THÈSE DE DOCTORAT DE

L'ÉCOLE CENTRALE DE NANTES
COMUE **U**NIVERSITÉ **B**RETAGNE **L**OIRE

École Doctorale N°601
Mathématiques et Sciences et Technologies de l'Information et de la Communication
Spécialité : Mathématiques et leurs Interactions

Par

« **Emmanuel Caron** »

« Comportement des estimateurs des moindres carrés du modèle linéaire dans un contexte dépendant »

«Étude asymptotique, implémentation, exemples»

Thèse présentée et soutenue à NANTES , le 17 octobre 2019
Unité de recherche : Laboratoire de Mathématiques Jean Leray UMR CNRS 6629

Rapporteurs avant soutenance :

Pierre Alquier, Research Scientist at RIKEN AIP
Jean-Marc Bardet, Professeur des Universités à l'Université Paris 1 - Panthéon-Sorbonne

Composition du jury :

Rapporteurs : Pierre Alquier, Research Scientist at RIKEN AIP
Jean-Marc Bardet, Professeur des Universités à l'Université Paris 1 - Panthéon-Sorbonne
Examinateurs : Fabienne Comte, Professeur des Universités à l'Université Paris Descartes
Anne Philippe, Professeur des Universités à l'Université de Nantes
Clémentine Prieur, Professeur des Universités à l'Université Grenoble Alpes

Dir. de thèse : Bertrand Michel, Professeur des Universités à l'École Centrale de Nantes

Co-dir. de thèse : Jérôme Dedecker, Professeur des Universités à l'Université Paris Descartes

REMERCIEMENTS

Étant d'une nature plutôt laconique, les remerciements seront sans doute assez courts. Soyez cependant tous assurés que mes pensées sont sincères et remplies de gratitude.

Je souhaite, en premier lieu, dire la profonde reconnaissance que j'éprouve à l'égard de mes directeurs de thèse, Jérôme Dedecker et Bertrand Michel, pour la patience, la gentillesse et la confiance qu'ils ont bien voulu m'accorder tout au long de ce travail. J'ai pu partager leur savoir et leurs brillantes intuitions. Qu'ils soient vivement remerciés de leur permanente disponibilité et des nombreux encouragements prodigués.

Je veux également remercier Olivier Bouaziz et Frédéric Lavancier qui, bien au-delà de leur rôle de « Comité de Suivi de Thèse », ont représenté de véritables soutiens.

Mes remerciements vont aussi à Pierre Alquier et Jean-Marc Bardet pour l'honneur qu'ils m'ont fait en acceptant d'être rapporteurs de cette thèse et pour le temps et les conseils qu'ils ont bien voulu me donner.

Tous mes remerciements vont également aux examinatrices, Fabienne Comte, Anne Philippe et Clémentine Prieur qui ont accepté de participer à mon jury de thèse.
Une pensée particulière pour Anne Philippe, qui a toujours été présente chaque fois que j'en avais besoin. Merci encore pour ses nombreux conseils et encouragements.

Aymeric Stamm a su être présent lorsqu'il a fallu compléter l'apprentissage du logiciel R et je lui en suis très reconnaissant.

Pendant l'exercice de mes fonctions à l'Ecole Centrale de Nantes, Françoise Foucher a toujours pu m'aider et m'encadrer à propos des enseignements que j'ai donnés ; qu'elle en soit vivement remerciée.

Je veux penser encore à tous mes camarades de thèse avec lesquels j'ai échangé et discuté à propos du travail qui nous motivait tous. Je ne ferai pas de liste : un grand merci pour votre bonne humeur et votre soutien.

Une pensée particulière va à mes parents et mes deux frères : je leur dois beaucoup et c'est grâce à eux si je peux écrire ces lignes aujourd'hui.

Mes derniers remerciements vont à Agnieszka qui a toujours su m'aider, me soutenir et me supporter dans tout ce que j'ai entrepris.

Que les nombreuses personnes qui ne sont pas citées mais qui ont été si précieuses pour m'accompagner durant toutes ces années sachent ici combien elles ont compté.

TABLE OF CONTENTS

INTRODUCTION

L'analyse des données est une composante des mathématiques dont l'ampleur ne cesse de croître devant l'afflux massif et quotidien de toutes sortes de données. Pour analyser ces dernières, le statisticien dispose d'une quantité de méthodes différentes.
L'une d'entre elles est particulièrement utilisée dans beaucoup de domaines scientifiques ; il s'agit du modèle de régression linéaire. Ce dernier cherche à établir une relation linéaire entre une variable, dite expliquée, et plusieurs variables, dites explicatives. L'hypothèse fondamentale faite sur ce modèle est que les erreurs sont des variables aléatoires i.i.d. [1] et le comportement asymptotique de l'estimateur des moindres carrés est bien connu dans ce cas.

Beaucoup d'auteurs ont approfondi les recherches sur ce modèle comme par exemple, Chatterjee et Hadi [27], Wolak [73], Bai, Rao et Wu [10], Davies [33], Drygas [40], Bassett et Koenker [14], Hoerl et Kennard [50], Tibshirani [72]. Nous pouvons aussi nous référer aux livres d'Azaïs et Bardet [8], ou encore celui d'Hocking [49], qui proposent une étude complète du modèle linéaire dans le cas i.i.d. Certains ont supprimé des hypothèses quant aux erreurs mais tout en conservant l'indépendance, comme Eicker [42] qui suppose des erreurs non forcément identiquement distribuées ou Chen [29] qui travaille avec des erreurs hétéroscédastiques.

Cependant beaucoup de données scientifiques montrent une dépendance temporelle significative ; cela a pour conséquence que l'hypothèse d'indépendance n'est pas vérifiée (voir par exemple Brockwell et Davis [23]). Cela est observé en astrophysique, géophysique, biostatistiques, climatologie et dans beaucoup d'autres domaines.

Illustrons notre propos avec un exemple et étudions un jeu de données fourni par l'Observatoire du Mona Loa (Hawaii). Il s'agit de la mesure moyenne mensuelle du taux de CO2 (partie par million : ppmv) dans l'atmosphère au large des côtes d'Hawaii. Les relevés ont été produits tous les mois entre 1959 et 1998, ce qui fait 468 données au total. Le graphe des données est affiché en Figure 1. Cet ensemble de données peut être modélisé par une série temporelle puisqu'il s'agit d'observations enregistrées à des temps réguliers.

De façon classique, une série temporelle se décompose en trois parties : une tendance m et une saisonnalité s qui sont des composantes déterministes, et les erreurs ϵ qui constituent la partie aléatoire du modèle. La tendance représente le comportement global de la série et la saisonnalité son comportement périodique. Formellement cela s'écrit :

$$Y_t = m_t + s_t + \epsilon_t,$$

où Y représente le taux de CO2 et t le temps. Pour la saisonnalité s_t, nous avons les contraintes usuelles $s_t = s_{t+12}$ et $\sum_{t=1}^{12} s_t = 0$. Pour estimer m_t et s_t, nous pouvons procéder de deux manières différentes : soit en utilisant un cadre non-paramétrique, soit en utilisant un cadre paramétrique. Dans

1. indépendantes et identiquement distribuées

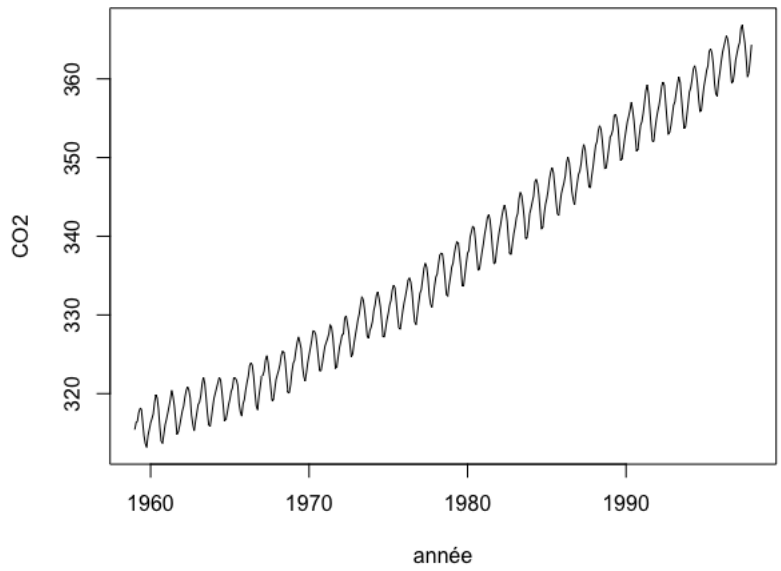


FIGURE 1 – Taux de CO2 (en ppmv) en fonction du temps.

le cas paramétrique, les deux composantes déterministes peuvent être regroupées dans une matrice, que nous notons X, et le modèle peut s'écrire sous la forme d'un modèle de régression linéaire :

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

Pour notre exemple, nous pouvons considérer un ajustement composé par un polynôme de degré 3 et un polynôme trigonométrique avec des fréquences bien choisies. Cet ajustement est affiché à gauche en Figure 2.

L'estimateur des moindres carrés est généralement utilisé pour estimer le paramètre β. Grâce à lui nous obtenons les valeurs ajustées des données par le modèle et les résidus, qui sont le résultat de la différence entre les données et ces valeurs approchées, affichés à droite dans la Figure 2. Intéressons-nous maintenant à la fonction d'autocorrélation du processus des résidus, qui est affichée en Figure 3. Si le processus était indépendant, seule la première barre au lag 0 devrait prendre la valeur 1 et tout le reste devrait être proche de 0. Or la Figure 3 nous montre que la corrélation entre les résidus n'est pas nulle, ce qui signifie qu'ils ne sont pas indépendants. Si nous considérons que les composantes déterministes sont bien estimées par notre modèle, de sorte que les résidus soient « proches » des erreurs non observées, il est donc plus que vraisemblable que ces erreurs ne soient pas indépendantes. Cependant, la plupart des résultats sur le modèle linéaire, comme les intervalles de confiance ou les procédures de test pour le paramètre β, sont construits en faisant une hypothèse d'indépendance sur les erreurs. Cela laisse donc supposer que ces résultats ne sont pas corrects si le processus des erreurs sous-jacent n'est pas indépendant.

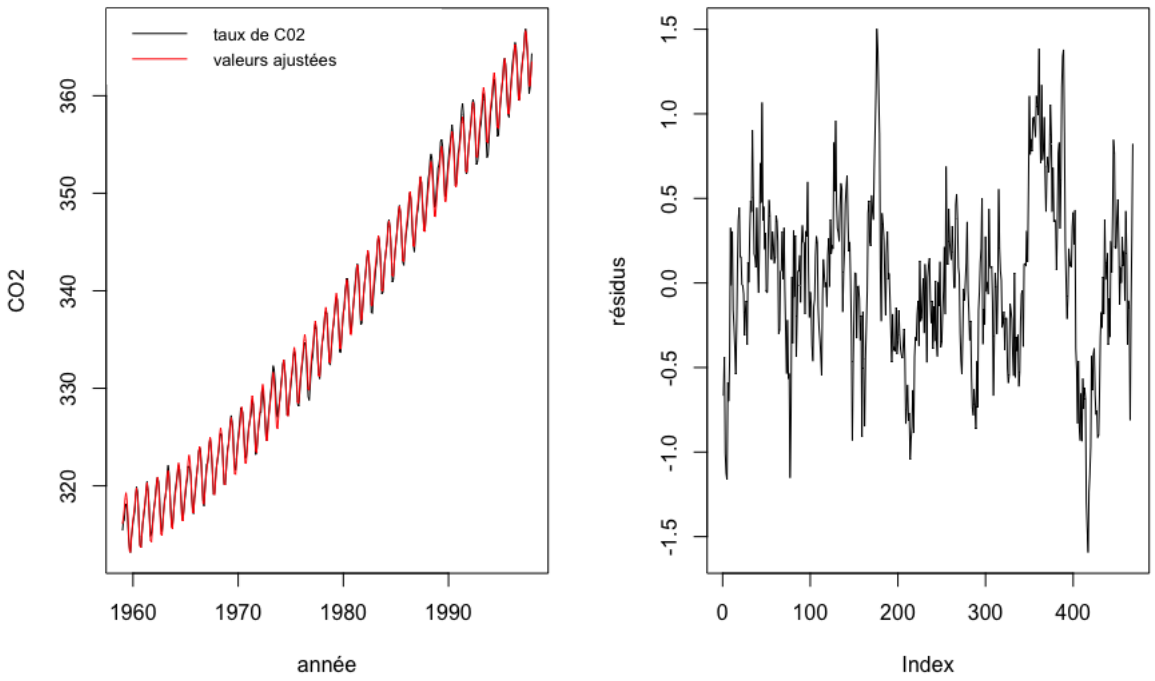


FIGURE 2 – Ajustement sur le taux de CO2 (à gauche) et les résidus (à droite).

Face à ce problème, certains auteurs ont étudié le modèle de régression linéaire avec erreurs dépendantes. Ainsi Pagan et Nicholls [65] ont considéré le cas où les erreurs suivent un processus MA(q), Krämer [53] le cas où elles suivent un processus AR(1), et Chib et Greenberg [30] le cas où les erreurs forment un processus ARMA(p,q). Les travaux de Wu [76] sur les processus stationnaires permettent de considérer un cadre plus général ; ainsi Wu [75] propose d'étudier le modèle linéaire où le processus des erreurs peut être non-linéaire ou non fortement mélangeant.

Nous proposons d'étudier dans cette thèse le modèle de régression linéaire avec erreurs dépendantes dans le cadre très général d'Hannan [47]. En effet, Hannan a montré un Théorème Limite Central pour l'estimateur des moindres carrés qui est vérifié pour la plupart des processus stationnaires à mémoire courte. Commençons par définir le cadre mathématique de notre étude.

Cadre mathématique

Nous considérons le modèle de régression linéaire :

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

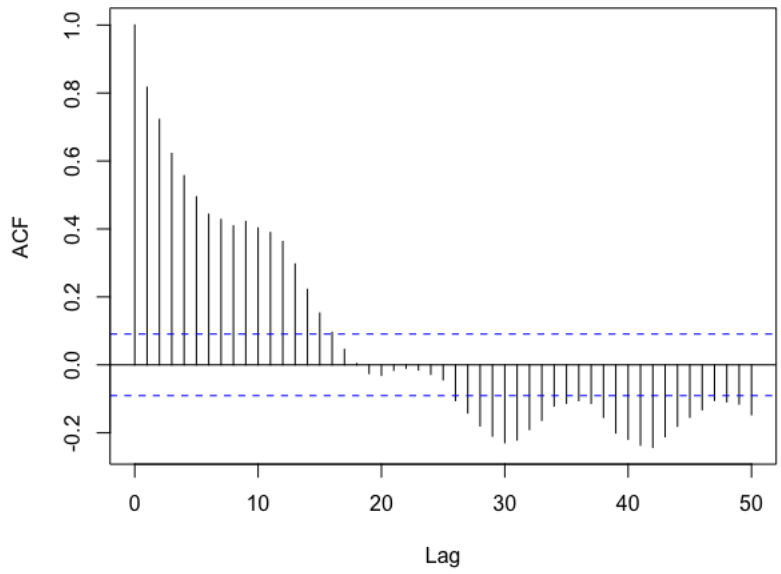


FIGURE 3 – Fonction d'autocorrélation du processus des résidus.

où le design X est la matrice des variables explicatives de taille $n \times p$, et ϵ est un processus d'erreurs défini sur un espace de probabilité $(\Omega, \mathcal{F}, \mathbb{P})$. Le processus ϵ sera toujours supposé indépendant du design X. Nous notons $X_{.,j}$ la j-ème colonne de la matrice X et $x_{i,j}$ le nombre réel à l'intersection de la ligne i et de la colonne j, où j appartient à $\{1, \dots, p\}$ et i à $\{1, \dots, n\}$. Les vecteurs aléatoires Y et ϵ appartiennent à $\mathbb{R}^n$ et β est le vecteur de taille p des paramètres inconnus du modèle. La norme euclidienne usuelle sur l'espace $\mathbb{R}^n$ sera notée $\|.\|_2$ et la norme $\mathbb{L}^p$ sur Ω sera notée $\|.\|_{\mathbb{L}^p}$.

Le processus des erreurs $(\epsilon_i)_{i\in\mathbb{Z}}$ est supposé strictement stationnaire, de moyenne nulle et, pour tout i dans $\mathbb{Z}$, ϵ_i appartient à $\mathbb{L}^2(\Omega)$. La définition de la stricte stationnarité est rappelée ici :

Définition 0.1 (Stricte stationnarité [23])**.** *Un processus stochastique $(\epsilon_i)_{i\in\mathbb{Z}}$ est dit strictement stationnaire si les lois jointes de $(\epsilon_{t_1}, \dots, \epsilon_{t_k})$ et $(\epsilon_{t_1+h}, \dots, \epsilon_{t_k+h})$ sont les mêmes pour tout $k \in \mathbb{N}^*$ et pour tout $t_1, \dots, t_k, h \in \mathbb{Z}$.*

Afin de couvrir un champ de processus le plus large possible, nous définissons le processus des erreurs en utilisant l'écriture suivante pour tout i dans $\mathbb{Z}$:

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

où $\mathbb{T} : \Omega \to \Omega$ est une fonction bijective bimesurable préservant la mesure de probabilité $\mathbb{P}$. Remarquons que tout processus strictement stationnaire peut s'écrire de cette manière.

Nous définissons ensuite la filtration $(\mathcal{F}_i)_{i\in\mathbb{Z}}$, qui est non-décroissante et construite de la manière

suivante pour tout i :

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

où $\mathcal{F}_0$ est une sous-tribu de $\mathcal{F}$ telle que $\mathcal{F}_0 \subseteq \mathbb{T}^{-1}(\mathcal{F}_0)$. Cela nous permet de considérer des tribus qui ne sont pas forcément adaptées au processus ϵ. En outre nous supposons toujours que la tribu $\mathcal{F}_{-\infty} = \cap_{i\in\mathbb{Z}}\mathcal{F}_i$ est la tribu triviale et que ϵ_0 est $\mathcal{F}_\infty$-mesurable.

La fonction d'autocovariance du processus ϵ est définie pour tout k, m dans $\mathbb{Z}$ par :

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

et sa matrice de covariance Γ_n est la matrice de Toeplitz définie par :

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

Enfin nous introduisons la densité spectrale f du processus ϵ définie sur l'intervalle $[-\pi, \pi]$ et qui est l'unique fonction dont les covariances sont les coefficients de Fourier :

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

Les résultats développés dans cette thèse sont fondés sur un Théorème Limite Central d'Hannan [47], qui a été démontré pour l'estimateur des moindres carrés dans le cas stationnaire, sous des conditions très faibles sur le processus des erreurs et sur le design.

Théorème d'Hannan

Nous allons donc présenter le théorème d'Hannan et expliciter les conditions requises pour l'appliquer. Notons que nous travaillerons toujours conditionnellement au design X puisque ce dernier peut être aléatoire. L'estimateur des moindres carrés sera noté $\hat{\beta}$ et est défini par :

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

Soit $(P_j)_{j\in\mathbb{Z}}$ une famille d'opérateurs de projections définie pour tout j dans $\mathbb{Z}$ et pour tout Z dans $\mathbb{L}^2(\Omega)$ par :

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

Le processus des erreurs doit satisfaire la condition suivante, que nous appellerons la condition d'Hannan :

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

Remarquons que cette condition implique que le processus étudié est à mémoire courte, c'est-à-dire

que la série des covariances est finie (voir par exemple Dedecker, Merlevède et Volný [38]) :

$$\sum_{k\in\mathbb{Z}} |\gamma(k)| < \infty.$$

L'intérêt du théorème d'Hannan est qu'il est vérifié pour la plupart des processus stationnaires à mémoire courte. Présentons une liste non-exhaustive des processus satisfaisant cette condition :

- Dedecker, Merlevède et Volný [38] ont montré qu'une grande classe de processus stationnaires vérifiant la condition d'Hannan est celle des fonctions de processus linéaires générés par des variables aléatoires i.i.d., de la forme :

$$\epsilon_k = F\left(\sum_{i\in\mathbb{Z}} a_i\eta_{k-i}\right) - \mathbb{E}\left(F\left(\sum_{i\in\mathbb{Z}} a_i\eta_{k-i}\right)\right),$$

  où $(\eta_i, i \in \mathbb{Z})$ est une suite de variables aléatoires i.i.d. Cela contient évidemment toute la classe des processus linéaires et donc tous les processus de type ARMA (autorégressif et moyenne mobile).
- Dedecker, Merlevède et Volný [38] ont aussi montré que si un processus vérifie des conditions de type différence de martingales, ou conditions « à la Gordin » [45], alors ce même processus vérifie la condition d'Hannan. Ainsi la Proposition 5 dans [38] montre que la condition d'Hannan est satisfaite si le processus d'erreur satisfait les conditions suivantes :

$$\sum_{k=1}^{\infty} \frac{1}{\sqrt{k}} \left\|\mathbb{E}(\epsilon_k|\mathcal{F}_0)\right\|_{\mathbb{L}^2} < \infty$$

$$\sum_{k=1}^{\infty} \frac{1}{\sqrt{k}} \left\|\epsilon_{-k} - \mathbb{E}(\epsilon_{-k}|\mathcal{F}_0)\right\|_{\mathbb{L}^2} < \infty.$$

- Une autre grande classe de processus, pour laquelle la condition d'Hannan est vérifiée, est la classe des processus faiblement mélangeants au sens de Dedecker et Prieur [39]. Par exemple si le processus ϵ est $\tilde{\phi}$-mélangeant, appartient à $\mathbb{L}^p$ pour $p \in [2, +\infty[$ et vérifie que $\sum_{k=1}^{\infty} \frac{1}{\sqrt{k}} \tilde{\phi}(k)^{\frac{p-1}{p}}$ converge, alors il satisfait la condition (C1). Sous des conditions similaires nous pouvons aussi montrer que si le processus est $\tilde{\alpha}$-mélangeant, alors il vérifie la condition d'Hannan. Nous pouvons noter que ce que nous avons écrit pour les suites faiblement mélangeantes est encore vrai pour les processus fortement mélangeants, par exemple α-mélangeant au sens de Rosenblatt [70].
- Enfin Wu a montré que la propriété de « 2-strong stability » introduite dans [74] est plus restrictive que la condition (C1). Ainsi, si un processus vérifie les conditions imposées par Wu, alors il vérifiera (C1).

La condition d'Hannan fournit donc un cadre très général pour les processus stationnaires. Des détails sur ces classes de processus sont disponibles en Section 1.4 de cette thèse ou dans la Section 4 de l'article de Caron et Dede [25].

Ensuite, pour appliquer le théorème d'Hannan, certaines conditions sur le design X sont requises.

Encore une fois, nous allons voir qu'elles sont très faibles et facilement vérifiables. Commençons par définir la norme euclidienne de la colonne j du design :

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

Cela nous permet de construire la matrice diagonale de normalisation $D(n)$, où $d_j(n)$ est le j-ème terme pour $j \in \{1, \dots, p\}$. Les conditions sur les colonnes du design X sont au nombre de trois :

$$\forall j \in \{1, \dots, p\}, \qquad \lim_{n\to\infty} d_j(n) = \infty, \quad p.s.^2. \tag{C2}$$

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

Enfin, la troisième condition indique que les limites suivantes doivent exister :

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

Les conditions (C2) et (C3) sont les mêmes que celles que nous devons vérifier dans le cas i.i.d. lorsque nous voulons obtenir un Théorème Limite Central pour l'estimateur des moindres carrés (conditions de Lindeberg). Dans le cas dépendant, nous avons en plus besoin de la condition (C4). Les coefficients ρ représentent les corrélations entre les colonnes du design. Ainsi $\rho_{j,l}$ est le coefficient de corrélation entre les colonnes j et l de la matrice X.

Les coefficients ρ nous permettent de construire la matrice $R(k)$, de taille $p \times p$, qui contient les coefficients $\rho_{j,l}(k)$ et qui est définie presque sûrement par :

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

où F_X est la mesure spectrale associée à la matrice $R(k)$. Grâce à cette mesure spectrale, nous pouvons définir les deux matrices F et G qui permettront d'obtenir la matrice de covariance asymptotique du théorème d'Hannan :

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

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

Enfin la dernière condition à satisfaire est que la matrice $R(0)$ doit être définie positive, ce que nous notons par :

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

Sous toutes ces conditions, nous pouvons écrire le Théorème Limite Central d'Hannan pour l'estimateur des moindres carrés $\hat{\beta}$:

2. p.s. : presque sûrement

Théorème 0.1 (Théorème d'Hannan)**.** *Soit $(\epsilon_i)_{i\in\mathbb{Z}}$ un processus stationnaire de moyenne nulle. Supposons que $\mathcal{F}_{-\infty}$ est la tribu triviale, que ϵ_0 est $\mathcal{F}_\infty$-mesurable et que le processus $(\epsilon_i)_{i\in\mathbb{Z}}$ satisfait la condition d'Hannan* (C1)*. Supposons que le design X satisfait, presque sûrement, les conditions* (C2)*,* (C3)*,* (C4) *et* (C5)*. Alors, pour toute fonction f continue bornée :*

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

où la loi de Z sachant X est une loi gaussienne, de moyenne nulle et de matrice de covariance égale à $F^{-1}GF^{-1}$. En outre nous avons la convergence du moment d'ordre 2 :

$$\mathbb{E}\left(D(n)(\hat{\beta}-\beta)(\hat{\beta}-\beta)^t D(n)^t\Big|X\right) \xrightarrow[n\to\infty]{} F^{-1}GF^{-1}, \quad p.s.^{3} \tag{2}$$

En pratique nous n'aurons jamais besoin de calculer la matrice de covariance asymptotique $F^{-1}GF^{-1}$. Pour les applications nous allons procéder autrement en considérant le moment d'ordre 2 :
$\mathbb{E}\left(D(n)(\hat{\beta}-\beta)(\hat{\beta}-\beta)^t D(n)^t\Big|X\right)$.
Dorénavant la matrice de covariance asymptotique $F^{-1}GF^{-1}$ sera notée C et $c_{j,l}$ seront ses coefficients pour tout j, l dans $\{1,\dots,p\}$.

Estimation de la matrice de covariance

Dans le but de calculer des régions de confiance ou d'effectuer des procédures de test pour le paramètre du modèle β, nous avons besoin d'un estimateur de la matrice de covariance asymptotique. Grâce au théorème d'Hannan, nous avons la convergence du moment d'ordre 2 établi en (2). Étant donné que nous travaillons conditionnellement au design X, le moment d'ordre 2 peut s'écrire :

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

où Γ_n est la matrice de covariance des erreurs de type Toeplitz définie en (1). Avec cette écriture, nous remarquons que la seule quantité inconnue est Γ_n. Nous avons donc juste besoin d'un estimateur de la matrice de covariance des erreurs pour avoir un estimateur de la matrice de covariance asymptotique. Nous allons de ce fait proposer un estimateur et prouver sa consistance sous les conditions d'Hannan.

Commençons en considérant la matrice aléatoire :

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

où les coefficients $\hat{\gamma}_k$ sont définis par :

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

3. La transposée d'une matrice X est notée X^t.

La fonction K est un noyau et doit satisfaire les trois conditions suivantes :

- K est positif, symétrique et $K(0) = 1$,
- K est à support compact,
- La transformée de Fourier de K est intégrable.

La suite de réels positifs h_n est la fenêtre de notre estimateur à noyau et peut s'interpréter comme le nombre de termes de covariances qu'il faut garder pour avoir une bonne estimation de la matrice de covariance des erreurs. La fenêtre h_n doit tendre vers l'infini et $\frac{h_n}{n}$ tendre vers 0 quand n tend vers l'infini.

Il n'est en revanche pas possible de travailler directement avec $\widehat{\Gamma}_{n,h_n}$. En effet, dans le contexte de la régression linéaire, les erreurs $(\epsilon_i)_{1\leq i\leq n}$ ne sont pas observées. Seuls les résidus sont des quantités observables puisque nous n'avons accès qu'aux données Y et au design X. Nous rappelons que les résidus sont définis par :

$$\hat{\epsilon}_j = Y_j - (x_j)^t\hat{\beta} = Y_i - \sum_{j=1}^{p} x_{i,j}\hat{\beta}_j.$$

Pour pallier ce problème, nous allons travailler avec la même matrice aléatoire mais nous allons faire un "plug-in" des résidus afin de remplacer les erreurs théoriques ϵ. En conséquence, nous considérons l'estimateur suivant pour Γ_n :

$$\widehat{\Gamma}^*_{n,h_n} = \left[K\left(\frac{j-l}{h_n}\right)\hat{\gamma}^*_{j-l}\right]_{1\leq j,l\leq n}, \tag{3}$$

où les estimateurs des covariances des erreurs sont définis par :

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

Cet estimateur est une version tronquée de la pleine matrice $\widehat{\Gamma}^*_n = \left[\hat{\gamma}^*_{j-l}\right]_{1\leq j,l\leq n}$, préservant la diagonale et certaines sous-diagonales. L'intérêt de lisser le spectre de covariances vient du fait que, pour un grand k, soit $\gamma(k)$ est proche de 0, soit $\hat{\gamma}^*_k$ n'est pas un estimateur fiable pour $\gamma(k)$. Lisser le spectre peut donc apporter une économie computationnelle considérable et rendre plus efficaces les procédures de simulations, si les petits ou trop éloignés $\hat{\gamma}^*_k$ sont laissés en dehors des calculs.

En conséquence, pour estimer la matrice de covariance asymptotique C, nous utilisons l'estimateur :

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

Les coefficients de la matrice C_n seront notés $c_{n,(j,l)}$, pour tout j,l dans $\{1,\dots,p\}$.

Un des résultats principaux de cette thèse est le résultat de consistance suivant pour l'estimateur de la matrice de covariance asymptotique C_n, dont nous montrons la convergence en norme $\mathbb{L}^1$ conditionnellement à X :

Théorème 0.2. *Soit h_n une suite de réels positifs telle que $h_n \to \infty$ quand $n \to \infty$, et :*

$$h_n\mathbb{E}\left(|\epsilon_0|^2\left(1\wedge\frac{h_n}{n}|\epsilon_0|^2\right)\right)\xrightarrow[n\to\infty]{} 0. \tag{4}$$

Sous les hypothèses du Théorème d'Hannan 0.1, l'estimateur C_n est consistant, c'est-à-dire que pour tout j, l dans $\{1, \ldots, p\}$:

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

L'intérêt de ce théorème est qu'il est vérifié sous des conditions très faibles. En effet la condition (4) n'est pas difficile à satisfaire en pratique, comme le prouvent les deux remarques suivantes :

Remarque 0.1. *Si ϵ_0 est de carré intégrable, alors il existe $h_n \to \infty$ telle que la condition* (4) *est vérifiée.*

Remarque 0.2. *Si $\mathbb{E}\left(|\epsilon_0|^{\delta+2}\right) < \infty$ avec δ dans $]0, 2]$, alors :*

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

Donc, si h_n satisfait $\frac{h_n^{1+\delta/2}}{n^{\delta/2}} \xrightarrow[n\to\infty]{} 0$, alors (4) *est vraie. En particulier, si la variable aléatoire ϵ_0 a un moment d'ordre 4, alors la condition sur h_n est $\frac{h_n^2}{n} \xrightarrow[n\to\infty]{} 0$.*

La preuve du théorème 0.2 s'appuie fortement sur un résultat concernant l'estimation de la densité spectrale. En effet dans le Chapitre 1 de la thèse (et dans Caron et Dede [25]), nous avons construit un estimateur de la densité spectrale et nous avons montré sa consistance sous les mêmes conditions que pour le théorème 0.2. Les propriétés de l'estimateur de la densité spectrale ont été discutées dans beaucoup de livres classiques sur les séries temporelles ; on peut citer par exemple, Anderson [4], Brillinger [22], Brockwell et Davis [23], Grenander et Rosenblatt [46], Priestley [67] et Rosenblatt [70]. Mais beaucoup de ces précédents résultats requièrent des conditions restrictives sur les processus sous-jacents (structure linéaire ou conditions de mélange fort). Wu et Liu [59] ont considéré le problème de l'estimation de la densité spectrale et ont étendu l'applicabilité de l'analyse spectrale aux processus non linéaires et/ou aux processus non fortement mélangeants. Ils ont proposé un estimateur consistant de la densité spectrale et ont donné des conditions sous lesquelles l'estimateur centré satisfait un Théorème Limite Central. Ces résultats sont fondés sur la théorie asymptotique des processus stationnaires développée par Wu [76]. Cependant, pour montrer ses résultats, Wu utilise une notion de dépendance qui est plus restrictive que celle d'Hannan. En effet la classe des processus satisfaisant la $\mathbb{L}^2$ « physical dependence measure » introduite par Wu est incluse dans la classe des processus satisfaisant la condition d'Hannan.

C'est pourquoi nous avons prouvé la consistance d'un estimateur de la densité spectrale sous la condition d'Hannan, c'est-à-dire pour la plupart des processus dépendants à mémoire courte. L'estimateur que nous avons utilisé est le suivant, défini pour tout λ dans $[-\pi, \pi]$:

$$f_n^*(\lambda) = \frac{1}{2\pi}\sum_{|k|\leq n-1} K\left(\frac{|k|}{c_n}\right)\hat{\gamma}_k^* e^{ik\lambda},$$

où :

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

Le théorème de consistance pour l'estimateur de la densité spectrale est le suivant :

Théorème 0.3. *Soit h_n une suite de réels positifs telle que $h_n \to \infty$ quand $n \to \infty$, et :*

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

Alors, sous les hypothèses du Théroème 0.1 :

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

Ce théorème a été démontré dans la Section 1.6.2 de cette thèse (ou dans Caron et Dede [25]) pour un design X déterministe. Mais une lecture de la preuve indique que le résultat reste vrai si nous travaillons conditionnellement à X avec un design aléatoire.

Pour terminer cette section nous allons établir les corollaires qui nous permettront d'avoir tous les outils nécessaires pour construire des intervalles de confiance ou effectuer des procédures de test. Ainsi, à partir du théorème 0.2, nous obtenons la convergence non-conditionnelle en probabilité pour l'estimateur C_n :

Corollaire 0.1. *Soit h_n une suite satisfaisant* (4)*. Alors l'estimateur C_n converge en probabilité vers C quand n tend vers l'infini.*

Finalement, en combinant le Théorème 0.1 et le Théorème 0.2, nous obtenons le corollaire suivant :

Corollaire 0.2. *Sous les hypothèses du Théorème 0.1 et du Théorème 0.2, nous avons :*

$$C_n^{-\frac{1}{2}} \left(D(n)(\hat{\beta} - \beta) \right) \xrightarrow[n\to\infty]{\mathcal{L}} \mathcal{N}(0, I_p), \tag{5}$$

où I_p est la matrice identité de taille p.

Vers une approche non-asymptotique

Dans la deuxième partie de cette thèse, nous nous intéressons particulièrement à une adaptation des tests de Fisher. En effet, grâce au Corollaire 0.2, nous pouvons établir une nouvelle statistique de test, afin que les tests sur le modèle linéaire aient toujours asymptotiquement un bon niveau et cela même lorsque le processus d'erreurs sous-jacent est dépendant.

Le test de Fisher consiste à tester $H_0 : \beta_{j_1} = \ldots = \beta_{j_{p_0}} = 0$, contre $H_1 : \exists\, j_z \in \{j_1, \ldots, j_{p_0}\}$ tel que $\beta_{j_z} \neq 0$. Nous rappelons que le niveau d'un test, noté α, est la probabilité de sélectionner l'hypothèse alternative H_1 alors que l'hypothèse nulle H_0 est vraie. En utilisant le corollaire 0.2, nous avons sous l'hypothèse H_0 :

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

où $C_{n_{p_0}}$ est la matrice de covariance C_n construite en supprimant les lignes et les colonnes qui n'appartiennent pas à l'ensemble discret $\{j_1, \ldots, j_{p_0}\}$. La matrice identité de taille p_0 est notée I_{p_0} et $0_{p_0 \times 1}$ et un vecteur de 0 de taille p_0. Nous définissons ainsi la statistique de test suivante :

$$\Xi = Z_{1,n}^2 + \ldots + Z_{p_0,n}^2.$$

Sous l'hypothèse H_0, la variable Ξ converge en loi vers une loi du χ^2 de paramètre p_0, et le test est donc asymptotiquement de niveau α.

En pratique, cependant, les jeux de données ont toujours un nombre d'observations n fini. Dès lors, afin d'appliquer ces résultats sur des cas pratiques, nous aimerions savoir quelle serait la valeur optimale pour h_n afin d'avoir des tests bien calibrés et un niveau non-asymptotique le plus proche possible du niveau α désiré. La solution naturelle à ce problème serait d'utiliser des techniques de statistique adaptative. La sélection de modèles a été beaucoup étudiée ces dernières années dans le cas des variables i.i.d. ; nous pouvons citer par exemple certains travaux de Barron, Birgé et Massart [13], de Birgé et Massart [17], [18], [20], ou encore de Massart [63], [64]. Pour la sélection de la fenêtre d'un estimateur à noyau dans le cas i.i.d., nous trouvons les travaux de Goldenshluger et Lepski [44] ou de Lacour, Massart et Rivoirard [54]. Nous pouvons aussi citer les travaux sur la validation croisée de Celisse et Arlot [6], qui est une technique très utilisée pour sélectionner un modèle ou régler un paramètre dans le cas de variables aléatoires i.i.d. Nous ne pouvons malheureusement pas utiliser ces méthodes dans notre cadre pour deux raisons. La première est que notre cible est le niveau d'un test, ce qui diffère des approches classiques où c'est le risque d'un estimateur qui est considéré. Nous ne pouvons jamais savoir à l'avance si nous sommes sous l'hypothèse nulle ou sous l'hypothèse alternative, ce qui rend impossible l'utilisation de certaines techniques, comme la validation croisée. La deuxième raison est que nous sommes non seulement dans un contexte de variables dépendantes, mais aussi dans le cadre très général d'Hannan dont le théorème s'applique pour la plupart des processus stationnaires à mémoire courte.

La sélection de modèles en présence de variables dépendantes a cependant été abordée par certains auteurs mais pour d'autres types de problèmes. Ainsi Comte [32] s'intéresse à une estimation data-driven de la densité spectrale d'une suite stationnaire gaussienne et Efromovich [41] prouve que l'estimateur avec noyau rectangulaire de la densité spectrale de certains processus stationnaires est asymptotiquement minimax. Il propose aussi des algorithmes basés sur les données pour estimer la valeur de la fenêtre du noyau. Nous pouvons aussi citer l'article de Lerasle [55] qui s'intéresse à l'estimation adaptative de la densité d'une suite de variables aléatoires β ou τ-mélangeante, les travaux de Alquier et Wintenberger [3] qui utilisent la sélection de modèles pour la prédiction des valeurs d'une série temporelle stationnaire, les travaux de Alquier et Doukhan [2] qui étudient le comportement des estimateurs ℓ_1-pénalisés dans un cadre d'observations dépendantes, ou encore ceux de Baraud, Comte et Viennet [11], [12] qui étudient le modèle de régression non-paramétrique sous certaines conditions de dépendance sur le design et les erreurs.

