n(\lambda)]_{j,j}[I_n(\lambda)]_{l,l}}{\hat{\gamma}_{j,j}(0)\hat{\gamma}_{l,l}(0)}} \\ &\leq \sqrt{[g_n(\lambda)]_{j,j}[g_n(\lambda)]_{l,l}} \leq \frac{1}{2}[g_n(\lambda)]_{j,j} + \frac{1}{2}[g_n(\lambda)]_{l,l}. \end{aligned}$$

Consequently, we have:

$$\int_{-\pi}^{\pi} |[g_n(\lambda)]_{j,l}|\, d\lambda \leq \frac{1}{2}\int_{-\pi}^{\pi} [g_n(\lambda)]_{j,j} d\lambda + \frac{1}{2}\int_{-\pi}^{\pi} [g_n(\lambda)]_{l,l} d\lambda \leq \frac{1}{2}[B_{0,n}]_{j,j} + \frac{1}{2}[B_{0,n}]_{l,l} \leq 1,$$

because $[B_{0,n}]_{j,j} = \hat{\rho}_{j,j}(0) = 1$ and $[B_{0,n}]_{l,l} = \hat{\rho}_{l,l}(0) = 1$.

We deduce that, for all j, l in $\{1, \ldots, p\}$:

$$\begin{aligned}
&\mathbb{E}\left(\left|\int_{-\pi}^{\pi} (f_n^*(\lambda) - f(\lambda))\, [g_n(\lambda)]_{j,l} d\lambda\right| \Big| X\right) \\
&\qquad \leq \sup_{\lambda\in[-\pi,\pi]} \mathbb{E}\left(|f_n^*(\lambda) - f(\lambda)| \Big| X\right) \int_{-\pi}^{\pi} |[g_n(\lambda)]_{j,l}|\, d\lambda \\
&\qquad\qquad \leq \sup_{\lambda\in[-\pi,\pi]} \mathbb{E}\left(|f_n^*(\lambda) - f(\lambda)| \Big| X\right).
\end{aligned}$$

Since we know that:

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

we have proved that, for all j, l in $\{1, \ldots, p\}$:

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

□

5.2. Corollary 3.1

Proof. We want to prove that, for all j, l in $\{1, \ldots, p\}$, $c_{n,(j,l)}$ converges in probability to $c_{j,l}$ as n tends to infinity, that is, for all $\epsilon > 0$:

$$\mathbb{E}\left(\mathbb{1}_{|c_{n,(j,l)} - c_{j,l}| > \epsilon}\right) \xrightarrow[n\to\infty]{} 0.$$

We have:

$$\mathbb{E}\left(\mathbb{1}_{|c_{n,(j,l)} - c_{j,l}| > \epsilon}\right) = \mathbb{E}\left(\mathbb{E}\left(\mathbb{1}_{|c_{n,(j,l)} - c_{j,l}| > \epsilon} \Big| X\right)\right).$$

Thanks to Theorem 3.1 and to Markov's inequality, we have almost surely:

$$\mathbb{E}\left(\mathbb{1}_{|c_{n,(j,l)} - c_{j,l}| > \epsilon} | X\right) \leq \frac{\mathbb{E}\left(\left|c_{n,(j,l)} - c_{j,l}\right| \Big| X\right)}{\epsilon} \xrightarrow[n\to\infty]{} 0.$$

Then, using the dominated convergence theorem, we get:

$$\mathbb{E}\left(\mathbb{E}\left(\mathbb{1}_{|c_{n,(j,l)} - c_{j,l}| > \epsilon} \Big| X\right)\right) \xrightarrow[n\to\infty]{} 0.$$

□

References

[1] G. J. Babu. Strong representations for lad estimators in linear models. Probability Theory and Related Fields, 83(4):547–558, 1989.

[2] Z. Bai, C. R. Rao, and Y. Wu. M-estimation of multivariate linear regression parameters under a convex discrepancy function. Statistica Sinica, pages 237–254, 1992.

[3] G. Bassett Jr and R. Koenker. Asymptotic theory of least absolute error regression. Journal of the American Statistical Association, 73(363):618–622, 1978.

[4] P. J. Bickel and E. Levina. Regularized estimation of large covariance matrices. The Annals of Statistics, pages 199–227, 2008.

[5] R. C. Bradley. Basic properties of strong mixing conditions. In Dependence in probability and statistics (Oberwolfach, 1985), volume 11 of Progr. Probab. Statist., pages 165–192. Birkhäuser Boston, Boston, MA, 1986.

[6] P. J. Brockwell and R. A. Davis. Time series: theory and methods. Springer Science & Business Media, 2013.

[7] E. Caron and S. Dede. Asymptotic distribution of least squares estimators for linear models with dependent errors: Regular designs. Mathematical Methods of Statistics, 27(4):268–293, 2018.

[8] S. Chib and E. Greenberg. Bayes inference in regression models with arma (p, q) errors. Journal of Econometrics, 64(1-2):183–206, 1994.

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

[10] J. Dedecker and C. Prieur. New dependence coefficients. examples and applications to statistics. Probability Theory and Related Fields, 132(2):203–236, 2005.

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

[12] X. He, Q.-M. Shao, et al. A general bahadur representation of m-estimators and its application to linear regression with nonstochastic designs. The Annals of Statistics, 24(6):2608–2630, 1996.

[13] R. A. Horn, R. A. Horn, and C. R. Johnson. Matrix analysis. Cambridge university press, 1990.

[14] A. Pagan and D. Nicholls. Exact maximum likelihood estimation of regression models with finite order moving average errors. The Review of Economic Studies, 43(3):383–387, 1976.

[15] W. B. Wu. Nonlinear system theory: Another look at dependence. Proceedings of the National Academy of Sciences of the United States of America, 102(40):14150–14154, 2005.

[16] W. B. Wu et al. M-estimation of linear models with dependent errors. The Annals of Statistics, 35(2):495–521, 2007.

[17] H. Xiao, W. B. Wu, et al. Covariance matrix estimation for stationary time series. The Annals of Statistics, 40(1):466–493, 2012.