Dans le Chapitre 3 de cette thèse (et dans Caron, Dedecker et Michel [26]), nous avons répondu partiellement au problème du réglage non-asymptotique des tests en construisant des méthodes empiriques basées sur les données. En partant de l'estimateur de la matrice de covariance C_n, nous

proposons une approche dite "plug-in" qui consiste à remplacer l'estimateur de la matrice de covariance des erreurs Γ_n. Ainsi nous introduisons l'estimateur suivant :

$$\widehat{C}(\widehat{\Gamma}_n) := D(n)(X^tX)^{-1}X^t\widehat{\Gamma}_nX(X^tX)^{-1}D(n), \tag{6}$$

et nous utilisons ce nouvel estimateur $\widehat{C}$ pour calculer les statistiques usuelles du modèle linéaire dont les tests. Nous avons défini différentes façons d'obtenir la matrice $\widehat{\Gamma}_n$: en adaptant un processus autorégressif sur le processus des résidus et en calculant les covariances théoriques de l'AR(p) obtenu, en utilisant l'estimateur à noyau défini en (3) avec une méthode bootstrap pour choisir la valeur de la fenêtre (Wu et Pourahmadi [77]), en utilisant un choix alternatif de la fenêtre dans le cas où serait utilisé un noyau rectangulaire (Efromovich [41]), ou encore en utilisant un estimateur adaptatif de la densité spectrale via une base d'histogrammes (Comte [32]). Ces méthodes sont décrites en détails dans le Chapitre 3, et des simulations ont été réalisées afin de comparer leurs performances.

Associé au Chapitre 3, nous avons développé un package R nommé `slm`. Ce-dernier reprend la structure et les méthodes de la fonction `lm` et modifie les sorties en prenant en compte la dépendance entre les données. Les méthodes décrites ci-dessus ont été implémentées et l'utilisateur choisit celle qu'il veut utiliser. Afin d'illustrer cela, reprenons l'exemple du taux de CO2 explicité au début de l'introduction. Nous procédons tout d'abord à une régression linéaire avec la fonction `lm` du logiciel R et un design composé d'un polynôme de degré 3 et d'un polynôme trigonométrique de degré 4, issus des fonctions : t, t^2, t^3, $\sin(2\pi t)$, $\cos(2\pi t)$, $\sin(4\pi t)$, $\cos(4\pi t)$, $\sin(6\pi t)$, $\cos(6\pi t)$, $\sin(8\pi t)$, $\cos(8\pi t)$. Si nous effectuons une sélection backward avec un niveau pour la p-value à 5% dans le but de supprimer les variables non-significatives, nous obtenons le modèle composé de tous les monômes précédents sauf $\cos(6\pi t)$ et $\cos(8\pi t)$, et l'ajustement correspondant est affiché à gauche en Figure 2. La fonction `lm` suppose que les erreurs sont indépendantes, mais nous avons vu en Figure 3 que les résidus sont fortement corrélés. En conséquence, nous ne pouvons pas faire confiance à la fonction `lm` pour prendre des décisions. Nous proposons de corriger cette régression linéaire en utilisant la fonction `slm` de notre package, avec la méthode qui consiste à adapter un processus autorégressif sur le processus des résidus. La méthode automatique utilise un critère AIC pour sélectionner l'ordre du processus AR. Nous procédons donc de la même manière qu'avec la fonction `lm`, en prenant d'abord le design complet présenté plus haut, puis en effectuant une sélection backward avec un seuil pour la p-value au niveau 5%. Nous obtenons avec `slm` le design final composé des monômes suivants : t, t^2, t^3, $\sin(2\pi t)$, $\cos(2\pi t)$, $\sin(4\pi t)$, $\cos(4\pi t)$, $\sin(6\pi t)$, $\cos(6\pi t)$, $\sin(8\pi t)$. Le seul monôme jugé non-significatif est ici $\cos(8\pi t)$. L'algorithme `slm` recommande de garder la variable $\cos(6\pi t)$, alors que `lm` jugeait qu'elle était non-significative. Devant l'évidente dépendance du processus des résidus, l'utilisateur devrait donc plutôt utiliser les conclusions de la fonction `slm`.

Nous terminons cette introduction en présentant le dernier chapitre de la thèse qui est un chapitre d'ouverture. Dans le Chapitre 1, nous avons développé une méthode d'estimation paramétrique par moindres carrés pour une fonction non-aléatoire dans le modèle de régression. Cela nous a par exemple été utile dans l'exemple avec le taux de CO2. Dans le Chapitre 4 nous nous intéressons à l'étude du modèle de régression dans le cas non-paramétrique via une approche par sélection de modèles, et en supposant que les erreurs forment un processus gaussien. Ce sujet a été traité par Birgé et Massart

dans le cas i.i.d. [18], [19], [20] et nous souhaitons l'étendre au cas où le processus d'erreurs ϵ est une suite gaussienne ; ainsi le processus ϵ suit une loi $\mathcal{N}(0, \Sigma)$ où Σ est la matrice de covariance de taille $n \times n$. La dépendance entre les erreurs, et donc entre les données, est ainsi conservée. Nous nous intéressons à l'estimation d'un vecteur non-aléatoire f^* de $\mathbb{R}^n$ dans le modèle :

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

En nous inspirant du cheminement établi dans le Chapitre 2 du livre de Giraud [43], l'objectif est d'expliciter la forme de la fonction de pénalité dans le cadre de variables gaussiennes dépendantes et d'établir une inégalité oracle pour l'estimateur de risque minimal parmi une collection de modèles.

Notons $\rho(\Sigma)$ le rayon spectral de la matrice de covariance Σ, c'est-à-dire sa plus grande valeur propre. Nous définissons une collection de modèles $\{S_m, m \in \mathcal{M}\}$ et une loi de probabilité associée $\{\pi_m, m \in \mathcal{M}\}$ sur $\mathcal{M}$ telle que $\sum_{m\in\mathcal{M}} e^{-\pi_m}$ converge. Pour chaque m dans $\mathcal{M}$, nous notons d_m la dimension de S_m, et nous associons à chaque espace l'estimateur des moindres carrés $\hat{f}_m$ de f^* dans S_m. L'estimateur $\hat{f}_{\hat{m}}$ est celui de risque minimal parmi la collection $\{\hat{f}_m, m \in \mathcal{M}\}$. Nous montrons entre autres que pour la pénalité :

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

nous obtenons l'inégalité oracle suivante :

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

où K et C_K sont des constantes strictement supérieures à 1.

CHAPTER 1

ASYMPTOTIC DISTRIBUTION OF LEAST SQUARES ESTIMATORS FOR LINEAR MODELS WITH DEPENDENT ERRORS : REGULAR DESIGNS

Le contenu de ce chapitre est issu d'un article écrit en collaboration avec Sophie Dede et publié dans la revue *Mathematical Methods of Statistics* [25].

Abstract

We consider the usual linear regression model in the case where the error process is assumed strictly stationary. We use a result from Hannan [47], who proved a Central Limit Theorem for the usual least squares estimator under general conditions on the design and on the error process. We show that for a large class of designs, the asymptotic covariance matrix is as simple as the i.i.d.[1] case. We then estimate the covariance matrix using an estimator of the spectral density whose consistency is proved under very mild conditions. As an application, we show how to modify the usual Fisher tests 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.

1.1 Introduction

We consider the usual fixed-design linear regression model:

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

where X is the fixed design matrix and $(\epsilon_i)_{i\in\mathbb{Z}}$ is a stationary process. This model is commonly used in time series regression.

Our work is based on the paper by Hannan [47], who proved a Central Limit Theorem for the usual least squares estimator under general conditions on the design and on the error process. Most of short-range dependent processes satisfy the 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

1. independent and identically distributed.

linear processes, and many processes under various mixing conditions (see for instance Dedecker, Merlevède, Volný [38], and also Dedecker [35] for the optimality of Hannan's condition).

In this paper, it is shown that for a large class of designs satisfying Hannan's conditions, the covariance matrix of the limit distribution of the least squares estimator is the same as in the i.i.d. case, up to the usual error variance term, which should be replaced by the covariance series of the error process. We shall refer to this very large class of designs as "regular designs" (see Section 1.2.3 for the precise definition). It includes many interesting examples, for instance the ANOVA type designs or the designs whose columns are regularly varying (such as the polynomial regression type designs).

For this class of regular designs, any consistent estimator of the covariance series of $(\epsilon_i)_{i\in\mathbb{Z}}$ may be used to obtain a Gaussian limit distribution with explicit covariance matrix for the normalized least squares estimator. Doing so, it is then possible to obtain confidence regions and test procedures for the unknown parameter β. In this paper, assuming only that Hannan's condition on (ϵ_i) is satisfied, we propose a consistent estimator of the spectral density of (ϵ_i) (as a byproduct, we get an estimator of the covariance series).

Wu and Liu [59] considered the problem of estimating the spectral density for a large class of short-range dependent processes. They proposed a consistent estimator for the spectral density, and gave some conditions under which the centered estimator satisfies a Central Limit Theorem. These results are based on the asymptotic theory of stationary processes developed by Wu [76]. This framework enables to deal with most of the statistical procedures from time series, including the estimation of the spectral density. However the class of processes satisfying the $\mathbb{L}^2$ "physical dependence measure" introduced by Wu is included in the class of processes satisfying Hannan's condition. In this paper, we prove the consistency of an estimator of the spectral density of the error process under Hannan's condition. Compared to Wu's precise results on the estimation of the spectral density (Central Limit Theorem, rates of convergence, deviation inequalities), our result is only a consistency result, but it holds under Hannan's condition, that is for most of short-range dependent processes.

Finally, we use these general results to modify the usual Fisher tests in cases where $(\epsilon_i)_{i\in\mathbb{Z}}$ and the design verify the conditions of Hannan, and we perform simulations with different models. For these simulations, we need to choose how many covariance terms have to be estimated. In this paper this number is chosen by considering only the autocovariance graph of the residuals. Developing a data driven criterion would be more satisfying. This is probably a very difficult question in such a general context; for this reason it is left out of the scope of the present paper.

The paper is organized as follows. In Section 1.2, we recall Hannan's Central Limit Theorem for the least squares estimator and define the class of "regular designs" (we also give many examples of such designs). In Section 1.3, we focus on the estimation of the spectral density of the error process under Hannan's condition. In Section 1.4, some examples of stationary processes satisfying Hannan's condition are presented. Finally, Section 1.5 is devoted to the correction of the usual Fisher tests in our dependent context, and some simulations are realized.

1.2 Hannan's theorem and regular design

1.2.1 Notations and definitions

Let us recall the equation of the linear regression model:

$$Y = X\beta + \epsilon, \tag{1.1}$$

where X is a deterministic design matrix and ϵ is an error process defined on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$. 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 $\{1, \dots, p\}$ and i in $\{1, \dots, n\}$. 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 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.*

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

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

1.2.2 Hannan's Central Limit Theorem

Let $\hat{\beta}$ be the usual least squares estimator for the unknown vector β. Hannan [47] 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:

$$\sum_{i\in\mathbb{Z}} \|P_0(\epsilon_i)\|_{\mathbb{L}^2} < +\infty. \tag{1.C1}$$

Note that this condition implies that:

$$\sum_{k\in\mathbb{Z}} |\gamma(k)| < \infty, \tag{1.2}$$

(see for instance Dedecker, Merlevède and Volný [38]).

Hannan's condition provides a very general framework for stationary processes. The hypothesis (1.C1) is a sharp condition to have a Central Limit Theorem for the partial sum sequence (see the paper of Dedecker, Merlevède and Volný [38] for more details). Notice that the condition (1.2) implies that the error process is short-range dependent. However, Hannan's condition is satisfied for most short-range dependent stationary processes. In particular, it is less restrictive that the well-known condition of Gordin [45]. Moreover the property of 2-strong stability introduced by Wu [74] is more restrictive than Hannan's condition. This property of 2-strong stability will be recalled in Section 1.4, where large classes of examples will be fully described.

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 the following conditions:

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

and:

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

Moreover, we assume that the following limits exist:

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

Notice that there is a misprint in Hannan's paper (the supremum is missing on condition (1.C3)). Note that Conditions (1.C2) and (1.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 (1.C4).

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

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

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

$$R(0) > 0. \tag{1.C5}$$

Let then F and G be the matrices:

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

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

The Central Limit Theorem for the regression parameter, due to Hannan [47], can be stated as follows:

Theorem 1.1. *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* (1.C1)*. Assume that the design X satisfies the conditions* (1.C2)*,* (1.C3)*,* (1.C4) *and* (1.C5)*. Then:*

$$D(n)(\hat{\beta} - \beta) \xrightarrow[n\to\infty]{\mathcal{L}} \mathcal{N}(0, F^{-1}GF^{-1}). \tag{1.3}$$

Furthermore, there is the convergence of second order moments:[2]

$$\mathbb{E}\left(D(n)(\hat{\beta} - \beta)(\hat{\beta} - \beta)^t D(n)^t\right) \xrightarrow[n\to\infty]{} F^{-1}GF^{-1}. \tag{1.4}$$

1.2.3 Regular design

Theorem 1.1 is very general because it includes a very large class of designs. In this paper, we will focus on the case where the design is regular in the following sense:

Definition 1.2 (Regular design)**.** *A fixed design X is called regular if, for any j, l in $\{1, \ldots, p\}$, the coefficients $\rho_{j,l}(k)$ do not depend on k.*

A large class of regular designs is the one for which the columns are regularly varying sequences. Let us recall the definition of regularly varying sequences :

Definition 1.3 (Regularly varying sequence [71])**.** *A sequence $S(\cdot)$ is regularly varying if and only if it can be written as:*

$$S(i) = i^\alpha L(i),$$

2. The transpose of a matrix X is denoted by X^t.

where $-\infty < \alpha < \infty$ and $L(\cdot)$ is a slowly varying sequence.

This includes the case of polynomial regression, where the columns are of the form: $x_{i,j} = i^j$.

Proposition 1.1. *Assume that each column $X_{.,j}$ is regularly varying with parameter α_j. If the parameters α_j are all strictly greater than $-\frac{1}{2}$, then Conditions* (1.C2)*,* (1.C3) *and* (1.C4) *on the design are satisfied. Moreover, for all j and l in $\{1,\dots,p\}$, the coefficients $\rho_{j,l}(k)$ do not depend on k and are equal to $\frac{\sqrt{2\alpha_j+1}\sqrt{2\alpha_l+1}}{\alpha_j+\alpha_l+1}$. Thereby, the design is regular, and* (1.C5) *is satisfied provided $\alpha_j \neq \alpha_l$ for any distinct j,l in $\{1,\dots,p\}$.*

An other important class of regular designs are the ANOVA type designs. An ANOVA design is represented by a matrix whose column vectors are orthogonal to one another. Each coordinate of a column is either 0 or 1, with consecutive sequences of 1's. The number of 0's and 1's in each column tends to infinity as n tends to infinity.

Note that a design whose columns are either ANOVA or regularly varying is again a regular design.

1.2.4 The asymptotic covariance matrix for regular design

For regular design, the asymptotic covariance matrix is easy to compute. Actually, we shall see that it is the same as in the case where the errors are independent up to a multiplicative factor. More precisely, the usual variance term $\sigma^2 = \mathbb{E}(\epsilon_0^2)$ should be replaced by the sum of covariances : $\sum_k \gamma(k)$.

Since the coefficients $\rho_{j,l}(k)$ are constant, the spectral measure F_X is the product of a Dirac mass at 0, denoted δ_0, with the matrix $R(k)$; consequently the spectral measure F_X is equal to $\delta_0 R(0)$. Notice that, in the case of regular design, the matrix $R(k) = [\rho_{j,l}(k)]$ is equal to $R(0) = [\rho_{j,l}(0)]$.

Thereby the matrix F and G can be computed explicitly:

$$F = \frac{1}{2\pi}\int_{-\pi}^{\pi} F_X(d\lambda) = \frac{1}{2\pi}\int_{-\pi}^{\pi} R(0)\delta_0(d\lambda) = \frac{1}{2\pi}R(0), \tag{1.5}$$

$$G = \frac{1}{2\pi}\int_{-\pi}^{\pi} F_X(d\lambda)\otimes f(\lambda) = \frac{1}{2\pi}\int_{-\pi}^{\pi} R(0)\otimes f(\lambda)\delta_0(d\lambda) = \frac{1}{2\pi}R(0)\otimes f(0) = f(0)F. \tag{1.6}$$

Thus, using (1.5) and (1.6), the covariance matrix can be written as:

$$F^{-1}GF^{-1} = f(0)F^{-1}.$$

The connection between the spectral density and the autocovariance function is known:

$$f(\lambda) = \frac{1}{2\pi}\sum_{k=-\infty}^{\infty}\gamma(k)e^{-ik\lambda}, \qquad \lambda\in[-\pi,\pi].$$

and at the point 0:

$$f(0) = \frac{1}{2\pi}\sum_{k=-\infty}^{\infty}\gamma(k).$$

Thereby the covariance matrix can be written:

$$f(0)F^{-1} = \left(\frac{1}{2\pi}\sum_{k=-\infty}^{\infty}\gamma(k)\right)F^{-1} = \left(\sum_{k=-\infty}^{\infty}\gamma(k)\right)R(0)^{-1},$$

since $F = \frac{R(0)}{2\pi}$ and $F^{-1} = 2\pi R(0)^{-1}$.

In conclusion, for regular design the following corollary holds:

Corollary 1.1. *Under the assumptions of Theorem 1.1, if moreover the design X is regular, then:*

$$D(n)(\hat{\beta}-\beta) \xrightarrow[n\to\infty]{\mathcal{L}} \mathcal{N}\left(0, \left(\sum_{k=-\infty}^{\infty}\gamma(k)\right)R(0)^{-1}\right), \tag{1.7}$$

and we have the convergence of the second order moment:

$$\mathbb{E}\left(D(n)(\hat{\beta}-\beta)(\hat{\beta}-\beta)^t D(n)^t\right) \xrightarrow[n\to\infty]{} \left(\sum_{k=-\infty}^{\infty}\gamma(k)\right)R(0)^{-1}. \tag{1.8}$$

One can see that, in the case of regular design, the asymptotic covariance matrix is similar to the one in the case where the random variables (ϵ_i) are i.i.d.; the variance term σ^2 is replaced by the series of covariances. Actually the matrix $R(0)^{-1}$ is the normalised limit of the matrix $(X^tX)^{-1}$. It is formed by the coefficients $\rho_{j,l}(0)$, which are, in this case, the limit of the normalised scalar products between the columns of the design.

Thus, to obtain confidence regions and tests for β, an estimator of the covariance matrix is needed. More precisely, it is necessary to estimate the quantity:

$$\sum_{k=-\infty}^{\infty}\gamma(k). \tag{1.9}$$

1.3 Estimation of the series of covariances

The properties of spectral density estimates have been discussed in many classical textbooks on time series; see, for instance, Anderson [4], Brillinger [22], Brockwell and Davis [23], Grenander and Rosenblatt [46], Priestley [67] and Rosenblatt [70] among others. But many of the previous results require restrictive conditions on the underlying processes (linear structure or strong mixing conditions). Wu [59] has developed an asymptotic theory for the spectral density estimate $f_n(\lambda)$, defined at (1.10), which extends the applicability of spectral analysis to nonlinear and/or non-strong mixing processes. In particular, he also proved a Central Limit Theorem and deviation inequalities for $f_n(\lambda)$. However, to show his results, Wu uses a notion of dependence that is more restrictive than Hannan's.

In this section, we propose an estimator of the spectral density under Hannan's dependence condition. Here, contrary to the precise results of Wu (Central Limit Theorem, deviation inequalities), we shall

only focus on the consistency of the estimator.

Let us first consider a preliminary random function defined as follows, for λ in $[-\pi, \pi]$:

$$f_n(\lambda) = \frac{1}{2\pi} \sum_{|k| \leq n-1} K\left(\frac{|k|}{h_n}\right) \hat{\gamma}_k e^{ik\lambda}, \tag{1.10}$$

where:

$$\hat{\gamma}_k = \frac{1}{n} \sum_{j=1}^{n-|k|} \epsilon_j \epsilon_{j+|k|}, \qquad 0 \leq |k| \leq (n-1), \tag{1.11}$$

and K is the kernel defined by:

$$\begin{cases} K(x) &= 1 & if\ |x| \leq 1 \\ K(x) &= 2 - |x| & if\ 1 \leq |x| \leq 2 \\ K(x) &= 0 & if\ |x| > 2. \end{cases}$$

The sequence of positive real numbers h_n is such that h_n tends to infinity and $\frac{h_n}{n}$ tends to 0 as n tends to infinity.

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

$$\hat{\epsilon}_i = Y_i - (x_i)^t \hat{\beta} = Y_i - \sum_{j=1}^{p} x_{i,j} \hat{\beta}_j,$$

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

$$f_n^*(\lambda) = \frac{1}{2\pi} \sum_{|k| \leq n-1} K\left(\frac{|k|}{h_n}\right) \hat{\gamma}_k^* e^{ik\lambda}, \qquad \lambda \in [-\pi, \pi], \tag{1.12}$$

where:

$$\hat{\gamma}_k^* = \frac{1}{n} \sum_{j=1}^{n-|k|} \hat{\epsilon}_j \hat{\epsilon}_{j+|k|}, \qquad 0 \leq |k| \leq (n-1). \tag{1.13}$$

Theorem 1.2 concludes this section:

Theorem 1.2. *Let h_n be a sequence of positive numbers such that $h_n \to \infty$ as $n \to \infty$ and:*

$$h_n \mathbb{E}\left(|\epsilon_0|^2 \left(1 \wedge \frac{h_n}{n} |\epsilon_0|^2\right)\right) \xrightarrow[n \to \infty]{} 0. \tag{1.14}$$

Then, under the assumptions of Theorem 1.1:

$$\sup_{\lambda \in [-\pi, \pi]} \|f_n^*(\lambda) - f(\lambda)\|_{\mathbb{L}^1} \xrightarrow[n \to \infty]{} 0. \tag{1.15}$$

Remark 1.1. *If ϵ_0 is in $\mathbb{L}^2$, then there exists $h_n \to \infty$ such that* (1.14) *holds.*

Remark 1.2. *Let us suppose that the random variable ϵ_0 is such that $\mathbb{E}\left(|\epsilon_0|^{\delta+2}\right) < \infty$, with $\delta \in]0, 2]$.*

Since for all real x, $1 \wedge |x|^2 \leq |x|^\delta$, *we have:*

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

Thus if h_n *satisfies* $\frac{h_n^{1+\delta/2}}{n^{\delta/2}} \xrightarrow[n\to\infty]{} 0$, *then* (1.14) *holds. In particular, if the random variable* ϵ_0 *has a fourth order moment, then the condition on* h_n *is* $\frac{h_n^2}{n} \xrightarrow[n\to\infty]{} 0$.

Theorem 1.1 implies the following result:

Corollary 1.2. *Under the assumptions of Corollary 1.1, and if* $f(0) > 0$, *then:*

$$\frac{R(0)^{\frac{1}{2}}}{\sqrt{2\pi f_n^*(0)}} D(n)(\hat{\beta} - \beta) \xrightarrow[n\to\infty]{\mathcal{L}} \mathcal{N}(0, I_p), \tag{1.16}$$

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

1.4 Examples of stationary processes

In this section, we present some classes of stationary processes satisfying Hannan's condition.

1.4.1 Functions of Linear processes

A large class of stationary processes for which one can check Hannan's condition is the class of smooth functions of linear processes generated by i.i.d. random variables.

Let us take $\Omega = \mathbb{R}^{\mathbb{Z}}$ and $\mathbb{P} = \mu^{\otimes\mathbb{Z}}$, where μ is a probability measure on $\mathbb{R}$. Let $(\eta_i, i \in \mathbb{Z})$ be a sequence of i.i.d. random variables with marginal distribution μ. Let $(a_i)_{i\in\mathbb{Z}}$ be a sequence of real numbers in ℓ^1, and assume that $\sum_{i\in\mathbb{Z}} a_i\eta_i$ is defined almost surely. The random variable ϵ_0 is square integrable and is regular with respect to the σ-algebras : $\mathcal{F}_i = \sigma(\eta_j, j \leq i)$. We focus on functions of real-valued linear processes:

$$\epsilon_k = F\left(\sum_{i\in\mathbb{Z}} a_i\eta_{k-i}\right) - \mathbb{E}\left(F\left(\sum_{i\in\mathbb{Z}} a_i\eta_{k-i}\right)\right).$$

Let us define the modulus of continuity of F on the interval $[-M, M]$ by:

$$\omega_{\infty,F}(h, M) = \sup_{|t|\leq h, |x|\leq M, |x+t|\leq M} |F(x+t) - F(x)|.$$

Let $(\eta_i')_{i\in\mathbb{Z}}$ be an independent copy of $(\eta_i)_{i\in\mathbb{Z}}$, and let:

$$M_k = \max\left\{\left|\sum_{i\in\mathbb{Z}} a_i\eta_i'\right|, \left|a_k\eta_0 + \sum_{i\neq k} a_i\eta_i'\right|\right\}.$$

According to Section 5 in the paper of Dedecker, Merlevède, Volný [38], if the following condition

holds:

$$\sum_{k\in\mathbb{Z}} \left\| \omega_{\infty,F}(|a_k||\eta_0|, M_k) \wedge \|\epsilon_0\|_\infty \right\|_{\mathbb{L}^2} < \infty, \tag{1.17}$$

then Hannan's condition holds. We have an interesting application if the function F is γ-Hölder on any compact set; if $\omega_{\infty,F}(h, M) \leq Ch^\gamma M^\alpha$ for some $C > 0$, $\gamma \in]0, 1]$ and $\alpha \geq 0$, then (1.17) holds as soon as $\sum |a_k|^\gamma < \infty$ and $\mathbb{E}(|\eta_0|^{2(\alpha+\gamma)}) < \infty$.

1.4.2 2-strong stability

Let us recall in this section the framework used by Wu. We consider stationary processes of the form:

$$\epsilon_i = H(\ldots, \eta_{i-1}, \eta_i),$$

where η_i, i in $\mathbb{Z}$, are i.i.d. random variables and H is a measurable function. Assume that ϵ_0 belongs to $\mathbb{L}^2$, and let η_0' be distributed as η_0 and independent of (η_i). Let us define the physical dependence measure in $\mathbb{L}^2$ [76], for $j \geq 0$:

$$\delta_2(j) = \left\| \epsilon_j - \epsilon_j^* \right\|_{\mathbb{L}^2},$$

where ϵ_j^* is a coupled version of ϵ_j with η_0 in the latter being replaced by η_0':

$$\epsilon_j^* = H(\ldots, \eta_{-1}, \eta_0', \eta_1, \ldots, \eta_{j-1}, \eta_j).$$

The sequence $(\epsilon_i)_{i\in\mathbb{Z}}$ is said to be 2-strong stable if:

$$\Delta_2 = \sum_{j=0}^{\infty} \delta_2(j) < \infty.$$

As a consequence of Theorem 1, $(i) - (ii)$ of Wu [74], we infer that if $(\epsilon_i)_{i\in\mathbb{Z}}$ is 2-strong stable, then it satisfies Hannan's condition with respect to the filtration $\mathcal{F}_i = \sigma(\eta_j, j \leq i)$. Many examples of 2-strong stable processes are presented in the paper by Wu [74]. We also refer to [76] for other examples.

1.4.3 Conditions in the style of Gordin

According to Proposition 5 of Dedecker, Merlevède and Volný [38], Hannan's condition holds if the error process satisfies the two following conditions:

$$\sum_{k=1}^{\infty} \frac{1}{\sqrt{k}} \left\| \mathbb{E}(\epsilon_k | \mathcal{F}_0) \right\|_{\mathbb{L}^2} < \infty \tag{1.18}$$

$$\sum_{k=1}^{\infty} \frac{1}{\sqrt{k}} \left\| \epsilon_{-k} - \mathbb{E}(\epsilon_{-k} | \mathcal{F}_0) \right\|_{\mathbb{L}^2} < \infty. \tag{1.19}$$

These conditions are weaker than the well-known conditions of Gordin [45], under which a martingale + coboundary decomposition holds in $\mathbb{L}^2$. An application is given in the next subsection.

1.4.4 Weak dependent coefficients

Hannan's condition holds if the error process is weakly dependent. In this case, the $(\epsilon_i)_{i\in\mathbb{Z}}$ process is $\mathcal{F}$-adapted and Condition (1.19) is always true.

Let us recall the definitions of weak dependence coefficients, introduced by Dedecker and Prieur [39]; for all integer $k \geq 0$:

$$\tilde{\phi}(k) = \tilde{\phi}(\mathcal{F}_0, \epsilon_k) = \sup_{t\in\mathbb{R}} \|\mathbb{P}(\epsilon_k \leq t|\mathcal{F}_0) - \mathbb{P}(\epsilon_k \leq t)\|_\infty ,$$

and:

$$\tilde{\alpha}(k) = \tilde{\alpha}(\mathcal{F}_0, \epsilon_k) = \sup_{t\in\mathbb{R}} \|\mathbb{P}(\epsilon_k \leq t|\mathcal{F}_0) - \mathbb{P}(\epsilon_k \leq t)\|_{\mathbb{L}^1} .$$

If $(\epsilon_i)_{i\in\mathbb{Z}}$ is $\tilde{\phi}$-dependent and is in $\mathbb{L}^p$ with $p \in [2, +\infty[$, then by Hölder's inequality:

$$\|\mathbb{E}(\epsilon_k|\mathcal{F}_0)\|_{\mathbb{L}^2} \leq \|\mathbb{E}(\epsilon_k|\mathcal{F}_0)\|_{\mathbb{L}^p} \leq \sup_{Z\in B_{\frac{p}{p-1}}(\mathcal{F}_0)} \mathbb{E}(Z\epsilon_k) \leq 2\tilde{\phi}(k)^{\frac{p-1}{p}} \|\epsilon_0\|_{\mathbb{L}^p} ,$$

where for all $q \in]1, 2]$, $B_q(\mathcal{F}_0)$ is the set of random variables Z, $\mathcal{F}_0$-measurable such that $\|Z\|_{\mathbb{L}^q} \leq 1$.

Consequently, if:

$$\sum_{k=1}^{\infty} \frac{1}{\sqrt{k}} \tilde{\phi}(k)^{\frac{p-1}{p}} < \infty, \tag{1.20}$$

then the condition (1.18) holds, and Hannan's condition is satisfied.

Now we look at the $\tilde{\alpha}$-weakly dependent sequence. We denote Q_ϵ the generalized inverse function of $x \to \mathbb{P}(|\epsilon| > x)$. If $(\epsilon_i)_{i\in\mathbb{Z}}$ is $\tilde{\alpha}$-mixing and verifies that there exists $r \in]2, +\infty[$, such that $\mathbb{P}(|\epsilon| \geq t) \leq t^{-r}$, then, by Cauchy-Schwarz's inequality and Rio's inequality (Theorem 1.1 [68]), we get:

$$\|\mathbb{E}(\epsilon_k|\mathcal{F}_0)\|_{\mathbb{L}^2} = \sup_{Z\in B_2(\mathcal{F}_0)} \mathbb{E}(Z\epsilon_k) \leq 2\left(\int_0^{\tilde{\alpha}(k)} Q^2_{\epsilon_k}(u)du\right)^{\frac{1}{2}}.$$

But:

$$\int_0^{\tilde{\alpha}(k)} Q^2_{\epsilon_k}(u)du \leq \int_0^{\tilde{\alpha}(k)} \frac{1}{u^{\frac{2}{r}}}du \leq \tilde{\alpha}(k)^{1-\frac{2}{r}}.$$

Hence, if:

$$\sum_{k=1}^{\infty} \frac{\tilde{\alpha}(k)^{\frac{1}{2}-\frac{1}{r}}}{\sqrt{k}} < \infty, \tag{1.21}$$

then (1.18) is true, and Hannan's condition is satisfied.

Notice that all we have written for $\tilde{\alpha}$-dependent sequences is also true for α-mixing processes in the sense of Rosenblatt [70].

1.5 Tests and Simulations

We consider the linear regression model (1.1), and we assume that Hannan's condition (1.C1) as well as the conditions (1.C2) to (1.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 Fisher tests can be modified and adapted to the case where the errors are short-range dependent.

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 (1.1) holds).

In the case of regular design, thanks to Corollary 1.2, the usual Fisher tests to test H_0 versus H_1, can be corrected by replacing the estimator of $\sigma^2 = \mathbb{E}(\epsilon_0^2)$ by an estimator of: $\sum_k \gamma(k)$.

Recall that if the errors are i.i.d. Gaussian random variables, the test statistic is:

$$F = \frac{1}{p - p_0} \times \frac{RSS_0 - RSS}{\hat{\sigma}_\epsilon^2}. \tag{1.22}$$

In this expression, the integer p_0 is the dimension of the model under the H_0-hypothesis, RSS is the sum of the squares of the residuals for the complete model (1.1) (equal to $\|\hat{\epsilon}\|_2^2$), RSS_0 is the corresponding quantity under H_0, and $\hat{\sigma}_\epsilon^2$ is the estimator of the variance of ϵ_0 (equal to $\frac{RSS}{n-p}$). Under H_0, the quantity F follows a Fisher distribution with parameters $(p - p_0, n - p)$.

In the case where the design satisfies Hannan's conditions, if the random variables (ϵ_i) are i.i.d. but do not necessarily follow a gaussian distribution, the test statistic is the same as (1.22) and converges to a χ^2-distribution under the H_0-hypothesis:

$$F \xrightarrow[n\to\infty]{\mathcal{L}} \frac{\chi^2(p - p_0)}{p - p_0}.$$

Now if the error process $(\epsilon_i)_{i\in\mathbb{Z}}$ is stationary, the test statistic must be corrected as follows:

$$\tilde{F}_c = \frac{1}{p - p_0} \times \frac{RSS_0 - RSS}{2\pi f_n^*(0)}, \tag{1.23}$$

where f_n^* is defined in (1.12). Thanks to Corollary 1.2, it converges to a χ^2-distribution:

$$\tilde{F}_c \xrightarrow[n\to\infty]{\mathcal{L}} \frac{\chi^2(p - p_0)}{p - p_0}.$$

In practice, we shall only estimate a finite number of $\gamma(k)$, say a_n. For the simulations, we shall use the graph of the empirical autocovariance of the residuals to choose a_n, and instead of (1.23), we shall consider the statistics:

$$F_c = \frac{1}{p - p_0} \times \frac{RSS_0 - RSS}{\hat{\gamma}_0^* + 2\sum_{k=1}^{a_n} \hat{\gamma}_k^*}, \tag{1.24}$$

with $\hat{\gamma}_k^*$ defined in (1.13).

1.5.1 Example 1: A non-mixing autoregressive process

The process $(\epsilon_1, \ldots, \epsilon_n)$ is simulated, according to the AR(1) equation:

$$\epsilon_{k+1} = \frac{1}{2}(\epsilon_k + \eta_{k+1}),$$

where ϵ_1 is uniformly distributed over $[-\frac{1}{2}, \frac{1}{2}]$, and $(\eta_i)_{i\geq 2}$ is a sequence of i.i.d. random variables, independent of ϵ_1, such that $\mathbb{P}(\eta_i = -\frac{1}{2}) = \mathbb{P}(\eta_i = \frac{1}{2}) = \frac{1}{2}$. In this example, $\mathcal{F}_i = \sigma(\eta_k, k \leq i)$, and the σ-algebra $\mathcal{F}_{-\infty}$ is trivial.

The transition kernel of the chain $(\epsilon_i)_{i\geq 1}$ is:

$$K(f)(x) = \frac{1}{2}\left(f\left(\frac{x}{2} + \frac{1}{4}\right) + f\left(\frac{x}{2} - \frac{1}{4}\right)\right),$$

and the uniform distribution on $[-\frac{1}{2}, \frac{1}{2}]$ is the unique invariant distribution by K. Hence, the chain $(\epsilon_i)_{i\geq 1}$ is strictly stationary. Furthermore, it is not α-mixing in the sense of Rosenblatt [21], but it is $\tilde{\phi}$-dependent. Indeed, one can prove that the coefficient $\tilde{\phi}$ of the chain $(\epsilon_i)_{i\geq 1}$ decreases geometrically [39]:

$$\tilde{\phi}(k) \leq 2^{-k}.$$

Consequently Hannan's conditions are satisfied and the Fisher tests can be corrected as indicated above.

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

$$Y_i = \beta_0 + \beta_1 i + 10\epsilon_i.$$

The random variables ϵ_i are multiplied by 10 to increase the variance. The coefficient β_0 is chosen equal to 3. We test the hypothesis H_0: $\beta_1 = 0$, against the hypothesis H_1: $\beta_1 \neq 0$.

The estimated level of the Fisher test will be studied for different choices of n and a_n, which is the number of covariance terms considered. Under the hypothesis H_0, the same Fisher 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 $a_n = 0$ (no correction):

n	200	400	600	800	1000
Estimated level	0.2745	0.2655	0.2615	0.2845	0.2445

Here, since $a_n = 0$, 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 quantities a_n may be chosen by analyzing the graph of the empirical autocovariances, Figure 1.1, obtained with $n = 600$. For this example, this graph suggests a choice of $a_n = 2$ or 3.

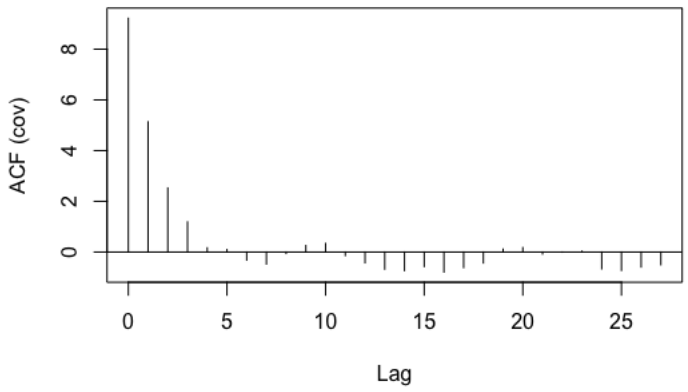


Figure 1.1 – Empirical autocovariances for the first model of Example 1, n = 600.

• Case $\beta_1 = 0$, $a_n = 2$:

n	200	400	600	800	1000
Estimated level	0.0805	0.086	0.0745	0.0675	0.077

As suggested by the graph of the empirical autocovariances, the choice $a_n = 2$ gives a better estimated level than $a_n = 0$.

• Case $\beta_1 = 0$, $a_n = 3$:

n	200	400	600	800	1000
Estimated level	0.078	0.0725	0.074	0.059	0.0625

Here, we see that the choice $a_n = 3$ works well also, and seems even slightly better than the choice $a_n = 2$. If one increases the size of the samples n, and the number of estimated covariance terms a_n, we are getting closer to the estimated level 5 %. If $n = 5000$ and $a_n = 4$, the estimated level is around 0.05.

• Case $\beta_1 = 0.005$, $a_n = 3$:

In this example, H_0 is not satisfied. We choose β_1 equal to 0.005, 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	0.2255	0.728	0.9945	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 0.2255, 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 $\{1, ..., n\}$:

$$Y_i = \beta_0 + \beta_1 i + \beta_2 i^2 + 10\epsilon_i.$$

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

• Case $\beta_1 = \beta_2 = 0$ and $a_n = 0$ (no correction):

n	200	400	600	800	1000
Estimated level	0.402	0.378	0.385	0.393	0.376

As for the first simulation, if $a_n = 0$ the test will reject the null hypothesis too often.

As suggested by the graph of the estimated autocovariances Figure 1.2, the choice $a_n = 4$ should give a better result for the estimated level.

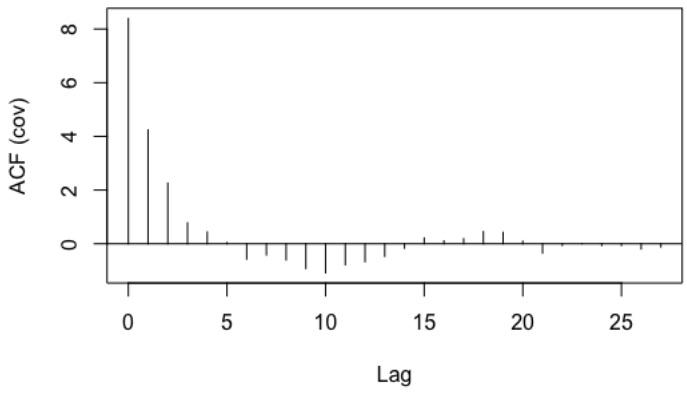


Figure 1.2 – Empirical autocovariances for the second model of Example 1, n = 600.

• Case $\beta_1 = \beta_2 = 0$, $a_n = 4$:

n	200	400	600	800	1000
Estimated level	0.103	0.076	0.069	0.056	0.063

Here, we see that the choice $a_n = 4$ works well. For $n = 1000$, the estimated level is around 0.06. If $n = 2000$ and $a_n = 4$, the estimated level is around 0.05.

• Case $\beta_1 = 0.005$, $\beta_2 = 0$, $a_n = 4$:

Now, we study the estimated power of the test. The coefficient β_1 is chosen equal to 0.005 and β_2 is zero.

n	200	400	600	800	1000
Estimated power	0.2145	0.634	0.9855	1	1

As expected, the estimated power increases with the size of the samples, and it is around 1 as soon as $n = 800$.

The third model that we consider is the following linear regression model, for all i in $\{1, ..., n\}$:

$$Y_i = \beta_0 + \beta_1\sqrt{i} + \beta_2 \log(i) + 10\epsilon_i.$$

We test again the hypothesis H_0: $\beta_1 = \beta_2 = 0$ against H_1: $\beta_1 \neq 0$ or $\beta_2 \neq 0$. The coefficient β_0 is equal to 3. The conditions of the simulation are the same as above except for the size of the samples. Indeed, for this model, the size of the samples n must be greater than previously to have an estimated level close to 5% with the correction.

• Case $\beta_1 = \beta_2 = 0$ and $a_n = 0$ (no correction):

n	500	1000	2000	3000	4000	5000
Estimated level	0.4435	0.4415	0.427	0.3925	0.397	0.4075

As for the first and second simulation, if $a_n = 0$ the test will reject the null hypothesis too often.

As suggested by the graph of the estimated autocovariances Figure 1.3, the choice $a_n = 4$ should give a better result for the estimated level.

• Case $\beta_1 = \beta_2 = 0$, $a_n = 4$:

n	500	1000	2000	3000	4000	5000
Estimated level	0.106	0.1	0.078	0.072	0.077	0.068

For $a_n = 4$ and $n = 5000$, the estimated level is around 0.07. If $n = 10000$, it is around 5%.

Then, we study the estimated power of the test for β_0 or β_1 non equal to 0.

• Case $\beta_1 = 0$, $\beta_2 = 0.2$, $a_n = 4$:

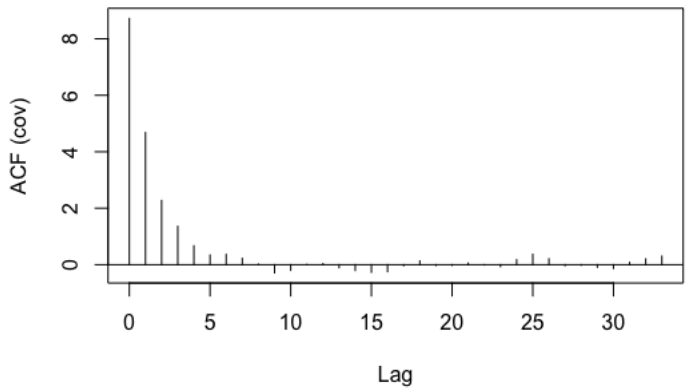


Figure 1.3 – Empirical autocovariances for the third model of Example 1, n = 2000.

n	500	1000	2000	3000	4000	5000
Estimated power	0.2505	0.317	0.4965	0.6005	0.725	0.801

As expected, the estimated power increases with the size of the samples, and it is around 0.8 as soon as $n = 5000$.

1.5.2 Example 2: Intermittent maps

For γ in $]0, 1[$, we consider the intermittent map θ_γ from $[0, 1]$ to $[0, 1]$, introduced by Liverani, Saussol and Vaienti [60]:

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

It follows from [60] that there exists a unique absolutely continuous θ_γ-invariant probability measure ν_γ, with density h_γ.

Let us briefly describe the Markov chain associated with θ_γ, and its properties. Let first K_γ be the Perron-Frobenius operator of θ_γ with respect to ν_γ, defined as follows: for any functions u, v in $\mathbb{L}^2([0, 1], \nu_\gamma)$:

$$\nu_\gamma(u \cdot v \circ \theta_\gamma) = \nu_\gamma(K_\gamma(u) \cdot v).$$

The operator K_γ is a transition kernel, and ν_γ is invariant by K_γ. Let now $(\xi_i)_{i\geq 1}$ be a stationary Markov chain with invariant measure ν_γ and transition kernel K_γ. It is well-known that on the probability space $([0, 1], \nu_\gamma)$, the random vector $(\theta_\gamma, \theta_\gamma^2, \ldots, \theta_\gamma^n)$ is distributed as $(\xi_n, \xi_{n-1}, \ldots, \xi_1)$. Now it is proved in

Dedecker, Gouëzel, Merlevède [37] that there exists two positive constants A, B such that:

$$\frac{A}{(n+1)^{\frac{1-\gamma}{\gamma}}} \leq \tilde{\alpha}_\xi(n) \leq \frac{B}{(n+1)^{\frac{1-\gamma}{\gamma}}}$$

Moreover, the chain $(\xi_i)_{i\geq 1}$ is not α-mixing in the sense of Rosenblatt [69].

In the following simulations, we consider linear regression models, where $\epsilon_i = \theta^i_\gamma$. But, in our context, the coefficient γ must belong to $]0, \frac{1}{2}[$. Indeed, if γ is lower than $\frac{1}{2}$, then Condition (1.21) is verified. Consequently, Hannan's condition is satisfied and we can apply our results. Note that if γ is greater than $\frac{1}{2}$, then the chain (ξ_i) is long-range dependent (see the introduction in Dedecker, Dehling and Taqqu [36]).

Recall that our results apply only in the short-range dependent case, so we shall only consider the case where $\gamma < \frac{1}{2}$. For the simulations, the coefficient γ is chosen equal to $\frac{1}{4}$. Consequently, $\tilde{\alpha}(n)$ is of order n^{-3}, which is quite slow. In addition, if $\mathcal{F}_i = \sigma(\xi_k, k \leq i)$ then $\mathcal{F}_{-\infty}$ is trivial (see for instance [37]).

Note that, in this example, the mean of the errors is not equal to 0, but this is not an issue because, it will only modified the intercept term in our different models.

For the first simulation, we consider the following linear regression model, for all i in $\{1, \ldots, n\}$:

$$Y_i = \beta_0 + \beta_1 i + 10\epsilon_i,$$

where the hypothesis H_0 is: $\beta_1 = 0$, and the hypothesis H_1 is: $\beta_1 \neq 0$. Again the coefficient β_0 is equal to 3 and the random variables ϵ_i are multiplied by 10 to increase the variance.

We shall study the estimated level of the test for different choices of n and a_n, which is the number of covariance terms considered. With intermittent maps the convergence is slower; the coefficient $\tilde{\alpha}$ do not decrease geometrically. Thereby we consider larger samples ($n = 500$ to $n = 5000$, sometimes $n = 10000$ or 20000). Under the hypothesis H_0, the same Fisher test is carried out 2000 times. Then we look at the frequency of rejection of the test when we are under H_0 (i.e. the level of the test). Let us specify that we want an estimated level close to 5%.

• Case $\beta_1 = 0$ and $a_n = 0$ (no correction):

n	500	1000	2000	3000	4000	5000
Estimated level	0.361	0.365	0.3685	0.371	0.3645	0.349

Here, since $a_n = 0$, 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 quantities a_n may be chosen by analyzing the graph of the empirical autocovariances (see Figure 1.4). In the case of intermittent maps, the number a_n should be larger than for the previous example.

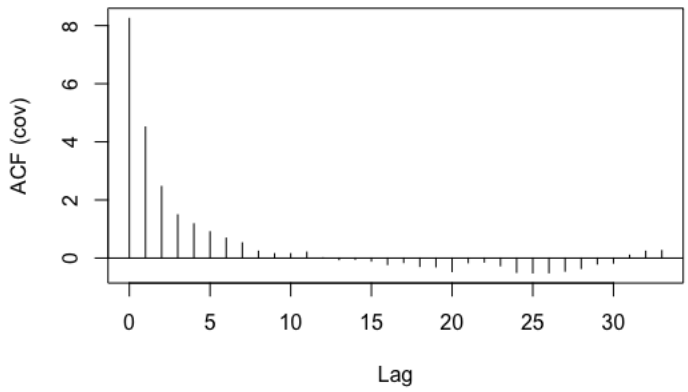


Figure 1.4 – Empirical autocovariances for the first model of Example 2, n = 2000.

• Case $\beta_1 = 0$, $a_n = 5$:

n	500	1000	2000	3000	4000	5000
Estimated level	0.101	0.0805	0.0755	0.073	0.0705	0.0805

• Case $\beta_1 = 0$, $a_n = 6$:

n	500	1000	2000	3000	4000	5000
Estimated level	0.086	0.076	0.0705	0.0635	0.066	0.0675

• Case $\beta_1 = 0$, $a_n = 7$:

n	500	1000	2000	3000	4000	5000
Estimated level	0.09	0.072	0.074	0.0585	0.061	0.06

For small samples ($n = 500$), a_n equal to 5 is enough. The estimated level does not change a lot, and is around 0.095. But for large samples, $a_n = 7$ is better. Indeed, with $n = 5000$ and $a_n = 7$, the estimated level is around 0.06, and if $n = 10000$, this is around 0.05. We see here that an automatic criterion to choose a_n would be useful.

Then, we study the estimated power of the test for β_1 non equal to 0.

• Case $\beta_1 = 0.0005$, $a_n = 6$:

n	500	1000	2000	3000	4000	5000
Estimated power	0.1195	0.1865	0.663	0.979	1	1

As one can see, the estimated power is always greater than 0.05. As expected, the estimated power increases with the size of the samples. For $n = 500$, the power of the test is around 0.12, and for $n = 4000$, the power is around 1. As soon as $n \geq 4000$, the test always rejects the H_0-hypothesis.

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

$$Y_i = \beta_0 + \beta_1 i + \beta_2 i^2 + 10\epsilon_i.$$

We test here the hypothesis H_0: $\beta_1 = \beta_2 = 0$ against H_1: $\beta_1 \neq 0$ or $\beta_2 \neq 0$.
The conditions of the simulation are the same as above.

• Case $\beta_1 = \beta_2 = 0$ and $a_n = 0$ (no correction):

n	500	1000	2000	3000	4000	5000
Estimated level	0.536	0.506	0.5275	0.5165	0.5055	0.4925

As for the first simulation, if $a_n = 0$ the test will reject the null hypothesis too often.

As suggested by the graph of the estimated autocovariances, the choice $a_n = 6$ or 7 should give a better result for the estimated level.

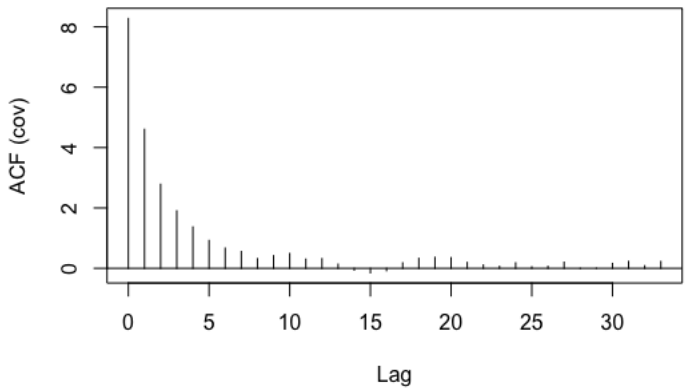


Figure 1.5 – Empirical autocovariances for the second model of Example 2, n = 2000.

• Case $\beta_1 = \beta_2 = 0$, $a_n = 5$:

n	500	1000	2000	3000	4000	5000
Estimated level	0.1265	0.0905	0.078	0.079	0.079	0.085

• Case $\beta_1 = \beta_2 = 0$, $a_n = 6$:

n	500	1000	2000	3000	4000	5000
Estimated level	0.1065	0.1	0.0795	0.08	0.0705	0.0685

• Case $\beta_1 = \beta_2 = 0$, $a_n = 7$:

n	500	1000	2000	3000	4000	5000
Estimated level	0.112	0.0815	0.071	0.07	0.0725	0.0615

As for the first example, for small samples, a_n equal to 5 is enough and it is not necessary to increase the value of a_n. But for large samples, larger values of a_n are required. So for $n = 5000$ and $a_n = 7$, the estimated level is around 0.06. If $n = 20000$ and $a_n = 9$, we approach the level 0.05.

Then, we study the estimated power of the test for β_0 or β_1 non equal to 0.

• Case $\beta_1 = 0.0005$, $\beta_2 = 0$, $a_n = 7$:

n	500	1000	2000	3000	4000	5000
Estimated power	0.13	0.1675	0.5685	0.964	1	1

As expected, the estimated power increases with the size of the samples, and it is around 1 as soon as $n \geq 4000$.

1.6 Proofs

1.6.1 Proposition 1.1

Proof. Let us define:

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

The condition (1.C2) is verified if:

$$\sum_{i=1}^{n} i^{2\alpha_j} L(i)^2 \to \infty. \tag{1.25}$$

When $2\alpha_j < -1$, it is known that (1.25) converges. However, for $2\alpha_j > -1$, thanks to Proposition $2.2.1$ of Pipiras and Taqqu [66], we have the following equivalence:

$$\sum_{i=1}^{n} i^{2\alpha_j} L(i)^2 \sim \frac{n^{2\alpha_j+1} L(n)^2}{2\alpha_j + 1},$$

and this quantity diverges as n tends to infinity. Thus the condition (1.C2) is satisfied if α_j is strictly greater than $-\frac{1}{2}$. We also immediately check that (1.C3) is satisfied.

Now let us compute the coefficients $\rho_{j,l}(k)$ and prove that they do not depend on k. For j, l belonging to $\{1, \dots, p\}$:

$$\sum_{m=1}^{n-k} \frac{x_{m,j} x_{m+k,l}}{d_j(n) d_l(n)} = \frac{\sum_{m=1}^{n-k} m^{\alpha_j} L(m)(m+k)^{\alpha_l} L'(m+k)}{\sqrt{\sum_{i=1}^{n} i^{2\alpha_j} L(i)^2} \sqrt{\sum_{q=1}^{n} q^{2\alpha_l} L'(q)^2}},$$

and we have:

$$\begin{aligned}
&\frac{\sum_{m=1}^{n-k} m^{\alpha_j} L(m)(m+k)^{\alpha_l} L'(m+k)}{\sqrt{\sum_{i=1}^{n} i^{2\alpha_j} L(i)^2} \sqrt{\sum_{q=1}^{n} q^{2\alpha_l} L'(q)^2}} \\
&\qquad = \frac{\sum_{m=1}^{n-k} (m^{\alpha_j}((m+k)^{\alpha_l} - m^{\alpha_l})) L(m) L'(m+k)}{\sqrt{\sum_{i=1}^{n} i^{2\alpha_j} L(i)^2} \sqrt{\sum_{q=1}^{n} q^{2\alpha_l} L'(q)^2}} \\
&\qquad\qquad + \frac{\sum_{m=1}^{n-k} m^{\alpha_j} m^{\alpha_l} L(m) L'(m+k)}{\sqrt{\sum_{i=1}^{n} i^{2\alpha_j} L(i)^2} \sqrt{\sum_{q=1}^{n} q^{2\alpha_l} L'(q)^2}}.
\end{aligned} \tag{1.26}$$

Let us deal with the first term of the right-hand side in (1.26). If $\alpha_l \geq 1$, we get:

$$\frac{\sum_{m=1}^{n-k}(m^{\alpha_j}((m+k)^{\alpha_l}-m^{\alpha_l}))L(m)L'(m+k)}{\sqrt{\sum_{i=1}^{n} i^{2\alpha_j}L(i)^2}\sqrt{\sum_{q=1}^{n} q^{2\alpha_l}L'(q)^2}}$$
$$\leq \frac{\sum_{m=1}^{n-k}(m^{\alpha_j}(k\alpha_l(m+k)^{\alpha_l-1}))L(m)L'(m+k)}{\sqrt{\sum_{i=1}^{n} i^{2\alpha_j}L(i)^2}\sqrt{\sum_{q=1}^{n} q^{2\alpha_l}L'(q)^2}}$$
$$\leq \frac{(k\alpha_l)\sum_{m=1}^{n-k} m^{\alpha_j}(m(1+\frac{k}{m}))^{\alpha_l-1}L(m)L'(m+k)}{\sqrt{\sum_{i=1}^{n} i^{2\alpha_j}L(i)^2}\sqrt{\sum_{q=1}^{n} q^{2\alpha_l}L'(q)^2}},$$

and since $\frac{k}{m} \leq k$:

$$\frac{(k\alpha_l)\sum_{m=1}^{n-k} m^{\alpha_j}(m(1+\frac{k}{m}))^{\alpha_l-1}L(m)L'(m+k)}{\sqrt{\sum_{i=1}^{n} i^{2\alpha_j}L(i)^2}\sqrt{\sum_{q=1}^{n} q^{2\alpha_l}L'(q)^2}}$$
$$\leq \frac{(k\alpha_l)\sum_{m=1}^{n-k} m^{\alpha_j}m^{\alpha_l-1}(1+k)^{\alpha_l-1}L(m)L'(m+k)}{\sqrt{\sum_{i=1}^{n} i^{2\alpha_j}L(i)^2}\sqrt{\sum_{q=1}^{n} q^{2\alpha_l}L'(q)^2}}$$
$$\leq \frac{(k\alpha_l)(1+k)^{\alpha_l-1}\sum_{m=1}^{n} m^{\alpha_j+\alpha_l-1}L(m)L'(m+k)}{\sqrt{\sum_{i=1}^{n} i^{2\alpha_j}L(i)^2}\sqrt{\sum_{q=1}^{n} q^{2\alpha_l}L'(q)^2}}.$$

Using again the proposition of Pipiras and Taqqu, we have:

$$\frac{(k\alpha_l)(1+k)^{\alpha_l-1}\sum_{m=1}^{n} m^{\alpha_j+\alpha_l-1}L(m)L'(m+k)}{\sqrt{\sum_{i=1}^{n} i^{2\alpha_j}L(i)^2}\sqrt{\sum_{q=1}^{n} q^{2\alpha_l}L'(q)^2}}$$
$$\sim \frac{(k\alpha_l)(1+k)^{\alpha_l-1}\frac{n^{\alpha_j+\alpha_l}}{\alpha_j+\alpha_l}L(n)L'(n+k)}{\sqrt{\frac{n^{2\alpha_j+1}}{2\alpha_j+1}L(n)^2}\sqrt{\frac{n^{2\alpha_l+1}}{2\alpha_l+1}L'(n)^2}}$$
$$\sim \frac{\sqrt{2\alpha_j+1}\sqrt{2\alpha_l+1}(k\alpha_l)(1+k)^{\alpha_l-1}}{\alpha_j+\alpha_l}\frac{1}{n}\frac{L'(n+k)}{L'(n)},$$

and this quantity tends to 0 as $n \to \infty$.

With the same idea, if $0 < \alpha_l < 1$ and again for the first term on the right-hand side in (1.26), we have:

$$\frac{\sum_{m=1}^{n-k} m^{\alpha_j}((m+k)^{\alpha_l}-m^{\alpha_l})L(m)L'(m+k)}{\sqrt{\sum_{i=1}^{n} i^{2\alpha_j}L(i)^2}\sqrt{\sum_{q=1}^{n} q^{2\alpha_l}L'(q)^2}}$$
$$\leq \frac{\sum_{m=1}^{n-k}(m^{\alpha_j}(k\alpha_l m^{\alpha_l-1}))L(m)L'(m+k)}{\sqrt{\sum_{i=1}^{n} i^{2\alpha_j}L(i)^2}\sqrt{\sum_{q=1}^{n} q^{2\alpha_l}L'(q)^2}}$$
$$\leq \frac{(k\alpha_l)\sum_{m=1}^{n} m^{\alpha_j+\alpha_l-1}L(m)L'(m+k)}{\sqrt{\sum_{i=1}^{n} i^{2\alpha_j}L(i)^2}\sqrt{\sum_{q=1}^{n} q^{2\alpha_l}L'(q)^2}}.$$

If $\alpha_j + \alpha_l > 0$, we can use the equivalence of Pipiras and Taqqu and show that it converges to 0:

$$\frac{(k\alpha_l)\sum_{m=1}^{n} m^{\alpha_j+\alpha_l-1}L(m)L'(m+k)}{\sqrt{\sum_{i=1}^{n} i^{2\alpha_j}L(i)^2}\sqrt{\sum_{q=1}^{n} q^{2\alpha_l}L'(q)^2}} \sim \frac{(k\alpha_l)\sqrt{2\alpha_j+1}\sqrt{2\alpha_l+1}}{\alpha_j+\alpha_l}\frac{1}{n}\frac{L'(n+k)}{L'(n)}.$$

If $\alpha_j + \alpha_l < 0$, the quantity converges to 0, because the numerator is summable and the denominator tends to infinity. Furthermore, if $\alpha_j + \alpha_l = 0$, the quantity converges to 0 too.

Finally, if $-\frac{1}{2} < \alpha_l < 0$, we have:

$$\begin{aligned}
\frac{\sum_{m=1}^{n-k}(m^{\alpha_j}((m+k)^{\alpha_l}-m^{\alpha_l}))L(m)L'(m+k)}{\sqrt{\sum_{i=1}^{n} i^{2\alpha_j}L(i)^2}\sqrt{\sum_{q=1}^{n} q^{2\alpha_l}L'(q)^2}} &\leq \frac{\sum_{m=1}^{n-k}(m^{\alpha_j}|(m+k)^{\alpha_l}-m^{\alpha_l}|)L(m)L'(m+k)}{\sqrt{\sum_{i=1}^{n} i^{2\alpha_j}L(i)^2}\sqrt{\sum_{q=1}^{n} q^{2\alpha_l}L'(q)^2}} \\
&\leq \frac{\sum_{m=1}^{n-k}(m^{\alpha_j}(k|\alpha_l|m^{\alpha_l-1}))L(m)L'(m+k)}{\sqrt{\sum_{i=1}^{n} i^{2\alpha_j}L(i)^2}\sqrt{\sum_{q=1}^{n} q^{2\alpha_l}L'(q)^2}} \\
&\leq \frac{(k|\alpha_l|)\sum_{m=1}^{n} m^{\alpha_j+\alpha_l-1}L(m)L'(m+k)}{\sqrt{\sum_{i=1}^{n} i^{2\alpha_j}L(i)^2}\sqrt{\sum_{q=1}^{n} q^{2\alpha_l}L'(q)^2}},
\end{aligned}$$

and we get the same results as above.

For the second term on the right-hand side in (1.26), we use again the proposition of Pipiras and Taqqu:

$$\begin{aligned}
\frac{\sum_{m=1}^{n-k} m^{\alpha_j+\alpha_l}L(m)L'(m+k)}{\sqrt{\sum_{i=1}^{n} i^{2\alpha_j}L(i)^2}\sqrt{\sum_{q=1}^{n} q^{2\alpha_l}L'(q)^2}} &\sim \frac{\frac{(n-k)^{\alpha_j+\alpha_l+1}}{\alpha_j+\alpha_l+1}L(n-k)L'(n)}{\sqrt{\frac{n^{2\alpha_j+1}}{2\alpha_j+1}L(n)^2}\sqrt{\frac{n^{2\alpha_l+1}}{2\alpha_l+1}L'(n)^2}} \\
&\sim \frac{\sqrt{2\alpha_j+1}\sqrt{2\alpha_l+1}}{\alpha_j+\alpha_l+1}\frac{(n-k)^{\alpha_j+\alpha_l+1}}{n^{\alpha_j+1/2}n^{\alpha_l+1/2}}\frac{L(n-k)}{L(n)},
\end{aligned}$$

and this quantity converges to $\frac{\sqrt{2\alpha_j+1}\sqrt{2\alpha_l+1}}{\alpha_j+\alpha_l+1}$.

Thereby the coefficients $\rho_{j,l}(k)$ are constants and equal to $\frac{\sqrt{2\alpha_j+1}\sqrt{2\alpha_l+1}}{\alpha_j+\alpha_l+1}$.

□

1.6.2 Theorem 1.2

Proof. The proof of Theorem 1.2 is splitted in two parts. Indeed, notice that:

$$\|f_n^*(\lambda) - f(\lambda)\|_{\mathbb{L}^1} \leq \|f_n^*(\lambda) - f_n(\lambda)\|_{\mathbb{L}^1} + \|f_n(\lambda) - f(\lambda)\|_{\mathbb{L}^1}$$

The proof is complete with Propositions 1.2 and 1.3:

Proposition 1.2. *Under the assumptions of Theorem 1.2, we have:*

$$\lim_{n\to\infty} \sup_{\lambda\in[-\pi,\pi]} \|f_n(\lambda) - f(\lambda)\|_{\mathbb{L}^1} = 0 \tag{1.27}$$

Proposition 1.3. *Under the assumptions of Theorem 1.2, we have:*

$$\lim_{n\to\infty} \sup_{\lambda\in[-\pi,\pi]} \|f_n^*(\lambda) - f_n(\lambda)\|_{\mathbb{L}^1} = 0 \tag{1.28}$$

□

Proposition 1.2

Proof. Without loss of generality, h_n is chosen such that $2h_n \leq n-1$. Let m be an integer such that: $1 \leq 2m \leq 2h_n \leq n-1$. For all $i \in \mathbb{Z}$, define:

$$\tilde{\epsilon}_{i,m} = \mathbb{E}(\epsilon_i|\mathcal{F}_{i+m}) - \mathbb{E}(\epsilon_i|\mathcal{F}_{i-m}). \tag{1.29}$$

and notice that $\mathbb{E}(\tilde{\epsilon}_{i,m}) = 0$. The associated spectral density estimate is defined as follows:

$$\tilde{f}_n^m(\lambda) = \frac{1}{2\pi} \sum_{|k|\leq n-1} K\left(\frac{|k|}{h_n}\right) \hat{\tilde{\gamma}}_{k,m} e^{ik\lambda}, \quad \lambda \in [-\pi,\pi],$$

where :

$$\hat{\tilde{\gamma}}_{k,m} = \frac{1}{n} \sum_{j=1}^{n-|k|} \tilde{\epsilon}_{j,m}\tilde{\epsilon}_{j+|k|,m}, \quad |k| \leq n-1.$$

By the triangle inequality, it follows that:

$$\begin{aligned} \|f_n(\lambda) - f(\lambda)\|_{\mathbb{L}^1} &\leq \left\|f_n(\lambda) - \tilde{f}_n^m(\lambda)\right\|_{\mathbb{L}^1} + \left\|\tilde{f}_n^m(\lambda) - \mathbb{E}(\tilde{f}_n^m(\lambda))\right\|_{\mathbb{L}^1} \\ &\quad + \left|\mathbb{E}(\tilde{f}_n^m(\lambda)) - \mathbb{E}(f_n(\lambda))\right| + \|\mathbb{E}(f_n(\lambda)) - f(\lambda)\|_{\mathbb{L}^1} \\ &\leq 2\left\|\tilde{f}_n^m(\lambda) - f_n(\lambda)\right\|_{\mathbb{L}^1} + \left\|\tilde{f}_n^m(\lambda) - \mathbb{E}(\tilde{f}_n^m(\lambda))\right\|_{\mathbb{L}^1} + \|\mathbb{E}(f_n(\lambda)) - f(\lambda)\|_{\mathbb{L}^1} \end{aligned}$$

because $\left|\mathbb{E}(\tilde{f}_n^m(\lambda)) - \mathbb{E}(f_n(\lambda))\right| \leq \left\|\tilde{f}_n^m(\lambda) - f_n(\lambda)\right\|_{\mathbb{L}^1}$.

The proof is complete using Lemmas 1.1, 1.2 and 1.3:

Lemma 1.1. *Under the assumptions of Theorem 1.2, we have:*

$$\lim_{n\to\infty} \sup_{\lambda\in[-\pi,\pi]} \|\mathbb{E}(f_n(\lambda)) - f(\lambda)\|_{\mathbb{L}^1} = 0 \tag{1.30}$$

Lemma 1.2. *Under the assumptions of Theorem 1.2, we have:*

$$\lim_{m\to\infty} \limsup_{n\to\infty} \sup_{\lambda\in[-\pi,\pi]} \left\|\tilde{f}_n^m(\lambda) - f_n(\lambda)\right\|_{\mathbb{L}^1} = 0 \tag{1.31}$$

Lemma 1.3. *Under the assumptions of Theorem 1.2, we have:*

$$\lim_{m\to\infty} \limsup_{n\to\infty} \sup_{\lambda\in[-\pi,\pi]} \left\| \tilde{f}_n^m(\lambda) - \mathbb{E}(\tilde{f}_n^m(\lambda)) \right\|_{\mathbb{L}^1} = 0 \tag{1.32}$$

□

Proof of Lemma 1.1. By the properties of expectation and by stationarity:

$$\mathbb{E}\left(f_n(\lambda)\right) = \frac{1}{2\pi} \sum_{|k|\leq n-1} K\left(\frac{|k|}{h_n}\right) \mathbb{E}(\hat{\gamma}_k) e^{ik\lambda} = \frac{1}{2\pi} \sum_{|k|\leq n-1} \left(\frac{n-|k|}{n}\right) K\left(\frac{|k|}{h_n}\right) \gamma_k e^{ik\lambda}.$$

Since $h_n \xrightarrow[n\to\infty]{} \infty$ and $\lim_{u\to 0} K(u) = 1$, thanks to dominated convergence theorem and because $\sum_k |\gamma(k)| < +\infty$, it is clear that (1.30) is true.

□

Proof of Lemma 1.2. Let S_n and $\tilde{S}_n^m$ be defined as:

$$S_n(\lambda) = \sum_{k=1}^{n} \epsilon_k e^{ik\lambda}$$

$$\tilde{S}_n^m(\lambda) = \sum_{k=1}^{n} \tilde{\epsilon}_{k,m} e^{ik\lambda}.$$

Since $(a+b)^2 \leq 2a^2 + 2b^2$, we have:

$$\begin{aligned}
\frac{1}{n}\left\|S_n(\lambda) - \tilde{S}_n^m(\lambda)\right\|_{\mathbb{L}^2}^2 &= \frac{1}{n}\left\| \sum_{k=1}^{n} \epsilon_k e^{ik\lambda} - \sum_{k=1}^{n} \tilde{\epsilon}_{k,m} e^{ik\lambda} \right\|_{\mathbb{L}^2}^2 \\
&= \frac{1}{n}\left\| \sum_{k=1}^{n} \epsilon_k e^{ik\lambda} - \left(\sum_{k=1}^{n} \mathbb{E}(\epsilon_k|\mathcal{F}_{k+m}) e^{ik\lambda} - \mathbb{E}(\epsilon_k|\mathcal{F}_{k-m}) e^{ik\lambda} \right) \right\|_{\mathbb{L}^2}^2 \\
&= \frac{1}{n}\left\| \sum_{k=1}^{n} (\epsilon_k - \mathbb{E}(\epsilon_k|\mathcal{F}_{k+m})) e^{ik\lambda} + \sum_{k=1}^{n} \mathbb{E}(\epsilon_k|\mathcal{F}_{k-m}) e^{ik\lambda} \right\|_{\mathbb{L}^2}^2 \\
&\leq \frac{2}{n}\left\| \sum_{k=1}^{n} (\epsilon_k - \mathbb{E}(\epsilon_k|\mathcal{F}_{k+m})) e^{ik\lambda} \right\|_{\mathbb{L}^2}^2 + \frac{2}{n}\left\| \sum_{k=1}^{n} \mathbb{E}(\epsilon_k|\mathcal{F}_{k-m}) e^{ik\lambda} \right\|_{\mathbb{L}^2}^2 .
\end{aligned} \tag{1.33}$$

We get for the first term of the right-hand side in (1.33):

$$\begin{aligned}
\frac{1}{n}\left\|\sum_{k=1}^{n}(\epsilon_k-\mathbb{E}(\epsilon_k|\mathcal{F}_{k+m}))e^{ik\lambda}\right\|_{\mathbb{L}^2}^2 &= \frac{1}{n}\left\|\sum_{k=1}^{n}\sum_{j=k+m+1}^{\infty}P_j(\epsilon_k)e^{ik\lambda}\right\|_{\mathbb{L}^2}^2 \\
&= \frac{1}{n}\left\|\sum_{j=m+2}^{\infty}\sum_{k=1}^{n}P_j(\epsilon_k)e^{ik\lambda}\mathbf{1}_{\{j\geq k+m+1\}}\right\|_{\mathbb{L}^2}^2 \\
&= \frac{1}{n}\sum_{j=m+2}^{\infty}\left\|\sum_{k=1}^{n}P_j(\epsilon_k)e^{ik\lambda}\mathbf{1}_{\{k-j\leq -(m+1)\}}\right\|_{\mathbb{L}^2}^2 \\
&\leq \frac{1}{n}\sum_{j=m+2}^{\infty}\left(\sum_{k=1}^{n}\|P_j(\epsilon_k)\|_{\mathbb{L}^2}\mathbf{1}_{\{k-j\leq -(m+1)\}}\right)^2,
\end{aligned}$$

using Pythagoras' theorem and the triangle inequality. It follows:

$$\begin{aligned}
\frac{1}{n}\sum_{j=m+2}^{\infty}\left(\sum_{k=1}^{n}\|P_j(\epsilon_k)\|_{\mathbb{L}^2}\mathbf{1}_{\{k-j\leq -(m+1)\}}\right)^2 &\leq \frac{1}{n}\sum_{j=m+2}^{\infty}\left(\sum_{k=1}^{n}\|P_0(\epsilon_{k-j})\|_{\mathbb{L}^2}\mathbf{1}_{\{k-j\leq -(m+1)\}}\right)^2 \\
&\leq \frac{1}{n}\sum_{j=m+2}^{\infty}\left(\sum_{r=-\infty}^{-(m+1)}\|P_0(\epsilon_r)\|_{\mathbb{L}^2}\mathbf{1}_{\{1-j\leq r\leq n-j\}}\right)^2 \\
&\leq \frac{1}{n}\sum_{j=m+2}^{\infty}\left(\mathbf{1}_{\{1-r\leq j\leq n-r\}}\sum_{r=-\infty}^{-(m+1)}\|P_0(\epsilon_r)\|_{\mathbb{L}^2}\right)^2 \\
&\leq \left(\sum_{r=-\infty}^{-(m+1)}\|P_0(\epsilon_r)\|_{\mathbb{L}^2}\right)^2. \qquad (1.34)
\end{aligned}$$

With the same arguments, the second term of the right-hand side in (1.33) satisfies the inequality:

$$\frac{1}{n}\left\|\sum_{k=1}^{n}\mathbb{E}(\epsilon_k|\mathcal{F}_{k-m})e^{ik\lambda}\right\|_{\mathbb{L}^2}^2 \leq \left(\sum_{r=m}^{\infty}\|P_0(\epsilon_r)\|_{\mathbb{L}^2}\right)^2. \qquad (1.35)$$

Consequently, combining (1.34) and (1.35), we obtain that:

$$\sup_{\lambda\in[-\pi,\pi]}\frac{1}{n}\left\|S_n(\lambda)-\tilde{S}_n^m(\lambda)\right\|_{\mathbb{L}^2}^2 \leq 2\left(\sum_{r=-\infty}^{-(m+1)}\|P_0(\epsilon_r)\|_{\mathbb{L}^2}\right)^2+2\left(\sum_{r=m}^{\infty}\|P_0(\epsilon_r)\|_{\mathbb{L}^2}\right)^2.$$

Then, since $\sum_{i=-\infty}^{\infty}\|P_0(\epsilon_i)\|_{\mathbb{L}^2}<+\infty$, we have this first result:

$$\lim_{m\to\infty}\limsup_{n\to\infty}\sup_{\lambda\in[-\pi,\pi]}\frac{1}{n}\left\|S_n(\lambda)-\tilde{S}_n^m(\lambda)\right\|_{\mathbb{L}^2}^2=0. \qquad (1.36)$$

Define now the two periodograms corresponding to the quantities S_n and $\tilde{S}_n^m$:

$$I_n(\lambda) = \frac{1}{2\pi n} |S_n(\lambda)|^2 = \frac{1}{2\pi} \sum_{k=1-n}^{n-1} \hat{\gamma}_k e^{ik\lambda} \tag{1.37}$$

$$\tilde{I}_n^m(\lambda) = \frac{1}{2\pi n} \left|\tilde{S}_n^m(\lambda)\right|^2 = \frac{1}{2\pi} \sum_{k=1-n}^{n-1} \hat{\tilde{\gamma}}_{k,m} e^{ik\lambda}. \tag{1.38}$$

By Cauchy-Schwarz's inequality and the triangle inequality:

$$\begin{aligned}
\left\|I_n(\lambda) - \tilde{I}_n^m(\lambda)\right\|_{\mathbb{L}^1} &= \left\| \frac{1}{2\pi n} |S_n(\lambda)|^2 - \frac{1}{2\pi n} \left|\tilde{S}_n^m(\lambda)\right|^2 \right\|_{\mathbb{L}^1} \\
&= \frac{1}{2\pi n} \left\| |S_n(\lambda)|^2 - \left|\tilde{S}_n^m(\lambda)\right|^2 \right\|_{\mathbb{L}^1} \\
&= \frac{1}{2\pi n} \left\| \left(|S_n(\lambda)| - \left|\tilde{S}_n^m(\lambda)\right|\right) \left(|S_n(\lambda)| + \left|\tilde{S}_n^m(\lambda)\right|\right) \right\|_{\mathbb{L}^1} \\
&\leq \frac{1}{2\pi n} \left\| |S_n(\lambda)| - \left|\tilde{S}_n^m(\lambda)\right| \right\|_{\mathbb{L}^2} \left\| |S_n(\lambda)| + \left|\tilde{S}_n^m(\lambda)\right| \right\|_{\mathbb{L}^2} \\
&\leq \frac{1}{2\pi} \frac{1}{\sqrt{n}} \left\|S_n(\lambda) - \tilde{S}_n^m(\lambda)\right\|_{\mathbb{L}^2} \left(\frac{\|S_n(\lambda)\|_{\mathbb{L}^2}}{\sqrt{n}} + \frac{\left\|\tilde{S}_n^m(\lambda)\right\|_{\mathbb{L}^2}}{\sqrt{n}} \right).
\end{aligned}$$

Thus, thanks to (1.36) and the following inequality for S_n and $\tilde{S}_n^m$:

$$\frac{1}{\sqrt{n}} \|S_n(\lambda)\|_{\mathbb{L}^2} \leq \sum_{k\in\mathbb{Z}} \|P_0(\epsilon_k)\|_{\mathbb{L}^2} < \infty, \tag{1.39}$$

we get:

$$\lim_{m\to\infty} \limsup_{n\to\infty} \sup_{\lambda\in[-\pi,\pi]} \left\|I_n(\lambda) - \tilde{I}_n^m(\lambda)\right\|_{\mathbb{L}^1} = 0. \tag{1.40}$$

Then, let $\hat{K}(.)$ be the Fourier transform of K:

$$\begin{aligned}
f_n(\lambda) - \tilde{f}_n^m(\lambda) &= \frac{1}{2\pi} \sum_{|k|\leq n-1} K\left(\frac{|k|}{h_n}\right) e^{ik\lambda} \left(\hat{\gamma}_k - \hat{\tilde{\gamma}}_{k,m}\right) \\
&= \frac{1}{2\pi} \sum_{|k|\leq n-1} \frac{1}{2\pi} \left(\int_{\mathbb{R}} \hat{K}(u) e^{iu\frac{k}{h_n}} du \right) e^{ik\lambda} \left(\hat{\gamma}_k - \hat{\tilde{\gamma}}_{k,m}\right) \\
&= \frac{1}{2\pi} \int_{\mathbb{R}} \hat{K}(u) \frac{1}{2\pi} \sum_{|k|\leq n-1} \left(\hat{\gamma}_k - \hat{\tilde{\gamma}}_k\right) e^{ik\left(\frac{u}{h_n}+\lambda\right)} du \\
&= \frac{1}{2\pi} \int_{\mathbb{R}} \hat{K}(u) \left(I_n\left(\frac{u}{h_n} + \lambda\right) - \tilde{I}_n^m\left(\frac{u}{h_n} + \lambda\right) \right) du,
\end{aligned}$$

using the definition of I_n and $\tilde{I}_n^m$ (see (1.37) and (1.38)). Hence, by the triangle inequality:

$$\begin{aligned}\left\|f_n(\lambda)-\tilde{f}_n^m(\lambda)\right\|_{\mathbb{L}^1} &= \left\|\frac{1}{2\pi}\int_{\mathbb{R}}\hat{K}(u)\left(I_n\left(\frac{u}{h_n}+\lambda\right)-\tilde{I}_n^m\left(\frac{u}{h_n}+\lambda\right)\right)du\right\|_{\mathbb{L}^1}\\ &\leq \frac{1}{2\pi}\int_{\mathbb{R}}\left|\hat{K}(u)\right|\left\|\left(I_n\left(\frac{u}{h_n}+\lambda\right)-\tilde{I}_n^m\left(\frac{u}{h_n}+\lambda\right)\right)\right\|_{\mathbb{L}^1}du\\ &\leq \frac{1}{2\pi}\sup_{\theta}\left\|I_n(\theta)-\tilde{I}_n^m(\theta)\right\|_{\mathbb{L}^1}\int_{\mathbb{R}}\left|\hat{K}(u)\right|du.\end{aligned}$$

Using (1.40) and the fact that $\hat{K}$ is integrable, Lemma 1.2 is proved.

□

Proof of Lemma 1.3. Without loss of generality, suppose $\theta=0$. We have:

$$\begin{aligned}\tilde{f}_n^m(0) &= \frac{1}{2\pi}\sum_{|k|\leq n-1}K\left(\frac{|k|}{h_n}\right)\hat{\tilde{\gamma}}_{k,m}\\ &= \frac{2}{2\pi}\sum_{k=1}^{n-1}K\left(\frac{k}{h_n}\right)\hat{\tilde{\gamma}}_{k,m}+\frac{1}{2\pi}\hat{\tilde{\gamma}}_{0,m}\\ &= \frac{2}{2\pi}\sum_{k=1}^{n-1}K\left(\frac{k}{h_n}\right)\frac{1}{n}\sum_{j=1}^{n-k}\tilde{\epsilon}_{j,m}\tilde{\epsilon}_{j+k,m}+\frac{1}{2\pi n}\sum_{j=1}^{n}\tilde{\epsilon}_{j,m}^2.\end{aligned}$$

By the triangle inequality again and a change of variables, we have:

$$\begin{aligned}&\left\|\tilde{f}_n^m(0)-\mathbb{E}\left(\tilde{f}_n^m(0)\right)\right\|_{\mathbb{L}^1}\\ =\ &\left\|\frac{2}{2\pi}\sum_{k=1}^{n-1}K\left(\frac{k}{h_n}\right)\frac{1}{n}\sum_{j=1}^{n-k}\tilde{\epsilon}_{j,m}\tilde{\epsilon}_{j+k,m}-\mathbb{E}(\tilde{\epsilon}_{j,m}\tilde{\epsilon}_{j+k,m})+\frac{1}{2\pi n}\sum_{j=1}^{n}\tilde{\epsilon}_{j,m}^2-\mathbb{E}(\tilde{\epsilon}_{j,m}^2)\right\|_{\mathbb{L}^1}\\ \leq\ &\frac{2}{2\pi}\left\|\sum_{k=1}^{n-1}K\left(\frac{k}{h_n}\right)\frac{1}{n}\sum_{j=1}^{n-k}\left(\tilde{\epsilon}_{j,m}\tilde{\epsilon}_{j+k,m}-\mathbb{E}(\tilde{\epsilon}_{j,m}\tilde{\epsilon}_{j+k,m})\right)\right\|_{\mathbb{L}^1}\\ &+\frac{1}{2\pi}\left\|\frac{1}{n}\sum_{i=1}^{n}\tilde{\epsilon}_{i,m}^2-\mathbb{E}(\tilde{\epsilon}_{0,m}^2)\right\|_{\mathbb{L}^1}\\ \leq\ &\frac{2}{2\pi}\left\|\frac{1}{n}\sum_{i=2}^{n}\sum_{j=(i-2h_n)\vee 1}^{i-1}K\left(\frac{i-j}{h_n}\right)\left(\tilde{\epsilon}_{i,m}\tilde{\epsilon}_{j,m}-\mathbb{E}(\tilde{\epsilon}_{i,m}\tilde{\epsilon}_{j,m})\right)\right\|_{\mathbb{L}^1}\\ &+\frac{1}{2\pi}\left\|\frac{1}{n}\sum_{i=1}^{n}\tilde{\epsilon}_{i,m}^2-\mathbb{E}(\tilde{\epsilon}_{0,m}^2)\right\|_{\mathbb{L}^1}.\end{aligned}$$

By the $\mathbb{L}^1$-ergodic theorem, it is known that, for m fixed:

$$\lim_{n\to\infty}\left\|\frac{1}{n}\sum_{i=1}^{n}\tilde{\epsilon}_{i,m}^2-\mathbb{E}(\tilde{\epsilon}_{0,m}^2)\right\|_{\mathbb{L}^1}=0.$$

Consequently, it remains to prove:

$$\lim_{m\to\infty} \limsup_{n\to\infty} \left\| \frac{1}{n} \sum_{i=2}^{n} \sum_{j=(i-2h_n)\vee 1}^{i-1} K\left(\frac{i-j}{h_n}\right) \left(\tilde{\epsilon}_{i,m}\tilde{\epsilon}_{j,m} - \mathbb{E}(\tilde{\epsilon}_{i,m}\tilde{\epsilon}_{j,m})\right) \right\|_{\mathbb{L}^1} = 0.$$

We know that:

$$\frac{1}{n} \sum_{i=2m+1}^{n} \sum_{j=(i-2h_n)\vee 1}^{i-2m} K\left(\frac{i-j}{h_n}\right) \mathbb{E}(\tilde{\epsilon}_{i,m}\tilde{\epsilon}_{j,m}) = 0, \tag{1.41}$$

Indeed,

$$\mathbb{E}(\tilde{\epsilon}_{i,m}\tilde{\epsilon}_{j,m}) = \mathbb{E}\left(\left(\mathbb{E}(\epsilon_j|\mathcal{F}_{j+m}) - \mathbb{E}(\epsilon_j|\mathcal{F}_{j-m})\right) \left(\mathbb{E}(\epsilon_i|\mathcal{F}_{i+m}) - \mathbb{E}(\epsilon_i|\mathcal{F}_{i-m})\right)\right).$$

But $\mathbb{E}(\epsilon_i|\mathcal{F}_{i+m}) - \mathbb{E}(\epsilon_i|\mathcal{F}_{i-m})$ is orthogonal to $\mathbb{L}^2(\mathcal{F}_{i-m})$, and $\mathbb{E}(\epsilon_j|\mathcal{F}_{j+m}) - \mathbb{E}(\epsilon_j|\mathcal{F}_{j-m})$ belongs to $\mathbb{L}^2(\mathcal{F}_{i-m})$ if $j+m \le i-m$. Thus $\mathbb{E}(\tilde{\epsilon}_{i,m}\tilde{\epsilon}_{j,m})$ is equal to zero if $j \le i-2m$ and (1.41) is true.

Thereby we have:

$$\begin{aligned}
&\left\| \frac{1}{n} \sum_{i=2}^{n} \sum_{j=(i-2h_n)\vee 1}^{i-1} K\left(\frac{i-j}{h_n}\right) \left(\tilde{\epsilon}_{i,m}\tilde{\epsilon}_{j,m} - \mathbb{E}(\tilde{\epsilon}_{i,m}\tilde{\epsilon}_{j,m})\right) \right\|_{\mathbb{L}^1} \\
\le\ &\left\| \frac{1}{n} \sum_{i=2}^{n} \sum_{j=(i-2m+1)\vee 1}^{i-1} K\left(\frac{i-j}{h_n}\right) \left(\tilde{\epsilon}_{i,m}\tilde{\epsilon}_{j,m} - \mathbb{E}(\tilde{\epsilon}_{i,m}\tilde{\epsilon}_{j,m})\right) \right\|_{\mathbb{L}^1} \\
&+ \left\| \frac{1}{n} \sum_{i=2m+1}^{n} \sum_{j=(i-2h_n)\vee 1}^{i-2m} K\left(\frac{i-j}{h_n}\right) \tilde{\epsilon}_{i,m}\tilde{\epsilon}_{j,m} \right\|_{\mathbb{L}^1} \\
\le\ &\left\| \frac{1}{n} \sum_{k=1}^{2m-1} \sum_{i=1}^{n-k} K\left(\frac{k}{h_n}\right) \left(\tilde{\epsilon}_{i,m}\tilde{\epsilon}_{i+k,m} - \mathbb{E}(\tilde{\epsilon}_{i,m}\tilde{\epsilon}_{i+k,m})\right) \right\|_{\mathbb{L}^1} \\
&+ \left\| \frac{1}{n} \sum_{i=2m+1}^{n} \sum_{j=(i-2h_n)\vee 1}^{i-2m} K\left(\frac{i-j}{h_n}\right) \tilde{\epsilon}_{i,m}\tilde{\epsilon}_{j,m} \right\|_{\mathbb{L}^1}.
\end{aligned} \tag{1.42}$$

For the first term on the right-hand side of (1.42), since the kernel K is bounded by 1, we have by the triangle inequality and the stationarity of the error process:

$$\left\| \frac{1}{n} \sum_{k=1}^{2m-1} \sum_{i=1}^{n-k} K\left(\frac{k}{h_n}\right) \left(\tilde{\epsilon}_{i,m}\tilde{\epsilon}_{i+k,m} - \mathbb{E}(\tilde{\epsilon}_{i,m}\tilde{\epsilon}_{i+k,m})\right) \right\|_{\mathbb{L}^1} \le \sum_{k=1}^{2m-1} \left\| \frac{1}{n} \sum_{i=1}^{n-k} (\tilde{\epsilon}_{i,m}\tilde{\epsilon}_{i+k,m} - \mathbb{E}(\tilde{\epsilon}_{0,m}\tilde{\epsilon}_{k,m})) \right\|_{\mathbb{L}^1}.$$

Using the $\mathbb{L}^1$-ergodic theorem, for all k fixed, we deduce that:

$$\sum_{k=1}^{2m-1} \left\| \frac{1}{n} \sum_{i=1}^{n-k} (\tilde{\epsilon}_{i,m}\tilde{\epsilon}_{i+k,m} - \mathbb{E}(\tilde{\epsilon}_{0,m}\tilde{\epsilon}_{k,m})) \right\|_{\mathbb{L}^1} \xrightarrow[n\to\infty]{} 0.$$

It remains to be shown that:

$$\lim_{m\to\infty}\limsup_{n\to\infty}\left\|\frac{1}{n}\sum_{i=2m+1}^{n}\sum_{j=(i-2h_n)\vee 1}^{i-2m}K\left(\frac{i-j}{h_n}\right)\tilde{\epsilon}_{i,m}\tilde{\epsilon}_{j,m}\right\|_{\mathbb{L}^1}=0.$$

We have:

$$\begin{aligned}&\left\|\frac{1}{n}\sum_{i=2m+1}^{n}\sum_{j=(i-2h_n)\vee 1}^{i-2m}K\left(\frac{i-j}{h_n}\right)\tilde{\epsilon}_{i,m}\tilde{\epsilon}_{j,m}\right\|_{\mathbb{L}^1}\\&=\left\|\frac{1}{n}\sum_{i=2m+1}^{2[n/2m]m}\sum_{j=(i-2h_n)\vee 1}^{i-2m}K\left(\frac{i-j}{h_n}\right)\tilde{\epsilon}_{i,m}\tilde{\epsilon}_{j,m}+\frac{1}{n}\sum_{i=2[n/2m]m+1}^{n}\sum_{j=(i-2h_n)\vee 1}^{i-2m}K\left(\frac{i-j}{h_n}\right)\tilde{\epsilon}_{i,m}\tilde{\epsilon}_{j,m}\right\|_{\mathbb{L}^1},\end{aligned}$$

then by triangle inequality:

$$\begin{aligned}&\left\|\frac{1}{n}\sum_{i=2m+1}^{2[n/2m]m}\sum_{j=(i-2h_n)\vee 1}^{i-2m}K\left(\frac{i-j}{h_n}\right)\tilde{\epsilon}_{i,m}\tilde{\epsilon}_{j,m}+\frac{1}{n}\sum_{i=2[n/2m]m+1}^{n}\sum_{j=(i-2h_n)\vee 1}^{i-2m}K\left(\frac{i-j}{h_n}\right)\tilde{\epsilon}_{i,m}\tilde{\epsilon}_{j,m}\right\|_{\mathbb{L}^1}\\&\qquad\leq\left\|\frac{1}{n}\sum_{i=2m+1}^{2[n/2m]m}\sum_{j=(i-2h_n)\vee 1}^{i-2m}K\left(\frac{i-j}{h_n}\right)\tilde{\epsilon}_{i,m}\tilde{\epsilon}_{j,m}\right\|_{\mathbb{L}^1}\\&\qquad\qquad+\left\|\frac{1}{n}\sum_{i=2[n/2m]m+1}^{n}\sum_{j=(i-2h_n)\vee 1}^{i-2m}K\left(\frac{i-j}{h_n}\right)\tilde{\epsilon}_{i,m}\tilde{\epsilon}_{j,m}\right\|_{\mathbb{L}^1},\end{aligned}$$

and using a change of variable:

$$\begin{aligned}&\left\|\frac{1}{n}\sum_{i=2m+1}^{2[n/2m]m}\sum_{j=(i-2h_n)\vee 1}^{i-2m}K\left(\frac{i-j}{h_n}\right)\tilde{\epsilon}_{i,m}\tilde{\epsilon}_{j,m}\right\|_{\mathbb{L}^1}\\&\qquad+\left\|\frac{1}{n}\sum_{i=2[n/2m]m+1}^{n}\sum_{j=(i-2h_n)\vee 1}^{i-2m}K\left(\frac{i-j}{h_n}\right)\tilde{\epsilon}_{i,m}\tilde{\epsilon}_{j,m}\right\|_{\mathbb{L}^1}\\&\qquad\leq\sum_{l=1}^{2m}\left\|\frac{1}{n}\sum_{r=1}^{[n/2m]-1}\tilde{\epsilon}_{2rm+l,m}\sum_{j=(2rm+l-2h_n)\vee 1}^{2(r-1)m+l}K\left(\frac{2rm+l-j}{h_n}\right)\tilde{\epsilon}_{j,m}\right\|_{\mathbb{L}^1}\\&\qquad\qquad+\left\|\frac{1}{n}\sum_{i=2[n/2m]m+1}^{n}\tilde{\epsilon}_{i,m}\sum_{j=(i-2h_n)\vee 1}^{i-2m}K\left(\frac{i-j}{h_n}\right)\tilde{\epsilon}_{j,m}\right\|_{\mathbb{L}^1}.\end{aligned}\tag{1.43}$$

For the second term of the right-hand side of (1.43), by the Cauchy-Schwarz inequality and by sta-

tionarity, we get:

$$\begin{aligned}\left\|\frac{1}{n}\sum_{i=2[n/2m]m+1}^{n}\tilde{\epsilon}_{i,m}\sum_{j=(i-2h_n)\vee 1}^{i-2m}K\left(\frac{i-j}{h_n}\right)\tilde{\epsilon}_{j,m}\right\|_{\mathbb{L}^1} &\leq \frac{1}{n}\sum_{i=2[n/2m]m+1}^{n}\sum_{j=(i-2h_n)\vee 1}^{i-2m}\left\|\tilde{\epsilon}_{i,m}\tilde{\epsilon}_{j,m}\right\|_{\mathbb{L}^1}\\ &\leq \frac{1}{n}\sum_{i=2[n/2m]m+1}^{n}\sum_{j=(i-2h_n)\vee 1}^{i-2m}\left\|\tilde{\epsilon}_{0,m}\right\|_{\mathbb{L}^2}^2\\ &\leq \frac{4}{n}\sum_{i=2[n/2m]m+1}^{n}\sum_{j=(i-2h_n)\vee 1}^{i-2m}\left\|\epsilon_0\right\|_{\mathbb{L}^2}^2\\ &\leq 16m\frac{h_n}{n}\left\|\epsilon_0\right\|_{\mathbb{L}^2}^2, \qquad (1.44)\end{aligned}$$

and (1.44) tends to 0 as $n\to\infty$.

Using ideas developed by Dedecker [34] (see the proof of his Theorem 1), we study the first term of the right-hand side of (1.43) and we shall prove that it is negligible. Let Z be:

$$Z(r,n,m)=\frac{1}{n}\sum_{r=1}^{[n/2m]-1}\tilde{\epsilon}_{2rm+l,m}\sum_{j=(2rm+l-2h_n)\vee 1}^{2(r-1)m+l}K\left(\frac{2rm+l-j}{h_n}\right)\tilde{\epsilon}_{j,m}. \qquad (1.45)$$

Let φ be the function defined by $\varphi'(0)=\varphi(0)=0$ and $\varphi''(t)=(1-|t|)\mathbf{1}_{\{|t|<1\}}$, that is the symmetric function such that, for all t greater or equal to 0, $\varphi(t)=\frac{1}{6}(1-t)^3\mathbf{1}_{\{t<1\}}+\frac{1}{2}t-\frac{1}{6}$.

Now, for all $\epsilon>0$, by the growth of φ, there exists a constant C such that:

$$\begin{aligned}\mathbb{E}(|Z(r,n,m)|) &= \mathbb{E}(|(Z(r,n,m)|\,\mathbf{1}_{\{|Z(r,n,m)|>\epsilon\}})+\mathbb{E}(|Z(r,n,m)|\,\mathbf{1}_{\{|Z(r,n,m)|<\epsilon\}})\\ &\leq C\mathbb{E}(\varphi(Z(r,n,m))\mathbf{1}_{\{|Z(r,n,m)|>\epsilon\}})+\mathbb{E}(|Z(r,n,m)|\,\mathbf{1}_{\{|Z(r,n,m)|<\epsilon\}})\\ &\leq C\mathbb{E}(\varphi(Z(r,n,m)))+\epsilon.\end{aligned}$$

because the function φ is positive.

We conclude the proof using Lemma 1.4.

□

Lemma 1.4. *In the conditions developed at the end of the previous proof, for all fixed m:*

$$\lim_{n\to\infty}\mathbb{E}(\varphi(Z(r,n,m)))=0. \qquad (1.46)$$

Proof of Lemma 1.4. To prove that (1.46) holds, the two following results are needed:

Lemma 1.5. *The following inequality holds:*

$$|\varphi(x+h)-\varphi(x)-h\varphi'(x)|\leq\psi(h),$$

where:

$$\psi(h) = |h|^2 \mathbf{1}_{\{|h|\leq 1\}} + (2|h| - 1)\mathbf{1}_{\{|h|>1\}}.$$

Proof. The function φ is continuous and differentiable in the neighborhood of 0. Using the Taylor formula, we have the following bounds:

$$|\varphi(x+h) - \varphi(x) - h\varphi'(x)| \leq \frac{|h|^2}{2} \sup_{u\in\mathbb{R}} |\varphi''(u)| \leq \frac{|h|^2}{2}.$$

Then, by the triangle inequality:

$$|\varphi(x+h) - \varphi(x) - h\varphi'(x)| \leq |\varphi(x+h) - \varphi(x)| + |h|\,|\varphi'(x)| \leq 2\,|h| \sup_{u\in\mathbb{R}} |\varphi'(u)| \leq |h|\,.$$

The proof is complete.

□

Lemma 1.6. *For all real x in $\mathbb{R}$, we have:*

$$|x|(1 \wedge |x|) \leq \psi(x) \leq 2|x|(1 \wedge |x|).$$

The proof of Lemma 1.6, being elementary, is left to the reader.

So we get:

$$\begin{aligned}
\mathbb{E}(\varphi(Z(r,n,m))) = \sum_{i=1}^{[n/2m]-1} \mathbb{E}\Bigg(\Bigg[\varphi\Bigg(\frac{1}{n}\sum_{q=1}^{i} \tilde{\epsilon}_{2qm+l,m} \sum_{j=(2qm+l-2h_n)\vee 1}^{2(q-1)m+l} K\left(\frac{2qm+l-j}{h_n}\right)\tilde{\epsilon}_{j,m}\Bigg) \\
- \varphi\Bigg(\frac{1}{n}\sum_{q=1}^{i-1} \tilde{\epsilon}_{2qm+l,m} \sum_{j=(2qm+l-2h_n)\vee 1}^{2(q-1)m+l} K\left(\frac{2qm+l-j}{h_n}\right)\tilde{\epsilon}_{j,m}\Bigg)\Bigg]\Bigg) \\
\leq \sum_{i=1}^{[n/2m]-1} \Bigg|\mathbb{E}\Bigg(\varphi\Bigg(\frac{1}{n}\sum_{q=1}^{i} \tilde{\epsilon}_{2qm+l,m} \sum_{j=(2qm+l-2h_n)\vee 1}^{2(q-1)m+l} K\left(\frac{2qm+l-j}{h_n}\right)\tilde{\epsilon}_{j,m}\Bigg) \\
- \varphi\Bigg(\frac{1}{n}\sum_{q=1}^{i-1} \tilde{\epsilon}_{2qm+l,m} \sum_{j=(2qm+l-2h_n)\vee 1}^{2(q-1)m+l} K\left(\frac{2qm+l-j}{h_n}\right)\tilde{\epsilon}_{j,m}\Bigg)\Bigg)\Bigg|.
\end{aligned}$$

Then applying Taylor's expansion, with :

$$\begin{aligned} x &= \frac{1}{n}\sum_{q=1}^{i-1}\tilde{\epsilon}_{2qm+l,m}\sum_{j=(2qm+l-2h_n)\vee 1}^{2(q-1)m+l} K\left(\frac{2qm+l-j}{h_n}\right)\tilde{\epsilon}_{j,m} \\ A(i,m) &= \frac{1}{n}\tilde{\epsilon}_{2im+l,m}\sum_{j=(2im+l-2h_n)\vee 1}^{2(i-1)m+l} K\left(\frac{2im+l-j}{h_n}\right)\tilde{\epsilon}_{j,m} \\ x+A(i,m) &= \frac{1}{n}\sum_{q=1}^{i}\tilde{\epsilon}_{2qm+l,m}\sum_{j=(2qm+l-2h_n)\vee 1}^{2(q-1)m+l} K\left(\frac{2qm+l-j}{h_n}\right)\tilde{\epsilon}_{j,m}, \end{aligned}$$

we have:

$$\begin{aligned} \mathbb{E}(\varphi(Z(r,n,m))) \leq \sum_{i=1}^{[n/2m]-1}\Bigg|\mathbb{E}\Bigg(\varphi'\Bigg(\frac{1}{n}\sum_{q=1}^{i-1}\tilde{\epsilon}_{2qm+l,m}\sum_{j=(2qm+l-2h_n)\vee 1}^{2(q-1)m+l} K\left(\frac{2qm+l-j}{h_n}\Bigg)\tilde{\epsilon}_{j,m}\Bigg) \\ \times\frac{1}{n}\tilde{\epsilon}_{2im+l,m}\sum_{j=(2im+l-2h_n)\vee 1}^{2(i-1)m+l} K\left(\frac{2im+l-j}{h_n}\right)\tilde{\epsilon}_{j,m}+\psi(A(i,m))\Bigg)\Bigg|. \end{aligned}$$

Then by the triangle inequality, we obtain:

$$\begin{aligned} \mathbb{E}(\varphi(Z(r,n,m))) &\leq \sum_{i=1}^{[n/2m]-1}\Bigg|\mathbb{E}\Bigg(\varphi'\Bigg(\frac{1}{n}\sum_{q=1}^{i-1}\tilde{\epsilon}_{2qm+l,m}\sum_{j=(2qm+l-2h_n)\vee 1}^{2(q-1)m+l} K\left(\frac{2qm+l-j}{h_n}\Bigg)\tilde{\epsilon}_{j,m}\Bigg) \\ &\quad\times\frac{1}{n}\tilde{\epsilon}_{2im+l,m}\sum_{j=(2im+l-2h_n)\vee 1}^{2(i-1)m+l} K\left(\frac{2im+l-j}{h_n}\right)\tilde{\epsilon}_{j,m}\Bigg)\Bigg| \\ &\quad+\sum_{i=1}^{[n/2m]-1}\left|\mathbb{E}\left(|A(i,m)|^2\mathbf{1}_{\{|A(i,m)|\leq 1\}}+(2|A(i,m)|-1)\mathbf{1}_{\{|A(i,m)|>1\}}\right)\right| \\ &\leq \sum_{i=1}^{[n/2m]-1}\Bigg|\mathbb{E}\Bigg(\varphi'\Bigg(\frac{1}{n}\sum_{q=1}^{i-1}\tilde{\epsilon}_{2qm+l,m}\sum_{j=(2qm+l-2h_n)\vee 1}^{2(q-1)m+l} K\left(\frac{2qm+l-j}{h_n}\Bigg)\tilde{\epsilon}_{j,m}\Bigg) \\ &\quad\times\frac{1}{n}\tilde{\epsilon}_{2im+l,m}\sum_{j=(2im+l-2h_n)\vee 1}^{2(i-1)m+l} K\left(\frac{2im+l-j}{h_n}\right)\tilde{\epsilon}_{j,m}\Bigg)\Bigg| \\ &\quad+\sum_{i=1}^{[n/2m]-1}\mathbb{E}\left(|A(i,m)|^2\mathbf{1}_{\{|A(i,m)|\leq 1\}}+(2|A(i,m)|-1)\mathbf{1}_{\{|A(i,m)|>1\}}\right). \end{aligned}$$

By definition, $(\tilde{\epsilon}_{i,m})_{i\in\mathbb{Z}}$ satisfies:

$$\mathbb{E}(\tilde{\epsilon}_{2im+l,m}|\mathcal{F}_{2im+l-m})=0.$$

Hence:

$$\mathbb{E}(\varphi(Z(r,n,m))) \leq \sum_{i=1}^{[n/2m]-1} \mathbb{E}\left(|A(i,m)|^2 \mathbf{1}_{\{|A(i,m)|\leq 1\}} + (2|A(i,m)|-1)\mathbf{1}_{\{|A(i,m)|>1\}}\right)$$
$$= \sum_{i=1}^{[n/2m]-1} \mathbb{E}(\psi(|A(i,m)|).$$

For this term, put:

$$B(i,j,m,l) = \frac{[(2(i-1)m+l) - ((2im+l-2h_n)\vee 1)+1]}{n} K\left(\frac{2im+l-j}{h_n}\right)\tilde{\epsilon}_{2im+l,m}\tilde{\epsilon}_{j,m}.$$

We recall here the Lemma 3 of Dedecker [34]:

Lemma 1.7. *Let X_1, X_2, X_3, X_4 be identically distributed real random variables. Then:*

$$\mathbb{E}\left(|X_1X_2|\left(1\wedge 2\,|X_3X_4|\right)\right) \leq 2\mathbb{E}\left(X_1^2\left(1\wedge X_1^2\right)\right).$$

Using the convexity of ψ and Lemma 1.7, we have that:

$$\mathbb{E}(\psi(A(i,m)))$$
$$\leq \frac{1}{[(2(i-1)m+l) - ((2im+l-2h_n)\vee 1)+1]} \sum_{j=(2im+l-2h_n)\vee 1}^{2(i-1)m+l} \mathbb{E}\left(\psi\left(B(i,j,m,l)\right)\right).$$

Then:

$$\frac{1}{[(2(i-1)m+l) - ((2im+l-2h_n)\vee 1)+1]} \sum_{j=(2im+l-2h_n)\vee 1}^{2(i-1)m+l} \mathbb{E}\left(\psi\left(B(i,j,m,l)\right)\right)$$
$$\leq \frac{2}{[(2(i-1)m+l) - ((2im+l-2h_n)\vee 1)+1]} \sum_{j=(2im+l-2h_n)\vee 1}^{2(i-1)m+l} \mathbb{E}\left(\frac{2h_n}{n}|\tilde{\epsilon}_{0,m}|^2\left(1\wedge\frac{2h_n}{n}|\tilde{\epsilon}_{0,m}|^2\right)\right),$$

and:

$$\frac{2}{[(2(i-1)m+l) - ((2im+l-2h_n)\vee 1)+1]} \sum_{j=(2im+l-2h_n)\vee 1}^{2(i-1)m+l} \mathbb{E}\left(\frac{2h_n}{n}|\tilde{\epsilon}_{0,m}|^2\left(1\wedge\frac{2h_n}{n}|\tilde{\epsilon}_{0,m}|^2\right)\right)$$
$$\leq \frac{8h_n}{n}\mathbb{E}\left(|\tilde{\epsilon}_{0,m}|^2\left(1\wedge\frac{h_n}{n}|\tilde{\epsilon}_{0,m}|^2\right)\right).$$

Thus we can conclude if, for m fixed:

$$\lim_{n\to\infty} h_n\mathbb{E}\left(|\tilde{\epsilon}_{0,m}|^2\left(1\wedge\frac{h_n}{n}|\tilde{\epsilon}_{0,m}|^2\right)\right) = 0. \tag{1.47}$$

To prove (1.47), notice that:

$$\begin{aligned} h_n\mathbb{E}\left(|\tilde{\epsilon}_{0,m}|^2\left(1\wedge\frac{h_n}{n}|\tilde{\epsilon}_{0,m}|^2\right)\right) &\leq 4h_n\mathbb{E}\left(|\mathbb{E}(\epsilon_0|\mathcal{F}_m)|^2\left(1\wedge\frac{h_n}{n}|\mathbb{E}(\epsilon_0|\mathcal{F}_m)|^2\right)\right) \\ &+ 4h_n\mathbb{E}\left(|\mathbb{E}(\epsilon_0|\mathcal{F}_m)|^2\left(1\wedge\frac{h_n}{n}|\mathbb{E}(\epsilon_0|\mathcal{F}_{-m})|^2\right)\right) \\ &+4h_n\mathbb{E}\left(|\mathbb{E}(\epsilon_0|\mathcal{F}_{-m})|^2\left(1\wedge\frac{h_n}{n}|\mathbb{E}(\epsilon_0|\mathcal{F}_m)|^2\right)\right) \\ &+4h_n\mathbb{E}\left(|\mathbb{E}(\epsilon_0|\mathcal{F}_{-m})|^2\left(1\wedge\frac{h_n}{n}|\mathbb{E}(\epsilon_0|\mathcal{F}_{-m})|^2\right)\right). \quad (1.48) \end{aligned}$$

For the first term and for the last term of (1.48), we use the convexity of ψ:

$$\begin{aligned} h_n\mathbb{E}\left(|\mathbb{E}(\epsilon_0|\mathcal{F}_m)|^2\left(1\wedge\frac{h_n}{n}|\mathbb{E}(\epsilon_0|\mathcal{F}_m)|^2\right)\right) &\leq n\mathbb{E}\left(\psi\left(\mathbb{E}\left(\frac{h_n}{n}|\epsilon_0|^2|\mathcal{F}_m\right)\right)\right) \\ &\leq n\mathbb{E}\left(\mathbb{E}\left(\psi\left(\frac{h_n}{n}|\epsilon_0|^2\right)|\mathcal{F}_m\right)\right) \\ &\leq n\mathbb{E}\left(\psi\left(\frac{h_n}{n}|\epsilon_0|^2\right)\right) \\ &\leq 2h_n\mathbb{E}\left(|\epsilon_0|^2\left(1\wedge\frac{h_n}{n}|\epsilon_0|^2\right)\right). \quad (1.49) \end{aligned}$$

With the same idea, for the last term, we show that:

$$h_n\mathbb{E}\left(|\mathbb{E}(\epsilon_0|\mathcal{F}_{-m})|^2\left(1\wedge\frac{h_n}{n}|\mathbb{E}(\epsilon_0|\mathcal{F}_{-m})|^2\right)\right)\leq 2h_n\mathbb{E}\left(|\epsilon_0|^2\left(1\wedge\frac{h_n}{n}|\epsilon_0|^2\right)\right). \quad (1.50)$$

For the second term of (1.48), by convexity of ψ, we have:

$$\begin{aligned} n\mathbb{E}\left(\frac{h_n}{n}|\mathbb{E}(\epsilon_0|\mathcal{F}_m)|^2\left(1\wedge\frac{h_n}{n}|\mathbb{E}(\epsilon_0|\mathcal{F}_{-m})|^2\right)\right) &\leq n\mathbb{E}\left(\frac{h_n}{n}\mathbb{E}(|\epsilon_0|^2|\mathcal{F}_m)\left(1\wedge\frac{h_n}{n}\mathbb{E}(|\epsilon_0|^2|\mathcal{F}_{-m})\right)\right) \\ &\leq n\mathbb{E}\left(\psi\left(\mathbb{E}\left(\frac{h_n}{n}|\epsilon_0|^2|\mathcal{F}_{-m}\right)\right)\right) \\ &\leq 2h_n\mathbb{E}\left(|\epsilon_0|^2\left(1\wedge\frac{h_n}{n}|\epsilon_0|^2\right)\right). \quad (1.51) \end{aligned}$$

Since $g: x\to 1\wedge x$ is a concave function on $\mathbb{R}_+^*$ and ψ is a convex function, we obtain for the third term of (1.48) that:

$$
\begin{aligned}
h_n\mathbb{E}\left(|\mathbb{E}(\epsilon_0|\mathbb{F}_{-m})|^2\left(1\wedge\frac{h_n}{n}|\mathbb{E}(\epsilon_0|\mathcal{F}_m)|^2\right)\right) &\leq n\mathbb{E}\left(\frac{h_n}{n}\mathbb{E}(|\epsilon_0|^2|\mathbb{F}_{-m})\mathbb{E}\left(g\left(\frac{h_n}{n}\mathbb{E}(|\epsilon_0|^2|\mathcal{F}_m)\right)|\mathcal{F}_{-m}\right)\right)\\
&\leq n\mathbb{E}\left(\frac{h_n}{n}\mathbb{E}(|\epsilon_0|^2|\mathbb{F}_{-m})g\left(\mathbb{E}\left(\frac{h_n}{n}\mathbb{E}(|\epsilon_0|^2|\mathcal{F}_m)|\mathcal{F}_{-m}\right)\right)\right)\\
&\leq n\mathbb{E}\left(\frac{h_n}{n}\mathbb{E}(|\epsilon_0|^2|\mathbb{F}_{-m})\left(1\wedge\frac{h_n}{n}\mathbb{E}(|\epsilon_0|^2|\mathcal{F}_{-m})\right)\right)\\
&\leq n\mathbb{E}\left(\psi\left(\frac{h_n}{n}\mathbb{E}(|\epsilon_0|^2|\mathbb{F}_{-m})\right)\right)\\
&\leq 2h_n\mathbb{E}\left(|\epsilon_0|^2\left(1\wedge\frac{h_n}{n}|\epsilon_0|^2\right)\right). \qquad (1.52)
\end{aligned}
$$

Using (1.48) to (1.52), we deduce that (1.47) is verified as soon as (1.14) is true.

□

Proposition 1.3

Proof. Recall that:

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

where:

$$\hat{\gamma}_k=\frac{1}{n}\sum_{j=1}^{n-|k|}\epsilon_j\epsilon_{j+|k|}=\frac{1}{n}\sum_{j=1}^{n-|k|}\left(Y_j-\sum_{l=1}^{p}x_{j,l}\beta_l\right)\left(Y_{j+|k|}-\sum_{l=1}^{p}x_{j+|k|,l}\beta_l\right),\qquad 0\leq|k|\leq(n-1),$$

and:

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

where:

$$\hat{\gamma}_k^*=\frac{1}{n}\sum_{j=1}^{n-|k|}\hat{\epsilon}_j\hat{\epsilon}_{j+|k|}=\frac{1}{n}\sum_{j=1}^{n-|k|}\left(Y_j-\sum_{l=1}^{p}x_{j,l}\hat{\beta}_l\right)\left(Y_{j+|k|}-\sum_{l=1}^{p}x_{j+|k|,l}\hat{\beta}_l\right),\qquad 0\leq|k|\leq(n-1).$$

Thus we have:

$$
\begin{aligned}
\|f_n^*(\lambda)-f_n(\lambda)\|_{\mathbb{L}^1} &= \left\|\frac{1}{2\pi}\sum_{|k|\leq n-1}K\left(\frac{|k|}{h_n}\right)\hat{\gamma}_k^* e^{ik\lambda}-\frac{1}{2\pi}\sum_{|k|\leq n-1}K\left(\frac{|k|}{h_n}\right)\hat{\gamma}_k e^{ik\lambda}\right\|_{\mathbb{L}^1}\\
&= \left\|\frac{1}{2\pi}\sum_{|k|\leq 2h_n}K\left(\frac{|k|}{h_n}\right)\left[\hat{\gamma}_k^*-\hat{\gamma}_k\right]e^{ik\lambda}\right\|_{\mathbb{L}^1}\\
&\leq \frac{1}{2\pi}\sum_{|k|\leq 2h_n}\|\hat{\gamma}_k^*-\hat{\gamma}_k\|_{\mathbb{L}^1}. \qquad (1.53)
\end{aligned}
$$

Since $\frac{h_n}{n}$ tends to 0 as $n \to \infty$, it remains to prove that:

$$\sup_{|k|\leq 2h_n} \|\hat{\gamma}_k^* - \hat{\gamma}_k\|_{\mathbb{L}^1} = \mathcal{O}\left(\frac{1}{n}\right). \tag{1.54}$$

Lemma 1.8. *The following inequality is verified:*

$$\begin{aligned}
\|\hat{\gamma}_k^* - \hat{\gamma}_k\|_{\mathbb{L}^1} &= \left\| \frac{1}{n}\sum_{j=1}^{n-|k|} \left(Y_j - \sum_{l=1}^{p} x_{j,l}\hat{\beta}_l\right)\left(Y_{j+|k|} - \sum_{l=1}^{p} x_{j+|k|,l}\hat{\beta}_l\right) \right. \\
&\quad \left. - \frac{1}{n}\sum_{j=1}^{n-|k|} \left(Y_j - \sum_{l=1}^{p} x_{j,l}\beta_l\right)\left(Y_{j+|k|} - \sum_{l=1}^{p} x_{j+|k|,l}\beta_l\right) \right\|_{\mathbb{L}^1} \\
&\leq \frac{1}{2n}\sum_{l=1}^{p}\sum_{l'=1}^{p} \left\| \left(\beta_l - \hat{\beta}_l\right)^2 \sum_{j=1}^{n-|k|} x_{j,l}^2 \right\|_{\mathbb{L}^1} + \frac{1}{2n}\sum_{l=1}^{p}\sum_{l'=1}^{p} \left\| \left(\beta_{l'} - \hat{\beta}_{l'}\right)^2 \sum_{j=1}^{n-|k|} x_{j+|k|,l'}^2 \right\|_{\mathbb{L}^1} \\
&\quad + \frac{1}{n}\sum_{l=1}^{p} \left\| \sum_{j=1}^{n-|k|} \epsilon_j x_{j+|k|,l} \left(\beta_l - \hat{\beta}_l\right) \right\|_{\mathbb{L}^1} + \frac{1}{n}\sum_{l=1}^{p} \left\| \sum_{j=1}^{n-|k|} \epsilon_{j+|k|} x_{j,l} \left(\beta_l - \hat{\beta}_l\right) \right\|_{\mathbb{L}^1}.
\end{aligned} \tag{1.55}$$

The proof of this lemma will be given in Section 1.6.3.

It remains to calculate these four terms. For the first term on the right-hand side, for all l, l' fixed and for all k, we have:

$$\left\| \left(\beta_l - \hat{\beta}_l\right)^2 \sum_{j=1}^{n-|k|} x_{j,l}^2 \right\|_{\mathbb{L}^1} \leq \left\| \left(\beta_l - \hat{\beta}_l\right)^2 \sum_{j=1}^{n} x_{j,l}^2 \right\|_{\mathbb{L}^1},$$

and:

$$\left\| \left(\beta_l - \hat{\beta}_l\right)^2 \sum_{j=1}^{n} x_{j,l}^2 \right\|_{\mathbb{L}^1} = \left\| d_l(n)^2 \left(\beta_l - \hat{\beta}_l\right)^2 \right\|_{\mathbb{L}^1} = d_l(n)^2 \mathbb{E}\left(\left(\beta_l - \hat{\beta}_l\right)^2\right).$$

Hannan has proved in his paper [47] a Central Limit Theorem (1.7) with the convergence of the second order moments (1.8). Consequently, we have:

$$\left\| d_l(n)^2 \left(\beta_l - \hat{\beta}_l\right)^2 \right\|_{\mathbb{L}^1} = \mathcal{O}(1),$$

hence:

$$\sup_{|k|\leq 2h_n} \left(\left\| \left(\beta_l - \hat{\beta}_l\right)^2 \sum_{j=1}^{n-|k|} x_{j,l}^2 \right\|_{\mathbb{L}^1} \right) \leq d_l(n)^2 \mathbb{E}\left(\left(\hat{\beta}_l - \beta_l\right)^2\right) = \mathcal{O}(1).$$

So we can conclude:

$$\sup_{|k|\leq 2h_n} \left(\frac{1}{2n}\sum_{l=1}^{p}\sum_{l'=1}^{p} \left\| \left(\beta_l - \hat{\beta}_l\right)^2 \sum_{j=1}^{n-|k|} x_{j,l}^2 \right\|_{\mathbb{L}^1} \right) = \mathcal{O}\left(\frac{1}{n}\right).$$

For the second term of (1.55), the same arguments are used, because $\sum_{j=1}^{n-|k|} x_{j+|k|,l}^2 \leq \sum_{j=1}^{n} x_{j,l}^2$. Hence:

$$\sup_{|k|\leq 2h_n} \left(\frac{1}{2n} \sum_{l=1}^{p} \sum_{l'=1}^{p} \left\| \left(\beta_{l'} - \hat{\beta}_{l'}\right)^2 \sum_{j=1}^{n-|k|} x_{j+|k|,l'}^2 \right\|_{\mathbb{L}^1} \right) = \mathcal{O}\left(\frac{1}{n}\right).$$

For the third term, for all l fixed, by the Cauchy-Schwarz inequality, we get:

$$\left\| \sum_{j=1}^{n-|k|} \epsilon_j x_{j+|k|,l} \left(\beta_l - \hat{\beta}_l\right) \right\|_{\mathbb{L}^1} \leq \left\| \sum_{j=1}^{n-|k|} \epsilon_j x_{j+|k|,l} \right\|_{\mathbb{L}^2} \left\| \beta_l - \hat{\beta}_l \right\|_{\mathbb{L}^2}.$$

Then, we have:

$$\begin{aligned} \left\| \sum_{j=1}^{n-|k|} \epsilon_j x_{j+|k|,l} \right\|_{\mathbb{L}^2}^2 &= \sum_{j=1}^{n-|k|} \sum_{i=1}^{n-|k|} \gamma_{i-j} x_{i+|k|,l} x_{j+|k|,l} \\ &= \sum_{i=1}^{n-|k|} \sum_{j=i}^{n-|k|} \gamma_{j-i} x_{i+|k|,l} x_{j+|k|,l} + \sum_{i=1}^{n-|k|} \sum_{j=1}^{i-1} \gamma_{j-i} x_{i+|k|,l} x_{j+|k|,l}. \end{aligned} \tag{1.56}$$

For the first term of the right-hand side in (1.56), it follows with the change of variables $r = j - i$ that:

$$\begin{aligned} \sum_{i=1}^{n-|k|} \sum_{j=i}^{n-|k|} \gamma_{j-i} x_{i+|k|,l} x_{j+|k|,l} &= \sum_{i=1}^{n-|k|} \sum_{r=0}^{n-|k|-i} \gamma_r x_{i+|k|,l} x_{i+|k|+r,l} \\ &\leq \sum_{i=1}^{n-|k|} \sum_{r=0}^{n-|k|-i} |\gamma_r| |x_{i+|k|,l}| |x_{i+|k|+r,l}| \\ &\leq \sum_{i=1}^{n-|k|} \sum_{r=0}^{n-|k|-i} |\gamma_r| (x_{i+|k|,l}^2 + x_{i+|k|+r,l}^2) \\ &\leq \sum_{i=1}^{n-|k|} \sum_{r=0}^{n-|k|-i} |\gamma_r| x_{i+|k|,l}^2 + \sum_{i=1}^{n-|k|} \sum_{r=0}^{n-|k|-i} |\gamma_r| x_{i+|k|+r,l}^2. \end{aligned}$$

Since $r \leq n - |k| - i$, we have $i \leq n - |k| - r$, and it follows that:

$$\begin{aligned} &\sum_{i=1}^{n-|k|} \sum_{r=0}^{n-|k|-i} |\gamma_r| x_{i+|k|,l}^2 + \sum_{i=1}^{n-|k|} \sum_{r=0}^{n-|k|-i} |\gamma_r| x_{i+|k|+r,l}^2 \\ &\qquad \leq \sum_{i=1}^{n-|k|} x_{i+|k|,l}^2 \sum_{r=0}^{n-|k|-i} |\gamma_r| + \sum_{r=0}^{n-|k|} |\gamma_r| \sum_{i=1}^{n-|k|-r} x_{i+|k|+r,l}^2 \\ &\qquad \leq \sum_{i=1}^{n-|k|} x_{i+|k|,l}^2 \sum_{r} |\gamma_r| + \sum_{r} |\gamma_r| \sum_{i=1}^{n-|k|-r} x_{i+|k|+r,l}^2. \end{aligned}$$

Since $\sum_k |\gamma(k)| < \infty$:

$$\begin{aligned}\sum_{i=1}^{n-|k|} x_{i+|k|,l}^2 \sum_r |\gamma_r| + \sum_r |\gamma_r| \sum_{i=1}^{n-|k|-r} x_{i+|k|+r,l}^2 &\leq M\left(\sum_{i=1}^{n-|k|} x_{i+|k|,l}^2 + \sum_{i=1}^{n-|k|-r} x_{i+|k|+r,l}^2\right)\\ &\leq M\left(\sum_{i=1}^{n} x_{i,l}^2 + \sum_{i=1}^{n} x_{i,l}^2\right)\\ &\leq M'\sum_{i=1}^{n} x_{i,l}^2.\end{aligned}$$

With the same idea, for the second term of the right-hand side of (1.56), we have:

$$\sum_{i=1}^{n-|k|}\sum_{j=1}^{i-1} \gamma_{j-i} x_{i+|k|,l} x_{j+|k|,l} \leq M'\sum_{j=1}^{n} x_{j,l}^2,$$

thus:

$$\sup_{|k|\leq 2h_n} \left\| \sum_{j=1}^{n-|k|} \epsilon_j x_{j+|k|,l} \right\|_{\mathbb{L}^2}^2 \leq 2M'\sum_{j=1}^{n} x_{j,l}^2 = M'' d_l(n)^2.$$

In conclusion, for the third term of (1.55):

$$\begin{aligned}\left\| \sum_{j=1}^{n-|k|} \epsilon_j x_{j+|k|,l} \left(\beta_l - \hat{\beta}_l\right) \right\|_{\mathbb{L}^1} &\leq \left\| \sum_{j=1}^{n-|k|} \epsilon_j x_{j+|k|,l} \right\|_{\mathbb{L}^2} \left\|\beta_l - \hat{\beta}_l\right\|_{\mathbb{L}^2}\\ &\leq C d_l(n)\sqrt{\mathbb{E}\left((\beta_l - \hat{\beta}_l)^2\right)}\\ &\leq C\sqrt{d_l(n)^2\mathbb{E}\left((\beta_l - \hat{\beta}_l)^2\right)} = \mathcal{O}(1),\end{aligned}$$

hence:

$$\sup_{|k|\leq 2h_n} \left\| \sum_{j=1}^{n-|k|} \epsilon_j x_{j+|k|,l} \left(\beta_l - \hat{\beta}_l\right) \right\|_{\mathbb{L}^1} = \mathcal{O}(1),$$

thereby:

$$\sup_{|k|\leq 2h_n} \left(\frac{1}{n}\sum_{l=1}^{p} \left\| \sum_{j=1}^{n-|k|} \epsilon_j x_{j+|k|,l} \left(\beta_l - \hat{\beta}_l\right) \right\|_{\mathbb{L}^1} \right) = \mathcal{O}\left(\frac{1}{n}\right).$$

The same idea is used for the fourth term of the right-hand side of (1.55). Thus (1.54) is verified and consequently (1.28) is true.

□

1.6.3 Proof of Lemma 1.8

We start by developing the term Y_j:

$$\begin{aligned}\left\|\hat{\gamma}_k^* - \hat{\gamma}_k\right\|_{\mathbb{L}^1} &= \left\|\frac{1}{n}\sum_{j=1}^{n-|k|}\left(Y_j - \sum_{l=1}^{p} x_{j,l}\hat{\beta}_l\right)\left(Y_{j+|k|} - \sum_{l=1}^{p} x_{j+|k|,l}\hat{\beta}_l\right)\right. \\ &\qquad \left. - \frac{1}{n}\sum_{j=1}^{n-|k|}\left(Y_j - \sum_{l=1}^{p} x_{j,l}\beta_l\right)\left(Y_{j+|k|} - \sum_{l=1}^{p} x_{j+|k|,l}\beta_l\right)\right\|_{\mathbb{L}^1} \\ &= \left\|\frac{1}{n}\sum_{j=1}^{n-|k|}\left(\sum_{l=1}^{p} x_{j,l}\left(\beta_l - \hat{\beta}_l\right) + \epsilon_j\right)\left(\sum_{l=1}^{p} x_{j+|k|,l}\left(\beta_l - \hat{\beta}_l\right) + \epsilon_{j+|k|}\right)\right. \\ &\qquad \left. - \frac{1}{n}\sum_{j=1}^{n-|k|}\left(Y_j - \sum_{l=1}^{p} x_{j,l}\beta_l\right)\left(Y_{j+|k|} - \sum_{l=1}^{p} x_{j+|k|,l}\beta_l\right)\right\|_{\mathbb{L}^1}.\end{aligned}$$

Since ϵ_j is equal to $Y_j - \sum_{l=1}^{p} x_{j,l}\beta_l$, we have:

$$\begin{aligned}&\left\|\frac{1}{n}\sum_{j=1}^{n-|k|}\left(\sum_{l=1}^{p} x_{j,l}\left(\beta_l - \hat{\beta}_l\right) + \epsilon_j\right)\left(\sum_{l=1}^{p} x_{j+|k|,l}\left(\beta_l - \hat{\beta}_l\right) + \epsilon_{j+|k|}\right)\right. \\ &\qquad \left. - \frac{1}{n}\sum_{j=1}^{n-|k|}\left(Y_j - \sum_{l=1}^{p} x_{j,l}\beta_l\right)\left(Y_{j+|k|} - \sum_{l=1}^{p} x_{j+|k|,l}\beta_l\right)\right\|_{\mathbb{L}^1} \\ &\quad = \left\|\frac{1}{n}\sum_{j=1}^{n-|k|}\left(\sum_{l=1}^{p} x_{j,l}\left(\beta_l - \hat{\beta}_l\right)\sum_{l=1}^{p} x_{j+|k|,l}\left(\beta_l - \hat{\beta}_l\right)\right.\right. \\ &\qquad \left.\left. + \epsilon_j\sum_{l=1}^{p} x_{j+|k|,l}\left(\beta_l - \hat{\beta}_l\right) + \sum_{l=1}^{p} x_{j,l}\left(\beta_l - \hat{\beta}_l\right)\epsilon_{j+|k|}\right)\right\|_{\mathbb{L}^1}.\end{aligned}$$

Using the triangle inequality, we obtain:

$$\begin{aligned}&\left\|\frac{1}{n}\sum_{j=1}^{n-|k|}\left(\sum_{l=1}^{p} x_{j,l}\left(\beta_l - \hat{\beta}_l\right)\sum_{l=1}^{p} x_{j+|k|,l}\left(\beta_l - \hat{\beta}_l\right)\right.\right. \\ &\qquad \left.\left. + \epsilon_j\sum_{l=1}^{p} x_{j+|k|,l}\left(\beta_l - \hat{\beta}_l\right) + \sum_{l=1}^{p} x_{j,l}\left(\beta_l - \hat{\beta}_l\right)\epsilon_{j+|k|}\right)\right\|_{\mathbb{L}^1} \\ &\leq \left\|\frac{1}{n}\sum_{j=1}^{n-|k|}\left(\sum_{l=1}^{p} x_{j,l}\left(\beta_l - \hat{\beta}_l\right)\sum_{l=1}^{p} x_{j+|k|,l}\left(\beta_l - \hat{\beta}_l\right)\right)\right\|_{\mathbb{L}^1} + \left\|\frac{1}{n}\sum_{j=1}^{n-|k|}\left(\epsilon_j\sum_{l=1}^{p} x_{j+|k|,l}\left(\beta_l - \hat{\beta}_l\right)\right)\right\|_{\mathbb{L}^1} \\ &\qquad + \left\|\frac{1}{n}\sum_{j=1}^{n-|k|}\left(\sum_{l=1}^{p} x_{j,l}\left(\beta_l - \hat{\beta}_l\right)\epsilon_{j+|k|}\right)\right\|_{\mathbb{L}^1},\end{aligned}$$

then we swap the sums between them:

$$\begin{aligned}
&\left\| \frac{1}{n} \sum_{j=1}^{n-|k|} \left(\sum_{l=1}^{p} x_{j,l} \left(\beta_l - \hat{\beta}_l \right) \sum_{l=1}^{p} x_{j+|k|,l} \left(\beta_l - \hat{\beta}_l \right) \right) \right\|_{\mathbb{L}^1} \\
&+ \left\| \frac{1}{n} \sum_{j=1}^{n-|k|} \left(\epsilon_j \sum_{l=1}^{p} x_{j+|k|,l} \left(\beta_l - \hat{\beta}_l \right) \right) \right\|_{\mathbb{L}^1} + \left\| \frac{1}{n} \sum_{j=1}^{n-|k|} \left(\sum_{l=1}^{p} x_{j,l} \left(\beta_l - \hat{\beta}_l \right) \epsilon_{j+|k|} \right) \right\|_{\mathbb{L}^1} \\
\leq \quad &\left\| \frac{1}{n} \sum_{j=1}^{n-|k|} \left(\sum_{l=1}^{p} x_{j,l} \left(\beta_l - \hat{\beta}_l \right) \sum_{l=1}^{p} x_{j+|k|,l} \left(\beta_l - \hat{\beta}_l \right) \right) \right\|_{\mathbb{L}^1} \\
&+ \left\| \frac{1}{n} \sum_{l=1}^{p} \sum_{j=1}^{n-|k|} \epsilon_j x_{j+|k|,l} \left(\beta_l - \hat{\beta}_l \right) \right\|_{\mathbb{L}^1} + \left\| \frac{1}{n} \sum_{l=1}^{p} \sum_{j=1}^{n-|k|} \epsilon_{j+|k|} x_{j,l} \left(\beta_l - \hat{\beta}_l \right) \right\|_{\mathbb{L}^1},
\end{aligned}$$

and using again the triangle inequality:

$$\begin{aligned}
&\left\| \frac{1}{n} \sum_{j=1}^{n-|k|} \left(\sum_{l=1}^{p} x_{j,l} \left(\beta_l - \hat{\beta}_l \right) \sum_{l=1}^{p} x_{j+|k|,l} \left(\beta_l - \hat{\beta}_l \right) \right) \right\|_{\mathbb{L}^1} \\
&+ \left\| \frac{1}{n} \sum_{l=1}^{p} \sum_{j=1}^{n-|k|} \epsilon_j x_{j+|k|,l} \left(\beta_l - \hat{\beta}_l \right) \right\|_{\mathbb{L}^1} + \left\| \frac{1}{n} \sum_{l=1}^{p} \sum_{j=1}^{n-|k|} \epsilon_{j+|k|} x_{j,l} \left(\beta_l - \hat{\beta}_l \right) \right\|_{\mathbb{L}^1} \\
\leq \quad &\left\| \frac{1}{n} \sum_{j=1}^{n-|k|} \left(\sum_{l=1}^{p} x_{j,l} \left(\beta_l - \hat{\beta}_l \right) \sum_{l'=1}^{p} x_{j+|k|,l'} \left(\beta_{l'} - \hat{\beta}_{l'} \right) \right) \right\|_{\mathbb{L}^1} \\
&+ \frac{1}{n} \sum_{l=1}^{p} \left\| \sum_{j=1}^{n-|k|} \epsilon_j x_{j+|k|,l} \left(\beta_l - \hat{\beta}_l \right) \right\|_{\mathbb{L}^1} + \frac{1}{n} \sum_{l=1}^{p} \left\| \sum_{j=1}^{n-|k|} \epsilon_{j+|k|} x_{j,l} \left(\beta_l - \hat{\beta}_l \right) \right\|_{\mathbb{L}^1}. \qquad (1.57)
\end{aligned}$$

For the first term on the right-hand side of (1.57), we have:

$$\begin{aligned}
&\left\| \frac{1}{n} \sum_{j=1}^{n-|k|} \left(\sum_{l=1}^{p} x_{j,l} \left(\beta_l - \hat{\beta}_l \right) \sum_{l'=1}^{p} x_{j+|k|,l'} \left(\beta_{l'} - \hat{\beta}_{l'} \right) \right) \right\|_{\mathbb{L}^1} \\
&\qquad = \left\| \frac{1}{n} \sum_{l=1}^{p} \sum_{l'=1}^{p} \sum_{j=1}^{n-|k|} x_{j,l} \left(\beta_l - \hat{\beta}_l \right) x_{j+|k|,l'} \left(\beta_{l'} - \hat{\beta}_{l'} \right) \right\|_{\mathbb{L}^1},
\end{aligned}$$

then by triangle inequality:

$$\begin{aligned}
&\left\| \frac{1}{n} \sum_{l=1}^{p} \sum_{l'=1}^{p} \sum_{j=1}^{n-|k|} x_{j,l} \left(\beta_l - \hat{\beta}_l \right) x_{j+|k|,l'} \left(\beta_{l'} - \hat{\beta}_{l'} \right) \right\|_{\mathbb{L}^1} \\
&\qquad \leq \frac{1}{n} \sum_{l=1}^{p} \sum_{l'=1}^{p} \left\| \sum_{j=1}^{n-|k|} x_{j,l} \left(\beta_l - \hat{\beta}_l \right) x_{j+|k|,l'} \left(\beta_{l'} - \hat{\beta}_{l'} \right) \right\|_{\mathbb{L}^1}.
\end{aligned}$$

Since $ab \leq \frac{1}{2}a^2 + \frac{1}{2}b^2$, we get:

$$\frac{1}{n}\sum_{l=1}^{p}\sum_{l'=1}^{p}\left\|\sum_{j=1}^{n-|k|} x_{j,l}\left(\beta_l - \hat{\beta}_l\right) x_{j+|k|,l'}\left(\beta_{l'} - \hat{\beta}_{l'}\right)\right\|_{\mathbb{L}^1} \leq \frac{1}{n}\sum_{l=1}^{p}\sum_{l'=1}^{p}\left\|\frac{1}{2}\sum_{j=1}^{n-|k|}\left(x_{j,l}\left(\beta_l - \hat{\beta}_l\right)\right)^2 + \frac{1}{2}\sum_{j=1}^{n-|k|}\left(x_{j+|k|,l'}\left(\beta_{l'} - \hat{\beta}_{l'}\right)\right)^2\right\|_{\mathbb{L}^1},$$

and by triangle inequality:

$$\frac{1}{n}\sum_{l=1}^{p}\sum_{l'=1}^{p}\left\|\frac{1}{2}\sum_{j=1}^{n-|k|}\left(x_{j,l}\left(\beta_l - \hat{\beta}_l\right)\right)^2 + \frac{1}{2}\sum_{j=1}^{n-|k|}\left(x_{j+|k|,l'}\left(\beta_{l'} - \hat{\beta}_{l'}\right)\right)^2\right\|_{\mathbb{L}^1} \leq \frac{1}{2n}\sum_{l=1}^{p}\sum_{l'=1}^{p}\left\|\left(\beta_l - \hat{\beta}_l\right)^2\sum_{j=1}^{n-|k|}x_{j,l}^2\right\|_{\mathbb{L}^1} + \frac{1}{2n}\sum_{l=1}^{p}\sum_{l'=1}^{p}\left\|\left(\beta_{l'} - \hat{\beta}_{l'}\right)^2\sum_{j=1}^{n-|k|}x_{j+|k|,l'}^2\right\|_{\mathbb{L}^1}.$$

In conclusion, we have:

$$\begin{aligned}\|\hat{\gamma}_k^* - \hat{\gamma}_k\|_{\mathbb{L}^1} \leq\ & \frac{1}{2n}\sum_{l=1}^{p}\sum_{l'=1}^{p}\left\|\left(\beta_l - \hat{\beta}_l\right)^2\sum_{j=1}^{n-|k|}x_{j,l}^2\right\|_{\mathbb{L}^1} + \frac{1}{2n}\sum_{l=1}^{p}\sum_{l'=1}^{p}\left\|\left(\beta_{l'} - \hat{\beta}_{l'}\right)^2\sum_{j=1}^{n-|k|}x_{j+|k|,l'}^2\right\|_{\mathbb{L}^1} \\ & + \frac{1}{n}\sum_{l=1}^{p}\left\|\sum_{j=1}^{n-|k|}\epsilon_j x_{j+|k|,l}\left(\beta_l - \hat{\beta}_l\right)\right\|_{\mathbb{L}^1} + \frac{1}{n}\sum_{l=1}^{p}\left\|\sum_{j=1}^{n-|k|}\epsilon_{j+|k|}x_{j,l}\left(\beta_l - \hat{\beta}_l\right)\right\|_{\mathbb{L}^1}.\end{aligned}$$

CHAPTER 2

ASYMPTOTIC DISTRIBUTION OF LEAST SQUARES ESTIMATORS FOR LINEAR MODELS WITH DEPENDENT ERRORS

Le contenu de ce chapitre est issu d'un article publié dans la revue *Statistics* [24].

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 [47], who proved a Central Limit Theorem for the usual least squares 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.

2.1 Introduction

The linear regression model is used in many domains of applied mathematics, and the asymptotic behavior of the least squares 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 Bassett and Koenker [14], Babu [9], Bai, Rao and Wu [10] and He and Shao [48] among others. However, many science and engineering data exhibit significant temporal dependence so that the assumption of independence is violated (see for instance Brockwell and Davis [23]). 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 [47]. 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 inter-dependence of the variables of the design. As in Hannan, we assume that the design 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 [47], who proved a Central Limit Theorem for the usual least squares 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 squares 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 [74]. We refer to our previous paper [25], 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 [65] consider the case where the errors follow a $MA(q)$ process, and Chib and Greenberg [30] the case where the errors are an $ARMA(p, q)$ process. A more general framework is used by Wu [75] for a class of short-range dependent processes. These results are based on the asymptotic theory of stationary processes developed by Wu in [74]. 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 squares estimators of the parameters. This estimator is derived from the estimator of the spectral density of the error process introduced in Caron and Dede [25]. 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 equal to 5%).

The paper is organized as follows. In Section 2.2, we recall Hannan's Central Limit Theorem for the least squares estimator. In Section 2.3, we focus on the estimation of the covariance matrix under Hannan's conditions. Finally, Section 2.4 is devoted to the correction of the usual Student and Fisher tests in our dependent context, and some simulations with different models are realized.

2.2 Hannan's theorem

2.2.1 Notations and definitions

Let us recall the equation of the linear regression model:

$$Y = X\beta + \epsilon, \tag{2.1}$$

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 $\{1, \dots, p\}$ and i in $\{1, \dots, n\}$. 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 (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:

$$\Gamma_n = [\gamma(j-l)]_{1\leq j,l\leq n}\,. \tag{2.2}$$

2.2.2 Hannan's Central Limit Theorem

Let $\hat{\beta}$ be the usual least squares estimator for the unknown vector β. Given the design X, Hannan [47] 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:

$$\sum_{k\in\mathbb{Z}} |\gamma(k)| < \infty, \tag{2.3}$$

(see for instance Dedecker, Merlevède and Volný [38]).

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 [38] for more details). Notice that the condition (2.3) 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 of Caron and Dede [25], 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 $R(k)$:

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

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 [47], 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:*

$$\mathbb{E}\left(f\left(D(n)(\hat{\beta}-\beta)\right)\Big|X\right) \xrightarrow[n\to\infty]{} \mathbb{E}\left(f(Z)\Big|X\right), \quad a.s. \tag{2.4}$$

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]

$$\mathbb{E}\left(D(n)(\hat{\beta}-\beta)(\hat{\beta}-\beta)^t D(n)^t\Big|X\right) \xrightarrow[n\to\infty]{} F^{-1}GF^{-1}, \quad a.s. \tag{2.5}$$

Remark 2.1. *Let us notice that, by the dominated convergence theorem, the property* (2.4) *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* (2.5)*. We shall see that, in fact,* (2.5) *holds under* (2.3) *and* (C4) - (C5) *(Hannan's condition* (C1)*, which implies* (2.3)*, is needed for* (2.4) *only). Moreover, this proof will serve as a preliminary to the proof of Theorem 2.2. 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 kth diagonal and 0 elsewhere.

Hence, we deduce that:

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

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:

$$B_{k,n} = [\hat{\rho}_{j,l}(k)] \quad \text{if } k \geq 0, \qquad B_{k,n} = [\hat{\rho}_{j,l}(-k)] \quad \text{if } k \leq -1, \tag{2.6}$$

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:*

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

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 (2.5)*.*

2.3 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:

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

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:

$$\widehat{\Gamma}^*_{n,h_n} = \left[K\left(\frac{j-l}{h_n}\right)\hat{\gamma}^*_{j-l}\right]_{1\leq j,l\leq n}, \tag{2.7}$$

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 Bickel and Levina [16], $\widehat{\Gamma}^*_{n,h_n}$ 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:

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

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

$$h_n\mathbb{E}\left(|\epsilon_0|^2\left(1\wedge\frac{h_n}{n}|\epsilon_0|^2\right)\right)\xrightarrow[n\to\infty]{}0. \tag{2.8}$$

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

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

Remark 2.3. *If ϵ_0 is square integrable, then there exists $h_n\to\infty$ such that* (2.8) *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* (2.8) *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 2.1. *Let* h_n *be a sequence satisfying* (2.8)*. Then the estimator* C_n *converges in probability to* C *as* n *tends to infinity.*

Remark 2.4. *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 [78], 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 Brockwell and Davis [23]). Consequently, by the Schur Product Theorem in matrix theory [51], 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 2.2, we obtain the following corollary, which is the main result of our paper:

Corollary 2.2. *Under the assumptions of Theorem 2.1 and Theorem 2.2, 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.*

2.4 Tests and simulations

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

2.4.1 Tests

Let us recall the assumptions. We consider the linear regression model (2.1), 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 (2.1) holds).

In order to test $H_0 : \beta_j = 0$ against $H_1 : \beta_j \neq 0$, for j in $\{1, \dots, p\}$, under the H_0-hypothesis and according to Corollary 2.2 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:

$$T_{j,n} = \frac{d_j(n)\hat{\beta}_j}{\sqrt{c_{n,(j,j)}}}. \tag{2.9}$$

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: $\exists j_z \in \{j_1, \dots, j_{p_0}\}$ such that $\beta_{j_z} \neq 0$. By Corollary 2.2, 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 $\{j_1, \dots, j_{p_0}\}$. 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:

$$\Xi = Z_{1,n}^2 + \cdots + Z_{p_0,n}^2. \tag{2.10}$$

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:

$$\begin{cases} K(x) &= 1 & if\ |x| < 0.8 \\ K(x) &= 5 - 5|x| & if\ 0.8 \leq |x| \leq 1 \\ K(x) &= 0 & if\ |x| > 1. \end{cases} \tag{2.11}$$

This kernel verifies the conditions defined at the beginning of Section 2.3, 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.

2.4.2 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 [21], but it is $\tilde{\phi}$-dependent in the sense of Dedecker and Prieur [39] (see also Caron and Dede [25], Section 4.4). Indeed, one can prove that the coefficients $\tilde{\phi}(k)$ of the chain $(Z_i)_{i\geq 1}$ decrease geometrically [39]: $\tilde{\phi}(k) \leq 2^{-k}$. 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 4.4 in Caron and Dede [25]). 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 $\{1, \ldots, n\}$:

$$Y_i = \beta_0 + \beta_1(i^2 + X_i) + \epsilon_i, \tag{2.12}$$

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 (2.9). 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 2.1. For this example, the shape of the empirical autocovariance suggests to keep only four terms. This leads to choose $h_n = 5$.

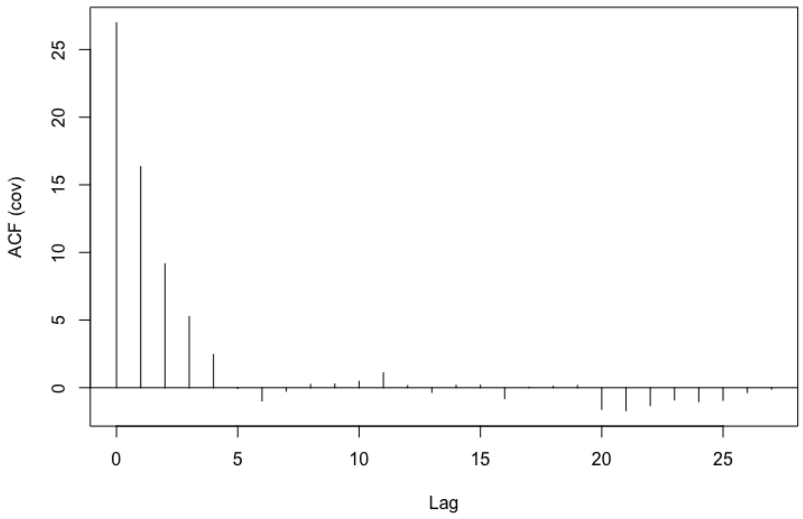


Figure 2.1 – Empirical autocovariance of the residuals of Model (2.12).

• 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 $\{1, \dots, n\}$:

$$Y_i = \beta_0 + \beta_1(\log(i) + \sin(i) + X_i) + \beta_2 i + \epsilon_i. \tag{2.13}$$

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 (2.10). 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 2.2, it suggests to keep only five terms of covariances. Given the kernel (2.11), if we want to keep five 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

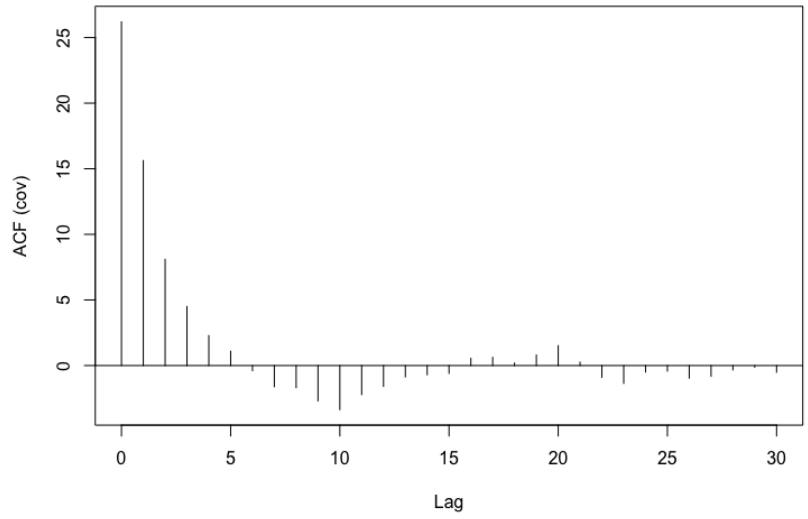


Figure 2.2 – Empirical autocovariance of the residuals of Model (2.13).

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 $n = 1000$.

2.5 Proofs

2.5.1 Theorem 2.2

Proof. In this proof, we use the notations introduced in Section 2.2 and Section 2.3. 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 $\{1, \dots, p\}$:

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

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

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

Then it remains to prove that:

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

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 (2.2), 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 (2.7). 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:

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

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:

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

and:

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

with:

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

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

and:

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

Let us compute:

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

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

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

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:

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

Theorem 3.1 of our paper [25] 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:

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

Furthermore, this result is still available for all kernel K verifying the conditions at the beginning of Section 2.3.

Thus it remains to find a bound for:

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

Let us recall (see (2.6)) 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:

$$\hat{\rho}_{j,l}(k) = \frac{\hat{\gamma}_{j,l}(k)}{\sqrt{\hat{\gamma}_{j,j}(0)\hat{\gamma}_{l,l}(0)}}. \tag{2.14}$$

For a multivariate time series, let us recall that the cross-periodogram is defined by, for all j, l in $\{1,\ldots,p\}$ [23]:

$$[I_n(\lambda)]_{j,l} = \frac{1}{2\pi} \sum_{k=-n+1}^{n-1} \hat{\gamma}_{j,l}(k) e^{ik\lambda}. \tag{2.15}$$

Combining (2.14) and (2.15), 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 [23], we get:

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

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 $\{1, \ldots, p\}$:

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

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 $\{1, \ldots, p\}$:

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

□

2.5.2 Corollary 2.1

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 2.2 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:

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

□

CHAPTER 3

LINEAR REGRESSION WITH STATIONARY ERRORS : THE R PACKAGE SLM

Le contenu de ce chapitre est issu d'un article écrit en collaboration avec Jérôme Dedecker et Bertrand Michel. L'article a été soumis à la revue *Journal of Statistical Software* et est disponible sur le site internet d'*arXiv*: `https://arxiv.org/abs/1906.06583`. Le package R associé `slm` est disponible sur le site internet du logiciel R: `https://cran.r-project.org/index.html`, ou sur *GitHub* via le lien suivant: `https://github.com/E-Caron/slm`.

Abstract

This paper introduces the R package `slm` which stands for Stationary Linear Models. The package contains a set of statistical procedures for linear regression in the general context where the error process is strictly stationary with short memory. We work in the setting of Hannan [47], who proved the asymptotic normality of the (normalized) least squares estimators (LSE) under very mild conditions on the error process. We propose different ways to estimate the asymptotic covariance matrix of the LSE, and then to correct the type-I error rates of the usual tests on the parameters (as well as confidence intervals). The procedures are evaluated through different sets of simulations, and two examples of real datasets are studied.

3.1 Introduction

We consider the usual linear regression model

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

where Y is the n-dimensional vector of observations, X is a (possibly random) $n \times p$ design matrix, β is a p-dimensional vector of parameters, and $\varepsilon = (\varepsilon_i)_{1\leq i\leq n}$ is the error process (with zero mean and independent of X). The standard assumptions are that the ε_i's are independent and identically distributed (i.i.d.) with zero mean and finite variance.

In this paper, we propose to modify the standard statistical procedures (tests, confidence intervals, ...) of the linear model in the more general context where the ε_i's are obtained from a strictly stationary process $(\varepsilon_i)_{i\in\mathbb{N}}$ with short memory. To be more precise, let $\hat{\beta}$ denote the usual least squares estimator of β. Our approach is based on two papers: the paper by Hannan [47] who proved the asymptotic normality of the least squares estimator $D(n)(\hat{\beta} - \beta)$ ($D(n)$ being the usual normalization) under very

mild conditions on the design and on the error process; and a recent paper by Caron [24] who showed that, under Hannan's conditions, the asymptotic covariance matrix of $D(n)(\hat{\beta} - \beta)$ can be consistently estimated.

Let us emphasize that Hannan's conditions on the error process are very mild and are satisfied for most of short-memory processes (see the discussion in Section 4.4 of Caron and Dede [25]). Putting together the two above results, we can develop a general methodology for tests and confidence regions on the parameter β, which should be valid for most of short-memory processes. This is of course directly useful for time-series regression (we shall present in Section 3.5.1 an application to the "Mona Loa" R data-set on CO2 concentration), but also in the more general context where the residuals of the linear model seem to be strongly correlated. More precisely, when checking the residuals of the linear model, if the autocorrelation function of the residuals shows significant correlations, and if the residuals can be suitably modeled by an ARMA process, then our methodology is likely to apply. We shall give an example of such a situation in Section 3.5.2 (Shangai pollution data-set).

Hence, the tools presented in the present paper can be seen from two different points of view:

- as appropriate tools for time series regression with short memory error process.
- as a way to robustify the usual statistical procedures when the residuals are correlated.

Let us now describe the organisation of the paper. In Section 3.2, we recall the mathematical background, the consistent estimator of the asymptotic covariance matrix introduced in Caron [24] and the modified Z-statistics and χ-square statistics for testing hypothesis on the parameter β. In Section 3.3 we present the `slm` package, and the different ways to estimate the asymptotic covariance matrix: by fitting an autoregressive process on the residuals (default procedure), by means of the kernel estimator described in Caron [24] (theoretically valid) with a bootstrap method to choose the bandwidth (Wu and Pourahmadi [77]), by using an alternative choice of the bandwidth for the rectangular kernel (Efromovich [41]), by means of an adaptative estimator of the spectral density via Histograms (Comte [32]). In Section 3.4, we estimate the level of a χ-square test for a linear model with random design, with different kind of error processes and for different estimation procedures. In Section 3.5, we present two different data sets "CO2 concentration", "Shangai pollution", and we compare the summary output of `slm` with the usual summary output of `lm`.

3.2 Linear regression with stationary errors

3.2.1 Asymptotic results for the kernel estimator

We start this section by giving a short presentation of linear regression with stationary errors, more details can be found for instance in Caron [24]. Let $\hat{\beta}$ be the usual least squares estimator for the unknown vector β. The aim is to provide hypothesis tests and confidence regions for β in the non i.i.d. context.

Let γ be the autocovariance function of the error process ε: for any integers k and m, let $\gamma(k) = \mathrm{Cov}(\varepsilon_m, \varepsilon_{m+k})$. We also introduce the covariance matrix

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

Hannan [47] has shown a Central Limit Theorem for $\hat{\beta}$ when the error process is strictly stationary, under very mild conditions on the design and the error process. Let us notice that the design can be random or deterministic. We introduce the normalization matrix $D(n)$ which is a diagonal matrix with diagonal term $d_j(n) = \|X_{.,j}\|_2$ for j in $\{1, \ldots, p\}$, where $X_{.,j}$ is the jth column of X. Roughly speaking, Hannan's result says in particular that, given the design X, the vector $D(n)(\hat{\beta} - \beta)$ converges in distribution to a centered Gaussian distribution with covariance matrix C. As usual, in practice the covariance matrix C is unknown and it has to be estimated. Hannan also showed the convergence of second order moment: [1]

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

where

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

In this paper we propose a general plug-in approach: for some given estimator $\widehat{\Gamma}_n$ of Γ_n, we introduce the plug-in 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 standardize the usual statistics considered for the study of linear regression.

Let us illustrate this plug-in approach with a kernel estimator which has been proposed in Caron [24]. For some K and a bandwidth h, the kernel estimator $\widehat{\Gamma}^*_{n,h}$ is defined by

$$\widehat{\Gamma}^*_{n,h} = \left[K\left(\frac{j-l}{h}\right)\hat{\gamma}^*_{j-l}\right]_{1\le j,l\le n}, \tag{3.1}$$

where the residual based empirical covariance coefficients are defined for $0 \le |k| \le n-1$ by

$$\hat{\gamma}^*_k = \frac{1}{n}\sum_{j=1}^{n-|k|} \hat{\varepsilon}_j \hat{\varepsilon}_{j+|k|}. \tag{3.2}$$

For a well-chosen kernel K and under mild assumptions on the design and the error process, it has been proved in Caron [24] that

$$(\widehat{C}^*_n)^{-1/2} D(n)(\hat{\beta} - \beta) \xrightarrow[n\to\infty]{\mathcal{L}} \mathcal{N}_p(0_p, I_p),$$

for the plug-in estimator $\widehat{C}^*_n := \widehat{C}(\widehat{\Gamma}^*_{n,h_n})$, for some suitable sequence of bandwidths (h_n).

More generally, in this paper we say that an estimator $\widehat{\Gamma}_n$ of Γ_n is *consistent for estimating the covariance matrix* C if $\widehat{C}(\widehat{\Gamma}_n)$ is positive definite and if it converges in probability to C. Note that such a property requires assumptions on the design, see Caron [24]. If $\widehat{C}(\widehat{\Gamma}_n)$ is consistent for estimating the covariance matrix C, then $\widehat{C}(\widehat{\Gamma}_n)^{-1/2} D(n)(\hat{\beta} - \beta)$ converges in distribution to a standard Gaussian vector.

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

3.2.2 Tests and confidence regions

We now present tests and confidence regions for arbitrary estimators $\widehat{\Gamma}_n$. The complete justifications are available for kernel estimators, see Caron [24].

Z-Statistics. We introduce the following univariate statistics:

$$Z_j = \frac{d_j(n)\hat{\beta}_j}{\sqrt{\widehat{C}_{(j,j)}}}, \tag{3.3}$$

where $\widehat{C} = \widehat{C}(\widehat{\Gamma}_n)$. If $\widehat{\Gamma}_n$ is consistent for estimating the covariance matrix C and if $\beta_j = 0$, the distribution of Z_j converges to a standard normal distribution when n tends to infinity. We directly derive an asymptotic hypothesis test for testing $\beta_j = 0$ against $\beta_j \neq 0$ as well as an asymptotic confidence interval for β_j.

Chi-square statistics. Let A be a $n \times k$ matrix with $\mathrm{rank}(A) = k$. Under Hannan's conditions, $D(n)(A\hat{\beta} - A\beta)$ converges in distribution to a centered Gaussian distribution with covariance matrix ACA^t. If $\widehat{\Gamma}_n$ is consistent for estimating the covariance matrix C, then $A\widehat{C}(\widehat{\Gamma}_n)$ converges in probability to AC. The matrix $A\widehat{C}(\widehat{\Gamma}_n)A^t$ being symmetric positive definite, this yields

$$W := (A\widehat{C}(\widehat{\Gamma}_n))^{-1/2} D(n) A(\hat{\beta} - \beta) \xrightarrow[n\to\infty]{\mathcal{L}} \mathcal{N}_k(0_k, I_k).$$

This last result provides asymptotical confidence regions for the vector $A\beta$. It also provides an asymptotic test for testing the hypothesis $H_0 : A\beta = 0$ against $H_1 : A\beta \neq 0$. Indeed, under the H_0-hypothesis, the distribution of $\|W\|_2^2$ converges to a $\chi^2(k)$-distribution.

The test can be used to simplify a linear model by testing that several linear combinations between the parameters β_j are zero, as we usually do for Anova and regression models. In particular, with $A = I_p$, the test corresponds to the test of overall significance.

3.3 Introduction to linear regression with the slm package

Using the `slm` package is very intuitive because the arguments and the outputs of `slm` are similar to those of the standard functions `lm`, `glm`, etc. The output of the main function `slm` is an object of class `slm`, a specific class that has been defined for linear regression with stationary processes. The `slm` class has methods `plot`, `summary`, `confint` and `predict`. Moreover, the class `slm` inherits from the `lm` class and thus provides the output of the classical `lm` function.

```
R> library(slm)
```

The statistical tools available in `slm` strongly depend on the choice of the covariance plug-in estimator $\widehat{C}(\widehat{\Gamma}_n)$ we use for estimating C. All the estimators $\widehat{\Gamma}_n$ proposed in `slm` are residual-based estimators, but

they rely on different approaches. In this section, we present the main functionality of `slm` together with the different covariance plug-in estimators.

For illustrating the package, we simulate synthetic data according to the linear model:

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

where Z is a gaussian autoregressive process of order 1, and ε is the Nonmixing process described in Section 3.4.1. We use the functions `generative_model` and `generative_process` respectively to simulate observations according to this regression design and with this specific stationary process. More details on the designs and the processes available with `generative_model` and `generative_process` are given in Section 3.4.1.

```
R> n = 500
R> eps = generative_process(n,"Nonmixing")
R> design = generative_model(n,"mod2")
R> design_sim = cbind(rep(1,n), as.matrix(design))
R> beta_vec = c(2,0.001,0.5)
R> Y = design_sim %*% beta_vec + eps
```

3.3.1 Linear regression via AR fitting on the residuals

A large class of stationary processes with continuous spectral density can be well approximated by AR processes, see for instance Corollary 4.4.2 in Brockwell and Davis [23]. The covariance structure of an AR process having a closed form, it is thus easy to derive an approximation $\widetilde{\Gamma}_{AR(p)}$ of Γ_n by fitting an AR process on the residual process.

The AR-based method for estimating C is the default version of `slm`. This method proceeds in four main steps:

1. Fit an autoregressive process on the residual process $\hat{\varepsilon}$;
2. Compute the theoretical covariances of the fitted AR process ;
3. Plug the covariances in the Toeplitz matrix $\widetilde{\Gamma}_{AR(p)}$;
4. Compute $\widehat{C} = \widehat{C}(\widetilde{\Gamma}_{AR(p)})$.

The `slm` function fits a linear regression of the vector Y on the design X and then fits an AR process on the residual process using the `ar` function from the `stats` package. The output of the `slm` function is an object of class `slm`. The order p of the AR process is set in the argument `model_selec`:

```
R> regslm = slm(Y ~ X1+X2, data = design, method_cov_st = "fitAR",
+          model_selec = 3)
```

The estimated covariance is recorded as a vector in the attribute `cov_st` of `regslm`, which is an object of class `slm`. The estimated covariance matrix can be computed by taking the Toeplitz matrix of `cov_st`, using the `toeplitz` function.

Summary method

As for `lm` objects, a summary of a `slm` object is given by

```
R> summary(regslm)

Call:
"slm(formula = myformula, data = data, x = x, y = y)"

Residuals:
     Min       1Q   Median       3Q      Max
-13.9086  -3.4586   0.1646   3.5025  13.7488

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept) 2.936183   0.855214   3.433 0.000596 ***
X1          0.084387   0.082371   1.024 0.305613
X2          0.492590   0.002738 179.938  < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 4.907
Multiple R-squared:  0.9953
chi2-statistic: 3.278e+04 on 2 DF,  p-value: < 2.2e-16
```

The coefficient table output by the summary provides the estimators of the β_j's, which are exactly the classical least squares estimators. The `z value` column provides the values of the Z_j statistics defined by (3.3). The `Std.Error` column gives an estimation of the standard errors of the $\hat{\beta}_j$'s, which are taken equal to $\frac{\sqrt{\widehat{C}_{(j,j)}}}{d_j(n)}$. As with the `lm` function, the p-value column is the p-value for testing $\beta_j = 0$ against $\beta_j \neq 0$. In this example, the small p-value for the second feature $X2$ is consistent with the value chosen for `beta_vec` at the beginning of the section. The `chi2-statistic` at the end of the summary is the χ^2 statistic for testing the significance of the model (see the end of Section 3.2.2). For this example, the p-value is very small, indeed the variable $X2$ has a significant effect on Y.

Plot argument and plot method

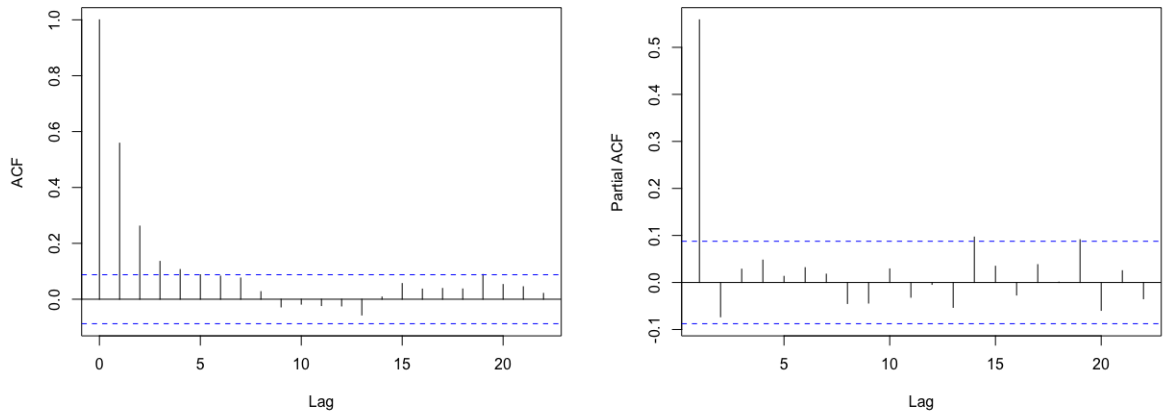
The `slm` function has a `plot` argument: with `plot=TRUE`, the function plots a figure which depends on the method chosen for estimating the covariance matrix C. Table 3.1 summarizes the plots for each method given in the argument `method_cov_st`. With the AR fitting method, the argument `plot=TRUE` outputs the ACF and the PACF of the residual process. The ACF and PACF are computed with the functions `acf` and `pacf` of the `stats` package. As usual, the ACF and PACF graphs should help the user to choose an appropriate order for the AR process.

method_cov_st=	plot
fitAR	ACF and PACF of the residual process
kernel	ACF of the residual process
kernel with model_selec = -1	Graph of the estimated risk and of the estimated $\gamma(k)$'s
spectralproj	Estimated spectral density
select	ACF of the residuals up to the selected order
efromovich	ACF of the residuals up to the selected order

Table 3.1 – Plot output for each method given in the `method_cov_st` of `slm`.

```
R> regslm = slm(Y ~ X1 + X2, data = design, method_cov_st = "fitAR",
+          model_selec = 2, plot = TRUE)
```

The plot output by the `slm` function for this example is given in Figure 3.1.



(a) ACF of the residual process.

(b) PACF of the residual process.

Figure 3.1 – Plots output by `slm` for the fitAR method.

Since the `slm` class inherits from the `lm` class, the former class comes with a `plot` method which is the same as for the `lm` class, namely the diagnostic analysis of the linear regression. The graphics are displayed using the command

```
R> plot(regslm)
```

Confidence intervals for the coefficients

The `confint` function computes the confidence intervals for the coefficients β_j estimated by `slm`. These intervals are computed according to the distribution of the Z_j statistics defined in (3.3).

```
R> confint(regslm, level = 0.90)
```

```
                    5 %      95 %
(Intercept)  1.56396552 4.3083996
X1          -0.05048587 0.2192597
X2           0.48821351 0.4969666
```

AR order selection

The order p of the AR process can be chosen at hand by setting `model_selec = p`, or automatically with the AIC criterion by setting `model_selec = -1`.

```
R> regslm = slm(Y ~ X1 + X2, data = design, method_cov_st = "fitAR",
+           model_selec = -1)
```

The order of the fitted AR process is recorded in the `model_selec` attribute of `regslm`:

```
R> regslm@model_selec
```

```
[1] 2
```

Here, the AIC criterion suggests to fit an AR(2) process on the residuals.

3.3.2 Linear regression via kernel estimation of the error covariance

The second method for estimating the covariance matrix C is the kernel estimation method (3.1) studied in Caron [24]. In short, this method estimates C via a smooth approximation of the covariance matrix Γ_n of the residuals. This estimation of Γ_n corresponds to the so-called tapered covariance matrix estimator in the literature, see for instance Xiao and Wu [78], or also to the "lag-window estimator" defined in Brockwell and Davis [23], page 330. It applies in particular for non negative symmetric kernels with compact support, with an integrable Fourier transform and such that $K(0) = 1$. Table 3.2 gives the list of the available kernels in the package `slm`.

`kernel_fonc =`	kernel definition
`rectangular`	$K(x) = \mathbb{1}_{\{\|x\|\leq 1\}}$
`triangle`	$K(x) = (1-\|x\|)\mathbb{1}_{\{\|x\|\leq 1\}}$
`trapeze`	$K(x) = \mathbb{1}_{\{\|x\|\leq \delta\}} + \frac{1}{1-\delta}(1-\|x\|)\mathbb{1}_{\{\delta\leq\|x\|\leq 1\}}$

Table 3.2 – Available kernel functions in `slm`.

It is also possible for the user to define his own kernel and to use it in the argument `kernel_fonc` of the `slm` function. Below we use the triangle kernel which assures that the covariance matrix is positive definite. The support of the kernel K in Equation (3.1) being compact, only the terms $\hat{\gamma}^*_{j-l}$ for small enough lag $j-l$ are kept and weighted by the kernel in the expression of $\widehat{\Gamma}^*_{n,h}$. Rather than setting the bandwidth h, we select the number of $\gamma(k)$'s that should be kept (the lag) with the argument `model_selec` in the `slm` function. Then the bandwidth h is calibrated accordingly, that is equal to `model_selec` $+ 1$.

```
R> regslm = slm(Y ~ X1 + X2, data = design, method_cov_st = "kernel",
+          model_selec = 5, kernel_fonc = triangle, plot = TRUE)
```

The plot output by the `slm` function is given in Figure 3.2.

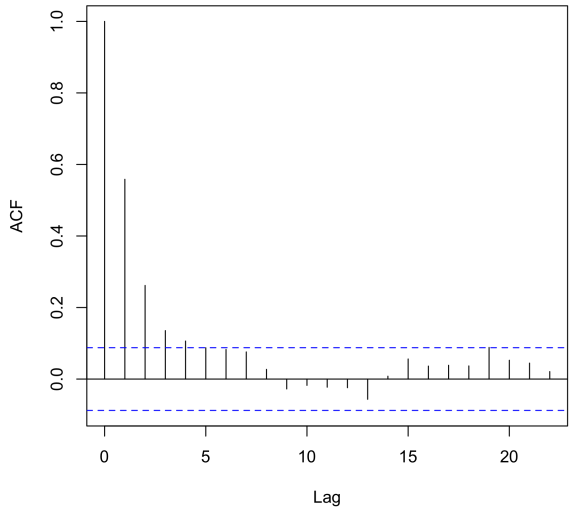


Figure 3.2 – ACF of the residual process.

Order selection via bootstrap

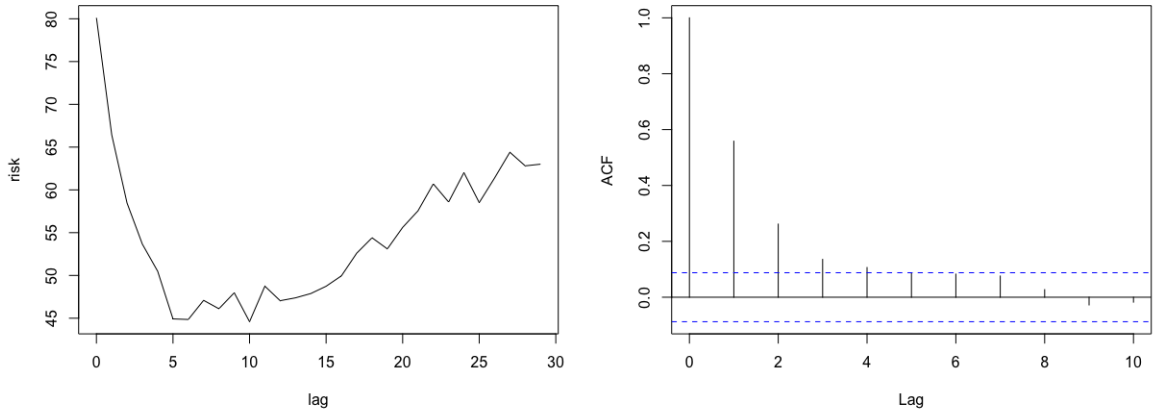
The order parameter can be chosen at hand as before or automatically by setting `model_selec = -1`. The automatic order selection is based on the bootstrap procedure proposed by Wu and Pourahmadi [77] for banded covariance matrix estimation. The `block_size` argument sets the size of bootstrap blocks and the `block_n` argument sets the number of blocks. The final order is chosen by taking the order which has the minimal risk. Figure 3.3 gives the plots of the estimated risk for the estimation of Γ_n (left) and the final estimated ACF (right).

```
R> regslm = slm(Y ~ X1 + X2, data = design, method_cov_st = "kernel",
+          model_selec = -1, kernel_fonc = triangle, model_max = 30,
+          block_size = 100, block_n = 100, plot = TRUE)
```

The selected order is recorded in the `model_selec` attribute of the `slm` object output by the `slm` function:

```
R> regslm@model_selec

[1] 10
```



(a) Estimated risk error via bootstrap.

(b) Estimated ACF for the selected order.

Figure 3.3 – Plots output by `slm` for the kernel method with bootstrap selection of the order.

Order selection by Efromovich's method (rectangular kernel)

An alternative method for choosing the bandwidth in the case of the rectangular kernel has been proposed in Efromovich [41]. For a large class of stationary processes with exponentially decaying autocovariance function (mainly the ARMA processes), Efromovich proved that the rectangular kernel is asymptotically minimax, and he proposed the following estimator:

$$\hat{f}_{J_{nr}}(\lambda) = \frac{1}{2\pi} \sum_{k=-J_{nr}}^{k=J_{nr}} \hat{\gamma}_k e^{ik\lambda},$$

with the lag

$$J_{nr} = \frac{\log(n)}{2r} \left[1 + (\log(n))^{-1/2}\right],$$

where r is a regularity index of the autocovariance index. In practice this parameter is unknown and is estimated thanks to the algorithm proposed in the section 4 of Efromovich [41]. As for the other methods, we use the residual based empirical covariances $\hat{\gamma}_k^*$ to compute $\hat{f}_{J_{nr}}(\lambda)$.

```
R> regslm = slm(Y ~ X1 + X2, data = design, method_cov_st = "efromovich",
+           model_selec = -1)
```

Positive definite projection

Depending of the method used, the matrix $\widehat{C}(\widehat{\Gamma}_n)$ may not always be positive definite. It is the case of the kernel method with rectangular or trapeze kernel. To overcome this problem, we make the projection of $\widehat{C}(\widehat{\Gamma}_n)$ into the cone of positive definite matrices by applying a hard thresholding on the spectrum of this matrix: we replace all eigenvalues lower or equal to zero with the smallest positive eigenvalue of

$\widehat{C}(\widehat{\Gamma}_n)$.

Note that this projection is useless for the triangle kernel because its Fourier transform is non-negative (leading to a positive definite matrix $\widehat{C}(\widehat{\Gamma}_n)$). Of course, it is also useless for the `fitAR` and `spectralproj` methods.

3.3.3 Linear regression via projection spectral estimation

The projection method relies on the ideas of Comte [32], where an adaptive nonparametric method has been proposed for estimating the spectral density of a stationary Gaussian process.

We use the residual process as a proxy for the error process and we compute the projection coefficients with the residual-based empirical covariance coefficients $\hat{\gamma}_k^*$, see Equation (3.2).

For some $d \in \mathbb{N}^*$, the estimator of the spectral density of the error process that we use is defined by computing the projection estimators for the residual process, on the basis of histogram functions

$$\phi_j^{(d)} = \sqrt{\frac{d}{\pi}} \mathbb{1}_{[\pi j/d, \pi(j+1)/d[}, \qquad j = 0, 1, \ldots, d-1.$$

The estimator is defined by

$$\hat{f}_d(\lambda) = \sum_{j=0}^{d-1} \hat{a}_j^{(d)} \phi_j^{(d)},$$

where the projection coefficients are

$$\hat{a}_j^{(d)} = \sqrt{\frac{d}{\pi}} \left(\frac{\hat{\gamma}_0^*}{2d} + \frac{1}{\pi} \sum_{r=1}^{n-1} \frac{\hat{\gamma}_r^*}{r} \left[\sin\left(\frac{\pi(j+1)r}{d} \right) - \sin\left(\frac{\pi j r}{d} \right) \right] \right).$$

The Fourier coefficients of the spectral density are equal to the covariance coefficients. Thus, for $k = 1, \ldots, n-1$ it yields

$$\begin{aligned} \gamma_k &= c_k \\ &= \frac{2}{k} \sqrt{\frac{d}{\pi}} \sum_{j=0}^{d-1} \hat{a}_j^{(d)} \left[\sin\left(\frac{k\pi(j+1)}{d} \right) - \sin\left(\frac{k\pi j}{d} \right) \right], \end{aligned}$$

and for $k = 0$:

$$\begin{aligned} \gamma_0 &= c_0 \\ &= 2\sqrt{\frac{\pi}{d}} \sum_{j=0}^{d-1} \hat{a}_j^{(d)}. \end{aligned}$$

This method can be proceeded in the `slm` function by setting `method_cov_st = "spectralproj"`:

```
R> regslm = slm(Y ~ X1 + X2, data = design, method_cov_st = "spectralproj",
+          model_selec = 10, plot = TRUE)
```

The graph of the estimated spectral density can be plotted by setting `plot = TRUE` in the `slm` function,

see Figure 3.4.

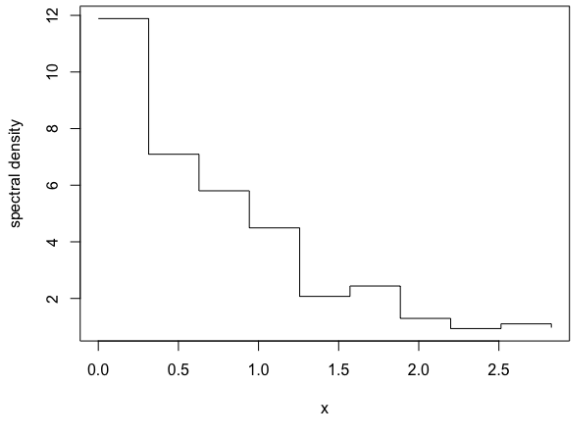


Figure 3.4 – Spectral density estimator by projection on the histogram basis.

Model selection

The Gaussian model selection method proposed in Comte [32] follows the ideas of Birgé and Massart, see for instance Massart [64]. It consists in minimizing the l_2 penalized criterion, see Section 5 in Comte [32]:

$$\text{crit}(d) := -\sum_{j=0}^{d-1} \left[\hat{a}_j^{(d)}\right]^2 + c\frac{d}{n}$$

where c is a multiplicative constant that in practice can be calibrated using the slope heuristic method, see Birgé and Massart [20]; Baudry, Maugis, and Michel [15] and the R package `Capushe`.

```
R> regslm = slm(Y ~ X1 + X2, data = design, method_cov_st = "spectralproj",
+          model_selec = -1, model_max = 50, plot = TRUE)
```

The selected dimension is recorded in the `model_selec` attribute of the `slm` object output by the `slm` function:

```
R> regslm@model_selec
```

```
[1] 8
```

The slope heuristic algorithm here selects an Histogram on a regular partition of size 8 over the interval $[0, \pi]$ to estimate the spectral density.

3.3.4 Linear regression via masked covariance estimation

This method is a full-manual method for estimating the covariance matrix C by only selecting covariance terms from the residual covariances $\hat{\gamma}_k^*$ defined by (3.2). Let I be a set of positive integers, then we consider

$$\hat{\gamma}_I(k) := \hat{\gamma}_k^* \mathbb{1}_{k \in I \cup \{0\}} \qquad 0 \leq |k| \leq n-1$$

and then we define the estimated covariance marix $\widehat{\Gamma}_I$ by taking the Toeplitz matrix of the vector $\hat{\gamma}_I$. This estimator is a particular example of masked sample covariance estimator, as introduced by Chen, Gittens, and Tropp [28], see also Levina and Vershynin [57]. Finally we derive from $\widehat{\Gamma}_I$ an estimator $\widehat{C}(\widehat{\Gamma}_I)$ for C.

The next instruction selects the coefficients 0, 1, 2 and 4 from the residual covariance terms:

```
R> regslm = slm(Y ~ X1 + X2, data = design, method_cov_st = "select",
+           model_selec = c(1,2,4))
```

The positive lags of the selected covariances are recordered in the `model_selec` argument. Let us notice that the variance γ_0 is automatically selected.

As for the kernel method, the resulting covariance matrix may not be positive definite. If it is the case, the positive definite projection method, described at the end of the section 3.3.2, is used.

3.3.5 Linear regression via manual plugged covariance matrix

This last method is a direct plug-in method. The user proposes his own vector estimator $\hat{\gamma}$ of γ and then the Toeplitz matrix $\widehat{\Gamma}_n$ of the vector $\hat{\gamma}$ is used for estimating C with $\widehat{C}(\widehat{\Gamma}_n)$.

```
R> v = rep(0,n)
R> v[1:10] = acf(epsilon, type = "covariance", lag.max = 9)$acf
R> regslm = slm(Y ~ X1 + X2, data = design, cov_st = v)
```

The user can also propose his own covariance matrix $\widehat{\Gamma}_n$ for estimating C.

```
R> v = rep(0,n)
R> v[1:10] = acf(epsilon, type = "covariance", lag.max = 9)$acf
R> V = toeplitz(v)
R> regslm = slm(Y ~ X1 + X2, data = design, Cov_ST = V)
```

Let us notice that the user must verify that the resulting covariance matrix is positive definite. The positive definite projection algorithm is not used with this method.

3.4 Numerical experiments and method comparisons

This section summarizes an extensive study which has been carried out to compare the performances of the different approaches presented before in the context of linear model with short range dependent stationary errors.

3.4.1 Description of the generative models

We first present the five generative models for the errors that we consider in the paper. We choose different kinds of processes to reflect the diversity of short-memory processes.

— **AR1 process.** The AR1 process is a gaussian AR(1) process defined by:

$$\varepsilon_i - 0.7\varepsilon_{i-1} = W_i,$$

where W_i is a standard gaussian distribution $\mathcal{N}(0,1)$.

— **AR12 process.** The AR12 process is a seasonal AR(12) process defined by:

$$\varepsilon_i - 0.5\varepsilon_{i-1} - 0.2\varepsilon_{i-12} = W_i,$$

where W_i is a standard gaussian distribution $\mathcal{N}(0,1)$. When studying monthly data-sets, one usually observes a seasonality of order 12. For example, when looking at climate data (such as the "CO2 concentration" dataset of Section 3.5), the data are often collected per month, and the same phenomenon tends to repeat every year. Even if the design integrates the deterministic part of the seasonality, a correlation of order 12 remains usually present in the residual process.

— **MA12 process.** The MA12 is also a seasonal process defined by:

$$\varepsilon_i = W_i + 0.5W_{i-2} + 0.3W_{i-3} + 0.2W_{i-12},$$

where the (W_i)'s are i.i.d. random variables following Student's distribution with 10 degrees of freedom.

— **Nonmixing process.** The three processes described above are basic ARMA processes, whose innovations have absolutely continuous distributions; in particular, they are strongly mixing in the sense of Rosenblatt [69], with a geometric decay of the mixing coefficients (in fact the MA12 process is even 12-dependent, which means that the mixing coefficient $\alpha(k) = 0$ if $k > 12$). Let us now describe a more complicated process: let $(Z_1, \ldots, Z_n)$ satisfying the AR(1) equation

$$Z_{i+1} = \frac{1}{2}(Z_i + \eta_{i+1}),$$

where Z_1 is uniformly distributed over $[0,1]$ and the η_i's are i.i.d. random variables with distribution $\mathcal{B}(1/2)$, independent of Z_1. The process $(Z_i)_{i\geq 1}$ is a strictly stationary Markov chain, but it is not α-mixing in the sense of Rosenblatt (see Bradley [21]). Let now Q_{0,σ^2} be the inverse of the cumulative distribution function of a centered Gaussian distribution with variance σ^2 (for the simulations below, we choose $\sigma^2 = 25$). The Nonmixing process is then defined by

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

The sequence $(\varepsilon_i)_{i\geq 1}$ is also a stationary Markov chain (as an invertible function of a stationary Markov chain). By construction, ε_i is $\mathcal{N}(0,\sigma^2)$-distributed, but the sequence $(\varepsilon_i)_{i\geq 1}$ is not a Gaussian process (otherwise it would be mixing in the sense of Rosenblatt). Although it is not

obvious, one can prove that the process $(\varepsilon_i)_{i\geq 1}$ satisfies Hannan's condition (see Caron [24], Section 4.2).

— **Sysdyn process.** The four processes described above have the property of "geometric decay of correlations", which means that the $\gamma(k)$'s tend to 0 at an exponential rate. However, as already pointed out in the introduction, Hannan's condition is valid for most of short memory processes, even for processes with slow decay of correlations (provided that the $\gamma(k)$'s are summable). Hence, our last example will be a non-mixing process (in the sense of Rosenblatt), with an arithmetic decay of the correlations.
For $\gamma \in]0,1[$, the intermittent map $\theta_\gamma : [0,1] \mapsto [0,1]$ introduced in Liverani, Saussol and Vaienti [60] is defined by

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

It follows from Liverani, Saussol and Vaienti [60] that there exists a unique θ_γ-invariant probability measure ν_γ. The Sysdyn process is then defined by

$$\varepsilon_i = \theta_\gamma^i.$$

From Liverani, Saussol and Vaienti [60], we know that, on the probability space $([0,1],\nu_\gamma)$, the autocorrelations $\gamma(k)$ of the stationary process $(\varepsilon_i)_{i\geq 1}$ are exactly of order $k^{-(1-\gamma)/\gamma}$. Hence $(\varepsilon_i)_{i\geq 1}$ is a short memory process provided $\gamma \in]0,1/2[$. Moreover, it has been proved in Section 4.4 of Caron and Dede [25] that $(\varepsilon_i)_{i\geq 1}$ satisfies Hannan's condition in the whole short-memory range, that is for $\gamma \in]0,1/2[$. For the simulations below, we took $\gamma = 1/4$, which give autocorrelations $\gamma(k)$ of order k^{-3}.

The linear regression models simulated in the experiments all have the following form:

$$Y_i = \beta_1 + \beta_2(\log(i) + \sin(i) + Z_i) + \beta_3 i + \varepsilon_i, \qquad \text{for all } i \text{ in } \{1,\dots,n\}, \tag{3.4}$$

where Z is a gaussian autoregressive process of order 1 and ε is one of the stationary processes defined above. For the simulations, β_1 is always equal to 3. All the error processes presented above can be simulated with the `slm` package with the `generative_process` function. The design can be simulated with the `generative_model` function.

3.4.2 Automatic calibration of the tests

It is of course of first importance to provide hypothesis tests with correct significance levels or at least with correct asymptotical significance levels, which is possible if the estimator $\widehat{\Gamma}_n$ of the covariance matrix Γ_n is consistent for estimating C. For instance, the results of Caron [24] show that it is possible to construct statistical tests with correct asymptotical significance levels. However in practice such asymptotical results are not sufficient since they do not indicate how to tune the bandwidth on a given dataset. This situation makes the practice of linear regression with dependent errors really more difficult than linear regression with i.i.d. errors. This problem happens for several methods given before:

order choice for the `fitAR` method, bandwidth choice for the `kernel` method, dimension selection for the `spectralproj` method.

It is a tricky issue to design a data driven procedure for choosing test parameters in order to have to correct Type I Error. Note that unlike with supervised problems and density estimation, it is not possible to calibrate hypothesis tests in practice using cross validation approaches. We thus propose to calibrate the tests using well founded statistical procedures for risk minimization : AIC criterion for the `fitAR` method, bootstrap procedures for the `kernel` method and slope heuristics for the `spectralproj` method. These procedures are implemented in the `slm` function with the `model_selec = -1` argument, as detailed in the previous section.

Let us first illustrate the calibration problem with the AR12 process. For $T = 1000$ simulations, we generate an error process of size n under the null hypothesis: $H_0 : \beta_2 = \beta_3 = 0$. Then we use the `fitAR` method of the `slm` function with orders between 1 and 50 and we perform the model significance test. The procedure is repeated 1000 times and we estimate the true level of the test by taking the average of the estimated levels on the 1000 simulations for each order. The results are given on Figure 3.5 for $n = 1000$. A boxplot is also displayed to visualize the distribution of the order selected by the automatic criterion (AIC).

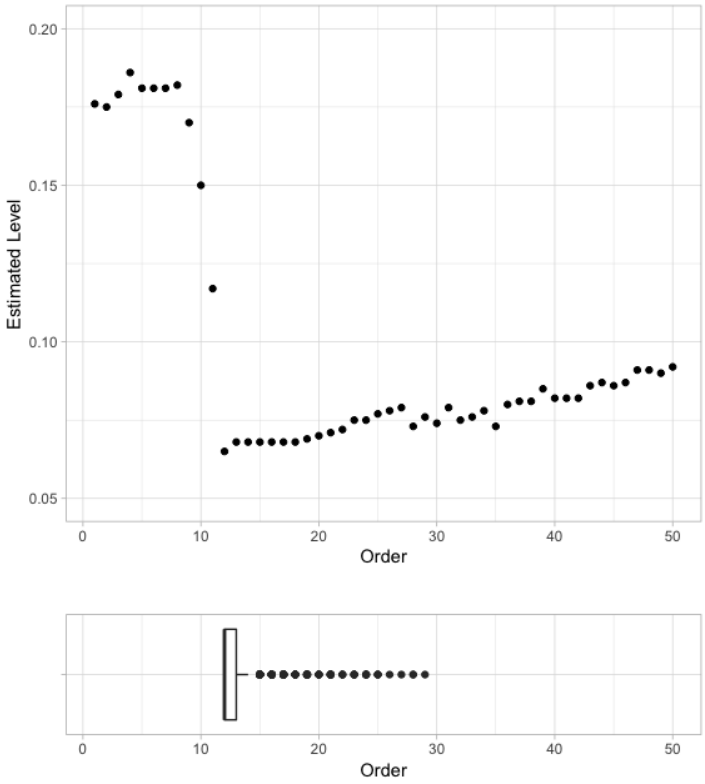


Figure 3.5 – Estimated level of the test according to the order of the fitted AR process on the residuals (top) and boxplot of the order selected by AIC, over 1000 simulations. The data has been simulated according to Model (3.4) with $\beta_1 = 3$ and $\beta_2 = \beta_3 = 0$, with $n = 1000$.

3.4.3 Non-Seasonal errors

We first study the case of non-Seasonal error processes. We simulate a n-error process according to the AR1, the Nonmixing or the Sysdyn processes. We simulate realizations of the linear regression model (3.4) under the null hypothesis: $H_0 : \beta_2 = \beta_3 = 0$. We use the automatic selection procedures for each method (`model_selec = -1`). The simulations are repeated 1000 times in order to estimate the true level of the model significance for each test procedure. We simulate either small samples ($n = 200$) or larger samples ($n = 1000, 2000, 5000$). The results of this experiments are summarized in Table 3.3.

n	Method / Process	Fisher i.i.d.	fitAR	spectralproj	efromovich	kernel
200	AR1 process	0.465	**0.097**	0.14	0.135	0.149
	NonMixing	0.298	**0.082**	0.103	0.096	0.125
	Sysdyn process	0.385	**0.105**	0.118	0.124	0.162
1000	AR1 process	0.418	0.043	**0.049**	**0.049**	0.086
	NonMixing	0.298	0.046	**0.05**	0.053	0.076
	Sysdyn process	0.393	**0.073**	0.077	0.079	0.074
2000	AR1 process	0.454	0.071	0.078	0.075	**0.067**
	NonMixing	0.313	**0.051**	0.053	0.057	0.067
	Sysdyn process	0.355	**0.063**	0.064	0.066	0.069
5000	AR1 process	0.439	0.044	**0.047**	**0.047**	**0.047**
	NonMixing	0.315	**0.053**	0.056	0.059	0.068
	Sysdyn process	0.381	0.058	0.061	**0.057**	0.064

Table 3.3 – Estimated levels for the non-seasonal processes.

For n large enough ($n \geq 1000$), all methods work well and the estimated level is around 0.05. However, for small samples ($n = 200$), it is more complicated. We can observe that the `fitAR` method works better than the others. The `kernel` method is slightly less effective. With this method, we must choose the size of the bootstrap blocks as well as the number of blocks and the test results are really sensitive to these parameters. In these simulations, we have chosen 100 blocks with a size of $n/2$. The results are expected to improve with a larger number of blocks.

Let us notice that for all methods and for all sample sizes, the estimated level is much better than if no correction is made (usual Fisher tests).

3.4.4 Seasonal errors

We now study the case of linear regression with seasonal errors. The experiment is exactly the same as before, except that we simulate AR12 or MA12 processes. The results of these experiments are summarized in Table 3.4.

We directly see that the case of seasonal processes is more complicated than for the non-seasonal processes especially for the AR12 process. For small samples size, the estimated level is between 0.17 and 0.24, which is clearly too large. It is however much better than the estimated level of the usual Fisher test, which is around 0.45. The `fitAR` method is the best method here for the AR12 process, because for $n \geq 1000$, the estimated level is between 0.06 and 0.07. For `efromovich` and `kernel` methods, a level

n	Method / Process	Fisher i.i.d.	fitAR	spectralproj	efromovich	kernel
200	AR12 process	0.436	**0.178**	0.203	0.223	0.234
	MA12 process	0.228	**0.113**	**0.113**	0.116	0.15
1000	AR12 process	0.468	**0.068**	0.183	0.181	0.124
	MA12 process	0.209	0.064	0.066	0.069	**0.063**
2000	AR12 process	0.507	**0.071**	0.196	0.153	0.104
	MA12 process	0.237	0.064	0.064	**0.058**	0.068
5000	AR12 process	0.47	**0.062**	0.183	0.1	0.091
	MA12 process	0.242	0.044	**0.048**	0.043	0.057

Table 3.4 – Estimated levels for the seasonal processes.

less than 0.10 is reached but for large samples only. The `spectralproj` method does not seem to work well for the AR12 process, although it remains much better than the usual Fisher tests (around 19% of rejection instead of 45%).

The case of the MA12 process seems easier to deal with. For n large enough ($n \geq 1000$), the estimated level is between 0.04 and 0.07 whatever the method. It is less effective for small sample size ($n = 200$), with an estimated level around 0.115 for `fitAR`, `spectralproj` and `efromovich` methods.

3.5 Application to real data

3.5.1 Data CO2

Let us introduce the first dataset that we want to study. It concerns the well-known dataset "co2", available in the package `datasets` of R:

```
R> data("co2")
```

This dataset is provided by the observatory of Mona Loa (Hawaii). It contains average monthly measurements of CO2 (parts per million: ppmv) in the atmosphere of the Hawaiian coast. Surveys were produced monthly between 1959 and 1998, giving a total of 468 measurements. The graph of the data is displayed in Figure 3.6. More information on this dataset is available in the R documentation.

We model the CO2 measurements with a time series. Typically, a time series can be decomposed into three parts: a trend m and a seasonality s, which are deterministic components, and the errors ε, which constitute the random part of the model. The trend represents the overall behavior of the series and seasonality its periodic behavior. Formally, we have:

$$Y_t = m_t + s_t + \varepsilon_t,$$

where Y_t represents the CO2 rate at time t, with the usual constraints $s_t = s_{t+12}$ and $\sum_{t=1}^{12} s_t = 0$. The two deterministic components can be grouped into a matrix X and the model can be rewritten into a linear regression model:

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

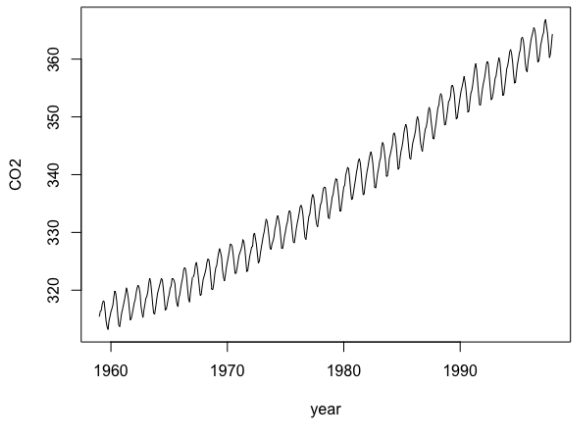


Figure 3.6 – CO2 rate as a function of time.

For this example, we fit a 3-degree polynomial for the trend and a trigonometric polynomial with well-chosen frequencies for the seasonality. Here the time t represents a month and t goes from 1 to 40 by step of length $1/12$. Let us perform a linear regression to fit the trend and the seasonality on the CO2 time series, using the `lm` function:

```
R> y = as.vector(co2)
R> t = as.vector(time(co2)) - 1958
R> regtrigo = lm(y ~ t + I(t^2) + I(t^3) + sin(2*pi*t) + cos(2*pi*t)
+           + sin(4*pi*t) + cos(4*pi*t) + sin(6*pi*t) + cos(6*pi*t)
+           + sin(8*pi*t) + cos(8*pi*t))
```

We obtain the following output:

```
R> summary.lm(regtrigo)

Call:
lm(formula = y ~ t + I(t^2) + I(t^3) + sin(2 * pi * t) + cos(2 *
    pi * t) + sin(4 * pi * t) + cos(4 * pi * t) + sin(6 * pi *
    t) + cos(6 * pi * t) + sin(8 * pi * t) + cos(8 * pi * t))

Residuals:
     Min       1Q   Median       3Q      Max
-1.54750 -0.32688  0.00233  0.28100  1.50295

Coefficients:
                  Estimate Std. Error  t value Pr(>|t|)
(Intercept)      3.157e+02  1.118e-01 2823.332  < 2e-16 ***
```

```
t                3.194e-01  2.306e-02   13.847  < 2e-16 ***
I(t^2)           4.077e-02  1.293e-03   31.523  < 2e-16 ***
I(t^3)          -4.562e-04  2.080e-05  -21.930  < 2e-16 ***
sin(2 * pi * t)  2.751e+00  3.298e-02   83.426  < 2e-16 ***
cos(2 * pi * t) -3.960e-01  3.296e-02  -12.015  < 2e-16 ***
sin(4 * pi * t) -6.743e-01  3.296e-02  -20.459  < 2e-16 ***
cos(4 * pi * t)  3.785e-01  3.296e-02   11.484  < 2e-16 ***
sin(6 * pi * t) -1.042e-01  3.296e-02   -3.161  0.00168 **
cos(6 * pi * t) -4.389e-02  3.296e-02   -1.332  0.18362
sin(8 * pi * t)  8.733e-02  3.296e-02    2.650  0.00833 **
cos(8 * pi * t)  2.559e-03  3.296e-02    0.078  0.93814
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.5041 on 456 degrees of freedom
Multiple R-squared:  0.9989,        Adjusted R-squared:  0.9989
F-statistic: 3.738e+04 on 11 and 456 DF,  p-value: < 2.2e-16
```

We see in the summary that two variables have no significant effect on the CO2 rate. Next, we perform a backward selection method with a p-value threshold equal to 0.05. This selects the following model:

```
R> regtrigo = lm(y ~ t + I(t^2) + I(t^3) + sin(2*pi*t) + cos(2*pi*t)
+          + sin(4*pi*t) + cos(4*pi*t) + sin(6*pi*t) + sin(8*pi*t))
```

with the corresponding summary

```
R> summary.lm(regtrigo)

Call:
lm(formula = y ~ t + I(t^2) + I(t^3) + sin(2 * pi * t) + cos(2 *
    pi * t) + sin(4 * pi * t) + cos(4 * pi * t) + sin(6 * pi *
    t) + sin(8 * pi * t))

Residuals:
     Min       1Q   Median       3Q      Max
-1.59287 -0.32364  0.00226  0.29884  1.50154

Coefficients:
                  Estimate Std. Error  t value Pr(>|t|)
(Intercept)      3.157e+02  1.118e-01 2824.174  < 2e-16 ***
t                3.196e-01  2.306e-02   13.861  < 2e-16 ***
I(t^2)           4.075e-02  1.293e-03   31.522  < 2e-16 ***
I(t^3)          -4.560e-04  2.080e-05  -21.927  < 2e-16 ***
sin(2 * pi * t)  2.751e+00  3.297e-02   83.446  < 2e-16 ***
```

```
cos(2 * pi * t) -3.960e-01  3.295e-02  -12.018  < 2e-16 ***
sin(4 * pi * t) -6.743e-01  3.295e-02  -20.464  < 2e-16 ***
cos(4 * pi * t)  3.785e-01  3.295e-02   11.487  < 2e-16 ***
sin(6 * pi * t) -1.042e-01  3.295e-02   -3.162  0.00167 **
sin(8 * pi * t)  8.734e-02  3.295e-02    2.651  0.00831 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.504 on 458 degrees of freedom
Multiple R-squared:  0.9989,        Adjusted R-squared:  0.9989
F-statistic: 4.57e+04 on 9 and 458 DF,  p-value: < 2.2e-16
```

The sum of the estimated trend and estimated tendency is displayed on the left plot of Figure 3.7, and the residuals are displayed on the right plot. The `lm` procedure assumes that the errors are independent,

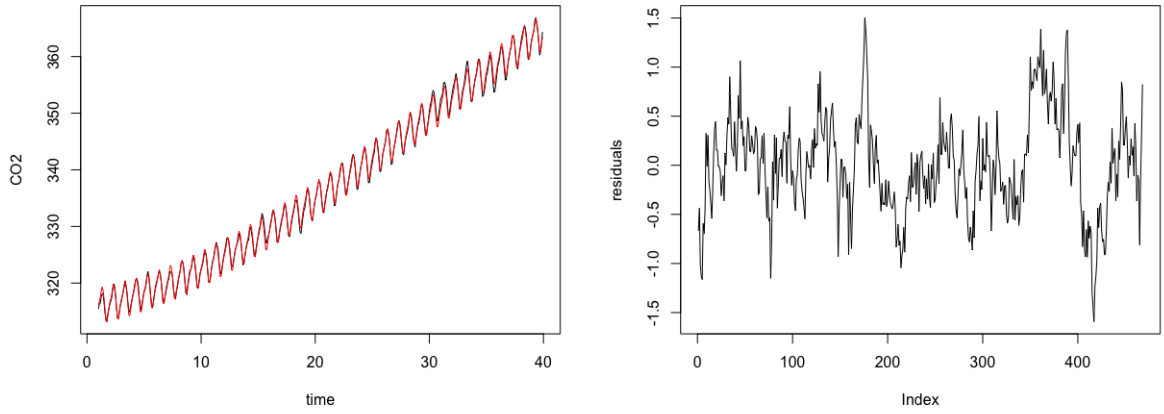


Figure 3.7 – CO2 adjustment (left) and residuals (right).

but if we look at the autocorrelation function of the residual process we clearly observe that the residuals are strongly correlated, see Figure 3.8. Consequently, the `lm` procedure may be unreliable in this context.

The autocorrelation function of the residuals decreases rather fast. Looking at the partial autocorrelation function, it seems reasonable to fit an AR process on the residuals. The automatic `fitAR` method selects an AR of order 14 and the residuals look like a white noise, see Figure 3.9.

We now use the `slm` function with the `fitAR` method with the following complete model

```
R> regtrigo = slm(y ~ t + I(t^2) + I(t^3) + sin(2*pi*t) + cos(2*pi*t)
+           + sin(4*pi*t) + cos(4*pi*t) + sin(6*pi*t) + cos(6*pi*t)
+           + sin(8*pi*t) + cos(8*pi*t), method_cov_st = "fitAR",
+           model_selec = -1)
```

Let us display the summary of the procedure:

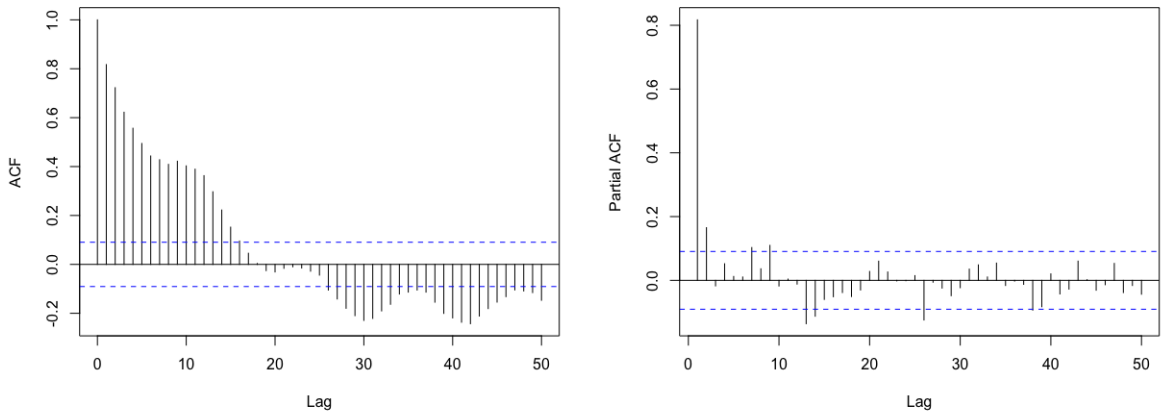


Figure 3.8 – Autocorrelation function (left) and partial autocorrelation function (right) of the residuals.

```
R> summary(regtrigo)

Call:
"slm(formula = myformula, data = data, x = x, y = y)"

Residuals:
     Min       1Q   Median       3Q      Max
-1.54750 -0.32688  0.00233  0.28100  1.50295

Coefficients:
                  Estimate Std. Error z value Pr(>|z|)
(Intercept)      3.157e+02  3.968e-01 795.646  < 2e-16 ***
t                3.194e-01  8.222e-02   3.884 0.000103 ***
I(t^2)           4.077e-02  4.619e-03   8.825  < 2e-16 ***
I(t^3)          -4.562e-04  7.430e-05  -6.140 8.23e-10 ***
sin(2 * pi * t)  2.751e+00  4.739e-02  58.054  < 2e-16 ***
cos(2 * pi * t) -3.960e-01  4.716e-02  -8.396  < 2e-16 ***
sin(4 * pi * t) -6.743e-01  2.051e-02 -32.875  < 2e-16 ***
cos(4 * pi * t)  3.785e-01  2.041e-02  18.548  < 2e-16 ***
sin(6 * pi * t) -1.042e-01  1.359e-02  -7.663 1.82e-14 ***
cos(6 * pi * t) -4.389e-02  1.359e-02  -3.228 0.001245 **
sin(8 * pi * t)  8.733e-02  1.246e-02   7.009 2.41e-12 ***
cos(8 * pi * t)  2.559e-03  1.252e-02   0.204 0.838038
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

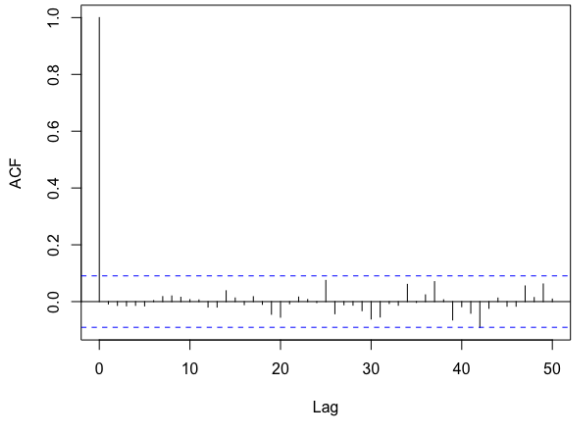


Figure 3.9 – Autocorrelation function of the residuals for the AR fitting.

```
Residual standard error: 0.5041
Multiple R-squared:  0.9989
chi2-statistic: 3.598e+04 on 11 DF,  p-value: < 2.2e-16
```

The last variable has no significant effect on the CO2. After performing a backward selection method with a p-value threshold equal to 0.05, we obtain the following model

```
R> regtrigo = slm(y ~ t + I(t^2) + I(t^3) + sin(2*pi*t) + cos(2*pi*t)
+           + sin(4*pi*t) + cos(4*pi*t) + sin(6*pi*t) + cos(6*pi*t)
+           + sin(8*pi*t), method_cov_st = "fitAR", model_selec = -1)
```

and the associated summary

```
R> summary(regtrigo)

Call:
"slm(formula = myformula, data = data, x = x, y = y)"

Residuals:
     Min       1Q   Median       3Q      Max
-1.54877 -0.32432  0.00187  0.28069  1.50168

Coefficients:
                  Estimate Std. Error z value Pr(>|z|)
(Intercept)      3.157e+02  3.969e-01 795.627  < 2e-16 ***
t                3.194e-01  8.223e-02   3.884 0.000103 ***
I(t^2)           4.077e-02  4.619e-03   8.825  < 2e-16 ***
```

```
I(t^3)          -4.562e-04  7.430e-05  -6.140 8.23e-10 ***
sin(2 * pi * t)  2.751e+00  4.738e-02  58.061  < 2e-16 ***
cos(2 * pi * t) -3.960e-01  4.716e-02  -8.397  < 2e-16 ***
sin(4 * pi * t) -6.743e-01  2.051e-02 -32.874  < 2e-16 ***
cos(4 * pi * t)  3.785e-01  2.041e-02  18.547  < 2e-16 ***
sin(6 * pi * t) -1.042e-01  1.359e-02  -7.664 1.80e-14 ***
cos(6 * pi * t) -4.389e-02  1.359e-02  -3.229 0.001244 **
sin(8 * pi * t)  8.733e-02  1.248e-02   6.998 2.60e-12 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.5036
Multiple R-squared:  0.9989
chi2-statistic: 3.596e+04 on 10 DF,  p-value: < 2.2e-16
```

There is a clear difference between the two backward procedures: `slm` keeps the variable $\cos(6\pi x)$, while `lm` does not. Given the obvious dependency of the error process, we recommend using `slm` instead of `lm` in this context.

3.5.2 PM2.5 Data of Shanghai

This dataset comes from a study about fine particle pollution in five Chinese cities. The data are available on the following website `https://archive.ics.uci.edu/ml/datasets/PM2.5+Data+of+Five+Chinese+Cities#`. We are interested here by the city of Shanghai. We study the regression of PM2.5 pollution in Xuhui District by other measurements of pollution in neighboring districts and also by meteorological variables. The dataset contains hourly observations between January 2010 and December 2015. More precisely it contains 52584 records of 17 variables: date, time of measurement, pollution and meteorological variables. More information on these data is available in the paper of Liang, Li, Zhang, Huang, and Chen [58].

We remove the lines that contain NA observations and we then extract the first 5000 observations. For simplicity, we will only consider pollution variables and weather variables. We start the study with the following 10 variables:

- PM_Xuhui: PM2.5 concentration in the Xuhui district (ug/m^3)
- PM_Jingan: PM2.5 concentration in the Jing'an district (ug/m^3)
- PM_US.Post: PM2.5 concentration in the U.S diplomatic post (ug/m^3)
- DEWP: Dew Point (Celsius Degree)
- TEMP: Temperature (Celsius Degree)
- HUMI: Humidity ($\%$)
- PRES: Pressure (hPa)
- Iws: Cumulated wind speed (m/s)
- precipitation: hourly precipitation (mm)
- Iprec: Cumulated precipitation (mm)

```
R> shan = read.csv("ShanghaiPM20100101_20151231.csv", header = TRUE,
+          sep = ",")
R> shan = na.omit(shan)
R> shan_complete = shan[1:5000,c(7,8,9,10,11,12,13,15,16,17)]
R> shan_complete[1:5,]
```

```
      PM_Jingan PM_US.Post PM_Xuhui DEWP  HUMI PRES TEMP Iws
26305        66         70       71   -5 69.00 1023    0  60
26306        67         76       72   -5 69.00 1023    0  62
26308        73         78       74   -4 74.41 1023    0  65
26309        75         77       77   -4 80.04 1023   -1  68
26310        73         78       80   -4 80.04 1023   -1  70
      precipitation Iprec
26305             0     0
26306             0     0
26308             0     0
26309             0     0
26310             0     0
```

The aim is to study the concentration of particles in Xuhui District according to the other variables. We first fit a linear regression with the `lm` function:

```
R> reglm = lm(shan_complete$PM_Xuhui ~ . ,data = shan_complete)
R> summary.lm(reglm)
```

```
Call:
lm(formula = shan_complete$PM_Xuhui ~ ., data = shan_complete)

Residuals:
     Min       1Q   Median       3Q      Max
-132.139   -4.256   -0.195    4.279  176.450

Coefficients:
               Estimate Std. Error t value Pr(>|t|)
(Intercept)   -54.859483  40.975948  -1.339 0.180690
PM_Jingan       0.596490   0.014024  42.533  < 2e-16 ***
PM_US.Post      0.375636   0.015492  24.246  < 2e-16 ***
DEWP           -1.038941   0.170144  -6.106 1.10e-09 ***
HUMI            0.291713   0.045799   6.369 2.07e-10 ***
PRES            0.025287   0.038915   0.650 0.515852
TEMP            1.305543   0.168754   7.736 1.23e-14 ***
Iws            -0.007650   0.002027  -3.774 0.000163 ***
precipitation   0.462885   0.132139   3.503 0.000464 ***
```

```
Iprec          -0.125456   0.039025  -3.215 0.001314 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 10.68 on 4990 degrees of freedom
Multiple R-squared:  0.9409,        Adjusted R-squared:  0.9408
F-statistic:  8828 on 9 and 4990 DF,  p-value: < 2.2e-16
```

The variable PRES has no significant effect on the PM_Xuhui variable. We then perform a backward selection procedure, which leads to select 9 significant variables:

```
R> shan_lm = shan[1:5000,c(7,8,9,10,11,13,15,16,17)]
R> reglm = lm(shan_lm$PM_Xuhui ~ . ,data = shan_lm)
R> summary.lm(reglm)

Call:
lm(formula = shan_lm$PM_Xuhui ~ ., data = shan_lm)

Residuals:
     Min       1Q   Median       3Q      Max
-132.122   -4.265   -0.168    4.283  176.560

Coefficients:
               Estimate Std. Error t value Pr(>|t|)
(Intercept)   -28.365506   4.077590  -6.956 3.94e-12 ***
PM_Jingan       0.595564   0.013951  42.690  < 2e-16 ***
PM_US.Post      0.376486   0.015436  24.390  < 2e-16 ***
DEWP           -1.029188   0.169471  -6.073 1.35e-09 ***
HUMI            0.285759   0.044870   6.369 2.08e-10 ***
TEMP            1.275880   0.162453   7.854 4.90e-15 ***
Iws            -0.007734   0.002023  -3.824 0.000133 ***
precipitation   0.462137   0.132127   3.498 0.000473 ***
Iprec          -0.127162   0.038934  -3.266 0.001098 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 10.68 on 4991 degrees of freedom
Multiple R-squared:  0.9409,        Adjusted R-squared:  0.9408
F-statistic:  9933 on 8 and 4991 DF,  p-value: < 2.2e-16
```

The autocorrelation of the residual process shows that the errors are clearly not i.i.d., see Figure 3.10. We thus suspect the `lm` procedure to be unreliable in this context.

The autocorrelation function decreases pretty fast, and the partial autocorrelation function suggests that fitting an AR process on the residuals should be an appropriate method in this case. The automatic

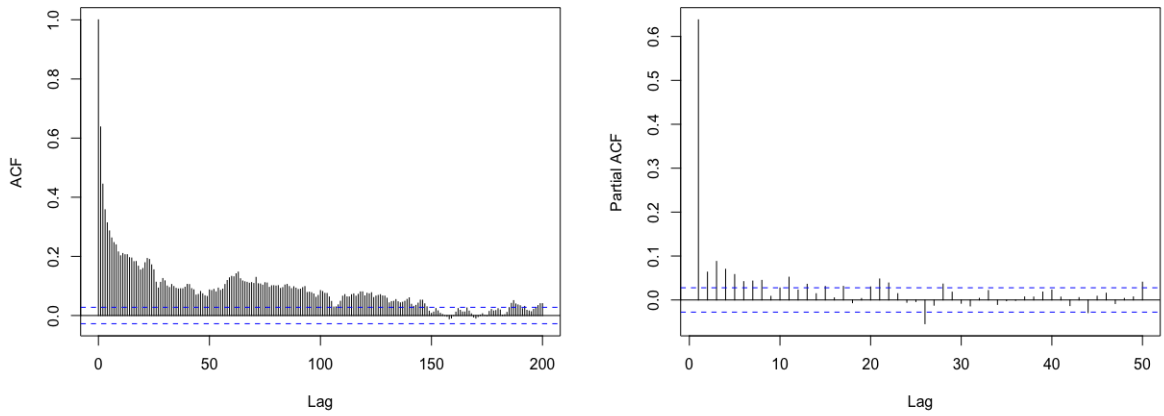


Figure 3.10 – Autocorrelation function (left) and partial autocorrelation function (right) of the residuals.

`fitAR` method of `slm` selects an AR process of order 28. The residuals of this AR fitting look like a white noise, as shown in Figure 3.11. Consequently, we propose to perform a linear regression with `slm`

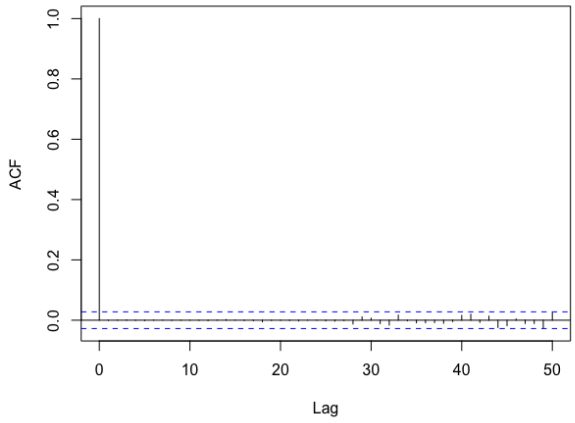


Figure 3.11 – Autocorrelation function of the residuals for the AR fitting.

function, using the `fitAR` method on the complete model

```
R> regslm = slm(shan_complete$PM_Xuhui ~ . ,data = shan_complete,
+          method_cov_st = "fitAR", model_selec = -1)
R> summary(regslm)

Call:
"slm(formula = myformula, data = data, x = x, y = y)"
```

```
Residuals:
     Min       1Q   Median       3Q      Max
-132.139   -4.256   -0.195    4.279  176.450

Coefficients:
                Estimate Std. Error z value Pr(>|z|)
(Intercept)   -54.859483 143.268399  -0.383 0.701783
PM_Jingan       0.596490   0.028467  20.953  < 2e-16 ***
PM_US.Post      0.375636   0.030869  12.169  < 2e-16 ***
DEWP           -1.038941   0.335909  -3.093 0.001982 **
HUMI            0.291713   0.093122   3.133 0.001733 **
PRES            0.025287   0.137533   0.184 0.854123
TEMP            1.305543   0.340999   3.829 0.000129 ***
Iws            -0.007650   0.005698  -1.343 0.179399
precipitation   0.462885   0.125641   3.684 0.000229 ***
Iprec          -0.125456   0.064652  -1.940 0.052323 .
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 10.68
Multiple R-squared:  0.9409
chi2-statistic:  8383 on 9 DF,  p-value: < 2.2e-16
```

Note that the variables show globally larger p-values than with the `lm` procedure, and more variables have no significant effect than with `lm`. After performing a backward selection we obtain the following results

```
R> shan_slm = shan[1:5000,c(7,8,9,10,11,13)]
R> regslm = slm(shan_slm$PM_Xuhui ~ . , data = shan_slm,
+          method_cov_st = "fitAR", model_selec = -1)
R> summary(regslm)

Call:
"slm(formula = myformula, data = data, x = x, y = y)"

Residuals:
     Min       1Q   Median       3Q      Max
-132.263   -4.341   -0.192    4.315  176.501

Coefficients:
             Estimate Std. Error z value Pr(>|z|)
(Intercept) -29.44924    8.38036  -3.514 0.000441 ***
```

```
PM_Jingan     0.60063    0.02911  20.636  < 2e-16 ***
PM_US.Post    0.37552    0.03172  11.840  < 2e-16 ***
DEWP         -1.05252    0.34131  -3.084 0.002044 **
HUMI          0.28890    0.09191   3.143 0.001671 **
TEMP          1.30069    0.32435   4.010 6.07e-05 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 10.71
Multiple R-squared:  0.9406
chi2-statistic:  8247 on 5 DF,  p-value: < 2.2e-16
```

The backward selection with `slm` only keeps 5 variables.

Acknowledgements

The authors are grateful to Anne Philippe and Aymeric Stamm for valuable discussions.

CHAPITRE 4

RÉGRESSION NON-PARAMÉTRIQUE GAUSSIENNE DÉPENDANTE

Introduction

Dans le Chapitre 1, nous avons développé une méthode d'estimation paramétrique par moindres carrés pour une fonction non-aléatoire dans le modèle de régression. À la fin de cette thèse, nous proposons d'étudier le modèle de régression dans le cas non-paramétrique via une approche par sélection de modèles, et en supposant que les erreurs forment une suite gaussienne. Ce sujet a déjà été largement traité notamment par Birgé et Massart quand les erreurs sont indépendantes et identiquement distribuées ; nous pouvons citer par exemple les articles [18], [19] et [20]. Baraud, Comte et Viennet se sont aussi intéressés à ce problème mais dans un cadre non-gaussien et en supposant certaines conditions de dépendance sur les variables explicatives du modèle et sur les erreurs [11], [12]. Dans notre cadre, les erreurs forment un processus gaussien et possèdent ainsi une structure de dépendance que nous prenons en compte dans nos résultats. Ce travail s'inspire beaucoup de la présentation qui a été faite dans le Chapitre 2 du livre de Giraud [43]. L'objectif est ici d'expliciter la forme de la fonction de pénalité dans le cadre de variables gaussiennes dépendantes et d'établir une inégalité oracle pour l'estimateur de risque minimal parmi une collection de modèles.

Contexte

Nous nous intéressons à l'estimation d'un vecteur non-aléatoire f^* appartenant à l'espace $\mathbb{R}^n$ dans le modèle :

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

où ϵ est un processus gaussien, c'est-à-dire $\epsilon \sim \mathcal{N}(0_{n\times 1}, \Sigma_{n\times n})$ avec Σ la matrice de covariance de taille n, et Y le vecteur des observations appartenant à $\mathbb{R}^n$. Nous définissons une collection d'espaces linéaires fini-dimensionnels $\{S_m, m \in \mathcal{M}\}$ que nous appelons des modèles. Pour chaque m dans $\mathcal{M}$, nous notons d_m la dimension de S_m et nous associons à chaque espace l'estimateur des moindres carrés $\hat{f}_m$ de f^* dans S_m défini par $\hat{f}_m = Proj_{S_m} Y$, où $Proj_{S_m}$ désigne l'opérateur de projection orthogonale sur l'espace S_m. Cet estimateur minimise la fonction de contraste des moindres carrés définie pour tous vecteurs t dans S_m par :

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

où $\|.\|_2$ est la norme euclidienne usuelle dans $\mathbb{R}^n$.

Pour quantifier la qualité d'un estimateur $\hat{f}_m$ de f^*, nous utilisons le risque ℓ^2 défini par :

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

Parmi la collection d'estimateurs $\{\hat{f}_m, m \in \mathcal{M}\}$, nous souhaitons sélectionner celui qui présente le risque minimal. Nous l'appelons l'estimateur oracle, noté $\hat{f}_{m_0}$, et il vérifie :

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

Le risque d'un estimateur se décompose généralement en deux parties : un terme de biais et un terme de variance. À partir de notre modèle $Y = f^* + \epsilon$, nous obtenons la décomposition suivante :

$$f^* - \hat{f}_m = (I - Proj_{S_m})f^* - Proj_{S_m}\epsilon.$$

En utilisant le théorème de Pythagore, nous pouvons écrire que :

$$\mathbb{E}\left[\left\|f^* - \hat{f}_m\right\|_2^2\right] = \|(I - Proj_{S_m})f^*\|_2^2 + \mathbb{E}\left[\|Proj_{S_m}\epsilon\|_2^2\right].$$

Le risque se décompose donc en un terme de biais $\|(I - Proj_{S_m})f^*\|^2$, qui reflète la qualité de l'approximation de f^* par sa projection sur l'espace S_m, et un terme de variance $\mathbb{E}\left[\|Proj_{S_m}\epsilon\|^2\right]$. Ces deux termes ont des comportements de nature opposés en fonction de la dimension. En effet, plus la dimension augmente plus le biais diminue, mais plus la variance augmente. Le but est donc de trouver la dimension qui équilibre le biais et la variance.

Étant donné que nous n'avons accès qu'aux données, nous ne pouvons malheureusement pas calculer directement les risques $R(\hat{f}_m)$ et utiliser l'estimateur $\hat{f}_{m_0}$. Nous prenons à la place le risque empirique qui est défini par :

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

Cependant nous ne pouvons pas utiliser seulement le risque empirique pour sélectionner un modèle. En effet ce critère aboutit systématiquement à sélectionner le modèle qui a la plus grande dimension, ce qui mène typiquement à des cas de surapprentissage. Pour compenser ce phénomène, il est nécessaire de pénaliser les plus grands modèles. L'idée d'utiliser une fonction de pénalisation n'est pas nouvelle ; cette idée remonte aux travaux précurseurs d'Akaike [1] et de Mallows [61]. Plus tard, Birgé et Massart ont développé une approche non-asymptotique de la sélection de modèles pénalisés [18], [19], [20]. Nous suivrons dans ce chapitre la stratégie développée par Birgé et Massart qui s'appuie sur un contrôle non-asymptotique des fluctuations du contraste empirique.

Nous souhaitons donc trouver l'estimateur $\hat{f}_{\hat{m}}$ tel que :

$$\hat{m} \in \operatorname{argmin}_{m\in\mathcal{M}}\left\{\left\|Y - \hat{f}_m\right\|_2^2 + pen(m)\right\}, \tag{4.1}$$

où $pen : \mathcal{M} \to \mathbb{R}^+$ est une fonction de pénalité. Une stratégie appropriée pour obtenir une bonne fonction

de pénalité est d'effectuer une analyse non-asymptotique du risque empirique et de choisir la pénalité de telle sorte que le risque de l'estimateur sélectionné soit le plus proche possible du risque oracle.

Pénalités et inégalités oracles

Pour commencer, nous associons à la collection de modèles $\{S_m, m \in \mathcal{M}\}$ une loi de probabilité $\pi = \{\pi_m, m \in \mathcal{M}\}$ sur $\mathcal{M}$, telle que $\sum_{m\in\mathcal{M}} e^{-\pi_m}$ converge. Nous rappelons que les erreurs ϵ forment une suite gaussienne de matrice de covariance Σ. Notons $(\lambda_i)_{\{1\leq i\leq n\}}$ les valeurs propres de Σ, et $\rho(\Sigma)$ son rayon spectral défini par :

$$\rho(\Sigma) = \max_{1\leq i\leq n} \lambda_i.$$

Selon le critère pénalisé défini par l'expression (4.1), pour une probabilité π et une constante $K > 1$ avec la pénalité :

$$pen_1(m) = K\left(\sqrt{\sum_{i=1}^{d_m}\lambda_i} + \sqrt{\rho(\Sigma)}\sqrt{2\log\left(\frac{1}{\pi_m}\right)}\right)^2, \tag{4.2}$$

le théorème ci-dessous montre que nous obtenons une inégalité oracle :

Théorème 4.1. *Pour l'estimateur $\hat{f}_{\hat{m}}$, il existe une constante $C_K > 1$, dépendant seulement de $K > 1$, telle que :*

$$\mathbb{E}\left[\left\|\hat{f}_{\hat{m}} - f^*\right\|_2^2\right] \leq C_K \min_{m\in\mathcal{M}}\left\{\mathbb{E}\left[\left\|\hat{f}_m - f^*\right\|_2^2\right] + \rho(\Sigma) + \left(\sqrt{\sum_{i=1}^{d_m}\lambda_i} + \sqrt{\rho(\Sigma)}\sqrt{2\log\left(\frac{1}{\pi_m}\right)}\right)^2\right\}. \tag{4.3}$$

Remarque 4.1. *Sous les mêmes conditions, la pénalité :*

$$pen_2(m) = K\rho(\Sigma)\left(\sqrt{d_m} + \sqrt{2\log\left(\frac{1}{\pi_m}\right)}\right)^2 \tag{4.4}$$

donne la borne oracle suivante :

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

Nous retrouvons une structure de pénalité très similaire à celle que nous trouvons dans Birgé et Massart [18], [19], pour le cas des erreurs gaussiennes i.i.d. Notamment pour la pénalité (4.4), nous remarquons que la seule différence est que la variance σ^2 a été remplacée par le rayon spectral $\rho(\Sigma)$. Si le rayon spectral est borné tout se passe donc, à une constante près, comme dans le cas i.i.d. En particulier, si la suite (ϵ_i) est stationnaire, nous avons la remarque suivante :

Remarque 4.2. *Si le processus gaussien ϵ est stationnaire et si la densité spectrale est bornée, alors le rayon spectral est borné et par conséquent la pénalité* (4.4) *est la même que dans le cas i.i.d., à une*

constante près.

La pénalité dépend de la probabilité π, que nous devons choisir de telle sorte que le terme $\sqrt{2\log(\frac{1}{\pi_m})}$ soit idéalement du même ordre de grandeur que $\sqrt{d_m}$. Ceci est possible si la collection de modèles n'est pas trop riche. Dans ce cas nous pouvons choisir une pénalité pen_2 proportionnelle à $\rho(\Sigma)d_m$, ou plus simplement encore proportionnelle à d_m :

$$pen_2(m) = K'\rho(\Sigma)d_m. \tag{4.5}$$

Heuristique de pente

Afin d'appliquer nos résultats, il reste encore à calibrer la constante de la pénalité directement à partir des données. Étant donné la grande similitude entre les pénalités dans le cas i.i.d. et dans le cas gaussien courte mémoire (voir Remarque 4.2), il est raisonnable d'utiliser la méthode de l'heuristique de pente, introduite par Birgé et Massart [18], [19], [20]. Des définitions de cette méthode sont notamment présentes dans l'article de Arlot et Massart [7], ou encore dans l'article de Arlot [5]. L'algorithme de l'heuristique de pente comporte deux versions principales dont l'une est appelée saut de dimension. Cette méthode est décrite en détails dans les Sections 2.3 et 2.4 de l'article de Arlot [5].

L'objectif est de calibrer la constante K' qui apparaît dans la pénalité pen_2, voir l'équation (4.5). Soit $\hat{m}(K')$ le modèle sélectionné par le critère pénalisé pour un choix de constante K' :

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

Dans notre cadre, l'algorithme du saut de dimension peut s'écrire de la façon suivante :

1. Calculer $(\hat{m}(K'))_{K'\geq 0}$,
2. Trouver $\widehat{K}'_{saut} > 0$ correspondant au « plus grand saut » de la fonction $K' \to d_{\hat{m}(K')}$,
3. Sélectionner $\hat{m} \in \operatorname{argmin}_{m\in\mathcal{M}} \left\{ \left\| Y - \hat{f}_m \right\|_2^2 + 2\widehat{K}'_{saut}\rho(\Sigma)d_m \right\}$.

Cet algorithme a été utilisé avec succès dans de nombreux contextes, et il s'appuie sur des justifications théoriques solides notamment dans le contexte de la régression gaussienne [20]. Nous nous référons aux articles de Arlot et Massart [7], de Baudry, Maugis et Michel [15] ou de Arlot [5] pour plus de détails sur les techniques de l'heuristique de pente ou du saut de dimension.

Simulations

Procédons maintenant à une simulation simple pour illustrer nos résultats en nous plaçant dans le cadre de la Remarque 4.2. Nous commençons par simuler les erreurs ϵ suivant le processus gaussien ARMA(2,1) suivant :

$$\epsilon_i - 0.4\epsilon_{i-1} - 0.2\epsilon_{i-2} = W_i + 0.3W_{i-1},$$

où W_i est une variable aléatoire suivant une loi gaussienne de moyenne 0 et de variance égale à 0.5. Nous choisissons la fonction déterministe f^* définie pour t appartenant à l'intervalle $[0,1]$ par :

$$3 - 0.1t + 0.5t^2 - t^3 + sin(8t),$$

et nous simulons un échantillon de données de taille $n = 1000$, défini pour tout i dans $\{1, \ldots, n\}$ par :

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

L'objectif est d'ajuster un régressogramme et de déterminer la meilleure partition régulière pour approcher la fonction f^*. Pour une dimension m allant de 1 à une dimension maximale, que nous prenons égale à 50, nous scindons l'intervalle $[0,1]$ en m intervalles et l'estimateur $\hat{f}_m$ est une fonction constante par morceaux, égale à la moyenne des Y_i sur chaque intervalle. Nous calculons ensuite le risque $\left\|\hat{f}_m - f^*\right\|_2^2$ pour m prenant les valeurs entières de 1 à 50. Cette simulation est répétée 100 fois et nous obtenons une courbe de risques moyens, affichée en Figure 4.1. Nous observons bien une courbe

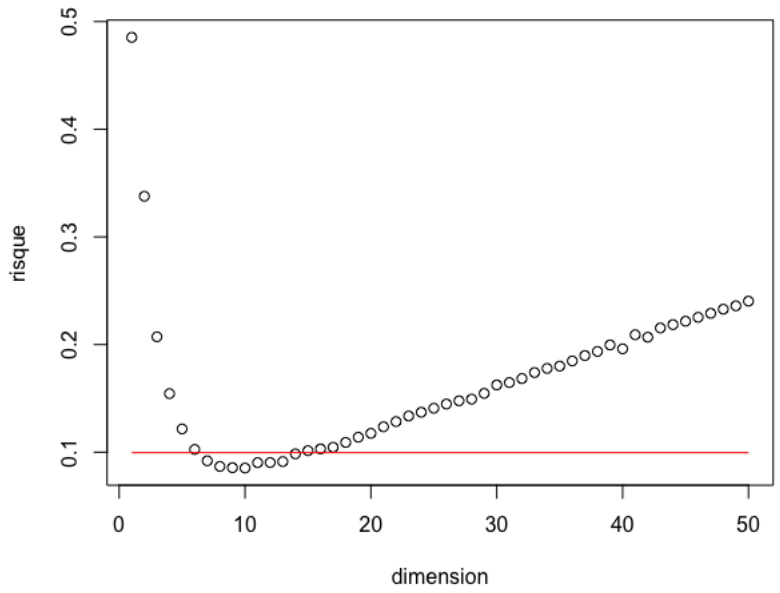


FIGURE 4.1 – Courbe de risques moyens sur 100 simulations, et risque moyen total de la procédure par heuristique de pente (ligne rouge).

qui décroît relativement vite au début lorsque le biais diminue, qui atteint un minimum vers la dimension 10 et enfin qui croît linéairement lorsque la variance domine.

Nous voulons maintenant évaluer les performances de l'algorithme du saut de dimension. Pour chaque simulation nous utilisons les données Y afin de calculer le risque $\left\|\hat{f}_m - Y\right\|_2^2$, puis nous utili-

sons la technique de l'heuristique de pente pour choisir la dimension qui présente le risque minimal. Cette opération est aussi répétée 100 fois.
La Figure 4.2 représente pour une simulation la fonction $K' \to d_{\hat{m}(K')}$ définie en page précédente et met en évidence la présence d'un « grand saut » (représenté par des petits traits bleus sur la figure). La Figure 4.3 est un boxplot représentant les dimensions choisies par l'algorithme. Ces dernières sont

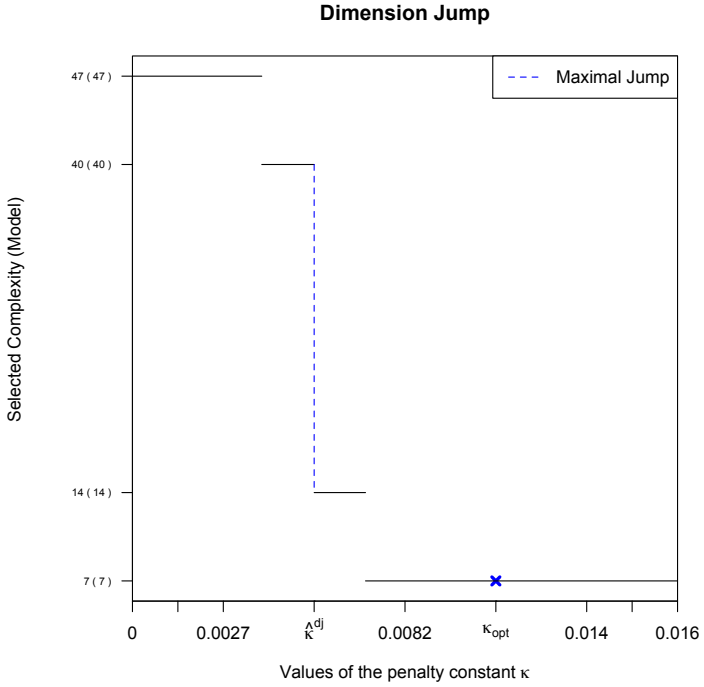


FIGURE 4.2 – Exemple d'un saut de dimension pour une simulation.

la plupart du temps autour de 10, ce qui correspond à la dimension ayant le risque minimal comme le montre la Figure 4.1. La ligne rouge en Figure 4.1 représente le risque moyen total sur les 100 simulations de la procédure en utilisant l'algorithme du saut de dimension. Nous pouvons remarquer que ce risque est assez proche du risque théorique minimal. Enfin la Figure 4.4 représente la fonction f^* ainsi que son estimation par un régressogramme de dimension 10, qui correspond à la dimension présentant le risque théorique moyen minimal.

Nous notons ici que les simulations ont été faites avec le logiciel R, en utilisant les fonctions du package R `Capushe`. Nous pouvons nous référer à l'article de Baudry, Maugis et Michel [15] pour plus d'informations sur la mise en pratique de l'heuristique de pente via le package `Capushe`.

En conclusion, il nous semble donc que la technique de l'heuristique de pente se comporte de façon raisonnable sur cette simulation. Des investigations supplémentaires seraient évidemment nécessaires pour étudier plus en détails le comportement de l'heuristique de pente dans le cas de variables aléatoires gaussiennes dépendantes.

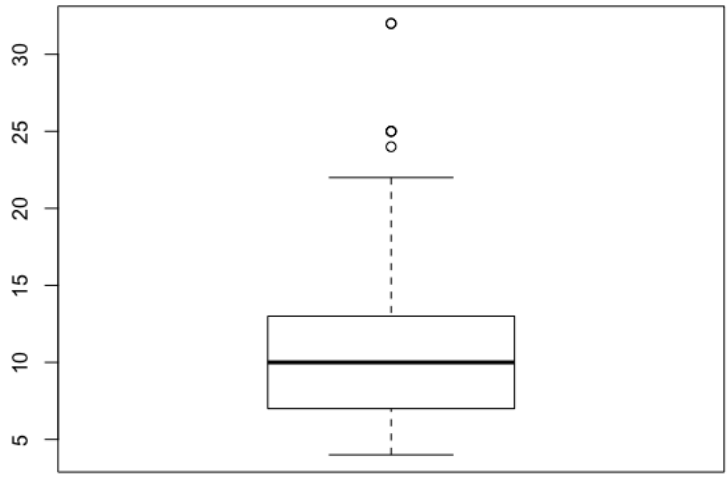


FIGURE 4.3 – Boxplot représentant les dimensions sélectionnées par l'algorithme du saut de dimension.

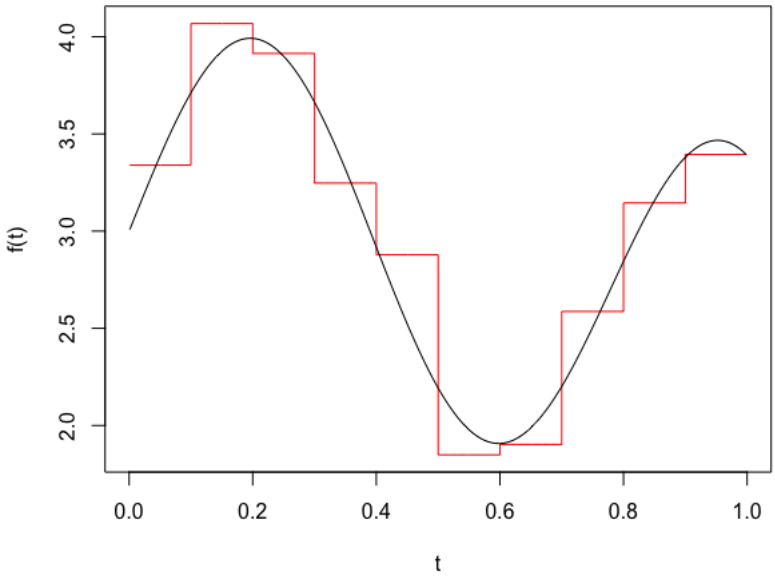


FIGURE 4.4 – Fonction f^* (en noir) et le régressogramme de dimension 10 (en rouge).

Preuve du Théorème 4.1

Nous suivons ici la preuve du Théorème 2.2 de Giraud [43]. Pour des raisons de clarté et de complétude nous reprenons toutes les étapes et nous mettons en évidence les endroits où la dépendance entre les variables intervient (via les valeurs propres de la matrice de covariance Σ et son rayon spectral).

Démonstration. En revenant à la définition de $\hat{m}$ (4.1), nous avons pour tout $m \in \mathcal{M}$:

$$\left\|Y - \hat{f}_{\hat{m}}\right\|^2 + pen(\hat{m}) \leq \left\|Y - \hat{f}_m\right\|^2 + pen(m),$$

$$\left\|\epsilon + (f^* - \hat{f}_{\hat{m}})\right\|^2 + pen(\hat{m}) \leq \left\|\epsilon + (f^* - \hat{f}_m)\right\|^2 + pen(m),$$

$$\|\epsilon\|^2 + \left\|f^* - \hat{f}_{\hat{m}}\right\|^2 + 2\langle\epsilon, (f^* - \hat{f}_{\hat{m}})\rangle + pen(\hat{m}) \leq \|\epsilon\|^2 + \left\|f^* - \hat{f}_m\right\|^2 + 2\langle\epsilon, (f^* - \hat{f}_m)\rangle + pen(m),$$

$$\left\|f^* - \hat{f}_{\hat{m}}\right\|^2 \leq \left\|f^* - \hat{f}_m\right\|^2 + 2\langle\epsilon, (f^* - \hat{f}_m)\rangle + 2\langle\epsilon, (\hat{f}_{\hat{m}} - f^*)\rangle + pen(m) - pen(\hat{m}).$$

Nous pouvons déjà simplifier l'inégalité ci-dessus en remarquant que la quantité $\mathbb{E}\left[\langle\epsilon, (f^* - \hat{f}_m)\rangle\right]$ est toujours négative. En effet :

$$\begin{aligned}\mathbb{E}\left[\langle\epsilon, (f^* - \hat{f}_m)\rangle\right] &= \mathbb{E}\left[\langle\epsilon, (f^* - Proj_{S_m}Y)\rangle\right] \\ &= \mathbb{E}\left[\langle\epsilon, (f^* - Proj_{S_m}(f^* + \epsilon))\rangle\right] = \mathbb{E}\left[\langle\epsilon, (f^* - Proj_{S_m}f^*)\rangle\right] - \mathbb{E}\left[\langle\epsilon, Proj_{S_m}\epsilon\rangle\right] \\ &= 0 - \mathbb{E}\left[\|Proj_{S_m}\epsilon\|^2\right] = -\mathbb{E}\left[\|Proj_{S_m}\epsilon\|^2\right],\end{aligned}$$

et cette quantité est toujours négative.

Nous avons le résultat principal suivant :

Proposition 4.1. *Pour des constantes $a > 1$ et $L_K \geq 0$, et pour la pénalité définie en* (4.2)*, il existe une variable aléatoire Z, avec $\mathbb{E}(Z) \leq L_K\rho(\Sigma)$, telle que :*

$$2\langle\epsilon, (\hat{f}_{\hat{m}} - f^*)\rangle - pen(\hat{m}) \leq a^{-1}\left\|\hat{f}_{\hat{m}} - f^*\right\|^2 + Z.$$

De ce fait, nous pouvons écrire :

$$\mathbb{E}\left[\left\|f^* - \hat{f}_{\hat{m}}\right\|^2\right] \leq \mathbb{E}\left[\left\|f^* - \hat{f}_m\right\|^2\right] + pen(m) + 2\mathbb{E}\left[\langle\epsilon, (f^* - \hat{f}_m)\rangle\right] + \mathbb{E}\left[2\langle\epsilon, (\hat{f}_{\hat{m}} - f^*)\rangle - pen(\hat{m})\right]$$

$$\mathbb{E}\left[\left\|f^* - \hat{f}_{\hat{m}}\right\|^2\right] \leq \mathbb{E}\left[\left\|f^* - \hat{f}_m\right\|^2\right] + pen(m) + a^{-1}\mathbb{E}\left[\left\|\hat{f}_{\hat{m}} - f^*\right\|^2\right] + \mathbb{E}(Z)$$

$$\frac{a-1}{a}\mathbb{E}\left[\left\|f^* - \hat{f}_{\hat{m}}\right\|^2\right] \leq \mathbb{E}\left[\left\|f^* - \hat{f}_m\right\|^2\right] + pen(m) + L_K\rho(\Sigma)$$

$$\mathbb{E}\left[\left\|f^* - \hat{f}_{\hat{m}}\right\|^2\right] \leq C_K\left(\mathbb{E}\left[\left\|f^* - \hat{f}_m\right\|^2\right] + \rho(\Sigma) + pen(m)\right)$$

avec $C_K = \max\left(\frac{a}{a-1}, \frac{aL_K}{a-1}\right)$, et $pen(m) = K\left(\sqrt{\sum_{i=1}^{d_m}\lambda_i} + \sqrt{\rho(\Sigma)}\sqrt{2\log\left(\frac{1}{\pi_m}\right)}\right)^2$. □

Preuve de la Proposition 4.1

Démonstration. Pour la preuve de cette proposition, nous avons besoin des deux lemmes suivants, qui font intervenir la dépendance du processus ϵ en remplaçant le terme de variance habituel σ^2 par le rayon spectral de la matrice de covariance $\rho(\Sigma)$:

Lemme 4.1. *Pour tout ensemble S et pour toute variable alétaoire $\eta \sim \mathcal{N}(0, I)$, la fonction $\eta \to \left\|Proj_S\sqrt{\Sigma}\eta\right\|$ est $\sqrt{\rho(\Sigma)}$-Lipschitzienne.*

Démonstration. Nous rappelons que $\|Proj_S z\| \leq \|z\|$. Ainsi nous obtenons :

$$\begin{aligned}\left\|Proj_S\sqrt{\Sigma}x - Proj_S\sqrt{\Sigma}y\right\| = \left\|Proj_S\sqrt{\Sigma}(x-y)\right\| &\leq \left\|\sqrt{\Sigma}(x-y)\right\| \\ &\leq |||\sqrt{\Sigma}||| \, \|x-y\| \leq \sqrt{\rho\left(\sqrt{\Sigma}\sqrt{\Sigma}^t\right)}\|x-y\| \leq \sqrt{\rho(\Sigma)}\|x-y\|.\end{aligned}$$

□

Lemme 4.2. *Pour tout $\epsilon \sim \mathcal{N}(0, \Sigma)$, nous avons l'inégalité de concentration suivante :*

$$\|Proj_S\epsilon\| \leq \mathbb{E}\|Proj_S\epsilon\| + \sqrt{\rho(\Sigma)}\sqrt{2\xi},$$

où ξ est une variable aléatoire suivant une loi exponentielle de paramètre 1.

Démonstration. Nous rappelons l'inégalité de concentration gaussienne de Cirel'son, Ibragimov et Sudakov [31], sous la forme énoncée dans l'annexe B du livre de Giraud [43] :

Théorème 4.2. *Supposons que $F : \mathbb{R}^d \to \mathbb{R}$ soit une fonction 1-Lipschitz et Z une loi gaussienne $\mathcal{N}(0, \sigma^2 I_d)$. Alors il existe une variable ξ suivant une loi exponentielle de paramètre 1 telle que :*

$$F(Z) \leq \mathbb{E}[F(Z)] + \sigma\sqrt{2\xi}.$$

Dans le cas d'une suite gaussienne nous avons $\epsilon \sim \mathcal{N}(0, \Sigma)$, ainsi ϵ peut s'écrire : $\epsilon = \sqrt{\Sigma}\eta$ avec $\eta \sim \mathcal{N}(0, I)$. En utilisant le théorème précédent avec la fonction $\eta \to \left\|Proj_S\sqrt{\Sigma}\eta\right\|$, nous obtenons :

$$\left\|Proj_S\sqrt{\Sigma}\eta\right\| \leq \mathbb{E}\left\|Proj_S\sqrt{\Sigma}\eta\right\| + \sqrt{\rho(\Sigma)}\sqrt{2\xi}.$$

□

Revenons à la preuve de la Proposition 4.1. Notons $< f^* >$ l'espace linéaire engendré par f^*, $\bar{S}_m$ l'espace $\bar{S}_m = S_m + < f^* >$ et $\tilde{S}_m$ l'orthogonal de $< f^* >$ dans $\bar{S}_m$. En particulier $\bar{S}_m$ est la

somme orthogonale de $< f^* >$ et $\tilde{S}_m$, qui est notée par $\bar{S}_m =< f^* > \oplus^{\perp} \tilde{S}_m$. En appliquant l'inégalité $2\langle x, y\rangle \leq a||x||^2 + ||y||^2/a$ pour $a > 1$, nous obtenons :

$$2\langle \epsilon, \hat{f}_{\hat{m}} - f^* \rangle - pen(\hat{m}) = 2\langle Proj_{\bar{S}_{\hat{m}}}\epsilon, \hat{f}_{\hat{m}} - f^* \rangle - pen(\hat{m})$$
$$\leq a \left\| Proj_{\bar{S}_{\hat{m}}}\epsilon \right\|^2 + a^{-1} \left\| \hat{f}_{\hat{m}} - f^* \right\|^2 - pen(\hat{m}),$$

et nous avons :

$$a \left\| Proj_{\bar{S}_{\hat{m}}}\epsilon \right\|^2 + a^{-1} \left\| \hat{f}_{\hat{m}} - f^* \right\|^2 - pen(\hat{m})$$
$$\leq a \left\| Proj_{<f^*>}\epsilon \right\|^2 + a \left\| Proj_{\tilde{S}_{\hat{m}}}\epsilon \right\|^2 + a^{-1} \left\| \hat{f}_{\hat{m}} - f^* \right\|^2 - pen(\hat{m})$$
$$\leq Z + a^{-1} \left\| \hat{f}_{\hat{m}} - f^* \right\|^2, \quad (4.6)$$

où Z est la variable aléatoire $a \left\| Proj_{<f^*>}\epsilon \right\|^2 + a \left\| Proj_{\tilde{S}_{\hat{m}}}\epsilon \right\|^2 - pen(\hat{m})$. Pour le premier terme $\left\| Proj_{<f^*>}\epsilon \right\|^2$, nous remarquons que :

$$\mathbb{E}\left(\left\| Proj_{<f^*>}\epsilon \right\|^2 \right) = \mathbb{E}\left[tr\left((Proj_{<f^*>}\epsilon)'(Proj_{<f^*>}\epsilon) \right) \right] = \mathbb{E}\left[tr\left(Proj_{<f^*>}\epsilon\epsilon' Proj'_{<f^*>} \right) \right],$$

car pour deux matrices A et B, la trace du produit AB est égale à celle du produit BA, et :

$$\mathbb{E}\left[tr\left(Proj_{<f^*>}\epsilon\epsilon' Proj'_{<f^*>} \right) \right] = tr\left(Proj_{<f^*>}\Sigma Proj'_{<f^*>} \right) \leq \rho(\Sigma),$$

car $\epsilon \sim \mathcal{N}(0, \Sigma)$. Là encore la dépendance intervient et le rayon spectral remplace le terme de variance.

Dans la suite nous choisissons $a = \frac{K+1}{2} > 1$, mais nous continuons à travailler avec la lettre a pour plus de clarté dans les calculs. Pour le deuxième terme nous avons :

$$\mathbb{E}\left[a \left\| Proj_{\tilde{S}_{\hat{m}}}\epsilon \right\|^2 - pen(\hat{m}) \right] \leq a\mathbb{E}\left[\max_{m \in \mathcal{M}} \left(\left\| Proj_{\tilde{S}_m}\epsilon \right\|^2 - \frac{1}{a} pen(m) \right) \right] \leq a \sum_{m \in \mathcal{M}} \mathbb{E}\left[\left(\left\| Proj_{\tilde{S}_m}\epsilon \right\|^2 - \frac{1}{a} pen(m) \right)_+ \right].$$

Grâce au lemme 4.2, nous pouvons contrôler la quantité $\left\| Proj_{\tilde{S}_m}\epsilon \right\|$:

$$\left\| Proj_{\tilde{S}_m}\epsilon \right\| \leq \mathbb{E}\left(\left\| Proj_{\tilde{S}_m}\epsilon \right\| \right) + \sqrt{\rho(\Sigma)}\sqrt{2\xi_m}. \quad (4.7)$$

Étant donné que $\mathbb{E}\left(\left\| Proj_{\tilde{S}_m}\epsilon \right\| \right) \leq \mathbb{E}\left(\left\| Proj_{\tilde{S}_m}\epsilon \right\|^2 \right)^{1/2}$, calculons :

$$\mathbb{E}\left[\left\| Proj_{\tilde{S}_m}\epsilon \right\|^2 \right] = tr\left(\mathbb{E}\left[\left\| Proj_{\tilde{S}_m}\epsilon \right\|^2 \right] \right) = \mathbb{E}\left[tr\left(\epsilon' Proj'_{\tilde{S}_m} Proj_{\tilde{S}_m}\epsilon \right) \right]$$
$$= \mathbb{E}\left[tr\left(Proj_{\tilde{S}_m}\epsilon\epsilon' Proj'_{\tilde{S}_m} \right) \right] = Proj_{\tilde{S}_m} tr\left(\mathbb{E}\left[\epsilon\epsilon' \right] \right) Proj'_{\tilde{S}_m} = tr\left(Proj_{\tilde{S}_m}\Sigma Proj_{\tilde{S}_m} \right) \leq \sum_{i=1}^{d_m} \lambda_i, \quad (4.8)$$

où les λ_i sont les valeurs propres de Σ telles que $\lambda_1 \geq \lambda_2 \geq \ldots \lambda_n$. Pour obtenir la dernière inégalité nous avons utilisé un résultat connu de l'Analyse en Composantes Principales. Le lecteur peut trouver

plus d'explications dans le livre de Jolliffe [52]. Nous observons ici que la dépendance du processus ϵ est représentée par les valeurs propres de la matrice Σ.

Ainsi nous obtenons :

$$\left\|Proj_{\tilde{S}_m}\epsilon\right\| \leq \sqrt{\sum_{i=1}^{d_m}\lambda_i} + \sqrt{\rho(\Sigma)}\sqrt{2\xi_m}.$$

Nous pouvons donc revenir à l'étude de la variable Z :

$$\begin{aligned}\mathbb{E}(Z) &= a\mathbb{E}\left(\left\|Proj_{<f^*>}\epsilon\right\|^2\right) + a\left(\mathbb{E}\left[\left\|Proj_{\tilde{S}_{\hat{m}}}\epsilon\right\|^2 - \frac{1}{a}pen(\hat{m})\right]\right)\\ &\leq a\rho(\Sigma) + a\sum_{m\in\mathcal{M}}\mathbb{E}\left[\left(\left\|Proj_{\tilde{S}_m}\epsilon\right\|^2 - \frac{1}{a}pen(m)\right)_+\right]\\ &\leq a\rho(\Sigma) + a\sum_{m\in\mathcal{M}}\mathbb{E}\left[\left(\left(\sqrt{\sum_{i=1}^{d_m}\lambda_i} + \sqrt{\rho(\Sigma)}\sqrt{2\xi_m}\right)^2 - \frac{1}{a}pen(m)\right)_+\right].\end{aligned}$$

Définissons la pénalité suivante :

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

alors :

$$\mathbb{E}(Z) \leq a\rho(\Sigma) + a\sum_{m\in\mathcal{M}}\mathbb{E}\left[\left(\left(\sqrt{\sum_{i=1}^{d_m}\lambda_i} + \sqrt{\rho(\Sigma)}\sqrt{2\xi_m}\right)^2 - \frac{K}{a}\left(\sqrt{\sum_{i=1}^{d_m}\lambda_i} + \sqrt{\rho(\Sigma)}\sqrt{2\log\left(\frac{1}{\pi_m}\right)}\right)^2\right)_+\right].$$

En utilisant l'inégalité $(x+y)^2 \leq (1+\alpha)x^2 + (1+\alpha^{-1})y^2$, avec $\alpha = \frac{K-a}{a}$, nous avons :

$$\begin{aligned}\left(\sqrt{\sum_{i=1}^{d_m}\lambda_i} + \sqrt{\rho(\Sigma)}\sqrt{2\xi_m}\right)^2 &\leq \left(\sqrt{\sum_{i=1}^{d_m}\lambda_i} + \sqrt{\rho(\Sigma)}\sqrt{2\log\left(\frac{1}{\pi_m}\right)} + \sqrt{\rho(\Sigma)}\sqrt{2\left(\xi_m - log\left(\frac{1}{\pi_m}\right)\right)_+}\right)^2\\ &\leq \frac{K}{a}\left(\sqrt{\sum_{i=1}^{d_m}\lambda_i} + \sqrt{\rho(\Sigma)}\sqrt{2\log\left(\frac{1}{\pi_m}\right)}\right)^2 + \frac{2K\rho(\Sigma)}{K-a}\left(\xi_m - log\left(\frac{1}{\pi_m}\right)\right)_+.\end{aligned}$$

Ainsi :

$$\begin{aligned}&\mathbb{E}\left[\left(\left(\sqrt{\sum_{i=1}^{d_m}\lambda_i} + \sqrt{\rho(\Sigma)}\sqrt{2\xi_m}\right)^2 - \frac{K}{a}\left(\sqrt{\sum_{i=1}^{d_m}\lambda_i} + \sqrt{\rho(\Sigma)}\sqrt{2\log\left(\frac{1}{\pi_m}\right)}\right)^2\right)_+\right]\\ &\qquad\leq \mathbb{E}\left[\frac{2K\rho(\Sigma)}{K-a}\left(\xi_m - log\left(\frac{1}{\pi_m}\right)\right)_+\right] \leq \frac{2K\rho(\Sigma)}{K-a}\pi_m,\end{aligned}$$

car $\mathbb{E}\left[\left(\xi_m - log\left(\frac{1}{\pi_m}\right)\right)_+\right] = \exp(-\log(\frac{1}{\pi_m})) = \pi_m$. En conséquence, puisque $\sum_{m\in\mathcal{M}} \pi_m = 1$:

$$\mathbb{E}(Z) \leq a\rho(\Sigma) + a\sum_{m\in\mathcal{M}} \frac{2K\rho(\Sigma)}{K-a}\pi_m \leq a\rho(\Sigma) + \frac{2aK\rho(\Sigma)}{K-a} \leq \frac{3aK-a^2}{K-a}\rho(\Sigma), \tag{4.9}$$

et nous pouvons noter $L_K = \frac{3aK-a^2}{K-a}$. Nous rappelons que nous avons choisi $a = \frac{K+1}{2}$, donc $L_K = \frac{5K^2+4K-1}{2K-2}$ qui est bien une quantité positive pour $K > 1$.

En combinant (4.6) et (4.9), nous obtenons la Proposition 4.1.

□

PERSPECTIVES DE RECHERCHE

Les sujets abordés dans cette thèse mènent naturellement vers des thématiques de recherche concernant l'étude des techniques de sélection de modèles pour les variables aléatoires dépendantes. Ce sujet est intéressant à étudier parce que beaucoup de jeux de données possèdent une structure de dépendance qui n'est malheureusement pas toujours prise en compte. Or la plupart des techniques de sélection de modèles ont été développées pour des variables aléatoires indépendantes.

Nous égrenons ici brièvement quelques pistes de recherche sur lesquelles nous souhaitons travailler à court terme.

Dans un premier temps, nous voulons approfondir le Chapitre 4 de cette thèse. Dans le cas de la régression non-paramétrique avec des erreurs stationnaires gaussiennes à courte mémoire, nous étudierons la vitesse de convergence de l'estimateur oracle ainsi que son adaptativité sur certains espaces de fonctions régulières. Ensuite nous désirons étudier ce même modèle mais dans le cas où les erreurs forment un processus gaussien à longue mémoire. Dans ce contexte, le rayon spectral de la matrice de covariance des erreurs n'est plus borné et une autre pénalité, ainsi qu'une nouvelle inégalité oracle, doivent être établies.

Dans un deuxième temps, nous souhaitons continuer l'étude du modèle de régression non-paramétrique mais en généralisant les résultats obtenus dans le cas précédent pour des erreurs dépendantes non gaussiennes. Les récents travaux de thèse d'Antoine Marchina [62] (inégalités de concentration pour les variables aléatoires indépendantes), qui sont susceptibles d'être étendus au cas dépendant, fournissent une piste de recherche solide permettant de s'intéresser à ce problème.

Ensuite, nous souhaiterions étudier la technique de l'heuristique de pente pour des variables aléatoires dépendantes, toujours dans le cadre de la régression non-paramétrique. Étant donné la grande similitude entre la pénalité que nous avons définie au Chapitre 4 et celle du cas i.i.d., nous nous sommes permis d'utiliser cette méthode dans le but de procéder à quelques applications. Cela semble bien fonctionner en pratique, mais il serait intéressant de justifier théoriquement l'utilisation de cette méthode. Nous pourrions nous inspirer des résultats de Lerasle [56] qui a prouvé que la méthode de l'heuristique de pente fonctionnait dans le cadre de l'estimation de densité pour des suites de variables aléatoires dépendantes vérifiant certaines conditions de mélange.

Enfin, il serait bon de s'intéresser à des problèmes de grande dimension dans le cadre du modèle de régression linéaire avec erreurs dépendantes.
En reprenant les idées du livre de Giraud [43], nous pourrions commencer par étudier le cas de la « coordinate sparse regression ».

BIBLIOGRAPHIE

[1] H. Akaike. Information theory and an extension of the maximum likelihood principle, proceedings of the 2nd international symposium on information, bn petrow, f. Czaki, Akademiai Kiado, Budapest, 1973.

[2] P. Alquier and P. Doukhan. Sparsity considerations for dependent variables. Electronic journal of statistics, 5 :750–774, 2011.

[3] P. Alquier and O. Wintenberger. Model selection for weakly dependent time series forecasting. Bernoulli, 18(3) :883–913, 2012.

[4] T. W. Anderson. The statistical analysis of time series, volume 19. John Wiley & Sons, 2011.

[5] S. Arlot. Minimal penalties and the slope heuristics : a survey. arXiv preprint arXiv :1901.07277, 2019.

[6] S. Arlot and A. Celisse. A survey of cross-validation procedures for model selection. Statistics surveys, 4 :40–79, 2010.

[7] S. Arlot and P. Massart. Data-driven calibration of penalties for least-squares regression. Journal of Machine learning research, 10(Feb) :245–279, 2009.

[8] J.-M. Azaïs and J.-M. Bardet. Le modèle linéaire par l'exemple-2e éd. : Régression, analyse de la variance et plans d'expérience illustrés avec R et SAS. Dunod, 2012.

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

[10] 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.

[11] Y. Baraud, F. Comte, and G. Viennet. Adaptive estimation in autoregression or-mixing regression via model selection. The Annals of Statistics, 29(3) :839–875, 2001.

[12] Y. Baraud, F. Comte, and G. Viennet. Model selection for (auto-) regression with dependent data. ESAIM : Probability and Statistics, 5 :33–49, 2001.

[13] A. Barron, L. Birgé, and P. Massart. Risk bounds for model selection via penalization. Probability theory and related fields, 113(3) :301–413, 1999.

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

[15] J.-P. Baudry, C. Maugis, and B. Michel. Slope heuristics : overview and implementation. Statistics and Computing, 22(2) :455–470, 2012.

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

[17] L. Birgé and P. Massart. From model selection to adaptive estimation. In Festschrift for lucien le cam, pages 55–87. Springer, 1997.

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

[19] L. Birgé and P. Massart. A generalized cp criterion for gaussian model selection. 2001.

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

[21] 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.

[22] D. R. Brillinger. Time series : data analysis and theory. SIAM, 2001.

[23] P. J. Brockwell and R. A. Davis. Time Series : theory and methods. Springer Science & Business Media, 1991.

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

[25] 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.

[26] E. Caron, J. Dedecker, and B. Michel. Linear regression with stationary errors : the R package slm. arXiv preprint arXiv :1906.06583, 2019.

[27] S. Chatterjee and A. S. Hadi. Influential observations, high leverage points, and outliers in linear regression. Statistical science, 1(3) :379–393, 1986.

[28] R. Y. Chen, A. Gittens, and J. A. Tropp. The masked sample covariance estimator : an analysis using matrix concentration inequalities. Information and Inference : A Journal of the IMA, 1(1) :2–20, 2012.

[29] S. X. Chen. On the accuracy of empirical likelihood confidence regions for linear regression model. Annals of the Institute of Statistical Mathematics, 45(4) :621–637, 1993.

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

[31] B. Cirel'son, I. Ibragimov, and V. Sudakov. Norms of gaussian sample functions. In Proceedings of the Third Japan–USSR Symposium on Probability Theory, volume 550 of Lecture Notes in Mathematics, pages 20–41. Springer-Verlag, Berlin, 1976.

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

[33] L. Davies. The asymptotics of s-estimators in the linear regression model. The Annals of Statistics, pages 1651–1675, 1990.

[34] J. Dedecker. A central limit theorem for stationary random fields. Probability Theory and Related Fields, 110(3) :397–426, 1998.

[35] J. Dedecker. On the optimality of Mcleish's conditions for the central limit theorem. Comptes Rendus Mathematique, 353(6) :557–561, 2015.

[36] J. Dedecker, H. Dehling, and M. S. Taqqu. Weak convergence of the empirical process of intermittent maps in $\mathbb{L}_2$ under long-range dependence. Stochastics and Dynamics, 15(02) :1550008, 2015.

[37] J. Dedecker, S. Gouëzel, and F. Merlevède. Some almost sure results for unbounded functions of intermittent maps and their associated markov chains. In Annales de l'institut Henri Poincaré (B), volume 46, pages 796–821, 2010.

[38] 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.

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

[40] H. Drygas. Weak and strong consistency of the least squares estimators in regression models. Probability Theory and Related Fields, 34(2) :119–127, 1976.

[41] S. Efromovich. Data-driven efficient estimation of the spectral density. Journal of the American Statistical Association, 93(442) :762–769, 1998.

[42] F. Eicker. Asymptotic normality and consistency of the least squares estimators for families of linear regressions. The Annals of Mathematical Statistics, 34(2) :447–456, 1963.

[43] C. Giraud. Introduction to high-dimensional statistics. Chapman and Hall/CRC, 2014.

[44] A. Goldenshluger and O. Lepski. Bandwidth selection in kernel density estimation : oracle inequalities and adaptive minimax optimality. The Annals of Statistics, 39(3) :1608–1632, 2011.

[45] M. I. Gordin. Central limit theorem for stationary processes. Doklady Akademii Nauk SSSR, 188(4) :739, 1969.

[46] U. Grenander and M. Rosenblatt. Statistical analysis of stationary time series, volume 320. American Mathematical Soc., 2008.

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

[48] X. He and Q.-M. Shao. 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.

[49] R. R. Hocking. Methods and applications of linear models : regression and the analysis of variance. John Wiley & Sons, 2013.

[50] A. E. Hoerl and R. W. Kennard. Ridge regression : Biased estimation for nonorthogonal problems. Technometrics, 12(1) :55–67, 1970.

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

[52] I. Jolliffe. Principal component analysis. Springer, 2002.

[53] W. Krämer. Finite sample efficiency of ordinary least squares in the linear regression model with autocorrelated errors. Journal of the American Statistical Association, 75(372) :1005–1009, 1980.

[54] C. Lacour, P. Massart, and V. Rivoirard. Estimator selection : a new method with applications to kernel density estimation. Sankhya A, 79(2) :298–335, 2017.

[55] M. Lerasle. Adaptive density estimation of stationary β-mixing and τ-mixing processes. Mathematical Methods of statistics, 18(1) :59–83, 2009.

[56] M. Lerasle et al. Optimal model selection for density estimation of stationary data under various mixing conditions. The Annals of Statistics, 39(4) :1852–1877, 2011.

[57] E. Levina and R. Vershynin. Partial estimation of covariance matrices. Probability theory and related fields, 153(3-4) :405–419, 2012.

[58] X. Liang, S. Li, S. Zhang, H. Huang, and S. X. Chen. Pm2.5 data reliability, consistency, and air quality assessment in five chinese cities. Journal of Geophysical Research : Atmospheres, 121(17) :10–220, 2016.

[59] W. Liu and W. B. Wu. Asymptotics of spectral density estimates. Econometric Theory, pages 1218–1245, 2010.

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

[61] C. L. Mallows. Some comments on cp. Technometrics, 15(4) :661–675, 1973.

[62] A. Marchina. Inégalités de concentration pour des fonctions de variables aléatoires indépendantes. PhD thesis, Paris Saclay, 2017.

[63] P. Massart. Some applications of concentration inequalities to statistics. In Annales de la Faculté des sciences de Toulouse : Mathématiques, volume 9, pages 245–303, 2000.

[64] P. Massart. Concentration inequalities and model selection, volume 1896 of Lecture Notes in Mathematics. Springer-Verlag Berlin Heidelberg, 2007.

[65] 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.

[66] V. Pipiras and M. Taqqu. Long-Range Dependence and Self-Similarity. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press, 2017.

[67] M. B. Priestley. Spectral analysis and time series. 1981.

[68] E. Rio. Théorie asymptotique des processus aléatoires faiblement dépendants, volume 31. Springer Science & Business Media, 1999.

[69] M. Rosenblatt. A central limit theorem and a strong mixing condition. Proceedings of the National Academy of Sciences, 42(1) :43–47, 1956.

[70] M. Rosenblatt. Stationary sequences and random fields. Springer Science & Business Media, 2012.

[71] E. Seneta. Regularly varying functions, volume 508. Springer, 2006.

[72] R. Tibshirani. Regression shrinkage and selection via the lasso. Journal of the Royal Statistical Society : Series B (Methodological), 58(1) :267–288, 1996.

[73] F. A. Wolak. An exact test for multiple inequality and equality constraints in the linear regression model. Journal of the American Statistical Association, 82(399) :782–793, 1987.

[74] 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.

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

[76] W. B. Wu. Asymptotic theory for stationary processes. Stat. Interface, 4(2) :207–226, 2011.

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

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

UNIVERSITE BRETAGNE LOIRE / MATHSTIC

CENTRALE NANTES

Titre : Comportement des estimateurs des moindres carrés du modèle linéaire dans un contexte dépendant : étude asymptotique, implémentation, exemples

Mot clés : Modèle linéaire, Processus stationnaires, Estimation paramétrique, Estimateur des moindres carrés, Normalité asymptotique, Densité spectrale, Tests d'hypothèses, Estimation non-paramétrique.

Resumé : Dans cette thèse, nous nous intéressons au modèle de régression linéaire usuel dans le cas où les erreurs sont supposées strictement stationnaires. Nous utilisons un résultat de Hannan (1973) qui a prouvé un Théorème Limite Central pour l'estimateur des moindres carrés sous des conditions très générales sur le design et le processus des erreurs. Pour un design et un processus d'erreurs vérifiant les conditions d'Hannan, nous définissons un estimateur de la matrice de covariance asymptotique de l'estimateur des moindres carrés et nous prouvons sa consistance sous des conditions très générales. Ensuite nous montrons comment modifier les tests usuels sur le paramètre du modèle linéaire dans ce contexte dépendant. Nous proposons différentes approches pour estimer la matrice de covariance afin de corriger l'erreur de première espèce des tests. Le paquet R `slm` que nous avons développé contient l'ensemble de ces méthodes statistiques. Les procédures sont évaluées à travers différents ensembles de simulations et deux exemples particuliers de jeux de données sont étudiés. Enfin, dans le dernier chapitre, nous proposons une méthode non-paramétrique par pénalisation pour estimer la fonction de régression dans le cas où les erreurs sont gaussiennes et corrélées.

Title : Behaviour of the least squares estimators of the linear model in a dependent context : asymptotic properties, implementation, examples

Keywords : Linear model, Stationary processes, Parametric estimation, Least squares estimators, Asymptotic normality, Spectral density, Hypothesis testing, Nonparametric estimation.

Abstract : In this thesis, we consider the usual linear regression model in the case where the error process is assumed strictly stationary. We use a result from Hannan (1973) who proved a Central Limit Theorem for the usual least squares estimator under general conditions on the design and on the error process. Whatever the design and the error process satisfying Hannan's conditions, we define an estimator of the asymptotic covariance matrix of the least squares estimator and we prove its consistency under very mild conditions. Then we show how to modify the usual tests on the parameter of the linear model in this dependent context. We propose various methods to estimate the covariance matrix in order to correct the type I error rate of the tests. The R package `slm` that we have developed contains all of these statistical methods. The procedures are evaluated through different sets of simulations and two particular examples of datasets are studied. Finally, in the last chapter, we propose a non-parametric method by penalization to estimate the regression function in the case where the errors are Gaussian and correlated.